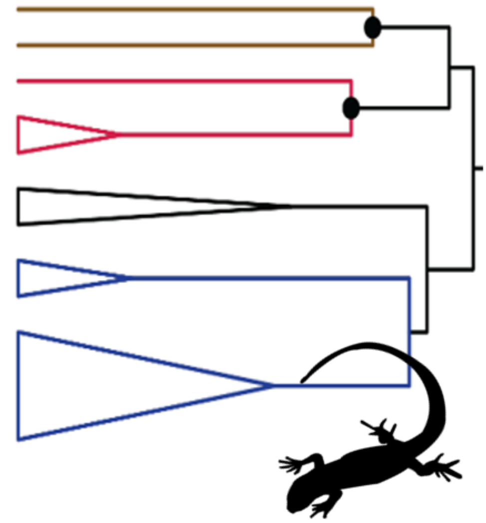
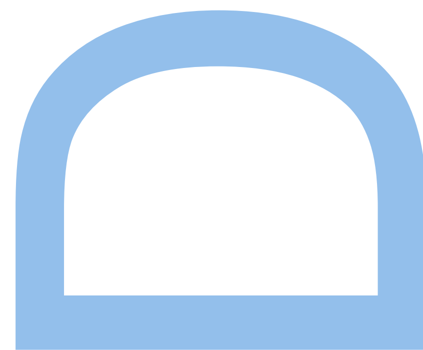
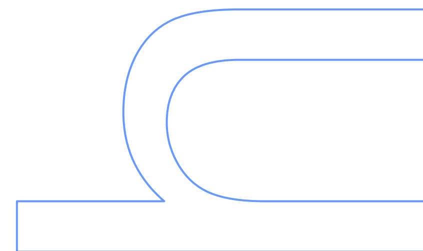


Evolutionary history of selected squamates: insights from nuclear genes and species tree with implications for biogeography and taxonomy



Joana da Silva Mendes

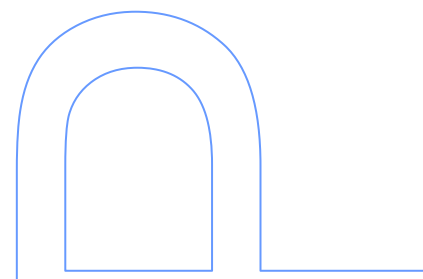
Doutoramento em Biodiversidade, Genética e Evolução
Departamento de Biologia
2017



Orientador
David James Harris, Professor Associado Convidado e Investigador, Faculdade de Ciências, Universidade do Porto

Co-orientador
Daniele Salvi, Professor Associado e Investigador
University of L'Aquila; CIBIO/InBIO (Universidade do Porto)

Co-orientador
Salvador Carranza, Investigador, Institute of Evolutionary Biology (CSIC - Universitat Pompeu Fabra)



NOTA PRÉVIA

Na elaboração desta tese, e nos termos do número 2 do Artigo 4º do Regulamento Geral dos Terceiros Ciclos de Estudos da Universidade do Porto e do Artigo 31º do Decreto-Lei 74/2006, de 24 de Março, com a nova redação introduzida pelo Decreto-Lei 230/2009, de 14 de Setembro, foi efetuado o aproveitamento total de um conjunto coerente de trabalhos de investigação já publicados ou submetidos para publicação em revistas internacionais indexadas e com arbitragem científica, os quais integram alguns dos capítulos da presente tese. Tendo em conta que os referidos trabalhos foram realizados com a colaboração de outros autores, o candidato esclarece que, em todos eles, participou ativamente na sua conceção, na obtenção, análise e discussão de resultados, bem como na elaboração da sua forma publicada. A instituição de origem da candidata foi a Faculdade de Ciências da Universidade do Porto, tendo o trabalho sido realizado sob orientação do Doutor David James Harris, Professor Associado Convidado no Departamento de Biologia da Faculdade de Ciências da Universidade do Porto e Investigador Principal do Centro de Investigação em Biodiversidade e Recursos Genéticos (CIBIO-InBio) e sob co-orientação do Doutor Daniele Salvi, Professor Associado na Universidade de L'Aquila e Investigador do CIBIO/InBIO. A instituição de acolhimento foi o Institut de Biología Evolutiva do Consejo Superior de Investigaciones Científicas – Universitat Pompeu Fabra (IBE-CSIC-UPF), sob a coorientação do Professor Doutor Salvador Carranza. O trabalho laboratorial foi realizado no CIBIO-InBio e no IBE-CSIC-UPF.

Este trabalho foi apoiado pela Fundação para a Ciência e Tecnologia (FCT) através da atribuição da bolsa de doutoramento (SFRH/BD/81528/2011).

Governo da República Portuguesa

UNIÃO EUROPEIA
Fundo Social Europeu

ACKNOWLEDGMENTS / AGRADECIMENTOS

The work accomplished in this thesis would have not been possible without many people and institutions and is the result of their efforts. They have help me grown as a young scientist and person. Here I express my sincere gratitude to all of them.

To my supervisor D. James Harris and co-supervisors Salvador Carranza and Daniele Salvi for having believed in me and helping drive this thesis to conclusion. Thank you for your encouraging ideas when I needed them most. I am extremely grateful for the possibility of working under your guidance, you have helped develop the ideas to work on and without you all this thesis would have not been accomplished.

To D. James Harris, thank you for accepting me in your remarkable team, for all the support along these long years, the great ideas to improve my work, which come from the amazing scientific career of yours, and for the opportunity to go twice to Morocco, which end up as being a great experience.

To Salvador Carranza, thank you for always generously welcoming me to your lab and to accept me as your student. That has made IBE a great place to always come back to, the experience of working there was very fruitful and fun. Your enthusiasm is inspiring, as well as your particularly energetic way of being. Thank you for all the encouragement throughout these years.

To Daniele Salvi, thank you. For your always present support and encouragement. For your brilliant ideas that have made the work in this thesis so much better. For sharing your experience and knowledge, not only about science but about life in general. I am thankful for your optimistic way of seeing life, for being mentored by you and for all the guidance, patience, and motivation, and for your friendship along all these years.

À Comissão Coordenadora e Científica do Programa Doutoral em Biodiversidade, Genética e Evolução por me terem aceitado no Programa Doutoral, na pessoa do Professor Doutor Nuno Ferrand de Almeida, do Professor Doutor Paulo Célio Alves e do Professor Doutor Paulo Alexandrino (em memória).

Agradeço a todo o pessoal do CIBIO por toda a ajuda burocrática e prática durante estes anos, nomeadamente à Sara Lemos Ferreira, Sandra Rodrigues e Maria Sant'Ana e também à Teresa e ao Sr. Bernardino. Agradeço aos membros da secção

de pós-graduação da Faculdade de Ciências pela ajuda e eficácia na resolução de questões burocráticas.

Agradeço à Fundação para a Ciência e Tecnologia pela concessão da bolsa de doutoramento que tornou todo este trabalho possível (SFRH/BD/81528/2011), sob o Programa Operacional Potencial Humano – Quadro de Referência Estratégico Nacional (POPH-QREN) do Fundo Social Europeu e do Ministério da Educação e Ciência.

Quero também agradecer ao pessoal do CIBIO, amigos e colegas que ao longo destes anos estiveram sempre presentes, desde o laboratório ao campo, dos barracões ao aquário. Obrigada à Iolanda Rocha e Beatriz Tomé pela ajuda em inúmeras ocasiões e também pelas conversas sobre tudo e sobre nada. Ao João Maia e ao Luís Rodrigues, camaradas de sushi, que me acompanharam no CIBIO e no IBE e em Barcelona em geral. À Margarida Lopes, Soraia Barbosa, Duarte Gonçalves, Liliana Farelo, Vânia Costa, Ricardo Castilho e Teresa Silva, companheiros no aquário. À Joana Santos, obrigada por encontrar aquela lagartixa no campo, Isabel Damas, Daniela Rosado, Fátima Jorge, Victoria Litsi, Amanda de Sousa e Diogo Silveira. Ao Mário Cunha e à Diana Carneiro, Verónica Gomes e Neftalí Sillero. Obrigada pela ajuda em diversas ocasiões, vinda de anos de experiência, à Anna Perera, Antigoni Kaliontzopoulou, Miguel Carretero, Michael Jowers, Zbyszek Boratynski e Angelica Crottini.

I would also like to thank to all my colleagues in the Institut de Biologia Evolutiva (CSIC-UPF) for brighten my time in Barcelona, in the lab and in the city. To Marc Simó Riudalbas and Joan Garcia-Porta for the scientific discussions and for helping me practice my catalan. To Karin Tamar, Raquel Vasconcelos, David Garcia, Santiago Montero, Helena Vizán, Marina Querejeta and Oliver Hawlitschek, getting to know Barcelona with you all was great. Thanks to Rita Arias and Emiliano Gonzalez for their support in administrative questions during my stays in IBE. To Margarita Metallinou for helping me when I first arrived at IBE, your liveliness and friendship will always be remembered.

To my Barcelonese friends outside of IBE, I am thankful for the discussions, for the scientific thoughts and about life in general. I must also thank you for the ever-present availability of a couch or bed in Barcelona, even when I arrived without warning. Thank you Clàudia Huertas, Eugènia Almacellas and the Andreus – Mata, Cera, Domènech and Ubach.

Agraeixo a la meva segona família de Menorca, per tot l'afecte i suport. Moltes gràcies Montse i Carles, el vostre suport ha estat molt important per mi. Gràcies als

amics que he fet a l'illa de la calma: Xavi, Xisco, Anna, Jara, Omar, Dani, Shikha, gràcies per tots els moments passats a la vora del mar i a les excursions

Um muito obrigada à minha família, sem a qual esta tese, como tantas outras coisas na vida, não teriam sido possíveis. Obrigada pelo suporte e carinho ao longo destes anos, quando estava fora e, principalmente, quando volto a casa.

Finalmente, obrigada Àlex! Obrigada por me acompanhares durante esta jornada, não tenho palavras para agradecer todo o amor, carinho, paciência e encorajamento que me deste durante estes anos. Gràcies per portar-me a Menorca, per la teva calma quan jo no la tinc, per les abraçades mentals quan no hi sóc, per tot. Moltes gràcies bitxo.

SUMMARY

The order Squamata is a species-rich group of reptiles including the groups Sauria, Serpentes and Amphisbaenia. Several members of these groups have been extensively used as model organisms for a great variety of studies of different fields including ecology, behaviour and medicine. Such studies require a well-established phylogenetic framework to trace the evolution and diversity of the studied traits in the squamate tree of life. However, the accuracy of phylogenetic inference in Squamata is unbalanced: while the position of the main squamate groups and high level relationships have been successfully and consistently estimated across phylogenetic studies, the phylogenetic relationships of some groups have provided inconsistent results over the years. The accurate estimation of relationships between these groups might have been hindered by methodological artefacts of the phylogenetic inference due to limited data and analytical tools.

For the last couple of decades, and even presently, phylogenetic inference has relied heavily on mitochondrial DNA as the main molecular marker as a result of its advantages, such as the lack of recombination, easy amplification and high evolutionary rate. While its use is based on sound reasons, some drawbacks and limitations have been highlighted, such as the overall small size, and the fact that all genes are linked thus it represents a single locus. To overcome this, the addition of nuclear DNA complements the use of mitochondrial DNA in phylogenetic inference. Within the nuclear DNA, some markers may present so little variation that it is insufficient to recover phylogenetic relationships, yet slow-evolving nuclear genes have been the most widely used markers in phylogenetic studies. On the other hand, it has been shown that the use of highly informative fast evolving nuclear genes, in combination with mitochondrial DNA, can provide resolution at different parts and depths of the evolutionary history of the species.

Phylogenetic analyses of multiple loci have historically been based on the concatenation approach, .i.e. the combination of the sequences into a single alignment that is then analysed to represent the species tree. This method, however, relies on the assumption that the most commonly occurring gene tree is equivalent to the true species tree of any taxa, which may not be true in all cases. Another limitation of the concatenation approach is the elevated percentage of missing data that it sometimes entails, especially in large taxon sets, meaning that the sequence representation is

uneven across taxa. To overcome the limitations of the concatenation approach and to account for the stochastic sorting of gene genealogies embedded within a species tree, new methods for estimating relationships between species have been developed, including the species tree approach based on the multispecies coalescent model.

The main aim of this thesis was to understand how the addition of the newly developed fast evolving nuclear markers and the coalescent species tree approach could improve the phylogenetic inference of selected groups of squamates from the families Lacertidae and Colubridae, which have provided contrasting phylogenetic hypotheses from different phylogenetic approaches.

The Lacertidae family is taxonomically divided in two families, Gallotiinae and Lacertinae. The Lacertinae comprises two tribes, the Lacertini, mainly distributed in the Palaearctic, and the Eremiadini, from Afro-tropical ecozones. The Lacertini tribe has been the subject of many phylogenetic studies which have proposed two very different phylogenetic hypotheses: a basal polytomy, based on concatenated data of mostly mitochondrial genes, and a fully resolved phylogeny based on a supermatrix approach with hundreds of taxa, both applying the same subset of genes for the Lacertini taxa. In this thesis, the application of the species tree multilocus approach including fast evolving nuclear genes supported the basal polytomy suggestive of a fast radiation but was also able to recover new clade relationships with the monotypic genera *Archaeolacerta*, *Zootoca*, *Teira* and *Scelarcis*, and to provide support for the monophyly of the genus *Algyroides* and support for relationships between other genera. This work demonstrates how the supermatrix approach may provide high support for incorrect nodes.

Within the Colubridae family, the Colubrinae sub-family includes many groups of snakes, including the ratsnakes. These were previously included in a single genus, *Elaphe*, but phylogenetic studies have led to this genus being divided into several genera from the Old and New World. Within the Old World ratsnakes, the genera *Zamenis* and *Rhinechis* have been extensively used in phylogenetic studies, which have proposed different phylogenetic relationships between them and within the genus *Zamenis*. In this study we obtained a fully-resolved and well-supported phylogeny which was consistent across molecular markers, taxon-sets and phylogenetic approaches. A strongly supported *Zamenis* and *Rhinechis* clade was recovered, while the monophyletic *Zamenis* clade received low statistical support. Topological tests rejected previous phylogenetic hypotheses based on the supermatrix approach, demonstrating that this method recovers wrong relationships. The biogeographic and molecular dating analyses suggested an origin of *Zamenis* and *Rhinechis* in the Aegean region during the Late

Miocene and supported a scenario of east-to-west diversification. The little morphological and phylogenetic evidence for the distinctiveness between *Rhinechis* and *Zamenis* supports a classification lumping which better reflects their evolutionary history and, based on the priority rule, *R. scalaris* is moved into the genus *Zamenis* and designated as *Zamenis scalaris* **comb. nov**.

Within the Gallotiinae subfamily, the combined use of fast nuclear markers and the species tree was fruitful for the inference of a robust phylogeny of the genus *Psammodromus*, distributed in the Iberian Peninsula and North Africa. In this study we analysed for the first time the six *Psammodromus* species to infer the association between the main cladogenetic events within this genus and the complex biogeographic dynamics across the Strait of Gibraltar. The inferred phylogeographical history suggests that *Psammodromus* probably originated in Iberia. The African species and the African lineage of *P. algirus* were a result of two over-sea dispersal events towards Africa that occurred 10 Ma and 1.5 Ma, while continental vicariance events might have shaped the diversification of the species within Iberia and Africa. These results, combined with previous literature, provide compelling evidence that major biotic exchanges occurred across the Strait of Gibraltar well before or long after the land connection during the Messinian Salinity Crisis (5.9 – 5.33 Ma). These findings suggest caution in the application of the relatively short event of the opening of the Strait of Gibraltar at the end of the Messinian Salinity Crisis as a cause for divergence in molecular clock calibrations, which is a common approach in literature.

At the species level, the use of fast evolving nuclear markers proved to be fruitful to understand the geographical pattern of phyletic diversification within the *Omanosaura* lizards belonging to the Eremiadini tribe, endemic to the Hajar Mountains in the Arabian Peninsula. Multilocus phylogenetic analyses recovered two highly divergent lineages within *O. cyanura* which are geographically associated to the northernmost and to the south and eastern regions of the Hajar Mountains. These lineages are reciprocally monophyletic at both mitochondrial and nuclear loci suggesting a long history of independent evolution and the need of a comprehensive taxonomic assessment. *Omanosaura cyanura* represents an additional case of cryptic diversity in the north Hajar Mountains, an area that has been demonstrated to hide high levels of genetic diversity in other reptile groups.

Overall, the work developed in this dissertation has demonstrated that the inclusion of fast evolving nuclear genes and the coalescent species tree approach was relevant for resolving challenging phylogenetic questions on selected squamate groups.

This approach allowed the recovery of new clades and provided support for old and recent relationships. These methods have further allowed the comparison between contrasting phylogenetic hypotheses on the Lacertini and Colubrinae, demonstrating that the inference based on the supermatrix approach used by many recent studies may provide high support for incorrect nodes. Future research directions include (i) the application of the phylogenomic approach based on Next Generation Sequence data to further resolve basal polytomy within the Lacertini and Colubrinae radiations; (ii) a taxonomical assessment of *Omanosaura*; (iii) a re-evaluation of molecular clock calibrations based on the end of the Messinian Salinity Crisis as general cause for divergence, in order to account for biotic exchanges by over-sea dispersal across the Mediterranean well before or long after the land connection during this period, as suggested by emerging literature and the study case on *Psammodromus*.

RESUMO

A ordem Squamata é um grupo rico em espécies de répteis que inclui os grupos Sauria, Serpentes e Amphisbaenia. Vários membros destes grupos têm sido utilizados como organismos modelo numa grande variedade de estudos em diferentes áreas, incluindo ecologia, comportamento e medicina. A realização deste tipo de estudos requer um contexto filogenético base para que seja possível seguir a evolução e diversidade das características sob estudo na árvore da vida dos Squamata. Contudo, a precisão da inferência filogenética neste grupo não está em equilíbrio: enquanto a posição dos principais grupos de Squamata e as relações de nível basal têm sido consistentemente estimadas com sucesso nos estudos filogenéticos, a inferência filogenética de alguns grupos tem sido inconstante durante os últimos anos. Uma estimativa exacta das relações filogenéticas entre estes grupos poderá ter sido influenciada por artefactos inerentes aos métodos de inferência filogenética.

Durante as últimas duas décadas, e ainda hoje, a filogenia tem dependido do DNA mitocondrial como fonte principal de marcadores moleculares. Isto é justificado pelas as vantagens que este genoma aporta, como por exemplo a fácil amplificação e a alta taxa evolutiva. Embora o uso destes marcadores moleculares esteja baseado em razões sólidas, algumas limitações têm-lhe sido apontadas, tais como o pequeno tamanho deste genoma e o facto de todos os genes estarem geneticamente ligados e portanto representarem um único locus. Para superar estas limitações foi adicionado DNA nuclear às análises filogenéticas para complementar os genes mitocondriais. Mas dentro do DNA nuclear existem genes que têm tão pouca variação que acabam por ser incapazes de inferir relações filogenéticas com sucesso. Ainda assim, estes genes têm sido maioritariamente utilizados em filogenia. Por outro lado, o uso de genes nucleares mais informativos, com uma taxa evolutiva mais elevada, em conjunto com os genes mitocondriais, oferecem resolução em diferentes partes e profundidades da história evolutiva das espécies.

A utilização de múltiplos loci nas análises filogenéticas levou ao desenvolvimento de métodos capazes de analisar estes conjuntos de dados. Entre estes, o método da concatenação, isto é, a junção das sequências num único alinhamento que é analisado para representar a *species tree*, tornou-se o mais utilizado para a análise de vários loci. No entanto, este método baseia-se na suposição de que a *gene tree* que ocorre mais comummente é equivalente à verdadeira *species tree* de qualquer taxa, o que pode não

ser verdade em muitos casos. Outra limitação deste método é a elevada percentagem de dados em falta que por vezes pode ocorrer, com uma representação de sequências de DNA desigual entre taxa. Para ultrapassar estas limitações da concatenação e para ter em conta a separação estocástica de cada gene, novos métodos para estimar a *species tree* foram desenvolvidos, entre os quais o método baseado no modelo coalescente de múltiplas espécies.

O objectivo principal desta tese foi compreender como a adição de novos genes nucleares com elevada taxa evolutiva e o método de *coalescent species tree* influenciaria ou melhoraria a inferência filogenética em determinados grupos de Squamata das famílias Lacertidae e Colubridae, os quais foram alvo de vários estudos moleculares que resultaram em diferentes hipóteses filogenéticas.

A família Lacertidae está taxonomicamente dividida em duas sub-famílias, Gallotiinae e Lacertinae. Esta última inclui duas tribos: Lacertini, distribuída maioritariamente no Paleárctico, e Eremiadini, distribuída na ecozona Afro-tropical. A tribo Lacertini tem sido alvo de vários estudos filogenéticos que resultaram em duas hipóteses bastante diferentes: uma politomia basal, baseada na concatenação de genes maioritariamente mitocondriais, e uma árvore filogenética completamente resolvida, obtida com o método de *supermatrix* utilizando centenas de taxa, ambas empregando os mesmos genes para as espécies da tribo Lacertini. Nesta tese, a utilização do método de *species tree* incluindo vários genes nucleares informativos suportou a politomia basal, indicando uma radiação rápida neste grupo mas também recuperou um novo grupo formado pelos géneros *Archaeolacerta*, *Zootoca*, *Teira* e *Scelarcis*, suportando também a monofilia do género *Algyroides* e relações entre outros géneros. Este trabalho demonstrou como a concatenação pode prover suporte elevado para relações filogeneticamente incorrectas.

Dentro da família Colubridae, a sub-família Colubrinae inclui vários grupos de cobras, tais como as *ratsnakes*. Originalmente, estas pertenciam a um só género, *Elaphe*, mas vários estudos filogenéticos dividiram este género em vários, incluindo géneros do Velho e do Novo Mundo. Dentro dos géneros do Velho Mundo, os géneros *Zamenis* e *Rhinechis* têm sido muito utilizados em vários estudos filogenéticos que propuseram diferentes hipóteses filogenéticas, com resultados incertos sobre a relação entre estes géneros e entre as espécies de *Zamenis*. Neste estudo obtemos uma árvore filogenética resolvida e bem suportada, consistente com a aplicação de diferentes marcadores moleculares, grupos de taxa e métodos filogenéticos. Os resultados revelaram que o clado *Zamenis* e *Rhinechis* é altamente suportado, enquanto o clado

monofilético de *Zamenis* recebeu baixo suporte estatístico. Testes de topologia rejeitaram as hipóteses filogenéticas obtidas para estes géneros com o método de *supermatrix*, demostrando assim que este método recupera relações filogenéticas incorrectas. As análises de biogeografia e datação molecular sugerem a origem de *Zamenis* e *Rhinechis* na região do Egeu durante o Mioceno e suportam o cenário de diversificação de este para oeste. A escassa evidência morfológica e filogenética da distinção entre *Rhinechis* e *Zamenis* suporta uma junção na classificação taxonómica que melhor reflecte a história evolutiva destes géneros e, baseados na regra de prioridade, *R. scalaris* é movida para o género *Zamenis* e designada *Zamenis scalaris* **comb. nov.**

Dentro da sub-família Gallotiinae, o uso combinado de genes nucleares informativos e a *species tree* foi eficaz para uma robusta estimativa da filogenia do género *Psammodromus*, que se encontra distribuído na Península Ibérica e Norte de África. Neste estudo, foram analisadas pela primeira vez as seis espécies de *Psammodromus* para compreender a associação entre os eventos cladogenéticos deste género e o complexo cenário biogeográfico em volta do Estreito de Gibraltar. A história filogeográfica estimada sugere que a origem de *Psammodromus* foi provavelmente na Península Ibérica. As espécies africanas e a linhagem africana da espécie *P. algirus* resultaram de dois eventos de dispersão transmarina para África que ocorreram há 10 e 1.5 milhões de anos. Eventos de vicariância continental possivelmente originaram a diversificação das espécies na Península Ibérica e no Norte de África. Estes resultados, em conjunto com a literatura, fornecem evidências convincentes de que grandes permutações bióticas ocorreram através do Estreito de Gibraltar muito antes e depois da conexão terrestre durante a Crise Salina do Messiniano (5.9-5.33 milhões de anos). Isto sugere caução aquando da utilização de um evento tão rápido como a abertura do Estreito de Gibraltar no final da Crise Salina do Messiniano como causa para divergência em calibrações do relógio molecular, o que constitui um método comum na literatura.

Ao nível de espécies, o uso de genes nucleares informativos foi útil para compreender o padrão geográfico da diversificação no género de lagartos *Omanosaura*, pertencente à tribo Eremiadini e endémico das montanhas Hajar, na Península Arábica. As análises filogenéticas com *multilocus* encontraram duas linhagens altamente divergentes dentro de *O. cyanura* que estão geograficamente associadas com as regiões do norte e sudeste das montanhas Hajar. Estas linhagens são reciprocamente monofiléticas nos genes mitocondriais e nucleares, o que sugere uma longa história de evolução independente e a necessidade de uma avaliação taxonómica. *Omanosaura cyanura* representa mais um caso de diversidade críptica no norte das montanhas Hajar,

uma área em que já foram demonstrados altos níveis de diversidade genética em outros grupos de répteis.

No geral, o trabalho desenvolvido nesta tese demonstrou que a adição de genes nucleares com elevada taxa de evolução, assim como a *species tree* baseada no modelo coalescente de múltiplas espécies foi relevante e benéfica para a resolução de questões filogenéticas desafiadoras nestes grupos de Squamata. Estes métodos permitiram a descoberta de novos clados e forneceram suporte a relações antigas e recentes. Permitiram também a comparação com hipóteses filogenéticas anteriores, demostrando que a inferência obtida com a concatenação pode fornecer suporte a relações incorrectas. Direcções de pesquisa futura incluem (i) a aplicação do método filogenómico baseado em dados da Sequenciação de Nova Geração para a resolução adicional da politomia basal dentro da radiação de Lacertini e Colubrinae; (ii) a avaliação taxonómica de *Omanosaura*; (iii) uma reavaliação de calibrações moleculares baseadas no final da Crise Salina do Messiniano como causa para divergência, para ter em conta permutações bióticas através de dispersão transmarina na região Mediterrânea muito antes ou depois desta época, tal como sugerido na literatura e com o estudo de *Psammodromus*.

KEYWORDS

Species-tree, gene tree, multilocus phylogeny, fast radiation, polytomy, concatenation, nuclear DNA, mitochondrial DNA, molecular clock calibration, Squamata, Lacertidae, Lacertini, *Psammodromus*, *Omanosaura*, Colubridae, Colubrinae, *Zamenis*, *Rhinechis*, Palaearctic, phylogeography, biogeography, dispersal, vicariance, cryptic diversity, diversification, evolutionary history

PALAVRAS-CHAVE

Species tree, gene tree, filogenia multilocus, radiação rápida, politomia, concatenação, DNA nuclear, DNA mitocondrial, calibração do relógio molecular, Squamata, Lacertidae, Lacertini, *Psammodromus*, *Omanosaura*, Colubridae, Colubrinae, *Zamenis*, *Rhinechis*, Paleárctico, filogeografia, biogeografia, dispersão, vicariância, diversidade críptica, diversificação, história evolutiva.

INDEX OF CONTENTS

Acknowledgments / Agradecimentos v

Summary ix

Resumo xiii

Keywords / Palavras-chave xvii

Abbreviations xxxiii

Chapter 1. General Introduction 1

1. Squamata, a diverse reptile group 3
    - Origin and Biogeography 8
    - Squamata as model organisms 10
    - Phylogenetic Inference in Squamata 10
2. Molecular tools of Phylogenetic Inference 17
    - Soft and Hard Polytomies 17
    - Mitochondrial and nuclear DNA 18
    - Gene trees, concatenation and species tree 22
    - The effect of missing data 26
3. Case studies in Squamata 27
    - The Lacertidae family 27
        - The tribe *Lacertini* 30
        - The genus *Psammodromus* 32
        - The genus *Omanosaura* 33
    - The Colubridae family 34

The genus *Zamenis* 36

4. Objectives and Thesis Framework 38

References 40

Chapter 2. Multilocus phylogeny of Lacertini 59

Article I. Evaluating the phylogenetic signal from mitogenomes, slow evolving nuclear genes, and the concatenation approach. New insights into the Lacertini radiation using fast evolving nuclear genes and species trees. 61

Chapter 3. Multilocus phylogeny of *Zamenis* ratsnakes 99

Article II. Evolution, biogeography and systematics of the Western Palaearctic ratsnakes *Zamenis*, with the designation of *Zamenis scalaris* comb. nov. 101

Chapter 4. Multilocus phylogeny of *Psammodromus* lizards 147

Article III. Biogeographic crossroad across the Pillars of Hercules: evolutionary history of *Psammodromus* lizards in space and time. 149

Chapter 5. Multilocus phylogeny of *Omanosaura* lizards 183

Article IV. Hidden in the Arabian mountains: multilocus phylogeny reveals cryptic diversity in the endemic *Omanosaura* lizards. 170

Chapter 6. General Discussion 217

The importance of the choice of molecular markers 220

Differences from phylogenetic tools: concatenation and species tree 222

Missing data 224

The role of next generation sequencing 224

Taxonomical implications 226

Concluding remarks 228

Future perspectives 230

References 232

Appendices 239

Article I 240

Article II 244

Article III 254

Article IV 259

LIST OF FIGURES

CHAPTER 1. General Introduction

Figure 1.1. Number of Squamata species over the last 250 years. 4

Figure 1.2. Distribution of all Squamata species and of Sauria, Serpentes and Amphisbaenia. 6

Figure 1.3. Paleogeographic representation of landmasses position from the Upper Triassic until the Oligocene. 9

Figure 1.4. Evolutionary history of Squamata as proposed by Camp in 1923. 12

Figure 1.5. Cladogram of Squamata based on morphological characters proposed by Estes in 1988. 13

Figure 1.6. Most recent phylogenetic tree of the relationships of the main groups of Squamata. 16

Figure 1.7. Graphical representation of a polytomy. 17

Figure 1.8. Methodological shifts and evolution of the fields of phylogeography and phylogenetics. 19

Figure 1.9. Distribution of the use of molecular markers in Sauria between 1980 and 2009. 20

Figure 1.10. Variability of nuclear protein-coding loci in squamate reptiles. 22

Figure 1.11. Incongruence between species tree and gene trees. 23

Figure 1.12. Publications and citations of phylogenetic studies in Lacertidae. 28

Figure 1.13. Percentage of phylogeographic studies by Sauria families until 2010. 28

Figure 1.14. Area cladogram of the Lacertidae distribution. 30

Figure 1.15. *Omanosaura jayakari* and *Omanosaura cyanura*. 34

Figure 1.16. Phylogenetic relationships of the group Serpentes. 35

Figure 1.17. Publications and citations of phylogenetic studies in Colubridae. 35

Figure 1.18. Phylogenetic inference of *Zamenis* species. 37

CHAPTER 2. Multilocus phylogeny of Lacertini

Figure 2.1. Phylogenetic relationships of Lacertini based on Bayesian analyses of concatenated mitochondrial DNA sequences. 77

Figure 2.2. Phylogenetic relationships of Lacertini based on Bayesian analyses of concatenated nuclear DNA sequences. 79

Figure 2.3. Phylogenetic relationships of Lacertini based on concatenated mitochondrial and nuclear DNA sequences. 80

Figure 2.4. Species tree of Lacertini inferred from mitochondrial and nuclear DNA sequences using the multispecies coalescent model. 82

CHAPTER 3. Multilocus phylogeny of *Zamenis* ratsnakes

Figure 3.1. Distribution of the species of the ratsnakes genera *Zamenis*, *Rhinechis* and *Coronella* in the Palaearctic. 104

Figure 3.2. Summary of previous phylogenetic hypotheses on *Zamenis* and *Rhinechis* based on molecular data and different taxon-sets. 105

Figure 3.3. Neighbor-Net haplotype networks for *Zamenis*, *Rhinechis*, *Coronella*, *Hierophis* and *Hemorrhois* species inferred from single locus DNA and individual nuclear loci. 118

Figure 3.4. Bayesian tree depicting the relationships between the genera *Zamenis*, *Rhinechis* and *Coronella* based on concatenated mitochondrial and nuclear DNA. 119

Figure 3.5. Bayesian tree depicting the phylogenetic relationships between *Zamenis*, *Rhinechis*, *Coronella* and representatives of five other ratsnakes' genera based on mitochondrial and nuclear DNA. 120

Figure 3.6. Species tree of *Zamenis*, *Rhinechis* and *Coronella* ratsnakes inferred from mitochondrial and nuclear DNA using the multispecies coalescent model. 121

Figure 3.7. Biogeographical reconstruction of the evolutionary history of *Zamenis* and *Rhinechis* according to *rase* and RASP analyses based on the species tree applying the time calibration I. 125

CHAPTER 4. Multilocus phylogeny of *Psammodromus* lizards

Figure 4.1. Paleogeological history of the Strait of Gibraltar from the middle Miocene until the Messinian Salinity Crisis. 153

Figure 4.2. Sampling locations of the *Psammodromus* specimens used in this study. Distribution of *Psammodromus* species in the Iberian Peninsula and North Africa. 155

Figure 4.3. Bayesian tree depicting the relationships between *Psammodromus* lizards based on mitochondrial and nuclear DNA, divergence time estimation and results from the ancestral range reconstruction with two areas located in North Africa and the Iberian Peninsula. 163

Figure 4.4. Species tree of *Psammodromus* lizards inferred from mitochondrial and nuclear DNA and time of divergence between species. 165

Figure 4.5. Dispersal and vicariance events in *Psammodromus* in Iberia and North Africa before and after the Messinian Salinity Crisis. 168

CHAPTER 5. Multilocus phylogeny of *Omanosaura* lizards

Figure 5.1. Sampling locations in the Hajar Mountains of the *Omanosaura* specimens used in this study. 189

Figure 5.2. Bayesian tree depicting phylogenetic relationships of *Omanosaura* and divergence time estimates between species and lineages. 200

Figure 5.3. Species tree of *Omanosaura* and divergence time estimates between species. 201

Figure 5.4. Median-joining haplotype network of *Omanosaura* species inferred from concatenated mitochondrial DNA. 201

Figure 5.5. Median-joining haplotype network of *Omanosaura* species inferred from nuclear genes *cmos* and *mc1r*. 204

LIST OF TABLES

CHAPTER 1. General Introduction

Table 1.1. Number of total Squamata, Amphisbaenia, Sauria and Serpentes species from 2008 to 2016. 3

Table 1.2. Squamata species, molecular markers and phylogenetic approaches used in previous phylogenetic studies in Squamata. 14

CHAPTER 2. Multilocus phylogeny of Lacertini

Table 2.1. Sample codes, species, sampling locality and GenBank accession numbers for the Lacertini samples and sequences. 70

Table 2.2. Primers and PCR protocols used for the amplification of the molecular markers used in this study. 74

Table 2.3. Number of sequences, fragment's length, models of sequence evolution and number of variable positions for genes amplified for Lacertini. 75

Table 2.4. Age, in million years (Mya) and node 95% highest posterior density (HPD) intervals for the major supported nodes in the species tree. 81

Table 2.5. Results of topological tests using three sets of topological constrains based on relationships recovered in previous studies with node support ≥85 or ≥90 or ≥95. 83

CHAPTER 3. Multilocus phylogeny of *Zamenis* ratsnakes

Table 3.1. Code, location coordinates and GenBank accession numbers of the colubrid species and sequences used in the study. 108

Table 3.2. Name, sequence and reference of the primers used in the study. 110

Table 3.3. Length of gene alignments, models of nucleotide substitution and number of variable positions for the datasets used in the study. 116

Table 3.4. Results of topological tests. 122

Table 3.5. Divergence time estimates and 95% HPD intervals based on rates for calibrations I and II within the *Zamenis* taxon-set. 123

CHAPTER 4. Multilocus phylogeny of *Psammodromus* lizards

Table 4.1. Species, code, map location, locality, coordinates and GenBank accession number of the *Psammodromus* specimens used in the study. 157

Table 4.2. Name, sequences and reference of the primers, and the amplification conditions for the genes used in the study. 159

Table 4.3. Length of gene fragments, models of nucleotide substitution and number of variable positions within *Psammodromus* and in *Psammodromus* plus outgroup. 162

CHAPTER 5. Multilocus phylogeny of *Omanosaura* lizards

Table 5.1. Code, geographical coordinates and GenBank accession numbers of the *Omanosaura* and outgroup species. 192

Table 5.2. Length, models of nucleotide substitution, number of intra- and interspecific variable positions and parsimony informative sites for each gene used for the study. 196

Table 5.3. Genetic pairwise *p*-distance values, range and standard error between and within *Omanosaura* species for the mitochondrial genes. 198

LIST OF APPENDICES

Supplementary Figures and Tables

Figure S2.1. Phylogenetic relationships of Lacertini based on Maximum Likelihood analyses of the concatenated tree based on mitochondrial DNA. 240

Figure S2.2. Phylogenetic relationships of Lacertini based on Maximum Likelihood analyses of the concatenated nuclear DNA sequences. 241

Figure S2.3. Phylogenetic relationships of Lacertini based on Maximum Likelihood analyses of concatenated mitochondrial and nuclear DNA sequences. 242

Figure S2.4. Distribution and phylogenetic relationships of the Lacertini genera. 243

Table S3.1. Divergence time estimates and 95% HPD intervals based on rates from calibrations I and II with the *Zamenis* + ratsnakes taxon-set. 244

Table S3.2. Geographic location of the ancestors and lineages of *Rhinechis* and *Zamenis* species for each time slice of the *rase* analyses based on the species tree applying the time calibration I. 245

Figure S3.1. Bayesian tree topology of the BEAST analyses with one individual per species and the Yule prior of speciation. 247

Figure S3.2. Median-joining haplotype networks for *Zamenis*, *Rhinechis*, *Coronella* and outgroups inferred with mitochondrial and phased nuclear DNA sequences. 248

Figure S3.3. Bayesian tree depicting the phylogenetic relationships between *Zamenis*, *Rhinechis* and *Coronella* based on concatenated nuclear DNA sequences. 249

Figure S3.4. Bayesian tree depicting the phylogenetic relationships between *Zamenis*, *Rhinechis* and *Coronella* based on concatenated nuclear DNA sequences using the reduced dataset. 250

Figure S3.5. Bayesian tree depicting the phylogenetic relationships between *Zamenis*, *Rhinechis* and *Coronella* based on concatenated mitochondrial and nuclear DNA sequences using the reduced dataset. 251

Figure S3.6. Bayesian tree depicting the phylogenetic relationships between *Zamenis*, *Rhinechis*, *Coronella* and representatives of five other ratsnakes genera based on concatenated nuclear DNA sequences. 252

Figure S3.7. Species tree of *Zamenis*, *Rhinechis*, *Coronella* and representatives of five other ratsnakes genera based on DNA sequences from mitochondrial and phased nuclear loci. 253

Figure S4.1. Bayesian tree of *Psammodromus* with one individual per species. 254

Figure S4.2. Bayesian tree of *Psammodromus* with the cross-validation of the substitution rates. 255

Figure S4.3. Bayesian tree depicting the phylogenetic relationships between the *Psammodromus* lizards based on mitochondrial and nuclear DNA sequences, and the results from the ancestral area reconstruction with three areas in Iberian Peninsula and North Africa. 256

Figure S4.4. Distribution of the described fossils of *Psammodromus* and *Gallotia* clade, with information detailed in Supplementary Table S4.1. 257

Table S4.1. Table of the described fossils of *Psammodromus* and *Gallotia* clade presented in Supplementary Fig. S4.3. 258

Table S5.1. Primers and PCR conditions used for the amplification of the genes in the *Omanosaura* study 259

Figure S5.1. Phylogenetic relationships of *Omanosaura* inferred from mtDNA and nucDNA by Maximum Likelihood. 261

Figure S5.2. Bayesian tree of *Omanosaura* inferred from concatenated mtDNA. 262

Figure S5.3. Bayesian tree of *Omanosaura* with the cross-validation of the substitution rates. 263

ABBREVIATIONS

acm4 – Acetylcholonergic receptor M4

AICc – corrected Akaike information criterion

AU – Approximately Unbiased

BA – Bayesian analyses

BBM – Bayesian Binary MCMC

BEAST – Bayesian Evolutionary Analysis Sampling Trees

BEAUTi – Bayesian Evolutionary Analysis Utility

BI – Bayesian Inference

BIC – Bayesian information criterion

bp – base pairs

BPP – Bayesian posterior probabilities

BS – Bootstrap support

cmos – Oocyte maturation factor

CIPRES – Cyberinfrastructure for Phylogenetic Research

cyt*b* – Cytochrome b

DEC – Dispersal-Extinction Cladogenesis

DNA – deoxyribonucleic acid

dnah3 – Dynein axonemal heavy chain 3

DIVA – Dispersal-Vicariance Analysis

ERDF – European Regional Development Fund

ESS – efective sample size

FCT – Fundação para a Ciência e Tecnologia

GTR – General Time Reversible

HKY – Hasegawa-Kishino-Yano

HPD – highest probability density

indel – insertion or deletion

IUCN – International Union for Conservation of Nature

IUPAC – International Union of Pure and Applied Chemestry

JC – Jukes-Cantor model

K80 – Kimura model

Ma – Million years ago

MaxChi – Maximum Chi square

mc1r – Melanocortin 1 receptor

MCC – Maximum clade credibility

MCMC – Markov Chain Monte Carlo

MEGA – Molecular Evolutionary Genetics Analysis

ML – Maximum Likelihood

MSC – Messinian Salinity Crisis

mtDNA – mitochondrial DNA

mt-nucDNA – mitochondrial and nuclear DNA

MUSCLE – Multiple Sequence Comparison by Log-Expectation

Mya – Million years ago

Myr – Million years

NA – North Africa

nd4 – NADH Dehydrogenase 4

NGS – next generation sequencing

NPCL – nuclear protein-coding loci

NSRF – National Strategic Reference Framework

nucDNA – nuclear DNA

ON.2 – North Portugal Regional Operational Program

PCR – Polymerase Chain Reaction

pdc – Phosducin

phi – Pairwise Homoplasy Index

pomc - Proopiomelanocortin

prlr – Prolactin receptor

rag1 – Recombination activating gene 1

rase – Range Ancestral State Estimation

RASP – Reconstruct Ancestral State in Phylogenies

RaxML – Randomized Axelerated Maximum Likelihood

RDP – Recombination Detection Program

reln – Reelin

rRNA – ribosomal ribonucleic acid

S-DIVA – Statistical Dispersal-Vicariance Analysis

SH – Shimodaira-Hasegawa

SHLaLRT – Shimodaira-Hasegawa-Like implementation of the approximate likelihood-ratio test

sptbn1 – Spectrin beta, non-erythrocytic 1

SVL – Snout-vent length

TMRCA – time to most recent common ancestor

vim – Vimentin

βfib – β fibrinogen

CHAPTER 1

General Introduction

1. SQUAMATA, A DIVERSE REPTILE GROUP

The order Squamata Oppel, 1811, from the Latin *squama*, scale, includes by far the highest number of reptiles in the world, and is the largest and most diverse terrestrial vertebrate radiation. Over ten thousand species belong to Squamata (Uetz, P., Freed, P., & Jirí Hošek (eds), The Reptile Database, http://www.reptile-database.org/, accessed April 2017). It currently comprises three suborders: Sauria, the diverse paraphyletic group of lizards, including 6263 species in 38 families, Amphisbaenia or worm lizards, with 196 Amphisbaenia species in 6 families, and the snakes' suborder Serpentes, with 3619 snake species in 26 families (Table 1.1). The number of squamate species is far from static, as it has been increasing at a stable rate for the last 60 years, with a higher rate over the last ten years. As an example of this growth, over 150 new species of Sauria and over 50 species of Serpentes were described each year since 2011 (Table 1.1, Fig. 1.1). These values are even higher than the yearly rates of the 18th and 19th centuries (e.g. year 1758: 118 new species; year 1854: 144 new species; Uetz, P., Freed, P., & Jirí Hošek (eds), The Reptile Database, http://www.reptile-database.org/, accessed April 2017). This increase in the number of Squamata species in the last decade can be seen as a direct result of an intensification in phylogenetic research in Squamata, with higher effort dedicated in including not only more individuals in the analyses, but also new and fast evolving molecular markers.

Table 1.1. Number of Amphisbaenia, Sauria, Serpentes and total Squamata. Some authors reject the idea of subspecies and either synonymize them with their parent species or consider them as valid species. Retrieved from the Reptile Database, accessed in March 2017.

	Feb 2008	**Jan 2011**	**Feb 2012**	**Feb 2013**	**Aug 2014**	**Aug 2015**	**Aug 2016**
Amphisbaenia	168	181	181	184	188	193	**196**
Sauria	5079	5461	5634	5796	5987	6145	**6263**
Serpentes	3149	3315	3378	3432	3496	3567	**3619**
Squamata total	8396	8957	9193	9412	9671	9905	**10078**

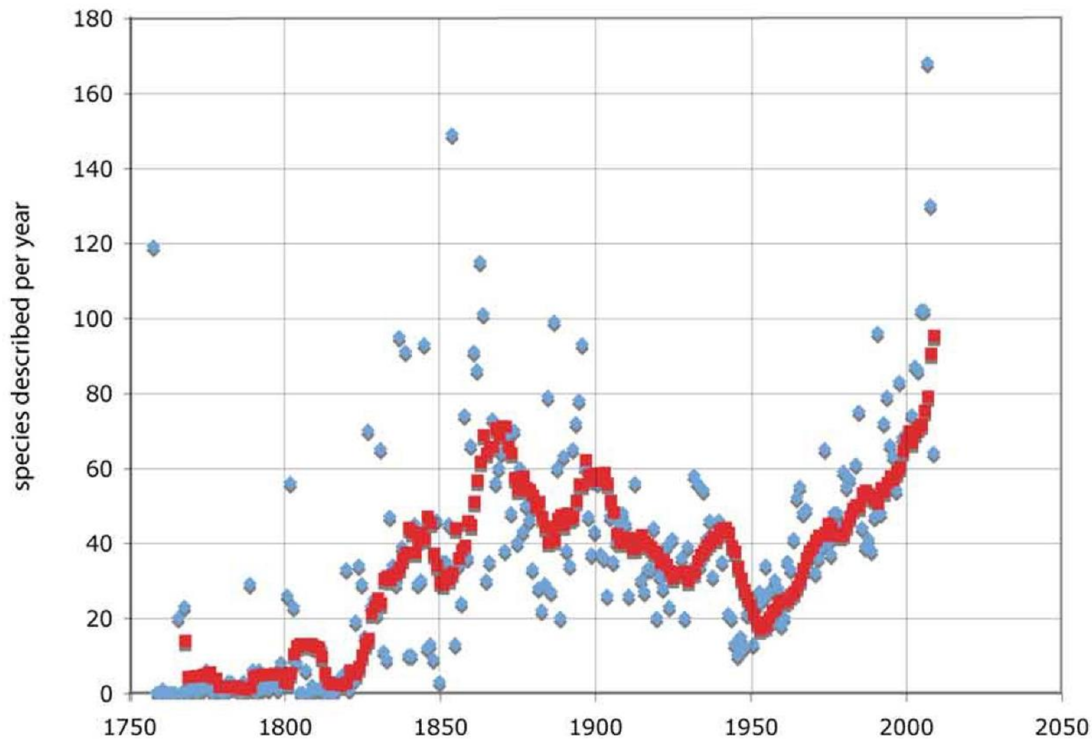
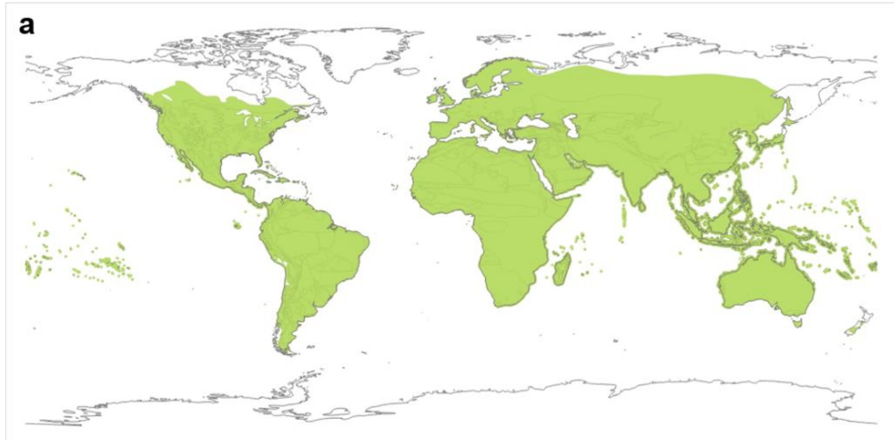


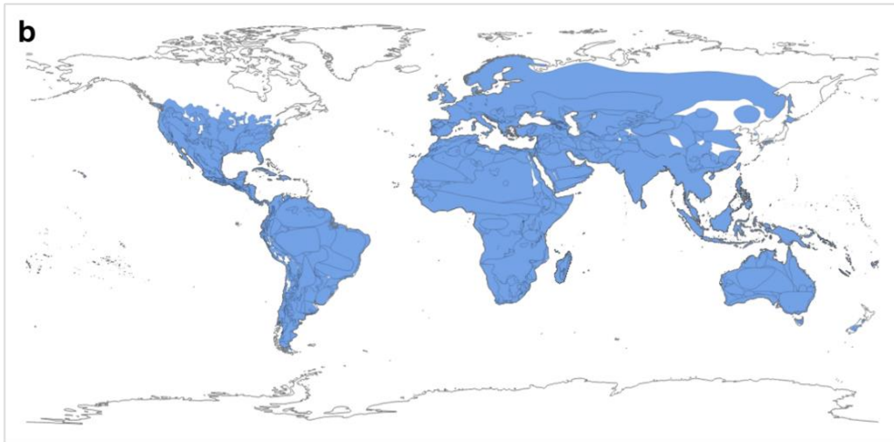
Figure 1.1. Species numbers during 250+ years of reptile taxonomy. Diamonds represent the number of species described per year; squares represent average species numbers over the previous 10 years. From Uetz (2010).

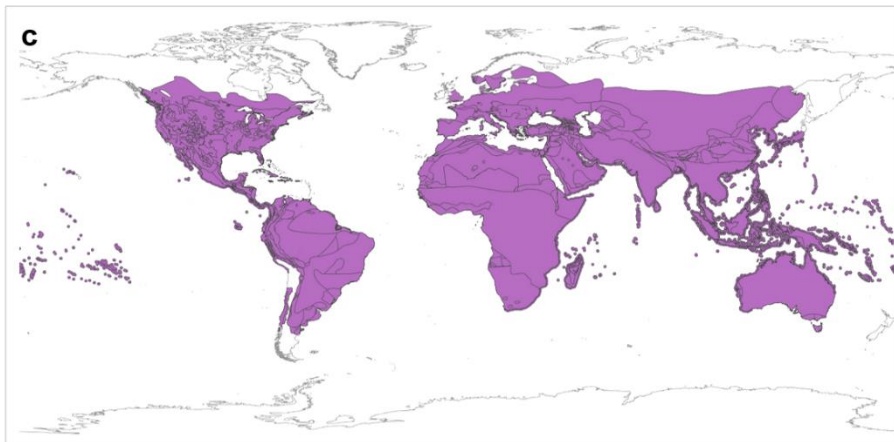
Morphological features that are common to all squamate reptiles and define them as a separate order include the scale-bearing skin, which gives them its name and is the most obvious morphological characteristic in the external part of the body. They also share, among many other morphological features, a specialized quadrate articulation that allows the movement of the upper jaw relative to the neurocranium, which is particularly visible in snakes (reviewed in Evans, 2003). A widely accepted division within Squamates, before more recent phylogenetic hypotheses, was based on morphology and separated the Iguania (Iguanidae, Agamidae, Chamaeleonidae) and the Scleroglossa (all other squamates) (Estes et al., 1988). This division was based on a major event in the evolution of squamates that was the switch from tongue prehension of food used by iguanians, to the teeth and jaw prehension of food by the scleroglossans, freeing in this way the tongue for chemoreception. This change would allow the origin of the exploration of new habitats and foraging modes unavailable to iguanians (Herrel et al., 2001; Vitt et al., 2003; Vidal & Hedges, 2005). The hypothesis of this morphologically based division, however, is not supported by molecular data (see section on Phylogenetic Inference in Squamata).

Despite these shared morphological traits, squamate reptiles display a striking morphological diversity associated with extensive ecological variation. Body size, for instance, ranges from a few millimetres (e.g. members of the Leaf Chamaeleon genus *Brookesia*, with mean snout-vent length, SVL, of 16 mm and the gecko *Sphaerodactylus ariasae* with mean SVL of 16 mm are the smallest known amniotes; Hedges & Thomas, 2001) to more than six meters, with the snake *Python reticulatus* reaching 650 cm and the world's largest lizard, the Komodo Dragon (*Varanus komodoensis*), reaching a total body size of over 300 cm. Variation in body form is also considerable, with different families having particularities in this aspect, but diversity in major body form can be exemplified by the lizard and snake body types. Concerning ecological and behavioural traits, the diversity is immense: there are gliders, specialist climbers, parachuters, swimmers, burrowers and sand swimmers, facultative bipedal runners and ground dwellers, among others. The diet is also diverse, with omnivores, insectivores, herbivores and durophagous reptiles. Some are active predators and may use venom, a unique feature among living tretapods. Reproductive strategies include both viviparity and oviparity and also, in some particular cases, parthenogenesis (squamata diversity reviewed in Evans, 2003).

Due to the extraordinary diversity of adaptations within squamate reptiles, these were able to colonise all the continents except Antarctica, and are also present in isolated oceanic islands (Fig. 1.2a). Of the three suborders, lizards and snakes have the widest distributions, with lizards occupying practically the whole world with the exception of Antarctica and northernmost latitudes (Fig. 1.2b) and snakes presenting a similar distribution, plus the Indian and South Pacific oceans, thanks to the existence of coral and sea snakes (Fig. 1.2c). The Amphisbaenia have a more restricted distribution in Southern Europe, Africa, South America and isolated parts of North America and the Middle East (Fig. 1.2d), what could be explained by the small number of species, when compared with lizards and snakes.


a


b


c

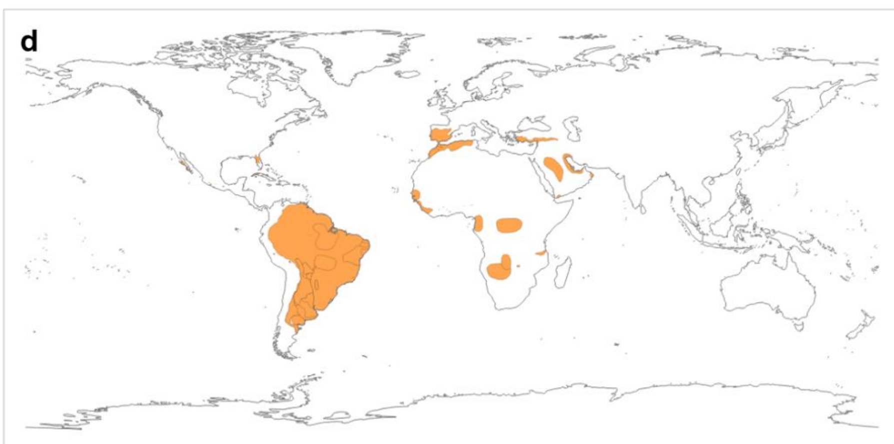

d

Figure 1.2. Distribution range of Squamata species. a) Distribution of all Squamata, b) distribution of Sauria species, c) distribution of Serpentes and d) distribution of Amphisbaenia. Data from the IUCN Red List of Threatened Species, accessed March 2017. The distribution of some species is not available in the IUCN website, and therefore the distribution areas are approximate.

At present, many squamate species face serious conservation pressures. The threats, as in other plant and animal groups, have both global and local origins. At the global scale threats are mostly human mediated and include climate change and habitat destruction (Ceballos et al., 2015). Climate change is particularly prone to affect squamate species due to their ectothermic condition, in addition to habitat modifications as a result of the climate influence in vegetation and water conditions (e.g. Meng et al., 2016; Minoli and Avila, 2017). Likewise, habitat reduction caused by human activities is a major risk factor for many reptiles, due to their limited dispersal ability and niche adaptation, sometimes to very strict levels (e.g. Abdala et al., 2017; Colli et al., 2016). The increase in illegal wildlife trade is also becoming an important issue for reptile conservation, with over 350 threatened species being intentionally targeted by collectors (Auliya et al., 2016). At a local scale, the effects of the arrival of invasive species, either from other squamates or other animal groups, are particularly harmful for endemic reptiles, especially those endemic to islands, which seem to be disproportionally more affected (e.g. Bellard et al., 2016; Li et al., 2014; Silva-Rocha et al., 2015; Thibault et al., 2017). Cultural myths can be menacing to squamata reptiles as well, with the best example given by snakes, villainized by old legends that incite the indiscriminate killing of these animals (e.g. Pinheiro et al., 2016). Undescribed diversity may also pose a threat for conservation (e.g. Carranza et al., 2016; Salvi et al., 2017), when cryptic species within species complexes may disappear even before being discovered. The description of new squamate species in recent years is biased toward small species from small geographic ranges, and thus, more likely to be threatened with population decline and extinction (Meiri, 2016).

Despite the elevated number and the diversity of threats, the extinction risk among squamates is not well understood when compared with mammals, birds and amphibians (Isaac et al., 2007, 2012; Jetz et al., 2014; Meiri & Chapple, 2016). Recent global analyses estimated high percentages of species extinction in the next decades (Sinervo et al., 2010), however, studies of extinction risks in squamates sample less than 15% of the squamate species (Böhm et al., 2013). The number of squamate reptiles evaluated by the IUCN Red List as of 2015 included approximately 70% of squamate genera. Recently, a study by Tonini et al. (2016) assessed the phylogenetic distribution

of evolutionary distinctiveness and threat status for squamates and their results suggested the critical need to assess extinction risks for close relatives, since threat status was phylogenetically clustered. This study also showed that species from degraded tropical regions such as Madagascar, Australia and India seem to be at higher risk of extinction, and the Sauria groups of geckos, iguanas and chameleons seem to be more threatened.

ORIGIN AND BIOGEOGRAPHY

Several different studies have attempted to estimate the age of the crown-group Squamata, with results that range from 170 to 280 million years ago (Ma) (reviewed in Jones et al., 2013). The last study, including over 4000 species, 52 genes and 13 fossils used for calibration of the divergence time, suggests a Squamata origin around 205 Ma, in the Late Triassic, very close to the boundary with the Early Jurassic (Zheng & Wiens, 2016). Most of the divergence time estimates performed in recent years (Kumar & Hedges, 1998; Vidal & Hedges, 2005; Wiens et al., 2006, 2012; Albert et al., 2009; Alfaro et al., 2009; Hipsley et al., 2009; Jones et al., 2013; Zheng & Wiens, 2016) rely on the fossil record of Squamata and of Rhynchocephalia, the sister order of Squamata, nowadays represented only by the genus *Sphenodon*, but with an extensive fossil record from the Triassic and Jurassic (Evans, 2003).

The earliest putative fossils of squamata are from the Early Jurassic (Evans et al., 2002). However, the presence of fossils of the sister order Rhynchocephalia in the Late Triassic, suggests that stem lineages representative of Squamata would be present in this timeframe as well. Within Squamata, the fossil record until the mid-Jurassic is scarce but, from this period on, the fossil record is relatively abundant (Carroll, 1975). Assemblages of squamate reptiles are known from the Late Cretaceous and fossils of lizards are relatively frequent in the northern continents that were part of Laurasia. In the southern continents that were part of Gondwana, the lizard fossil record is poorer, in a contrast with snake fossils, which are more common here than in the northern continents, for reasons not yet understood (reviewed in Evans, 2003).

Overall, the origin of Squamata, both from the fossil record and from time calibrated phylogenetic tree, is estimated between the Late Triassic and Early Jurassic, taking into account the wide intervals associated with uncertainties in the age estimation. Climatically, this was a period of general vegetation modification, in a clear indication of

changes towards warmer climates and greater continental aridity (Bonis & Kürschner, 2012). Geologically, at the end of the Triassic, all landmasses were part of the supercontinent Pangea, which broke apart a few million years later, around 200 Ma, in the Early Jurassic (Fig. 1.3b). If the origin of squamates is placed in the Late Triassic, the breaking of Pangea a few million years later indicates a vicariant origin of some major older squamate groups. Given that the divergence history within Squamata is spread in the Mesozoic and the beginning of the Cenozoic (200 – 50 Ma), following multiple divergences within this group were probably a result of both vicariance mediated by the separation of Laurasia and Gondwana during the Cretaceous (145-66 Ma; Fig. 1.3) (Blakey, 2008; Upchurch, 2008), and of dispersal, as a number of transoceanic dispersals would have also occurred, which has been hypothesised for some groups (Carranza et al., 2000; Calsbeek & Smith, 2003; De Queiroz, 2005; Vidal et al., 2008).

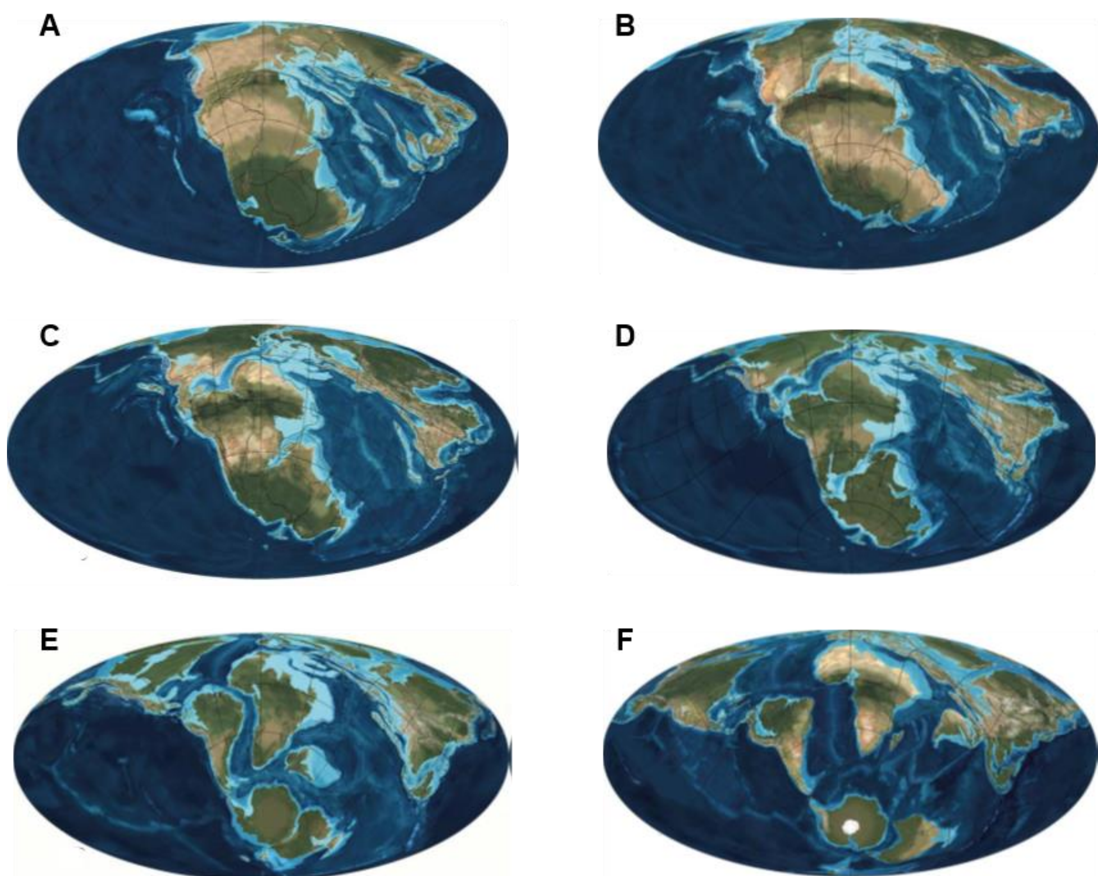


Figure 1.3. Paleogeography from the Upper Triassic until the Oligocene. A) Middle Triassic, ca. 240 Ma, B) Early Jurassic, ca. 200 Ma, C) Middle Jurassic, ca 170 Ma, D) Late Jurassic, ca. 150 Ma, E) Mid-Cretaceous, ca. 105 Ma, F) Oligocene, ca. 35 Ma. From Blakey (2008).

SQUAMATA AS MODEL ORGANISMS

Squamate reptiles are among the taxa broadly used as model organisms in a vast number of studies over the last decades, in fields including ecology, behaviour, phylogenetics, phylogeography and medicine. This is partly due to the high morphological and ecological diversity among squamates. The relatively ease in spotting and capturing the animals, as well as the tolerance in being manipulated for experimental purposes, has facilitated the use of these organisms in a variety of disciplines. Altogether, these characteristics have led, over recent decades, to a solid and long-term accumulation of knowledge in different areas, such as demographic, ecophysiological, and adaptive morphological studies (Camargo et al., 2010). The study of viviparity in lizards and snakes, for instance, has been around for over a century, in an attempt to understand how viviparity evolved in mammals and other vertebrates (Blackburn, 2006). Another focus of study concerns parthenogenesis, with 40 documented independent origins in one clade of snakes and seven major clades of lizards (Kearney et al., 2009). The evolution of body shape, particularly the transition to the elongated and limbless snake-form has also been studied, with 25 independent occurrences, sometimes repeatedly among closely related species (Wiens et al., 2006; Sites et al., 2011). Lizards of the Lacertidae family have also been widely used as models for studies on thermal biology (Castilla et al., 1999), locomotor performance (Kaliontzopoulou et al., 2010b), behaviour in the face of predation pressures (Cooper et al., 2014) and morphological differentiation related to habitat use (Vanhooydonck & Van Damme, 1999; Kaliontzopoulou et al., 2010a). Finally, because of the close relationship with human health due to the high number of human fatalities attributed to venomous snakes (between 20 000 and 94 000 every year; Kasturiratne et al., 2008), the venom and venom delivery systems have been a long-term interest of biologists (Sites et al., 2011 and references herein).

PHYLOGENETIC INFERENCE IN SQUAMATA

All the inherent diversity of squamate reptiles, in addition to the accumulation of knowledge in several different fields throughout recent decades makes this group extremely attractive for evolutionary studies, which in turn requires a strong phylogenetic framework. For example, the described biological and ecological studies on viviparity, parthenogenesis, venom and evolution of body shape have been conducted in a

phylogenetic framework, which makes it possible to trace the evolution of these traits and their diversity in the squamate tree of life. Likewise, phylogenetic inference has aided the assessment of extinction risk across squamates (Tonini et al., 2016). An accurate phylogenetic inference is fundamental and the cornerstone for all other comparative studies, not only applied to Squamata, but to the whole Tree of Life. Additionally, a steady increase in the phylogenetic studies over recent years, together with the application of new molecular tools has been the responsible of the growth in the number of squamate species, particularly through the discovery of cryptic diversity (Table 1.1, Fig. 1.1).

Phylogenetic studies among Squamata groups have a long history, starting early in the 20th century. Camp (1923) presented a branching diagram of squamate groups based on morphological characters of extant taxa and fossil groups that identified several groups that are still supported in modern analyses (Fig. 1.4). Later, the relationships inferred by Estes et al. (1988) from extant squamate groups remained the mostly widely used hypothesis of Squamata relationships until quite recently. In this work, the Squamata were divided in Iguania (including Iguanidae, Agamidae and Chamaeleonidae), which were considered the oldest Squamata members, and the Scleroglossa (all other squamates), based mainly on characters related to feeding behaviour, with tongue prehension in Iguania and teeth and jaw prehension in Scleroglossa (Fig. 1.5). Many other morphologically-based relationships were inferred in the last couple of decades, mainly drawn from the dataset used by Estes et al. (1988) albeit including fossils and different morphological characters in the analyses (Wu et al., 1996; Evans & Barbadillo, 1998, 1999; Lee, 1998; Caldwell, 1999; Lee & Caldwell, 2000; Evans & Wang, 2005; Conrad, 2008).

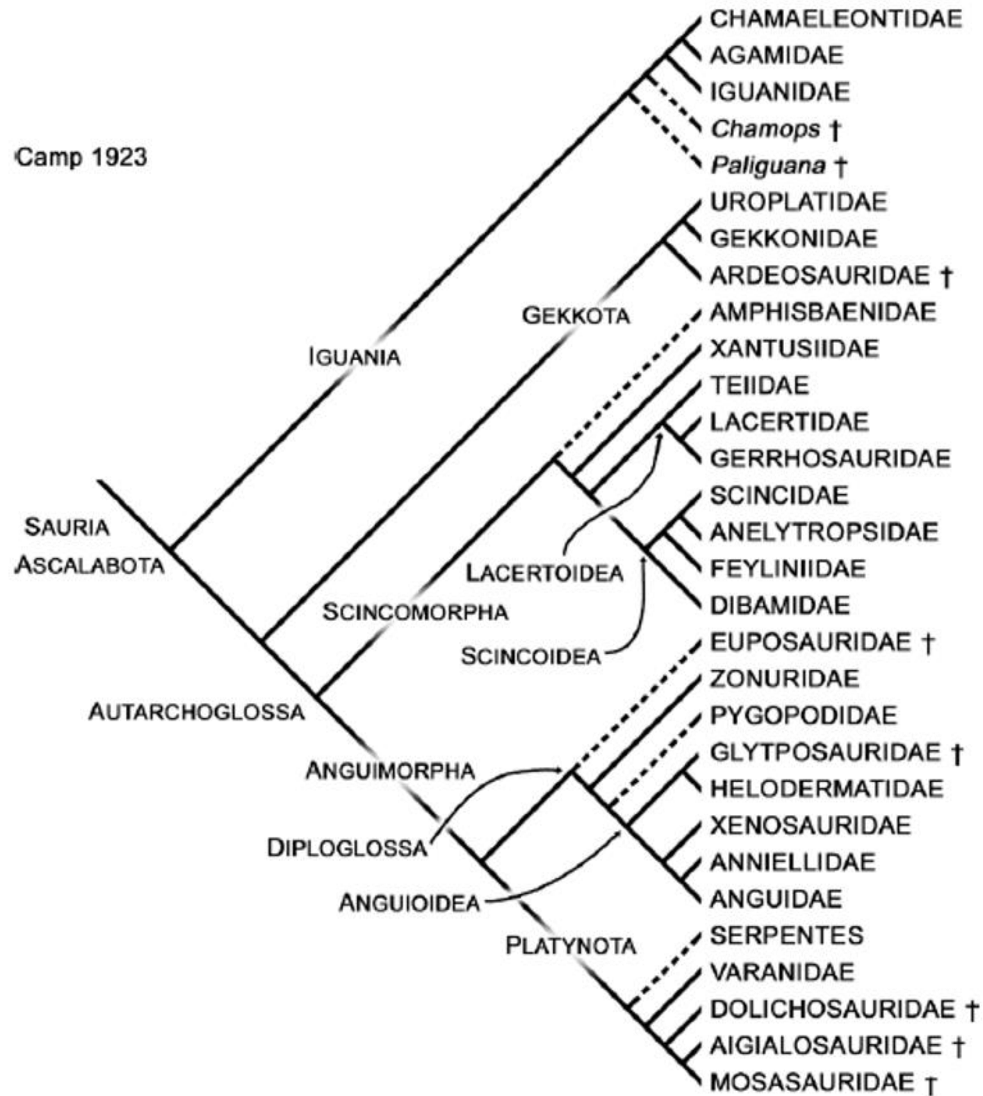


Figure 1.4. Evolutionary history of Squamata as proposed by Camp (1923). Fossil taxa represented with a cross. From Conrad (2008)

The great morphological diversity within Squamata and especially the high level of convergence, however, has hindered establishing high level phylogenetic relationships based on morphology. The advent of molecular-based phylogenetic inference has indicated that some morphological based groups may be artificial, particularly the division between Iguania and Scleroglossa. In recent years, several studies on high level squamate phylogenetic inference have been produced, with varying number of taxa and type of molecular markers and phylogenetic tools (Table 1.2). The relationships and placement of major groups of Squamata have not been static over the years, but an increase in the number of taxa and of molecular markers, both mitochondrial and nuclear, in the phylogenetic analyses has stabilized the position of

most major clades (Townsend et al., 2004; Vidal & Hedges, 2005; Douglas et al., 2006; Zhou et al., 2006; Wiens et al., 2006, 2012; Kumazawa, 2007; Albert et al., 2009; Mulcahy et al., 2012; Jones et al., 2013; Pyron et al., 2013; Zheng & Wiens, 2016).

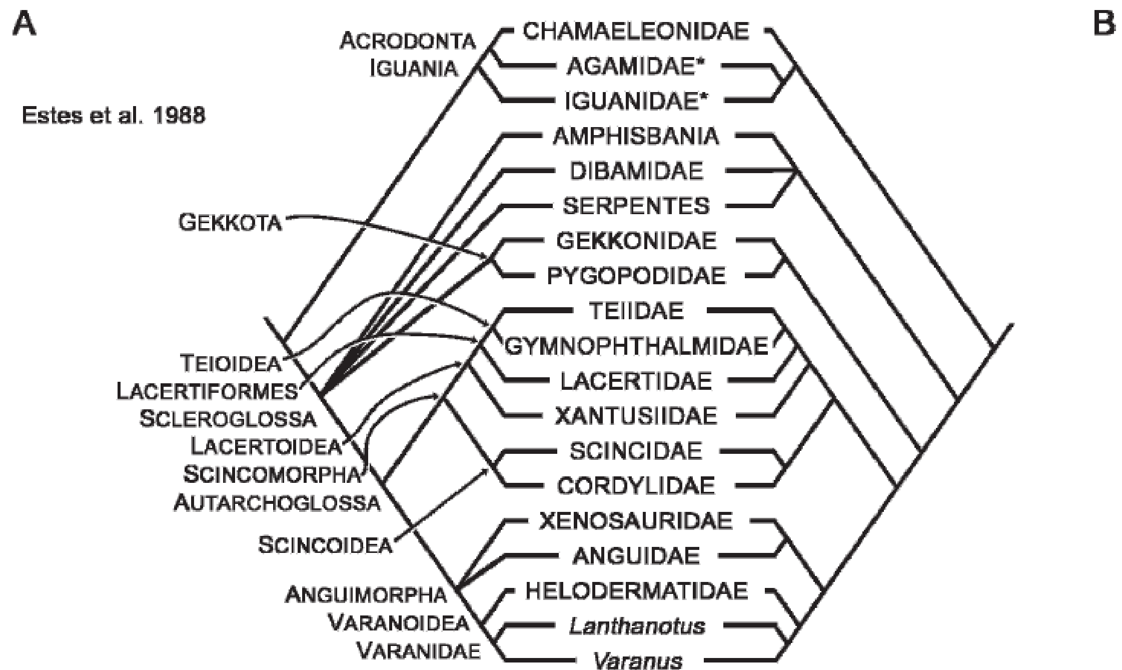


Figure 1.5. The cladogram on the left (A) represents the conservative cladogram of squamate relationships based on morphology as presented by Estes et al. (1988). The right cladogram (B) shows the relationships recovered when the data matrix by Estes et al. (1988) is run in PAUP. From Conrad (2008).

Following the development of new statistical tools and new technologies facilitating data generation over the last years, two main approaches for solving the phylogenetic relationships within the species-rich groups such as squamata have emerged: the phylogenomic and the supermatrix approaches. The first method is based on sampling many loci for few species and is mainly applied for solving higher level relationships (Chiari et al., 2012; Jarvis et al., 2014; Weigert et al., 2014), while the supermatrix approach involves including many species for fewer loci and is more directed to relationships inference at species level (Pyron & Wiens, 2011; Jetz et al., 2012; Pyron et al., 2013). The phylogenomic method was firstly applied in order to solve the position of major clades within Squamata (Table 1.2), mostly using complete mitogenomes (Douglas et al., 2006; Zhou et al., 2006; Kumazawa, 2007). Later, due to the increasing number of available sequences on GenBank, supermatrix approaches were applied in an attempt to solve both higher and species level relationships (Pyron et al., 2013). The

most recent phylogenetic work on Squamata is a combination of these two approaches (Zheng & Wiens, 2016; Fig. 1.6).

Table 1.2. Number of Squamata species, type of molecular markers and phylogenetic methods used in previous studies of phylogenetic inference on Squamata.

Reference	Taxa (Squamata)	Molecular markers	Phylogenetic method
Zheng & Wiens, 2016	4162 species	5 mtDNA loci 47 nucDNA loci	Phylogenomics Supermatrix
Pyron et al., 2013	4162 species	5 mtDNA loci 7 nucDNA loci	Supermatrix
Jones et al., 2013	62 species	1 nucDNA	Supermatrix
Mulcahy et al., 2012	64 species	25 nucDNA loci 1 mtDNA loci	Supermatrix
Wiens et al., 2012	161 species	44 nucDNA loci	Supermatrix Species tree
Albert et al., 2009	27 species	13 mtDNA loci	Supermatrix
Kumazawa, 2007	24 species	Mitogenome	Supermatrix
Wiens et al., 2006	Trees from previous studies	Trees from previous studies	Supertree
Zhou et al., 2006	15 species	Mitogenome	Supermatrix
Douglas et al., 2006	20 species	Mitogenome	Supermatrix
Vidal & Hedges, 2005	18 species	9 nucDNA loci	Supermatrix
Townsend et al., 2004	69 species	2 nucDNA loci 1 mtDNA loci	Supermatrix

Relationships recovered by this recently published phylogeny (Zheng & Wiens, 2016) are mainly congruent with previous molecular studies with a comparable taxon sampling (Townsend et al., 2004; Vidal & Hedges, 2005; Mulcahy et al., 2012; Wiens et al., 2012; Jones et al., 2013; Pyron et al., 2013) and include: the legless lizard family Dibamidae is sister to all other Squamata (except in Jones et al., 2013); the Gekkota is the second group to branch off; there is a major division between Scincoidea and the remaining Squamata (Episquamata); the latter is divided in Lacertoidea, including the families Teiidae, Gymnophthalmidae, Lacertidae and the Amphisbaenia group; Iguania (including Acrodonta and Pleurodonta), Anguimorpha and Serpentes. Although these

main clades are congruent between these studies, the relationships within the groups vary slightly, as for instance the position of Serpentes as sister to Anguimorpha (Townsend et al., 2004; Jones et al., 2013), as opposite to sister to the group composed by Iguania and Anguimorpha (Vidal & Hedges, 2005; Mulcahy et al., 2012; Pyron et al., 2013; Zheng & Wiens, 2016).

While the application of both phylogenomic and supermatrix approaches seems to have stabilised the high level phylogenetic position of the major groups of squamates with a relatively high support, further studies highlighted the possibility that these high levels of support might come from artefacts, such as the effect of missing data and incongruence between the gene trees of the loci used in these methods (see section 2 for more details). Additionally, the overconfidence of the support on the speciation nodes might hinder an accurate phylogenetic inference and discrimination in particularly challenging cases, as exemplified by the basal polytomies and fast radiation events found in some clades within the families Lacertidae and Colubridae, which have been used as case studies in this thesis.

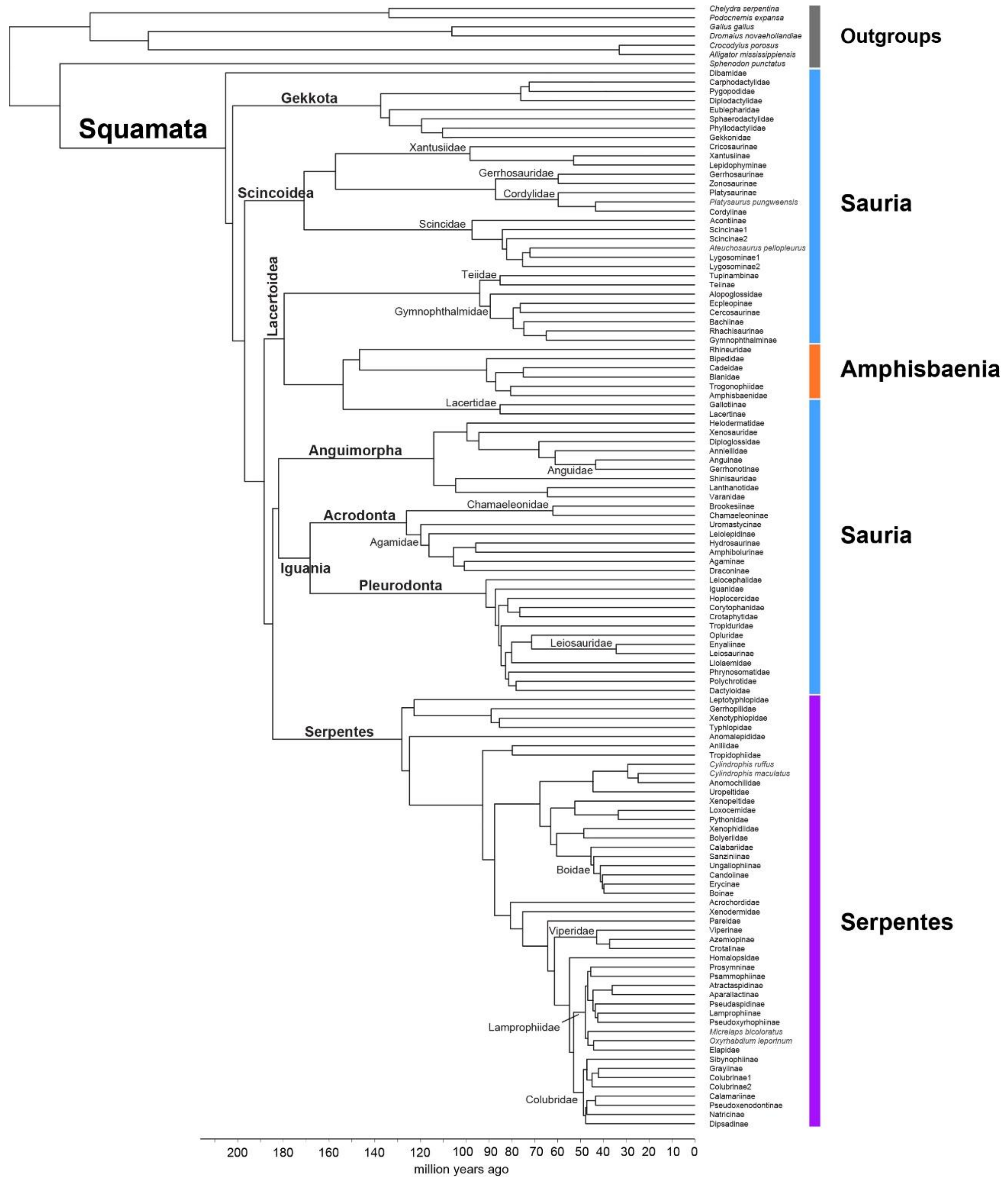


Figure 1.6. Phylogenetic relationships of Squamata. Adapted from Zheng & Wiens (2016).

2. MOLECULAR TOOLS OF PHYLOGENETIC INFERENCE

SOFT AND HARD POLYTOMIES

The interpretation of the results from the phylogenetic inference can sometimes be difficult due to the presence of a polytomy in the phylogenetic tree, instead of a dichotomy (Fig. 1.7). The basic definition of polytomies falls into two categories: soft and hard polytomies. A soft polytomy represents contradictory results due to insufficient data or idiosyncratic patterns of evolution of different markers, expressing an artefact created by the data. On the other hand, a hard polytomy represents the lack of available evidence to resolve branching order and may indicate a fast split of the forms during a short evolutionary period of time (Garland & Díaz-Uriarte, 1999).

Figure 1.7. Graphical representation of a polytomy. Adapted from Garland & Díaz-Uriarte (1999)

The lack of resolution associated to soft polytomies can be related with the nature and number of molecular markers used in the phylogenetic inference, as too much or too little variation can obscure the inference. On one side, highly variable mitochondrial DNA markers can lead to saturation over long time frames and subsequent shortening of branch length relative to the true divergence. On the other hand, too little variation of slowly evolving nuclear DNA markers limits phylogenetic inference due to the usual lack of phylogenetically informative characters provided by these markers. Given this, the

clarification of a soft polytomy is, in theory, easily achieved by the inclusion of additional lineages to the analyses, through for example multiple mitochondrial DNA and fast evolving nuclear DNA loci, a combination that should provide the inference with sufficient resolution and dissolve the polytomy (Kuhn et al., 2011). However, in the presence of an old and rapid radiation, represented by long terminal branches and short internodes (Whitfield & Lockhart, 2007), the addition of more DNA information might increase the length of the internal branches but, in consequence, increase the length of the terminal branches as well (Walsh et al., 1999). This could potentially lead to saturation, homoplasy and convergence between phylogenetically divergent lineages, also known as long-branch attraction, where artefactual relationships are supported.

The hard polytomies may not be so easily identified and here lies the challenge for an accurate phylogenetic inference. If a thorough inference based on multiple individuals and multiple mitochondrial and nuclear DNA loci is not enough to resolve the branching order of the phylogenetic tree, it may be due to the presence of a true rapid radiation. These cladogenic events have been documented in several groups of organisms including plants (Shaw et al., 2003), insects (Pena & Wahlberg, 2008), fish (Hollingsworth et al., 2013), amphibians (Mahoney, 2001), snakes (Wiens et al., 2008), birds (Barker et al., 2004) and mammals (Lin et al., 2002).

MITOCHONDRIAL AND NUCLEAR DNA

The use of mitochondrial DNA in phylogenetic studies ruled alone for decades. Its application is based on very sound reasons as it is easily amplified, it usually does not undergo recombination in animal taxa, has a high evolutionary rate when compared to nuclear DNA, and an effective population size of approximately one-quarter of the nuclear genome (reviewed in Ballard & Whitlock, 2004). Thus it was, and still remains, useful to assess the recent evolutionary history of taxa without an extensive sequencing effort. But although its application has been demonstrated successfully on many occasions, it has also been questioned and some disadvantages have been highlighted (Ballard & Whitlock, 2004; Brito & Edwards, 2009; Edwards & Bensch, 2009). The overall small size of the mitogenome and the possibility of integration on the nuclear genome creating pseudogenes that may still be amplified with conserved primers are good examples (Bensasson et al., 2001). Also, the fact that all genes are linked and do not undergo recombination means that the phylogenetic inference is based on a single locus, and therefore the resultant trees are not independent and represent only a small part of

the evolutionary history of taxa (Zhang & Hewitt, 2003). This tree could be different from the true species tree due to the effects of natural selection (e.g. Bazin et al., 2006), introgression (e.g. Melo-Ferreira et al., 2005) or simply to a large stochastic variance that comes from sampling gene trees from a set of populations or species (e.g. Hey & Machado, 2003). These obstacles have hampered the inference of more complex or ancient phylogenetic relationships (e.g. Gonçalves et al., 2007; Godinho et al., 2008).

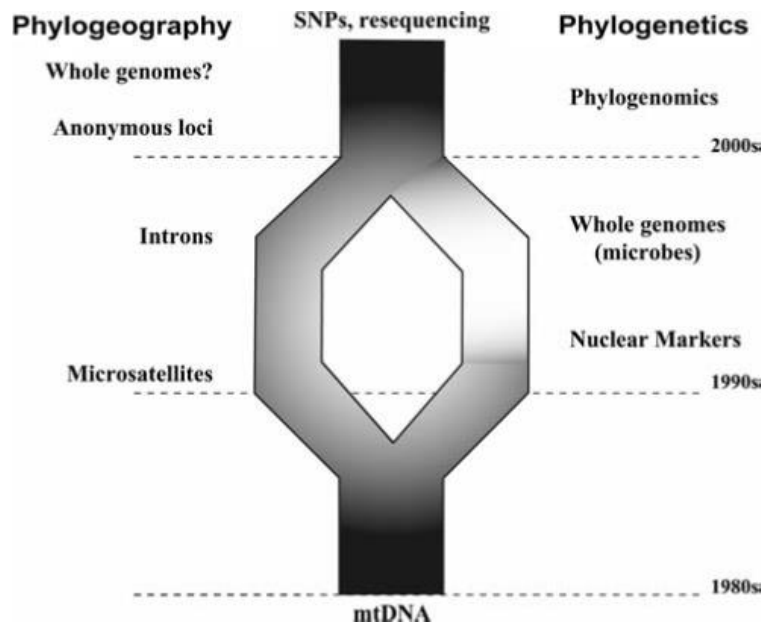


Figure 1.8. Methodological shifts and evolution of the fields of phylogeography and phylogenetics. From (Brito & Edwards, 2009)

The addition of nuclear DNA markers to the phylogenetic analyses came as the natural solution to some of the obstacles posed by the mitochondrial DNA and its use has been increasing in phylogenetic studies (Figs. 1.8 & 1.9). The fast and heterogeneous evolution of mitochondrial DNA can make them quite ineffective for the resolution of old and intermediate level relationships, while resolution power may be achieved by the addition of the generally more slowly evolving nuclear genes. Analysis of nuclear DNA possesses another particular advantage, multiplicity, as combining multiple nuclear loci can reveal a finer description of patterns than the single locus represented by the mitochondrial DNA (Edwards & Bensch, 2009). Even though nuclear DNA is less variable than mitochondrial, this cannot be generalised, as the proportion of polymorphisms in the nuclear genome can vary greatly due to factors such as

recombination, functional constraints, proportion of introns and exons, and levels of selection (Zhang & Hewitt, 2003).

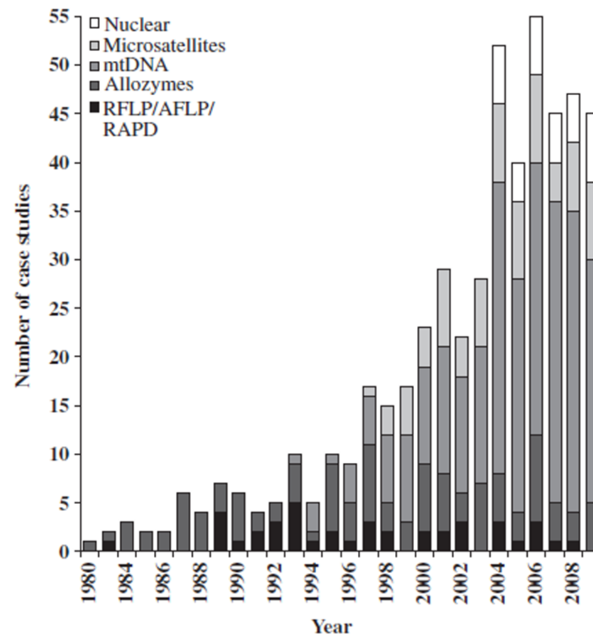


Figure 1.9. Distribution of the use of molecular markers in Sauria between 1980 and 2009. From Camargo et al. (2010).

The development of nuclear markers for phylogenetic analyses however, has not been as straightforward as with the mitochondrial genome. Eukaryotic cells include a single nucleus and therefore most nuclear genes present a single copy in each cell making the amplification more difficult compared to mitochondrial genes, which occur in multiple copies as a result of the presence of many mitochondrial cells within each Eukaryotic cell. At the beginning, the use of intronic regions, associated with a faster evolution rate seemed helpful for phylogenetic studies of closely related species. These regions, unfortunately are prone to high sequence length variation, making alignment and thus phylogenetic analyses more difficult, particularly for more divergent species. The nuclear protein-coding loci (NPCL) seemed an improved option, overcoming the alignment difficulties associated with introns. However, within NPCL, too little variation associated with the slow evolutionary rate of some nuclear genes can be an obstacle in solving phylogenetic relationships, since low information content in these genes might mean a very low capacity to recover speciation nodes. This means that within the nuclear DNA markers, some might be more useful than others. Until some years ago, and even

currently, phylogenetic studies in vertebrates have relied in the use of a few genes, for instance *cmos* (Oocyte Maturation Factor Mos) and *rag1* (Recombination Activating 1). The problem with these genes is the relatively low evolutionary rate, which might be helpful in resolving old relationships but presents difficulties in solving more recent or faster cladogenic events, that are not resolved by mitochondrial genes either. In recent years, fast-evolving informative nuclear genes have been developed (e.g. Townsend et al., 2008; Pinho et al., 2009; Portik et al., 2011; Fig. 1.10) and their use increased due to the higher levels of resolution in the phylogenetic inference. Indeed, the importance of nuclear markers with high levels of informative sites has been long recognised (Graybeal, 1994).

The combination of mitochondrial and fast evolving nuclear markers seems to be the most reasonable approach for the resolution of between species' phylogenetic relationships. The molecular markers reveal different parts and depths of the complex evolutionary history of taxa and, as has been demonstrated in the literature, both nuclear and mitochondrial markers complement each other very well in phylogenetic analyses (Espregueira Themudo et al., 2009; Fontanella et al., 2012; Vences et al., 2014; Cibois et al., 2017; De Silva et al., 2017; Roussel & Van Wormhoudt, 2017). Despite this, conflicting phylogenetic signal from mitochondrial and nuclear markers is still recovered (Crochet et al., 2003; Gonçalves et al., 2007; Salvi et al., 2017) and can be partially explained by several factors. For example, the very different rates of evolution from mitochondrial and slow evolving nuclear genes may provide very different phylogenies. The incomplete lineage sorting in the nuclear markers may also result in phylogenetic differences between nuclear and mitochondrial genes (Pinho et al., 2007). It can also be

explained by a combination of recent species origin and subsequent interspecific gene flow (Crochet et al., 2003).

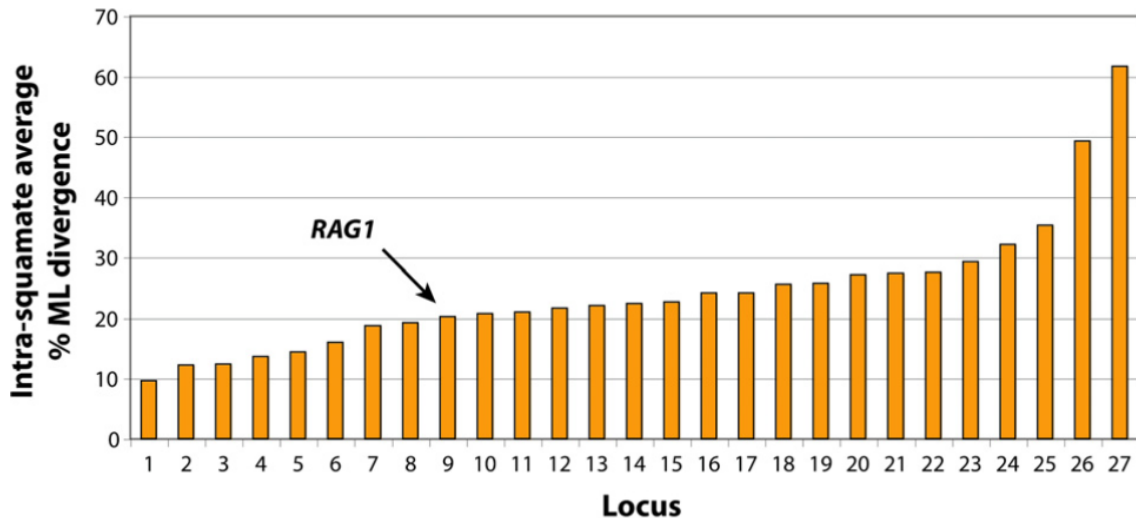


Figure 1.10. Variability of 27 NPCL in squamate reptiles. The commonly used locus RAG1 was used for comparison. 1) ZEB2, 2) BDNF, 3) FSTL5, 4) ZFP36L1, 5) ADNP, 6) BACH2, 7) PNN, 8) NGFB, 9) RAG1, 10) FSHR, 11) SLC301A, 12) SNCAIP, 13) TRAF6, 14) BMP2, 15) GPR37, 16) ECEL1, 17) PTGER4, 18) AHR, 19) MKL1, 20) DNAH3, 21) AKAP9, 22) REV3L, 23) NT3, 24) BACH1, 25) PTPN12, 26) UBN1, 27) PRLR. From Townsend et al. (2008)

GENE TREES, CONCATENATION AND SPECIES TREE

The increasing number of available sequences led to the increase of phylogenetic inference based on multilocus data, but still largely reliant on inference methods designed for single genes. Until around a decade ago, there was a general assumption that the most commonly occurring gene tree estimated from multilocus data ("democratic vote") was congruent with the true species tree (Degnan & Rosenberg, 2006; Kubatko & Degnan, 2007). However, this is not true in all situations, where there are discordances between the species tree and the gene trees due to the stochastic way that lineages sort during speciation (Fig. 1.11) (Degnan & Rosenberg, 2006). Causes of discordance among gene trees include incomplete lineage sorting or deep coalescence, gene duplication, horizontal gene transfer and branch length heterogeneity (Maddison, 1997; Kubatko & Degnan, 2007; Edwards, 2009). Indeed, an increasing number of studies discovered that gene tree approaches, such as the commonly used Maximum Likelihood and Bayesian Inference, may infer the genealogical path of individuals rather than the true evolutionary relationships among species, therefore leading to false inference of

relationships. As general examples, the work from Degnan & Rosenberg (2006) showed that for genealogies of five or more species, the branch lengths of the trees are more often discordant between gene trees. Further, the addition of multiple loci to the analyses increases the probability that the most frequently observed gene tree is an incorrect estimate of the species tree. In another study using simulations of the concatenation approach in conditions where the coalescent produces discordant gene trees (Kubatko & Degnan, 2007), the authors suggested that bootstrap support can be moderate to strong for incorrect genes trees.

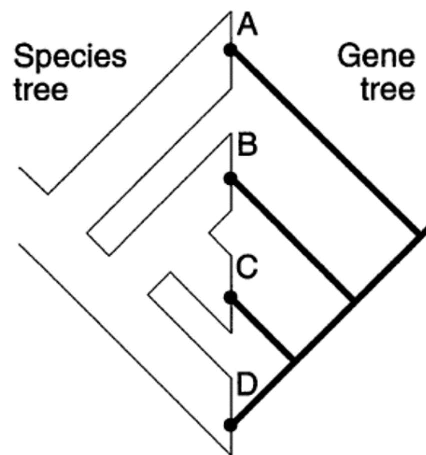


Figure 1.11. Incongruence between species tree and gene tree. Species B and C are sister species but their gene copies are not sister copies. From (Maddison, 1997)

Until a decade ago, two approaches had been primarily used for estimating species tree from multilocus gene trees analyses: the consensus and concatenation methods (Gadagkar et al., 2005; Degnan & Rosenberg, 2009). In the consensus method the gene trees with the same set of taxa are individually estimated and then summarized into a consensus tree. Similarly, a supertree approach is used when the input trees have overlapping but nonidentical sets of taxa (Bininda-Emonds, 2004). In the concatenation approach the sequences are combined into a supermatrix alignment that is then analysed to represent the species tree. The use of the concatenation arose from the "total evidence" approach (Kluge, 1989, 2004), a mostly philosophical idea that aims to include all available data in phylogenetic analyses. This method, however, came across some practical obstacles. The supermatrix concatenation method assumes that all the data evolved accordingly to a single evolutionary rate and depict the same pattern of

relationships, which is not the case with real multilocus data. This concern has been addressed by the development of the current analytical methods, such as likelihood and Bayesian algorithms, that are able to partition the data in the concatenated alignment and attribute different evolutionary rates to each partition. Although the models of nucleotide substitution do not completely describe the complexity of DNA sequence evolution, their application on phylogenetic analyses has dramatically increased the accuracy of the inference (Edwards, 2009). On the other hand, the conflicting signal from different genes has an unpredictable effect on the phylogenetic inferenced based on the supermatrix. Indeed, recent research into multilocus phylogenetics shows that the concatenation of sequences from multiple genes generates what has been called a 'wrong kind of average' between gene trees, leading to a poor estimation of the species tree (Degnan & Rosenberg, 2006; Heled & Drummond, 2010). Further, in some cases a multilocus phylogeny based on gene trees provides a statistically supported but incorrect tree (Kubatko & Degnan, 2007; Heled & Drummond, 2010). Additionally, the straightforward practice of using the estimated gene tree topology that occurs most often can be asymptotically guaranteed to produce the wrong estimate of the species tree in the so-called anomaly zone, a set of short internal branches in species trees that will generate gene trees that are discordant with the species tree more often than gene trees that are concordant (Degnan & Rosenberg, 2006).

The multilocus coalescence-based species tree approach appeared as an alternative to the consensus and concatenation of multilocus sequences, allowing the reconciliation of a set of gene trees embedded in a shared species phylogeny, under the multispecies coalescent model of molecular evolution. This has led to the development of maximum likelihood (Kubatko et al., 2009) and Bayesian (Drummond & Rambaut, 2007; Heled & Drummond, 2010) species tree inference approaches that provide with the true tree representing the species evolutionary history. According to the Bayesian method developed by Heled & Drummond (2010), a species tree specifies the tree topology of ancestral relationships, the relative divergence times between two species and the population size history for each species. The gene trees are embedded in the species tree by following the stochastic coalescent process back in time within each branch – multispecies coalescent.

Several comparative studies between concatenation and species tree methods have been performed (Edwards et al., 2007; Heled & Drummond, 2010; Bayzid & Warnow, 2012; Wiens et al., 2012; Hovmöller et al., 2013; Lambert et al., 2015). For example, in a comparison between concatenation and species tree in squamate reptiles performed by Wiens et al. (2012), the concatenated supermatrix yielded weak support

and great incongruence among genes on shorter branches that may increase with the addition of more loci. On the other hand, the coalescent-based species-tree approach was successful in resolving the shorter branches with strong support. Recently, another work reached a similar conclusion, that weakly supported relationships in the concatenated trees are more likely to have a different resolution in the species tree method (Lambert et al., 2015).

Sampling multiple taxa and loci can be useful for resolving species trees. While sampling multiple independent genes helps to control the stochastic forces that affect individual genes (Maddison & Knowles, 2006), sampling multiple individuals from the same species can increase the accuracy of the species tree, as well as the estimation of speciation times (Maddison & Knowles, 2006; Heled & Drummond, 2010). The balance between the number of individuals and genes can be related with the depth of the tree: in shallow trees, sampling more individuals is more effective, while in deep species trees, more loci are more helpful (Maddison & Knowles, 2006; McCormack et al., 2009). A recent study, however, suggested that while increasing the number of loci did substantially improve the resolution of the species tree, sampling more individuals had minor effects (Corl & Ellegren, 2013). The type of loci also seems to affect the resolution of the species tree, with the “informativeness” of each locus having great importance in the accuracy of the inference, particularly when the total number of loci in the analyses is low (Camargo et al., 2012).

Limitations of the species tree method are related with the high computational load required when the number of taxa and/or loci are very high. Now that more and more phylogenomic data are becoming available, this hinders the use of the species tree approach in solving deep phylogenetic relationships for a large number of taxa. Recently, Lambert et al. (2015) investigated the use of both concatenated and species tree methods in solving the relationships of a large group. The authors suggest that incorporating uncertainty in the concatenated tree may cover the differences between concatenation and species tree methods. They also performed analyses with different number of taxa and loci which indicate that taxon sampling may be more important than loci sampling when applying the species tree method to address deep phylogenetic questions.

THE EFFECT OF MISSING DATA

The amount and type of missing data, both of taxa and loci, and its affect in the inference of relationships between taxa are a controversial subject in phylogenetics (e.g. Huelsenbeck, 1991; Lemmon et al., 2009; Sanderson et al., 2010; Hovmöller et al., 2013; Roure et al., 2013; Jiang et al., 2014; Zheng & Wiens, 2015). The great increase of sequences deposited in GenBank is not evenly distributed across taxa or molecular markers. The number of mitochondrial markers is higher, compared with nuclear DNA genes and, within the latter, a handful of genes represent the majority of the nuclear sequences available. A similar situation is found for different taxa, while some species are very well represented in GenBank for a variety of sequences from different loci, others have only sequences for a pair of genes or are not represented at all. This situation, however, seems to be changing towards completeness of data due to the high number of new phylogenetic studies being published each year.

The majority of the supermatrix approaches use GenBank sequences in their inference and therefore inevitably have different levels of missing data in their datasets. The effects of missing data in phylogenetic inference at both species level and higher taxonomic levels has been the focus of some studies, as the potentially strong and negative effect of the missing data could hamper the accurate resolution of the phylogenetic relationships among taxa. The balance within big datasets is not easily achieved. Assumptions that missing data has a negative impact in the phylogenetic results inevitably leads to the exclusion of non-missing species or loci, together with the missing data. Several studies were conducted in order to evaluate the potential consequences of missing taxa and missing characters in branch length, node support, species tree reconstruction and divergence time estimation. A recent study aimed to understand the effects of genes with missing data in two well resolved phylogenies by conducting simulations with varying degrees of missing data in the incomplete genes (Jiang et al., 2014). Results from this study show that the costs of deleting the non-missing data associated with the removal of missing data are higher than the benefits of excluding the genes with missing data. It indicates that adding incomplete genes generally increases the accuracy of the phylogenetic inference and is particularly helpful for resolving poorly supported nodes. Another study by Hovmöller et al. (2013) concluded that the consequences of missing data in species tree reconstruction using different inference methods are overall negligible, but only for larger multilocus datasets with between 25 to 100 loci. In general, sampling more individuals per species had the strongest effect of improving the accuracy of the species tree reconstruction. Likewise,

the amount of missing data had little impact on divergence time estimates, even with a high percentage – 75% to 80% – of missing data, while the use of fewer fossil calibration points in the analyses resulted in much higher mean error for the estimates (Zheng & Wiens, 2015). On the other hand, many other studies suggest that missing data can strongly bias the phylogenetic inference (Lemmon et al., 2009; Sanderson et al., 2010). It is very difficult to compare such opposite results coming from different studies on the effect of missing data because all these studies are based either on specific empirical cases (which may not be representative of many other real dataset) or on simulations which strongly depend on the simulation design.

The studies reporting minor effects of missing data on the accuracy of phylogenetic inference, however, should not stimulate overconfidence in the use of supermatrices with a prevalent portion of missing entries. Several recent studies applying both supermatrix and phylogenomic approaches have a high percentage of missing data that is usually biased towards completeness of mitochondrial DNA genes relatively to the missing nuclear genes (e.g. Pyron et al., 2013; Figueroa et al., 2016). For some species, the missing data for nuclear genes can be higher than 95%, in some cases even 100%, while missing mitochondrial DNA data is very low. Independently of the total number of mitochondrial and nuclear genes used in the study as a whole, due to the missing nuclear information, for some taxa the phylogenetic inference relies mainly on mitochondrial DNA, meaning an overall single-locus approach. Effects of missing data between mitochondrial and nuclear DNA remain mainly unstudied.

3. CASE STUDIES IN SQUAMATA

THE LACERTIDAE FAMILY

Within Squamata, the lizard family Lacertidae Oppel, 1811 has been widely used as a model for studies in several different areas such as population genetics, ecophysiology and locomotor performance related with morphological variation (reviewed in Camargo et al., 2010). Indeed, regarding phylogeographical studies, for instance, of all Sauria, Lacertidae is the most studied family (Figs. 1.12 & 1.13), and of the ten most studied genera, four belong to Lacertidae (Camargo et al., 2010). Its position within Squamata is currently well established as the sister family of the worm lizards Amphisbaenia group

(Fig. 1.6). These families, together with the American Teiidae and Gymnophthalmidae form the major Lacertoidea group, an old clade in the Squamata tree (Fig. 1.6).

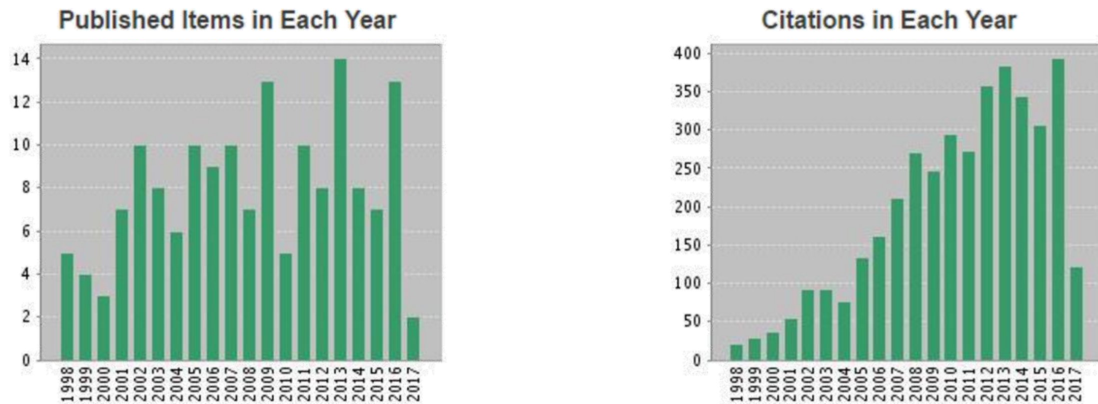


Figure 1.12. Number of publications and citation of phylogenetic studies with Lacertidae. From the Web of Science (June 2017).

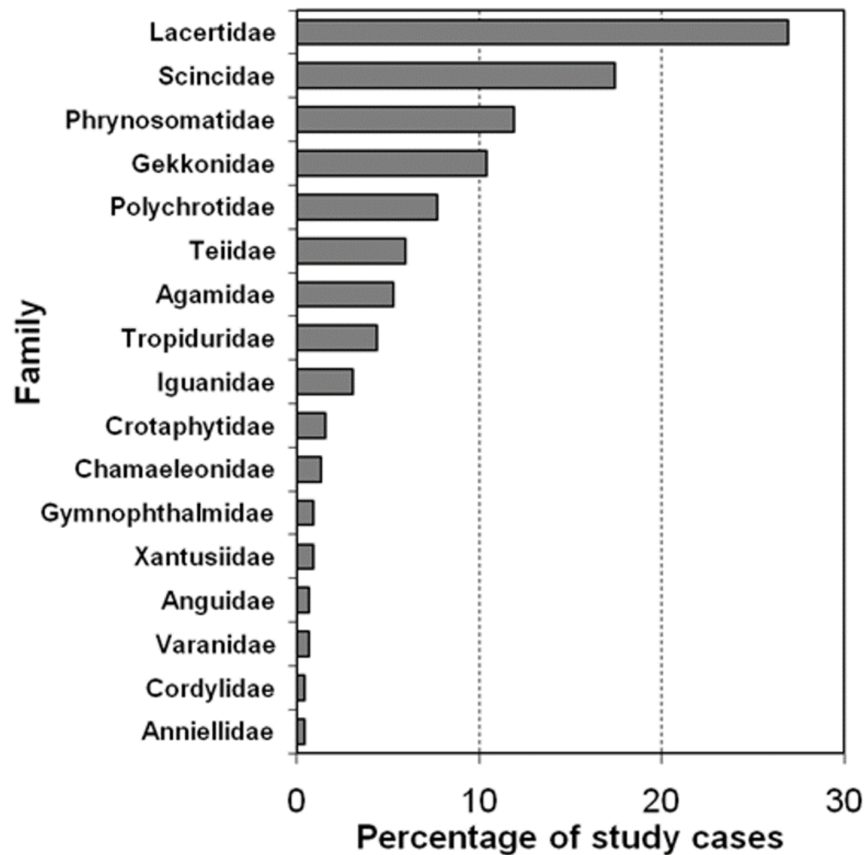


Figure 1.13. Percentage of phylogeographic studies by Sauria families until 2010. From (Camargo et al., 2010)

Lacertidae is a relatively small family with 323 currently described species compared with some other families of Sauria that have over 1000 species. It is distributed in the Palearctic and Afro-tropical ecozones. Lacertids are defined by a number of exclusive synapomorphies, including for instance lack of downgrowths on the parietal bone (Estes et al., 1988), usual presence of sexual variation in the number of presacral vertebrae and abdominal fat-bodies largely outside the peritoneum (Arnold, 1989; Arnold et al., 2007). With the exception of a few species whose body size is much larger than the mean (*Gallotia* species reaching over 40 cm SVL), lacertid lizards are relatively small, with less than 10 cm SVL. These are also remarkably similar in form, but with high levels of colour variation.

Taxonomically the Lacertidae family is divided into two subfamilies, the Gallotiinae and the Lacertinae. Gallotiinae is the oldest subfamily and is represented by two genera, *Gallotia* and *Psammodromus*, the first endemic to the Canary Islands and the second from the Western Mediterranean. Lacertinae is divided in two tribes: Eremiadini, mainly including Afrotropical, Saharan and Eurasian taxa, and Lacertini with Palaearctic and Oriental taxa (Fig. 1.14). In fact, the Lacertini are the most common reptiles in the Palearctic. This taxonomical division is supported by molecular data and also by morphological features (Arnold et al., 2007).

The Lacertidae probably originated in Europe, based on an extensive collection of lacertid fossils in this continent, extending back to the Paleocene, approximately 60 Ma (Augé, 2005). A recently discovered fossil of the *Gallotiinae* group has further confirmed the ancient origin of this group in Europe (Čerňanský et al., 2016). The separation between Gallotiinae and Lacertinae is estimated to be in Europe and relatively old, between 20 to 30 Ma or even older (Arnold et al., 2007; Hipsley et al., 2009; Mendes et al., 2016), followed by the colonisation of the Canary Islands by *Gallotia*. The divergence between Lacertini and Eremiadini is estimated to approximately 16 Ma (Arnold et al., 2007) and is possibly associated with movement to Africa of the Eremiadini ancestor, which would have been facilitated by the collision of Africa with Eurasia, between 15-19 Ma (Rögl, 1999). Phylogenetic structure within Eremiadini shows two main clades associated with the distribution areas of the species, an Ethiopian sub-clade and Saharo-Eurasian sub-clade (Mayer & Pavlicev, 2007). The branching of the Eremiadini is quite structured and supported in the phylogenetic tree. On the other hand, Lacertini shows an unresolved and bush-like topology, a basal polytomy.

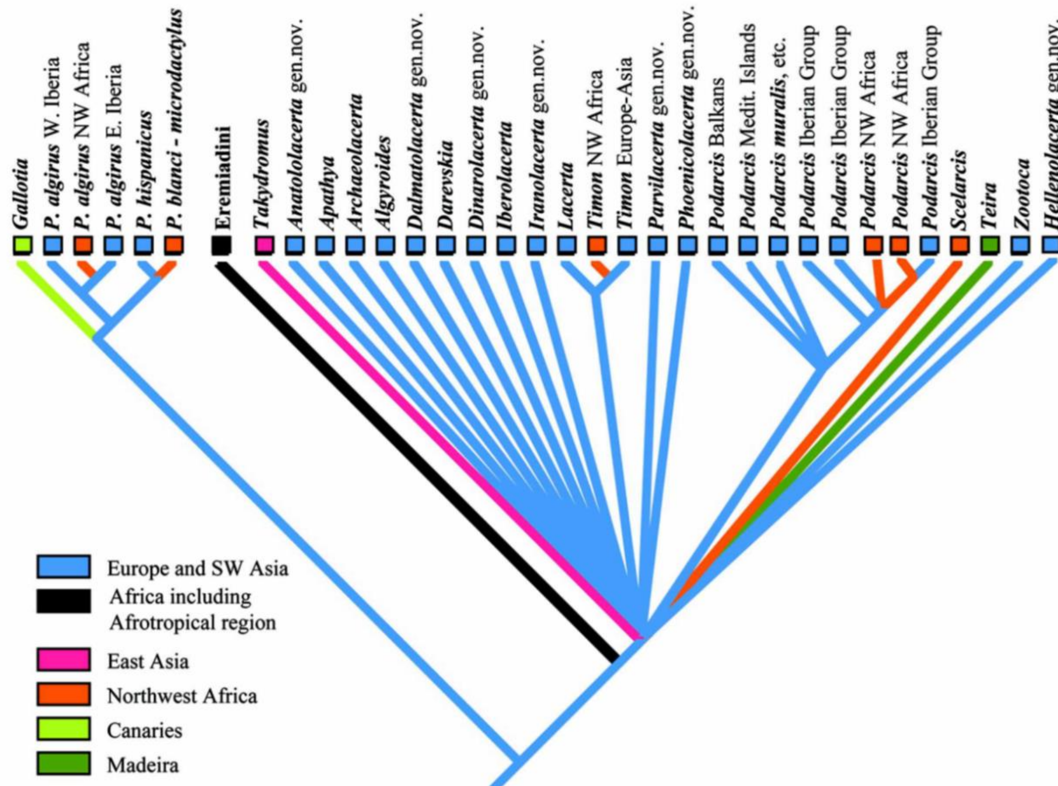


Figure 1.14. Area cladogram of the Lacertidae distribution. Parsimony analyses indicates the origin of the family in Europe. From (Arnold et al., 2007).

THE TRIBE LACERTINI

Lacertini currently includes 19 genera and 124 species but these numbers are likely to change since this group has had a relatively high taxonomic activity in the last decade, as exemplified by the description of eight new genera, previously included in *Lacerta* (Arnold et al., 2007) and the discovery and description of cryptic diversity within *Podarcis* (Geniez et al., 2007; Kaliontzopoulou et al., 2011).

This tribe includes three of the most studied lizard genera, including *Podarcis* which is the most used model organism in ecology and evolutionary studies among Lacertidae (Camargo et al., 2010). As already mentioned for the whole Squamata group, these assessments require an understanding of the phylogenetic history of the studied group in order to draw comparisons within an accurate phylogenetic framework.

The inference of phylogenetic relationships within Lacertini has been challenging. Numerous studies have attempted to assess the basal phylogenetic structure and

relationships between genera of this tribe, from morphological, bio-chemical and molecular studies including mitochondrial DNA and, lately, nuclear genes (Fu, 1998, 2000; Harris et al., 1998; Arnold et al., 2007; Mayer & Pavlicev, 2007; Hipsley et al., 2009; Pavlicev & Mayer, 2009; Pyron et al., 2013). The results from these studies have been only partially successful, with a few sister relationships between genera being consistently recovered through the majority of these studies, as exemplified by the monotypic genera *Teira* and *Scelarcis*, or *Lacerta* and *Timon*. The position of many other genera, however, has been inconsistent between studies, as for instance the Asian genus *Takydromus* that has been inferred to be the sister of Lacertini (Harris et al., 1998; Pavlicev & Mayer, 2009) or has been placed within Lacertini but with low statistical support (e.g. Arnold et al., 2007; Mayer & Pavlicev, 2007). Other major uncertainty within Lacertini is the monophyly of the genus *Algyroides*, which has been questioned by Pavlicev & Mayer (2009). Apart from the intergeneric relationships within Lacertini, the basal splitting pattern has been, since the beginning of the molecular studies within this group, its main unresolved question. Because this basal polytomy might be due to insufficient data, since the majority of studies were performed with mitochondrial DNA, Mayer & Pavlicev (2007) and Pavlicev & Mayer (2009) performed phylogenetic analyses including nuclear sequence data and an increasing taxon sampling. Their results, though, generated no improvements in the basal resolution of the phylogenetic tree and therefore discarded the hypothesis of a soft polytomy due to a methodological artefact. However, an alternative explanation for the lack of resolution in this last phylogenetic assessments may be the extremely slow-evolving genes (*cmos* and *rag*) applied, likely containing not enough information to recover the splitting pattern within Lacertini. In contrast, a study from Pyron et al. (2013), focusing on relationships between 4161 Squamata taxa, appears to have successfully resolved the internal branching within Lacertini recovering high statistical support from internal to tip nodes. Yet, in this study, the authors used mainly the same two slow evolving nuclear markers employed by Pavlicev & Mayer (2009) and mitochondrial information from previous studies. Therefore, the current state of knowledge on the evolutionary history of Lacertini has two contrasting phylogenetic hypotheses drawn from concatenated dataset using mostly the same DNA sequences from mitochondrial and slow evolving nuclear markers. As discussed in the molecular tools section, when single-gene trees are significantly different and incongruent, as it seems the case for Lacertini, the concatenation approach might lead to statistically inconsistent estimation of phylogenies, providing supported but incorrect phylogenetic trees. In cases such as Lacertini, the species tree based on the multispecies coalescent approach seems promising, because it allows the reconciliation of a set of gene trees embedded in a shared species phylogeny. The species tree

method, paired with the addition of fast evolving nuclear genes and a comprehensive taxon sampling, might be the appropriate tool to test the reliability of previous phylogenies based on mainly mitochondrial dataset and to dissect the very different phylogenetic estimates of Lacertini based on the concatenation approach.

THE GENUS *PSAMMODROMUS*

The genus *Psammodromus*, together with *Gallotia*, form the sub-family Gallotiinae, the sister taxon to all other extant members of the Lacertidae family. *Psammodromus* has six species, three in the Iberian Peninsula (*P. hispanicus*, *P. occidentalis* and *P. edwarsianus*), two in North Africa (*P. microdactylus* and *P. blanci*) and one distributed both in Iberia and North Africa (*P. algirus*). The phylogenetic inference attention given to this genus has been, to some extent, mainly divided between the Iberian species (Fitze et al., 2011, 2012) and *P. algirus* (Carranza et al., 2006; Verdú-Ricoy et al., 2010), for which there is a relatively good phylogenetic knowledge. However, the evolutionary history of *Psammodromus* as a whole is fairly unknown. Previous studies have based the phylogenetic inference mainly on concatenated mitochondrial genes, with very few data from slow-evolving nuclear markers (Fitze et al., 2011, 2012). Moreover, previous assessments have focused on the Iberian species and *P. algirus*, whereas little is known about the North African species. Indeed, the African endemic *P. microdactylus* has never been included before in any molecular study. Besides its phylogenetic position, biological and ecological traits of this species are relatively unknown, due to the elusive nature of these animals (in den Bosch, 2005). Similarly, the other African endemic species, *P. blanci*, has only three mitochondrial sequences of the same individual on GenBank, and has been included in very few phylogenetic studies.

A good estimate of the phylogeny of this genus, together with a robust estimate of the diversification time is not only valuable in its own right, to comprehend the evolutionary history of this genus, but also to provide important insights on the association between diversification and biogeographic processes. Indeed, *Psammodromus*, comprising taxa demonstrating both intra and interspecific genetic distributions on the European and African margins of the Strait of Gibraltar would provide an excellent model for a comprehensive vision of the dispersal and vicariance events between continents across the Strait of Gibraltar.

Assessing the evolutionary history of *Psammodromus* would require a comprehensive molecular dataset for all the species of this genus, including both

mitochondrial and fast evolving nuclear genes, as well as a coalescent species tree and estimation of divergence times of the main lineages and their ancestral geographical distribution.

THE GENUS *OMANOSAURA*

The genus *Omanosaura* includes the only endemic lacertid lizards from the Hajar Mountains in the Arabian Peninsula. It comprises only two species, morphologically very distinct since *O. jayakari* is a relatively large and robust lizard and *O. cyanura* is a small and delicate animal, with a distinctive electric blue tail (Fig. 1.15). Little is known about these species since they figure in very few studies, mostly in the phylogenetic analyses on Eremiadini, the lacertid tribe to which they belong (Harris et al., 1998; Harris, 1999; Arnold et al., 2007; Mayer & Pavlicev, 2007; Greenbaum et al., 2011), and on the supermatrices build to infer relationships of squamates (Pyron et al., 2013). Sequences in GenBank, however, are only available for one specimen of *O. cyanura* (three mitochondrial genes - Harris et al., 1998) and two specimens of *O. jayakari*, with one specimen sequenced for three mitochondrial genes (Harris et al., 1998) and a different specimen sequenced for two slow-evolving nuclear genes (Mayer & Pavlicev, 2007). Apart from the fact that these are sister species and are genetically distinct from other Eremiadini genera, nothing is known about genetic diversity within each member of *Omanosaura*. This may be due to the fact that these species are distributed in an area with difficult access and that, until very recently, received poor scientific attention. In the case of *Omanosaura*, the addition of multilocus analyses, including mitochondrial and nuclear genes, would be particularly useful to obtain a strong estimation of the phylogeny of this genus. It would also be of particular interest for the assessment of cryptic diversity, as has been discovered in other taxa from this region, including the gecko genera *Ptyodactylus* (Metallinou et al., 2015; Simó-Riudalbas et al., 2017) and *Asaccus* (Carranza et al., 2016). Concerning the existence of cryptic taxa, a multilocus dataset would provide a conclusive assessment of reciprocal monophyly between lineages and whether these are sorted at both mitochondrial and nuclear genes.

Figure 1.15. *Omanosaura jayakari* and *O. cyanura*. Photos from Salvador Carranza and Roberto Sindaco, respectively.

THE COLUBRIDAE FAMILY

The family Colubridae Oppel, 1811 includes the highest number of snake's species than any other Serpentes family, comprising roughly a third of extant snake species (currently 1876 described species (Uetz, P., Freed, P., & Jirí Hošek (eds), The Reptile Database, http://www.reptile-database.org/, accessed April 2017). Representatives of this family are found in all continents except Antarctica. In the Squamata tree, according to the most recent estimate (Zheng & Wiens, 2016) the Colubridae is sister to the family Lamprophiidae and this is a relatively recent clade within Serpentes, with origin in the Eocene, approximately 50 Ma (Fig. 1.6 & 1.16).

Similarly to the Lacertidae family, the number of phylogenetic studies in colubrids has been increasing over the last two decades (Fig. 1.17). Such increase is due to the challenging phylogenetic inference of colubrids and, up to date, several different studies, applying different phylogenetic tools, have tried to infer the relationships within this family (e.g. Nagy et al., 2004; Lawson et al., 2005; Vidal et al., 2007; Pyron et al., 2011, 2013; Figueroa et al., 2016; Zheng & Wiens, 2016). These studies have resulted in several *incertae sedis* groups and taxonomical revisions and species descriptions being suggested. After the later study with supermatrix and phylogenomics tools the current taxonomy of the Colubridae family includes seven sub-families: Calamariinae, Colubrinae, Grayiinae, Sibynophiinae, Dipsadinae, Pseudoxenodontinae and Natricinae (Fig. 1.16).

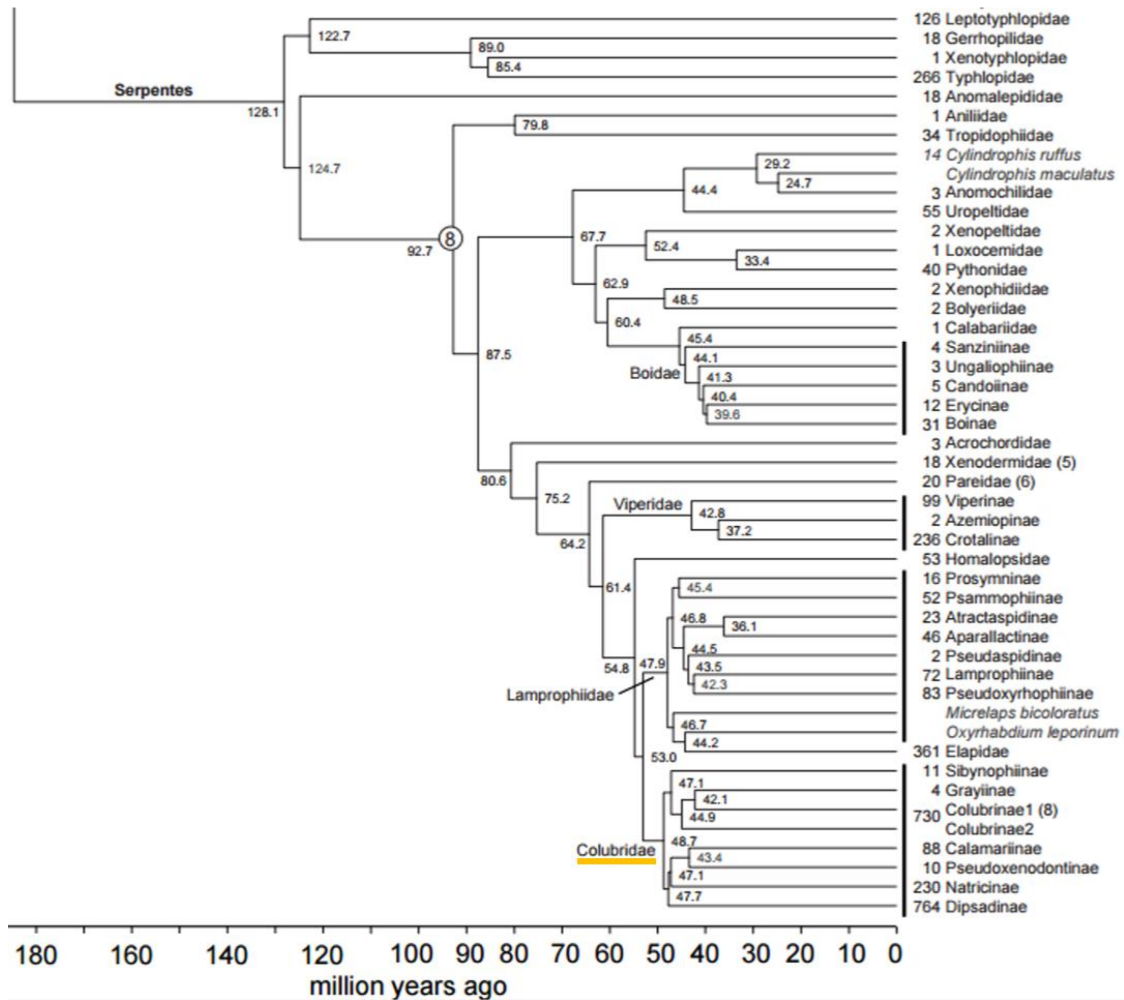


Figure 1.16. Phylogenetic relationships of the group Serpentes. Numbers next to sub-family names on the right indicate the number of species in each subfamily. Retrieved from The Reptile Database (http://www.reptile-database.org), and adapted from Zheng & Wiens (2016).

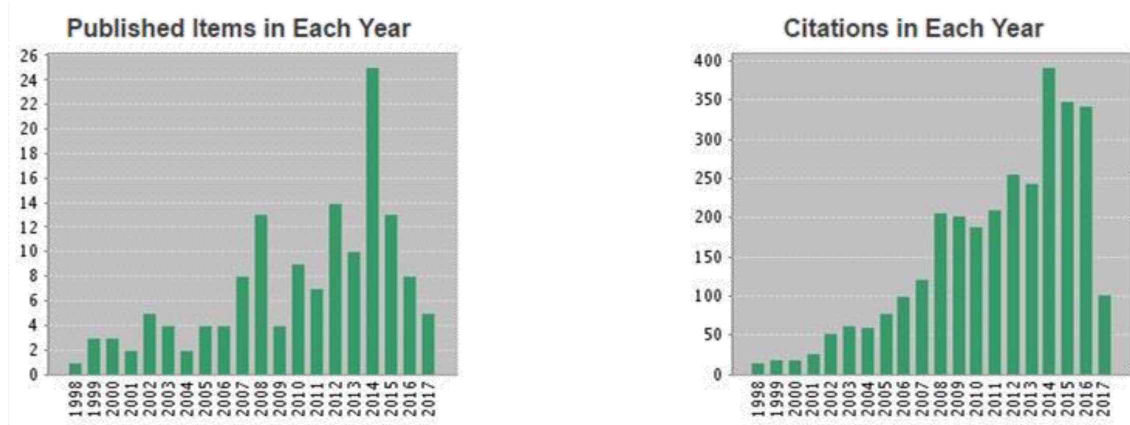


Figure 1.17. Number of publications and citation of phylogenetic studies with Colubridae. From the Web of Science (June 2017)

THE GENUS *ZAMENIS*

The Colubrinae sub-family includes nearly 100 genera and 680 species (Uetz, P., Freed, P., & Jirí Hošek (eds), The Reptile Database, http://www.reptile-database.org/, accessed April 2017) and includes members of ratsnakes, milksnakes and kingsnakes, among others. The ratsnake group originally comprised only one genus mostly distributed in the Northern Hemisphere, *Elaphe*. This genus was divided in Old and New World ratsnakes by phylogenetic inference based on mitochondrial DNA (Utiger et al., 2002) and relationships within both groups were further investigated, leading to the division of *Elaphe* to several different genera (Helfenberger, 2001; Lenk et al., 2001; Burbrink & Lawson, 2007). Currently, the European branch of the Old World ratsnakes includes the genera *Elaphe*, *Coronella*, *Rhinechis* and *Zamenis*. The ratsnake genus *Zamenis* includes five species with Palaearctic distribution: *Z. lineatus*, *Z. longissimus*, *Z. situla*, *Z. persica* and *Z. hohenackeri*. The genus *Rhinechis* is monotypic, with the species *Rhinechis scalaris*. The closest relative to these ratsnakes' genera is the Palearctic smooth snake genus *Coronella* (Nagy et al., 2004; Burbrink & Lawson, 2007).

The phylogenetic inference of *Zamenis* species and other ratsnakes has provided two different estimates. On one hand, the first phylogenetic studies focusing on relationships between ratsnakes and applying mostly mitochondrial DNA and one slow evolving nuclear gene, *cmos*, have resulted in the monophyly of *Zamenis*, sister to *Rhinechis* (Helfenberger, 2001; Lenk et al., 2001; Utiger et al., 2002; Burbrink & Lawson, 2007). On the other hand, later studies focusing on the relationships of colubrids or squamates in general, based on mostly the same DNA sequence data of *Zamenis*, have resulted in a paraphyletic *Zamenis*, with *Rhinechis* deeply nested within the same clade with high statistical support (Pyron et al., 2011, 2013; Figueroa et al., 2016; Zheng & Wiens, 2016).

Besides the question of the monophyly of *Zamenis*, the intra-generic relationships are differently recovered across most of the previous studies (Fig. 1.18). However, as a general pattern, the eastern species *Z. hohenackeri* and *Z. persicus* seem to be the first to diverge within the genus and *Z. situla*, *Z. lineatus* and *Z. longissimus* form a derived polytomy. An accurate estimate of the speciation events within *Zamenis* would allow to estimate the biogeographical scenario of the evolution of this genus. Since the oldest species (*Z. hohenackeri* and *Z. persicus*) are distributed in the Southwestern Asia areas of Anatolia/Caucasus, while the most recent species (*Z. situla*, *Z. lineatus* and *Z. longissimus*) have a more western Palaearctic distribution, an oriental origin of European

ratsnakes has been suggested by previous studies (Utiger et al., 2002; Burbrink & Lawson, 2007).

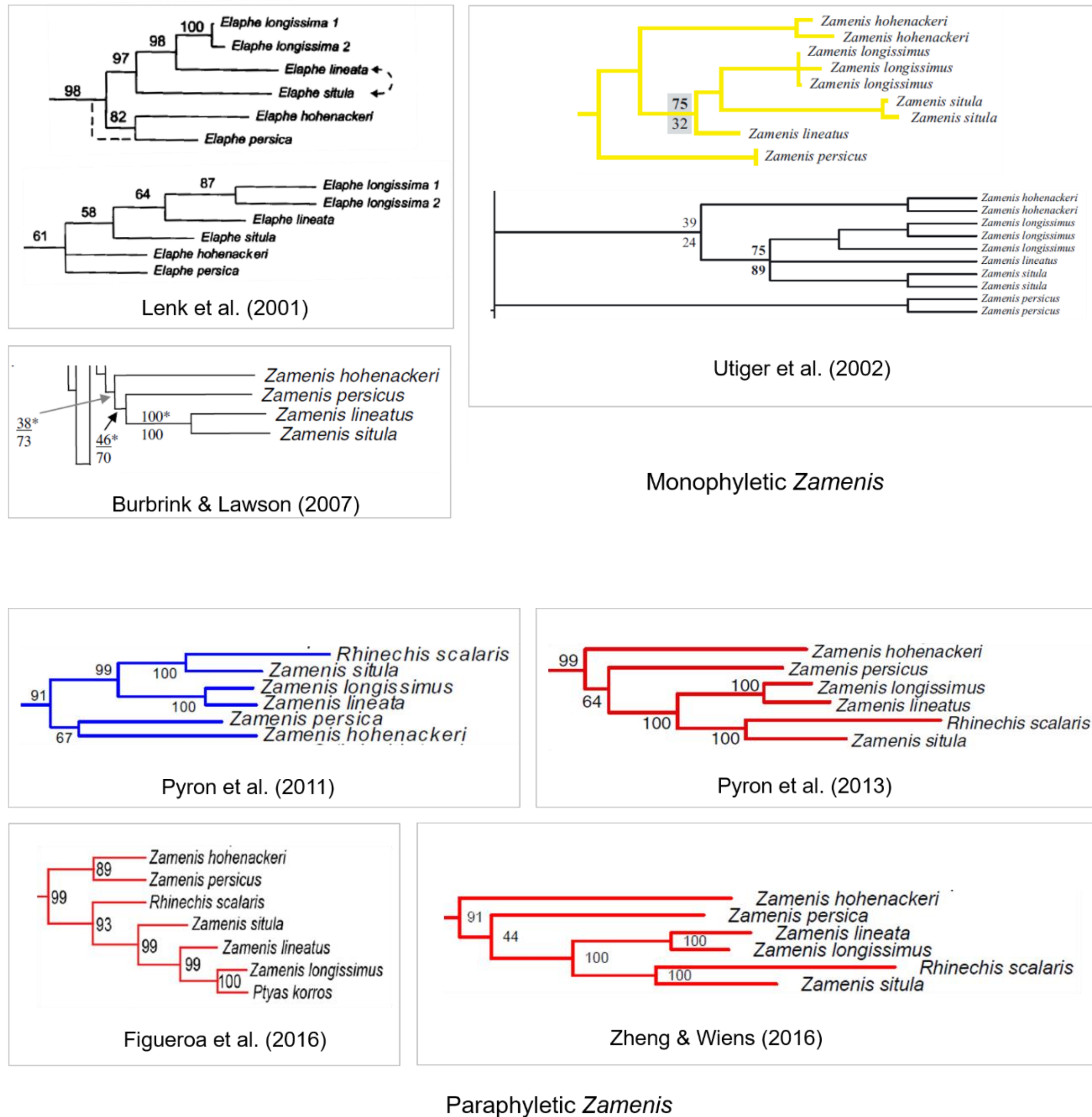


Figure 1.18. Phylogenetic inference of *Zamenis* species.

4. OBJECTIVES AND THESIS FRAMEWORK

The main aim of this dissertation was to investigate how the addition of novel molecular data, in particular fast evolving nuclear markers, and the use of recently developed species tree phylogenetic methods could improve the estimate of phylogenetic relationships within selected squamates groups covering different systematic levels from tribes, to genera and to species.

This thesis is organised in six Chapters, with **Chapter 1** as a General Introduction (the present chapter) and includes four chapters with scientific manuscripts developed with the purpose of answering different methodological, evolutionary and biogeographic questions based on a robust estimate of the phylogeny of four different organism groups through the application of fast-evolving nuclear data and the species tree based on the multispecies coalescent approach.

Firstly, the work developed in **Chapter 2** specifically aimed to test if the addition of fast evolving nuclear genes, combined with the species tree coalescent approach could resolve internal branching from a basal polytomy within the lizard tribe Lacertini. In addition, it aimed to provide resolution of relationships between genera and assess the monophyly of some genera, as well as discriminate between contrasting phylogenetic hypotheses from previous studies using concatenation methods. This chapter comprises one published paper entitled “Evaluating the phylogenetic signal limit from mitogenomes, slow evolving nuclear genes, and the concatenation approach. New insights into the Lacertini radiation using fast evolving nuclear genes and species trees” that has been published in the journal *Molecular Phylogenetics and Evolution*.

The work developed in **Chapter 3** aimed to test if the application of fast evolving nuclear genes, the species tree coalescent approach and a comprehensive taxon sampling could solve the question on whether the genus *Zamenis* is monophyletic and the branching patterns within this genus, as well as infer its biogeographical pattern across its distribution range in Europe. This chapter includes one paper currently in preparation entitled “Evolution, biogeography and systematics of the Western Palaearctic ratsnakes *Zamenis*, with the designation of *Zamenis scalaris* comb. nov.”.

The main aim of the work developed in **Chapter 4** was to frame the phylogeographical history of *Psammodromus,* by conducting analyses with all *Psammodromus* species and estimating a time-calibrated species tree with fast evolving nuclear markers to understand the biogeographic dynamics of this genus across the Strait of Gibraltar. This chapter is composed by the paper entitled “Biogeographic

crossroad across the Pillars of Hercules: evolutionary history of *Psammodromus* lizards in space and time", that has been published in the *Journal of Biogeography*.

The study from **Chapter 5** aimed to assess levels of inter- and intra- specific genetic diversity in the poorly studied genus *Omanosaura*, by performing phylogenetic inference with both mitochondrial and nuclear markers and with a comprehensive sampling of individuals of the two species from this genus. This work highlights the importance of a thorough phylogenetic inference in areas that are historically poorly studied but that can harbour high levels of cryptic diversity. This chapter comprises one paper entitled "Hidden in the Arabian mountains: multilocus phylogeny reveals cryptic diversity in the endemic *Omanosaura* lizards", published by the *Journal of Zoological Systematics and Evolutionary Research*.

Finally, the **Chapter 6** provides a General Discussion on all the subjects addressed in the previous chapters, emphasizing the main discoveries and general achievements of this dissertation, as well as implications and suggestions for further research that arose from this work.

REFERENCES

Abdala C.S., Semhan R.V., Laspiur A., & Acosta J.L. (2017) Rediscovery of *Liolaemus rabinoi* (Iguania: Liolaemidae) after 35 years: Redescription, biological and phylogentic information, and conservation challenges. *Salamandra*, **53**, ISSN 0036-3375.

Albert E.M., San Mauro D., García-París M., Rüber L., & Zardoya R. (2009) Effect of taxon sampling on recovering the phylogeny of squamate reptiles based on complete mitochondrial genome and nuclear gene sequence data. *Gene*, **441**, 12–21.

Alfaro M.E., Santini F., Brock C., Alamillo H., Dornburg A., Rabosky D.L., Carnevale G., & Harmon L.J. (2009) Nine exceptional radiations plus high turnover explain species diversity in jawed vertebrates. *Proceedings of the National Academy of Sciences*, **106**, 13410–13414.

Arnold E.N. (1989) Towards a phylogeny and biogegraphy of the Lacertidae: relationships within an Old-World family of lizards derived from morphology. *Bulletin of the British Museum (Natural History), Zoology*, **55**, 209–257.

Arnold E.N., Arribas O., & Carranza S. (2007) Systematics of the Palaearctic and Oriental lizard tribe Lacertini (Squamata: Lacertidae: Lacertinae), with descriptions of eight new genera. Magnolia Press. *Zootaxa*, **1430**, 1–86.

Augé M.L. (2005) *Évolution des lézards du Paléogène en Europe. Mémoires du Muséum national d'histoire naturelle, Tome 192.* Muséum National d'Histoire Naturelle, Paris.

Auliya M., Altherr S., Ariano-Sanchez D., Baard E.H., Brown C., Brown R.M., Cantu J.-C., Gentile G., Gildenhuys P., Henningheim E., Hintzmann J., Kanari K., Krvavac M., Lettink M., Lippert J., Luiselli L., Nilson G., Nguyen T.Q., Nijman V., Parham J.F., Pasachnik S.A., Pedrono M., Rauhaus A., Córdova D.R., Sanchez M.-E., Schepp U., van Schingen M., Schneeweiss N., Segniagbeto G.H., Somaweera R., Sy E.Y., Türkozan O., Vinke S., Vinke T., Vyas R., Williamson S., & Ziegler T. (2016) Trade in live reptiles, its impact on wild populations, and the role of the European market. *Biological Conservation*, **204**, 103–119.

Ballard J.W.O. & Whitlock M.C. (2004) The incomplete natural history of mitochondria. *Molecular Ecology*, **13**, 729–744.

Barker F.K., Cibois A., Schikler P., Feinstein J., & Cracraft J. (2004) Phylogeny and diversification of the largest avian radiation. *Proceedings of the National Academy of Sciences*, **101**, 11040–11045.

Bayzid M.S. & Warnow T. (2012) Estimating optimal species trees from incomplete gene trees under deep coalescence. *Journal of Computational Biology*, **19**, 591–605.

Bazin E., Glémin S., & Galtier N. (2006) Population size does not influence mitochondrial genetic diversity in animals. *Science*, **312**, 570–572.

Bellard C., Cassey P., & Blackburn T.M. (2016) Alien species as a driver of recent extinctions. *Biology Letters*, **12**, 20150623.

Bensasson D., Zhang D.X., Hartl D.L., & Hewitt G.M. (2001) Mitochondrial pseudogenes: Evolution's misplaced witnesses. *Trends in Ecology and Evolution*, **16**, 314–321.

Bininda-Emonds O.R.P. (2004) The evolution of supertrees. *Trends in Ecology and Evolution*, **19**, 315–322.

Blackburn D.G. (2006) Squamate reptiles as model organisms for the evolution of viviparity. *Herpetological Monographs*, **20**, 131–146.

Blakey R.C. (2008) Gondwana paleogeography from assembly to breakup - A 500 m.y. odyssey. *Geological Society of America Special Papers*, **441**, 1–28.

Böhm M., Collen B., Baillie J.E.M., Bowles P., Chanson J., Cox N., Hammerson G., Hoffmann M., Livingstone S.R., Ram M., Rhodin A.G.J., Stuart S.N., van Dijk P.P., Young B.E., Afuang L.E., Aghasyan A., García A., Aguilar C., Ajtic R., Akarsu F., Alencar L.R. V, Allison A., Ananjeva N., Anderson S., Andrén C., Ariano-Sánchez D., Arredondo J.C., Auliya M., Austin C.C., Avci A., Baker P.J., Barreto-Lima A.F., Barrio-Amorós C.L., Basu D., Bates M.F., Batistella A., Bauer A., Bennett D., Böhme W., Broadley D., Brown R., Burgess J., Captain A., Carreira S., Castañeda M. del R., Castro F., Catenazzi A., Cedeño-Vázquez J.R., Chapple D.G., Cheylan M., Cisneros-Heredia D.F., Cogalniceanu D., Cogger H., Corti C., Costa G.C., Couper P.J., Courtney T., Crnobrnja-Isailovic J., Crochet P.A., Crother B., Cruz F., Daltry J.C., Daniels R.J.R., Das I., de Silva A., Diesmos A.C., Dirksen L., Doan T.M., Dodd C.K., Doody J.S., Dorcas M.E., Duarte de Barros Filho J., Egan V.T., El Mouden E.H., Embert D., Espinoza R.E., Fallabrino A., Feng X., Feng Z.J., Fitzgerald L., Flores-Villela O., França F.G.R., Frost D., Gadsden H., Gamble T., Ganesh S.R., Garcia M.A., García-Pérez J.E., Gatus J., Gaulke M., Geniez P., Georges A., Gerlach J., Goldberg S., Gonzalez J.C.T., Gower D.J., Grant T.,

Greenbaum E., Grieco C., Guo P., Hamilton A.M., Hare K., Hedges S.B., Heideman N., Hilton-Taylor C., Hitchmough R., Hollingsworth B., Hutchinson M., Ineich I., Iverson J., Jaksic F.M., Jenkins R., Joger U., Jose R., Kaska Y., Kaya U., Keogh J.S., Köhler G., Kuchling G., Kumlutaş Y., Kwet A., La Marca E., Lamar W., Lane A., Lardner B., Latta C., Latta G., Lau M., Lavin P., Lawson D., LeBreton M., Lehr E., Limpus D., Lipczynski N., Lobo A.S., López-Luna M.A., Luiselli L., Lukoschek V., Lundberg M., Lymberakis P., Macey R., Magnusson W.E., Mahler D.L., Malhotra A., Mariaux J., Maritz B., Marques O.A. V, Márquez R., Martins M., Masterson G., Mateo J.A., Mathew R., Mathews N., Mayer G., McCranie J.R., Measey G.J., Mendoza-Quijano F., Menegon M., Métrailler S., Milton D.A., Montgomery C., Morato S.A.A., Mott T., Muñoz-Alonso A., Murphy J., Nguyen T.Q., Nilson G., Nogueira C., Núñez H., Orlov N., Ota H., Ottenwalder J., Papenfuss T., Pasachnik S., Passos P., Pauwels O.S.G., Pérez-Buitrago N., Pérez-Mellado V., Pianka E.R., Pleguezuelos J., Pollock C., Ponce-Campos P., Powell R., Pupin F., Quintero Díaz G.E., Radder R., Ramer J., Rasmussen A.R., Raxworthy C., Reynolds R., Richman N., Rico E.L., Riservato E., Rivas G., da Rocha P.L.B., Rödel M.O., Rodríguez Schettino L., Roosenburg W.M., Ross J.P., Sadek R., Sanders K., Santos-Barrera G., Schleich H.H., Schmidt B.R., Schmitz A., Sharifi M., Shea G., Shi H.T., Shine R., Sindaco R., Slimani T., Somaweera R., Spawls S., Stafford P., Stuebing R., Sweet S., Sy E., Temple H.J., Tognelli M.F., Tolley K., Tolson P.J., Tuniyev B., Tuniyev S., üzüm N., van Buurt G., Van Sluys M., Velasco A., Vences M., Veselý M., Vinke S., Vinke T., Vogel G., Vogrin M., Vogt R.C., Wearn O.R., Werner Y.L., Whiting M.J., Wiewandt T., Wilkinson J., Wilson B., Wren S., Zamin T., Zhou K., & Zug G. (2013) The conservation status of the world's reptiles. *Biological Conservation*, **157**, 372–385.

Bonis N.R. & Kürschner W.M. (2012) Vegetation history, diversity patterns, and climate change across the Triassic/Jurassic boundary. *Paleobiology*, **38**, 240–264.

Brito P.H. & Edwards S. V (2009) Multilocus phylogeography and phylogenetics using sequence-based markers. *Genetica*, **135**, 439–455.

Burbrink F.T. & Lawson R. (2007) How and when did Old World ratsnakes disperse into the New World? *Molecular Phylogenetics and Evolution*, **43**, 173–189.

Caldwell M.W. (1999) Squamate phylogeny and the relationships of snakes and mosasauroids. *Zoological Journal of the Linnean Society*, **125**, 115–147.

Calsbeek R. & Smith T.B. (2003) Ocean currents mediate evolution in island lizards. *Nature*, **426**, 552–555.

Camargo A., Avila L.J., Morando M., & Sites J.W. (2012) Accuracy and precision of species trees: effects of locus, individual, and base pair sampling on inference of species trees in lizards of the *Liolaemus darwinii* group (Squamata, Liolaemidae). *Systematic Biology*, **61**, 272–288.

Camargo A., Sinervo B., & Sites J.W. (2010) Lizards as model organisms for linking phylogeographic and speciation studies. *Molecular Ecology*, **19**, 3250–3270.

Camp C.L. (1923) Classification of the lizards. *Bulletin of the American Museum of Natural History*, **48**, 289–480.

Carranza S., Arnold E.N., Mateo J.A., & López-Jurado L.F. (2000) Long-distance colonization and radiation in gekkonid lizards, *Tarentola* (Reptilia: Gekkonidae), revealed by mitochondrial DNA sequences. *Proceedings of the Royal Society B Biological sciences*, **267**, 637–649.

Carranza S., Harris D.J., Arnold E.N., Batista V., & Gonzalez de la Vega J.P. (2006) Phylogeography of the lacertid lizard, *Psammodromus algirus*, in Iberia and across the Strait of Gibraltar. *Journal of Biogeography*, **33**, 1279–1288.

Carranza S., Simó-Riudalbas M., Jayasinghe S., Wilms T., & Els J. (2016) Microendemicity in the northern Hajar Mountains of Oman and the United Arab Emirates with the description of two new species of geckos of the genus *Asaccus* (Squamata: Phyllodactylidae). *PeerJ*, **4**, e2371.

Carroll R.L. (1975) Permo-Triassic “lizards” from the Karroo. *Palaeontologia Africana*, **18**, 71–87.

Castilla A.M., Van Damme R., & Bauwens D. (1999) Field body temperatures, mechanisms of thermoregulation and evolution of thermal characteristics in lacertid lizards. *Natura Croatica*, **8**, 253–274.

Ceballos G., Ehrlich P.R., Barnosky A.D., Garcia A., Pringle R.M., & Palmer T.M. (2015) Accelerated modern human-induced species losses: Entering the sixth mass extinction. *Science Advances*, **1**, e1400253.

Čerňanský A., Klembara J., & Smith K.T. (2016) Fossil lizard from central Europe resolves the origin of large body size and herbivory in giant Canary Island lacertids. *Zoological Journal of the Linnean Society*, **176**, 861–877.

Chiari Y., Cahais V., Galtier N., & Delsuc F. (2012) Phylogenomic analyses support the position of turtles as the sister group of birds and crocodiles (Archosauria). *BMC Biology*, **10**, 65.

Cibois A., Thibault J.C., Bonillo C., Filardi C.E., & Pasquet E. (2017) Phylogeny and biogeography of the imperial pigeons (Aves: Columbidae) in the Pacific Ocean. *Molecular Phylogenetics and Evolution*, **110**, 19–26.

Colli G.R., Fenker J., Tedeschi L.G., Barreto-Lima A.F., Mott T., & Ribeiro S.L.B. (2016) In the depths of obscurity: Knowledge gaps and extinction risk of Brazilian worm lizards (Squamata, Amphisbaenidae). *Biological Conservation*, **204**, 51–62.

Conrad J.L. (2008) Phylogeny and systematics of Squamata (Reptilia) based on morphology. *Bulletin of the American Museum of Natural History*, **310**, 1–182.

Cooper W.E., Pyron R.A., & Garland T. (2014) Island tameness: living on islands reduces flight initiation distance. *Proceedings of the Royal Society B Biological sciences* **281**, 20133019.

Corl A. & Ellegren H. (2013) Sampling strategies for species trees: the effects on phylogenetic inference of the number of genes, number of individuals, and whether loci are mitochondrial, sex-linked, or autosomal. *Molecular Phylogenetics and Evolution*, **67**, 358–366.

Crochet P.A., Chen J.Z., Pons J.-M., Lebreton J.D., Hebert P.D.N., & Bonhomme F. (2003) Genetic differentiation at nuclear and mitochondrial loci among large white-headed gulls: sex-biased interspecific gene flow? *Evolution*, **57**, 2865–2878.

Degnan J.H. & Rosenberg N.A. (2006) Discordance of species trees with their most likely gene trees. *PLoS Genetics*, **2**, e68.

Degnan J.H. & Rosenberg N.A. (2009) Gene tree discordance, phylogenetic inference and the multispecies coalescent. *Trends in Ecology & Evolution*, **24**, 332–340.

Douglas D.A., Janke A., & Arnason U. (2006) A mitogenomic study on the phylogenetic position of snakes. *Zoologica Scripta*, **35**, 545–558.

Drummond A.J. & Rambaut A. (2007) BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology*, **7**, 214.

Edwards S. & Bensch S. (2009) Looking forwards or looking backwards in avian phylogeography? A comment on Zink and Barrowclough 2008. *Molecular Ecology*, **18**, 2930–2933.

Edwards S.V (2009) Is a new and general theory of molecular systematics emerging? *Evolution* **63**, 1–19.

Edwards S.V, Liu L., & Pearl D.K. (2007) High-resolution species trees without

concatenation. *Proceedings of the National Academy of Sciences of the United States of America*, **104**, 5936–5941.

Espregueira Themudo G., Wielstra B., & Arntzen J.W. (2009) Multiple nuclear and mitochondrial genes resolve the branching order of a rapid radiation of crested newts (*Triturus*, Salamandridae). *Molecular Phylogenetics and Evolution*, **52**, 321–328.

Estes R., de Queiroz K., & Gauthier J. (1988) Phylogenetic relationships within Squamata. *Phylogenetic elationships of the lizard families* (ed. by R. Estes and G. Pregill), pp. 119–281. Stanford University Press, Palo Alto.

Evans S.E. (2003) At the feet of the dinosaurs: the early history and radiation of lizards. *Biological reviews of the Cambridge Philosophical Society*, **78**, 513–551.

Evans S.E. & Barbadillo L.J. (1998) An unusual lizard (Reptilia: Squamata) from the Early Cretaceous of Las Hoyas, Spain. *Zoological Journal of the Linnean Society*, **124**, 235–265.

Evans S.E. & Barbadillo L.J. (1999) A shortlimbed lizard from the Lower Cretaceous of Spain. *Special Papers in Palaeontology*, **60**, 73–85.

Evans S.E., Prasad G.V.R., & Manhas B.K. (2002) Fossil lizards from the Jurassic Kota Formation of India. *Journal of Vertebrate Paleontology*, **22**, 299–312.

Evans S.E. & Wang Y. (2005) The Early Cretaceous lizard Dalinghosaurus from China. *Acta Palaeontologica Polonica*, **50**, 725–742.

Figueroa A., McKelvy A.D., Grismer L.L., Bell C.D., & Lailvaux S.P. (2016) A species-level phylogeny of extant snakes with description of a new colubrid subfamily and genus. *PLoS ONE*, **11**, e0161070.

Fitze P.S., Gonzalez-Jimena V., San-Jose L.M., Mauro D.S., & Zardoya R. (2012) A new species of sand racer, *Psammodromus* (Squamata: Lacertidae), from the Western Iberian Peninsula. *Zootaxa*, **52**, 41–52.

Fitze P.S., Gonzalez-Jimena V., San-Jose L.M., San Mauro D., Aragón P., Suarez T., & Zardoya R. (2011) Integrative analyses of speciation and divergence in *Psammodromus hispanicus* (Squamata: Lacertidae). *BMC Evolutionary Biology*, **11**, 347.

Fontanella F.M., Olave M., Avila L.J., Sites Jr J.W., & Morando M. (2012) Molecular dating and diversification of the South American lizard genus *Liolaemus* (subgenus

Eulaemus) based on nuclear and mitochondrial DNA sequences. *Zoological Journal of the Linnean Society*, **164**, 825–835.

Fu J. (1998) Toward the phylogeny of the family Lacertidae: implications from mitochondrial DNA 12S and 16S gene sequences (Reptilia: Squamata). *Molecular Phylogenetics and Evolution*, **9**, 118–130.

Fu J. (2000) Toward the phylogeny of the family Lacertidae - Why 4708 base pairs of mtDNA sequences cannot draw the picture. *Biological Journal of the Linnean Society*, **71**, 203–217.

Gadagkar S.R., Rosenberg M.S., & Kumar S. (2005) Inferring species phylogenies from multiple genes: Concatenated sequence tree versus consensus gene tree. *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution*, **304**, 64–74.

Garland T. & Díaz-Uriarte R. (1999) Polytomies and phylogenetically independent contrasts: examination of the bounded degrees of freedom approach. *Systematic Biology*, **48**, 547–558.

Geniez P., Cluchier A., Sá-sousa P., Guillaume C.P., & Crochet P. (2007) Systematics of the *Podarcis hispanicus*-complex (Sauria, Lacertidae) I: Redefinition, morphology and distribution of the nominotypical taxon. *Herpetological Journal*, **17**, 69–80.

Godinho R., Crespo E.G., & Ferrand N. (2008) The limits of mtDNA phylogeography: Complex patterns of population history in a highly structured Iberian lizard are only revealed by the use of nuclear markers. *Molecular Ecology*, **17**, 4670–4683.

Gonçalves H., Martínez-Solano I., Ferrand N., & García-París M. (2007) Conflicting phylogenetic signal of nuclear vs mitochondrial DNA markers in midwife toads (Anura, Discoglossidae, *Alytes*): deep coalescence or ancestral hybridization? *Molecular Phylogenetics and Evolution*, **44**, 494–500.

Graybeal A. (1994) Evaluating the phylogenetic utility of genes: a search for genes informative about deep divergences among vertebrates. *Systematic Biology*, **43**, 174–193.

Greenbaum E., Villanueva C.O., Kusamba C., Aristote M.M., & Branch W.R. (2011) A molecular phylogeny of Equatorial African Lacertidae, with the description of a new genus and species from eastern Democratic Republic of the Congo. *Zoological Journal of the Linnean Society*, **163**, 913–942.

Harris D.J. (1999) Molecular systematics and evolution of lacertid lizards. *Natura*

Croatica, **8**, 161–180.

Harris D.J., Arnold E.N., & Thomas R.H. (1998) Relationships of lacertid lizards (Reptilia: Lacertidae) estimated from mitochondrial DNA sequences and morphology. *Proceedings of the Royal Society B Biological sciences*, **265**, 1939–1948.

Hedges S.B. & Thomas R. (2001) At the lower size limit in amniote vertebrates: A new diminutive lizard from the West Indies. *Caribbean Journal of Science*, **37**, 168–173.

Heled J. & Drummond A.J. (2010) Bayesian inference of species trees from multilocus data. *Molecular Biology and Evolution*, **27**, 570–580.

Helfenberger N. (2001) Phylogenetic relationships of Old World ratsnakes based on visceral organ topography, osteology, and allozyme variation. *Russian Journal of Herpetology*, **8**, 1–64.

Herrel A., Meyers J.J., Nishikawa K.C., & De Vree F. (2001) The evolution of feeding motor patterns in lizards: Modulatory complexity and possible constraints. *American Zoologist*, **41**, 1311–1320.

Hey J. & Machado C.A. (2003) The study of structured populations - new hope for a difficult and divided science. *Nature Reviews Genetics*, **4**, 535–543.

Hipsley C.A., Himmelmann L., Metzler D., & Müller J. (2009) Integration of Bayesian molecular clock methods and fossil-based soft bounds reveals early Cenozoic origin of African lacertid lizards. *BMC Evolutionary Biology*, **9**, 151.

Hollingsworth P.R., Simons A.M., Fordyce J.A., & Hulsey C. (2013) Explosive diversification following a benthic to pelagic shift in freshwater fishes. *BMC Evolutionary Biology*, **13**, 272.

Hovmöller R., Lacey Knowles L., & Kubatko L.S. (2013) Effects of missing data on species tree estimation under the coalescent. *Molecular Phylogenetics and Evolution*, **69**, 1057–1062.

Huelsenbeck J.P. (1991) When are fossils better than extant taxa in phylogenetic analysis? *Systemtic Zoology*, **40**, 458–469.

in den Bosch H.A.J. (2005) *Psammodromus microdactylus* (Boettger, 1881), a rare lizard species? *Podarcis*, **6**, 2–35.

Isaac N.J.B., Redding D.W., Meredith H.M., & Safi K. (2012) Phylogenetically-informed priorities for amphibian conservation. *PLoS ONE*, **7**, 1–8.

Isaac N.J.B., Turvey S.T., Collen B., Waterman C., & Baillie J.E.M. (2007) Mammals on the EDGE: Conservation priorities based on threat and phylogeny. *PLoS ONE*, **2**, e296.

Jarvis E.D., Mirarab S., Aberer A.J., Li B., Houde P., Li C., Ho S.Y.W., Faircloth B.C., Nabholz B., Howard J.T., Suh A., Weber C.C., da Fonseca R.R., Li J., Zhang F., Li H., Zhou L., Narula N., Liu L., Ganapathy G., Boussau B., Bayzid M.S., Zavidovych V., Subramanian S., Gabaldon T., Capella-Gutierrez S., Huerta-Cepas J., Rekepalli B., Munch K., Schierup M., Lindow B., Warren W.C., Ray D., Green R.E., Bruford M.W., Zhan X., Dixon A., Li S., Li N., Huang Y., Derryberry E.P., Bertelsen M.F., Sheldon F.H., Brumfield R.T., Mello C. V., Lovell P. V., Wirthlin M., Schneider M.P.C., Prosdocimi F., Samaniego J.A., Velazquez A.M. V., Alfaro-Nunez A., Campos P.F., Petersen B., Sicheritz-Ponten T., Pas A., Bailey T., Scofield P., Bunce M., Lambert D.M., Zhou Q., Perelman P., Driskell A.C., Shapiro B., Xiong Z., Zeng Y., Liu S., Li Z., Liu B., Wu K., Xiao J., Yinqi X., Zheng Q., Zhang Y., Yang H., Wang J., Smeds L., Rheindt F.E., Braun M., Fjeldsa J., Orlando L., Barker F.K., Jonsson K.A., Johnson W., Koepfli K.-P., O'Brien S., Haussler D., Ryder O.A., Rahbek C., Willerslev E., Graves G.R., Glenn T.C., McCormack J., Burt D., Ellegren H., Alstrom P., Edwards S. V., Stamatakis A., Mindell D.P., Cracraft J., Braun E.L., Warnow T., Jun W., Gilbert M.T.P., & Zhang G. (2014) Whole-genome analyses resolve early branches in the tree of life of modern birds. *Science*, **346**, 1320–1331.

Jetz W., Thomas G.H., Joy J.B., Hartmann K., & Mooers A.O. (2012) The global diversity of birds in space and time. *Nature*, **491**, 444–448.

Jetz W., Thomas G.H., Joy J.B., Redding D.W., Hartmann K., & Mooers A.O. (2014) Global distribution and conservation of evolutionary distinctness in birds. *Current Biology*, **24**, 919–930.

Jiang W., Chen S.-Y., Wang H., Li D.-Z., & Wiens J.J. (2014) Should genes with missing data be excluded from phylogenetic analyses? *Molecular Phylogenetics and Evolution*, **80**, 308–318.

Jones M.E., Anderson C., Hipsley C.A., Müller J., Evans S.E., & Schoch R.R. (2013) Integration of molecules and new fossils supports a Triassic origin for Lepidosauria (lizards, snakes, and tuatara). *BMC Evolutionary Biology*, **13**, 208.

Kaliontzopoulou A., Carretero M.A., & Llorente G.A. (2010a) Intraspecific ecomorphological variation: Linear and geometric morphometrics reveal habitat-related patterns within *Podarcis bocagei* wall lizards. *Journal of Evolutionary*

Biology, **23**, 1234–1244.

Kaliontzopoulou A., Carretero M.A., & Llorente G.A. (2010b) Sexual dimorphism in traits related to locomotion: Ontogenetic patterns of variation in *Podarcis* wall lizards. *Biological Journal of the Linnean Society*, **99**, 530–543.

Kaliontzopoulou A., Pinho C., Harris D.J., & Carretero M.A. (2011) When cryptic diversity blurs the picture: a cautionary tale from Iberian and North African *Podarcis* wall lizards. *Biological Journal of the Linnean Society*, **103**, 779–800.

Kasturiratne A., Wickremasinghe A.R., De Silva N., Gunawardena N.K., Pathmeswaran A., Premaratna R., Savioli L., Lalloo D.G., & De Silva H.J. (2008) The global burden of snakebite: A literature analysis and modelling based on regional estimates of envenoming and deaths. *PLoS Medicine*, **5**, 1591–1604.

Kearney M., Fujita M.K., & Ridenour J. (2009) Lost sex in reptiles: constraints and correlations. *Lost Sex* (ed. by I. Schön, K. Martens, and P. van Dijk), pp. 447–474. Springer, London.

Kluge A.G. (1989) A concern for evidence and a phylogenetic hypothesis of relationships among *Epicrates* (Boidae, Serpentes). *Systematic Zoology*, **38**, 7–25.

Kluge A.G. (2004) On total evidence: For the record. *Cladistics*, **20**, 205–207.

Kubatko L.S., Carstens B.C., & Knowles L.L. (2009) STEM: species tree estimation using maximum likelihood for gene trees under coalescence. *Bioinformatics (Oxford, England)*, **25**, 971–973.

Kubatko L.S. & Degnan J.H. (2007) Inconsistency of phylogenetic estimates from concatenated data under coalescence. *Systematic Biology*, **56**, 17–24.

Kuhn T.S., Mooers A., & Thomas G.H. (2011) A simple polytomy resolver for dated phylogenies. *Methods in Ecology and Evolution*, **2**, 427–436.

Kumar S. & Hedges S.B. (1998) A molecular timescale for vertebrate evolution. *Nature*, **392**, 917–920.

Kumazawa Y. (2007) Mitochondrial genomes from major lizard families suggest their phylogenetic relationships and ancient radiations. *Gene*, **388**, 19–26.

Lambert S.M., Reeder T.W., & Wiens J.J. (2015) When do species-tree and concatenated estimates disagree? An empirical analysis with higher-level scincid lizard phylogeny. *Molecular Phylogenetics and Evolution*, **82**, 146–155.

Lawson R., Slowinski J.B., Crother B.I., & Burbrink F.T. (2005) Phylogeny of the Colubroidea (Serpentes): New evidence from mitochondrial and nuclear genes. *Molecular Phylogenetics and Evolution*, **37**, 581–601.

Lee M.S.Y. (1998) Convergent evolution and character correlation in burrowing reptiles: towards a resolution of squamate relationships. *Biological Journal of the Linnean Society*, **65**, 369–453.

Lee M.S.Y. & Caldwell M.W. (2000) *Adriosaurus* and the affinities of mosasaurs, dolichosaurs, and snakes. *Journal of Paleontology*, **74**, 915–937.

Lemmon A.R., Brown J.M., Stanger-Hall K., & Lemmon E.M. (2009) The effect of ambiguous data on phylogenetic estimates obtained by maximum likelihood and bayesian inference. *Systematic Biology*, **58**, 130–145.

Lenk P., Joger U., & Wink M. (2001) Phylogenetic relationships among European rat snakes of the genus *Elaphe* Fitzinger based on mitochondrial DNA sequence comparisons. *Amphibia-Reptilia*, **22**, 329–339.

Li B., Belasen A., Pafilis P., Bednekoff P., & Foufopoulos J. (2014) Effects of feral cats on the evolution of anti-predator behaviours in island reptiles: insights from an ancient introduction. *Proceedings of the Royal Society B Biological sciences*, **281**, 20140339.

Lin Y.-H., McLenachan P.A., Gore A.R., Phillips M.J., Ota R., Hendy M.D., & Penny D. (2002) Four new mitochondrial genomes and the increased stability of evolutionary trees of mammals from improved taxon sampling. *Molecular Biology and Evolution*, **19**, 2060–2070.

Maddison W.P. (1997) Gene Trees in Species Trees. *Systematic Biology*, **46**, 523–536.

Maddison W.P. & Knowles L.L. (2006) Inferring phylogeny despite incomplete lineage sorting. *Systematic Biology*, **55**, 21–30.

Mahoney M.J. (2001) Molecular systematics of *Plethodon* and *Aneides* (Caudata: Plethodontidae: Plethodontini): Phylogenetic analysis of an old and rapid radiation. *Molecular Phylogenetics and Evolution*, **18**, 174–188.

Mayer W. & Pavlicev M. (2007) The phylogeny of the family Lacertidae (Reptilia) based on nuclear DNA sequences: convergent adaptations to arid habitats within the subfamily Eremiainae. *Molecular Phylogenetics and Evolution*, **44**, 1155–1163.

McCormack J.E., Huang H., & Knowles L.L. (2009) Maximum likelihood estimates of

species trees: How accuracy of phylogenetic inference depends upon the divergence history and sampling design. *Systematic Biology*, **58**, 501–508.

Meiri S. (2016) Small, rare and trendy: traits and biogeography of lizards described in the 21st century. *Journal of Zoology*, **299**, 251–261.

Meiri S. & Chapple D.G. (2016) Biases in the current knowledge of threat status in lizards, and bridging the "assessment gap." *Biological Conservation*, **204**, 6–15.

Melo-Ferreira J., Boursot P., Suchentrunk F., Ferrand N., & Alves P.C. (2005) Invasion from the cold past: Extensive introgression of mountain hare (*Lepus timidus*) mitochondrial DNA into three other hare species in northern Iberia. *Molecular Ecology*, **14**, 2459–2464.

Mendes J., Harris D.J., Carranza S., & Salvi D. (2016) Evaluating the phylogenetic signal limit from mitogenomes, slow evolving nuclear genes, and the concatenation approach. New insights into the Lacertini radiation using fast evolving nuclear genes and species trees. *Molecular Phylogenetics and Evolution*, **100**, 254–267.

Meng H., Carr J., Beraducci J., Bowles P., Branch W.R., Capitani C., Chenga J., Cox N., Howell K., Malonza P., Marchant R., Mbilinyi B., Mukama K., Msuya C., Platts P.J., Safari I., Spawls S., Shennan-Farpon Y., Wagner P., & Burgess N.D. (2016) Tanzania's reptile biodiversity: Distribution, threats and climate change vulnerability. *Biological Conservation*, **204**, 72–82.

Metallinou M., Červenka J., Crochet P.-A., Kratochvíl L., Wilms T., Geniez P., Shobrak M.Y., Brito J.C., & Carranza S. (2015) Species on the rocks: Systematics and biogeography of the rock-dwelling *Ptyodactylus* geckos (Squamata: Phyllodactylidae) in North Africa and Arabia. *Molecular Phylogenetics and Evolution*, **85**, 208–220.

Minoli I. & Avila L.J. (2017) Conservation assessments in climate change scenarios: Spatial perspectives for present and future in two *Pristidactylus* (Squamata: Leiosauridae) lizards from Argentina. *Zootaxa*, **4237**, 91–111.

Mulcahy D.G., Noonan B.P., Moss T., Townsend T.M., Reeder T.W., Sites J.W., & Wiens J.J. (2012) Estimating divergence dates and evaluating dating methods using phylogenomic and mitochondrial data in squamate reptiles. *Molecular Phylogenetics and Evolution*, **65**, 974–991.

Nagy Z.T., Lawson R., Joger U., & Wink M. (2004) Molecular systematics of racers, whipsnakes and relatives (Reptilia: Colubridae) using mitochondrial and nuclear

markers. *Journal of Zoological Systematics and Evolutionary Research*, **42**, 223–233.

Pavlicev M. & Mayer W. (2009) Fast radiation of the subfamily Lacertinae (Reptilia: Lacertidae): history or methodical artefact? *Molecular Phylogenetics and Evolution*, **52**, 727–734.

Pena C. & Wahlberg N. (2008) Prehistorical climate change increased diversification of a group of butterflies. *Biology Letters*, **4**, 274–278.

Pinheiro L.T., Rodrigues J.F.M., & Borges-Nojosa D.M. (2016) Formal education, previous interaction and perception influence the attitudes of people toward the conservation of snakes in a large urban center of northeastern Brazil. *Journal of Ethnobiology and Ethnomedicine*, **12**, 25.

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.

Pinho C., Rocha S., Carvalho B.M., Lopes S., Mourão S., Vallinoto M., Brunes T.O., Haddad C.F.B., Gonçalves H., Sequeira F., & Ferrand N. (2009) New primers for the amplification and sequencing of nuclear loci in a taxonomically wide set of reptiles and amphibians. *Conservation Genetics Resources*, **2**, 181–185.

Portik D.M., Wood P.L., Grismer J.L., Stanley E.L., & Jackman T.R. (2011) Identification of 104 rapidly-evolving nuclear protein-coding markers for amplification across scaled reptiles using genomic resources. *Conservation Genetics Resources*, **4**, 1–10.

Pyron R.A., Burbrink F.T., Colli G.R., de Oca A.N.M., Vitt L.J., Kuczynski C.A., & Wiens J.J. (2011) The phylogeny of advanced snakes (Colubroidea), with discovery of a new subfamily and comparison of support methods for likelihood trees. *Molecular Phylogenetics and Evolution*, **58**, 329–342.

Pyron R.A. & Wiens J.J. (2011) A large-scale phylogeny of Amphibia including over 2800 species, and a revised classification of extant frogs, salamanders, and caecilians. *Molecular Phylogenetics and Evolution*, **61**, 543–583.

Pyron R., Burbrink F.T., & Wiens J.J. (2013) A phylogeny and revised classification of Squamata, including 4161 species of lizards and snakes. *BMC Evolutionary Biology*, **13**, 93.

De Queiroz A. (2005) The resurrection of oceanic dispersal in historical biogeography. *Trends in Ecology and Evolution*, **20**, 68–73.

Rögl F. (1999) Mediterranean and Paratethys. Facts and hypothesis of an Oligocene to Miocene paleogeography (Short Overview). *Geologica Carpathica*, **50**, 339–349.

Roure B., Baurain D., & Philippe H. (2013) Impact of missing data on phylogenies inferred from empirical phylogenomic data sets. *Molecular Biology and Evolution*, **30**, 197–214.

Roussel V. & Van Wormhoudt A. (2017) The effect of Pleistocene climate fluctuations on distribution of European Abalone (*Haliotis tuberculata*), revealed by combined mitochondrial and nuclear marker analyses. *Biochemical Genetics*, **55**, 124–154.

Salvi D., Pinho C., & Harris D.J. (2017) Digging up the roots of an insular hotspot of genetic diversity: decoupled mito-nuclear histories in the evolution of the Corsican-Sardinian endemic lizard *Podarcis tiliguerta*. *BMC Evolutionary Biology*, **17**, 63.

Sanderson M.J., McMahon M.M., & Steel M. (2010) Phylogenomics with incomplete taxon coverage: the limits to inference. *BMC Evolutionary Biology*, **10**, 155.

Shaw A.J., Cox C.J., Goffinet B., Buck W.R., & Boles S.B. (2003) Phylogenetic evidence of a rapid radiation of pleurocarp mosses (Bryophyta). *Evolution*, **57**, 2226–2241.

Silva-Rocha I., Salvi D., Sillero N., Mateo J.A., & Carretero M.A. (2015) Snakes on the Balearic Islands: An invasion tale with implications for native biodiversity conservation. *PLoS ONE*, **10**, e0121026.

De Silva T.N., Peterson A.T., Bates J.M., Fernando S.W., & Girard M.G. (2017) Phylogenetic relationships of weaverbirds (Aves: Ploceidae): A first robust phylogeny based on mitochondrial and nuclear markers. *Molecular Phylogenetics and Evolution*, **109**, 21–32.

Simó-Riudalbas M., Metallinou M., de Pous P., Els J., Jayasinghe S., Péntek-Zakar E., Wilms T., Al-Saadi S., & Carranza S. (2017) Cryptic diversity in *Ptyodactylus* (Reptilia: Gekkonidae) from the northern Hajar Mountains of Oman and the United Arab Emirates uncovered by an integrative taxonomic approach. *PloS ONE*.

Sinervo B., Méndez-de-la-Cruz F., Miles D.B., Heulin B., Bastiaans E., Villagrán-Santa Cruz M., Lara-Resendiz R., Martínez-Méndez N., Calderón-Espinosa M.L., Meza-Lázaro R.N., Gadsden H., Avila L.J., Morando M., De la Riva I.J., Sepulveda P. V., Rocha C.F.D., Ibargüengoytía N., Puntriano C.A., Massot M., Lepetz V., Oksanen T.A., Chapple D.G., Bauer A.M., Branch W.R., Clobert J., & Sites J.W. (2010)

Erosion of lizard diversity by climate change and altered thermal niches. *Science*, **328**, 894–899.

Sites J.W., Reeder T.W., & Wiens J.J. (2011) Phylogenetic insights on evolutionary novelties in lizards and snakes: sex, birth, bodies, niches, and venom. *Annual Review of Ecology, Evolution, and Systematics*, **42**, 227–244.

Thibault M., Brescia F., Jourdan H., & Vidal E. (2017) Invasive rodents, an overlooked threat for skinks in a tropical island hotspot of biodiversity. *New Zealand Journal of Ecology*, **41**, 74–83.

Tonini J.F.R., Beard K.H., Ferreira R.B., Jetz W., & Pyron R.A. (2016) Fully-sampled phylogenies of squamates reveal evolutionary patterns in threat status. *Biological Conservation*, **204**, 23–31.

Townsend T., Larson A., Louis E., & Macey J. (2004) Molecular phylogenetics of Squamata: the position of snakes, amphisbaenians, and dibamids, and the root of the squamate tree. *Systematic Biology*, **53**, 735–757.

Townsend T.M., Alegre R.E., Kelley S.T., Wiens J.J., & Reeder T.W. (2008) Rapid development of multiple nuclear loci for phylogenetic analysis using genomic resources: An example from squamate reptiles. *Molecular Phylogenetics and Evolution*, **47**, 129–142.

Uetz P. (2010) The original descriptions of reptiles. *Zootaxa*, **68**, 59–68.

Upchurch P. (2008) Gondwanan break-up: legacies of a lost world? *Trends in Ecology and Evolution*, **23**, 229–236.

Utiger U., Helfenberger N., Schätti B., Schmidt C., Ruf M., & Ziswiler V. (2002) Molecular systematics and phylogeny of Old World and New World ratsnakes, *Elaphe* auct., and related genera (Reptilia, Squamata, Colubridae). *Russian Journal of Herpetology*, **9**, 105–124.

Vanhooydonck B. & Van Damme R. (1999) Evolutionary relationships between body shape and habitat use in lacertid lizards. *Evolutionary Ecology Research*, **1**, 785–803.

Vences M., Sanchez E., Hauswaldt J.S., Eikelmann D., Rodríguez A., Carranza S., Donaire D., Gehara M., Helfer V., Lötters S., Werner P., Schulz S., & Steinfartz S. (2014) Nuclear and mitochondrial multilocus phylogeny and survey of alkaloid content in true salamanders of the genus *Salamandra* (Salamandridae). *Molecular Phylogenetics and Evolution*, **73**, 208–216.

Verdú-Ricoy J., Carranza S., Salvador A., Busack S., & Díaz J. (2010) Phylogeography of *Psammodromus algirus* (Lacertidae) revisited: systematic implications. *Amphibia-Reptilia*, **31**, 576–582.

Vidal N., Azvolinsky A., Cruaud C., & Hedges S.B. (2008) Origin of tropical American burrowing reptiles by transatlantic rafting. *Biology Letters*, **4**, 115–118.

Vidal N., Delmas A.S., David P., Cruaud C., Couloux A., & Hedges S.B. (2007) The phylogeny and classification of caenophidian snakes inferred from seven nuclear protein-coding genes. *Comptes Rendus - Biologies*, **330**, 182–187.

Vidal N. & Hedges S.B. (2005) The phylogeny of squamate reptiles (lizards, snakes, and amphisbaenians) inferred from nine nuclear protein-coding genes. *Comptes rendus biologies*, **328**, 1000–1008.

Vitt L.J., Pianka E.R., Cooper J., & Schwenk K. (2003) History and the global ecology of squamate reptiles. *American Naturalist*, **162**, 44–60.

Walsh H.E., Kidd M.G., Moum T., & Friesen V.L. (1999) Polytomies and the power of phylogenetic inference. *Evolution*, **53**, 932–937.

Weigert A., Helm C., Meyer M., Nickel B., Arendt D., Hausdorf B., Santos S.R., Halanych K.M., Purschke G., Bleidorn C., & Struck T.H. (2014) Illuminating the base of the Annelid tree using transcriptomics. *Molecular Biology and Evolution*, **31**, 1391–1401.

Whitfield J.B. & Lockhart P.J. (2007) Deciphering ancient rapid radiations. *Trends in Ecology and Evolution*, **22**, 258–265.

Wiens J.J., Brandley M.C., & Reeder T.W. (2006) Why does a trait evolve multiple times within a clade? Repeated evolution of snakelike body form in squamate reptiles. *Evolution*, **60**, 123–141.

Wiens J.J., Hutter C.R., Mulcahy D.G., Noonan B.P., Townsend T.M., Sites J.W., & Reeder T.W. (2012) Resolving the phylogeny of lizards and snakes (Squamata) with extensive sampling of genes and species. *Biology Letters*, **8**, 1043–1046.

Wiens J.J., Kuczynski C.A., Smith S.A., Mulcahy D.G., Sites J.W., Townsend T.M., & Reeder T.W. (2008) Branch lengths, support, and congruence: testing the phylogenomic approach with 20 nuclear loci in snakes. *Systematic Biology*, **57**, 420–431.

Wu X.-C., Brinkman D.B., & Russel A.P. (1996) *Sineoamphisbaena hexatabularis*: an

amphisbaenian (Diapsida: Squamata) from the Upper Cretaceous redbeds at Bayan Mandahu (Inner Mongolia, People's Republic of China), and comments on the phylogenetic relationships of the Amphisbaenia. *Canadian Journal of Earth Sciences*, **33**, 541–577.

Zhang D.-X. & Hewitt G.M. (2003) Nuclear DNA analyses in genetic studies of populations: practice, problems and prospects. *Molecular Ecology*, **12**, 563–584.

Zheng Y. & Wiens J.J. (2015) Do missing data influence the accuracy of divergence-time estimation with BEAST? *Molecular Phylogenetics and Evolution*, **85**, 41–49.

Zheng Y. & Wiens J.J. (2016) Combining phylogenomic and supermatrix approaches, and a time-calibrated phylogeny for squamate reptiles (lizards and snakes) based on 52 genes and 4162 species. *Molecular Phylogenetics and Evolution*, **94**, 537–547.

Zhou K., Li H., Han D., Bauer A.M., & Feng J. (2006) The complete mitochondrial genome of *Gekko gecko* (Reptilia: Gekkonidae) and support for the monophyly of Sauria including Amphisbaenia. *Molecular Phylogenetics and Evolution*, **40**, 887–892.

CHAPTER 2

Multilocus phylogeny of Lacertini

Article I. Evaluating the phylogenetic signal limit from mitogenomes, slow evolving nuclear genes, and the concatenation approach. New insights into the Lacertini radiation using fast evolving nuclear genes and species trees.

Joana Mendes[1,2], D. James Harris[1], Salvador Carranza[2], Daniele Salvi[1*]

Molecular Phylogenetics and Evolution, 2016, 100: 254-267;

DOI: 10.1016/j.ympev.2016.04.016.

[1]CIBIO-InBIO, Centro de Investigação em Biodiversidade e Recursos Genéticos, Universidade do Porto, Campus Agrário de Vairão, 4485-661 Vairão, Portugal.

[2] Institute of Evolutionay Biology (CSIC-Universitat Pompeu Fabra), Passeig Marítim de la Barceloneta 37-49, 08003 Barcelona, Spain.

ABSTRACT

Estimating the phylogeny of lacertid lizards, and particularly the tribe Lacertini has been challenging, possibly due to the fast radiation of this group resulting in a hard polytomy. However this is still an open question, as concatenated data primarily from mitochondrial markers have been used so far whereas in a recent phylogeny based on a compilation of these data within a squamate supermatrix the basal polytomy seems to be resolved.

In this study we estimate phylogenetic relationships between all Lacertini genera using for the first time DNA sequences from five fast evolving nuclear genes (*acm4*, *mc1r*, *pdc*, *βfib* and *reln*) and two mitochondrial genes (*nd4* and *12S*). We generated a total of 529 sequences from 88 species and used Maximum Likelihood and Bayesian Inference phylogenetic methods based on concatenated multilocus dataset as well as a coalescent-based species tree approach and molecular dating with the aim of (i) shedding light on the basal relationships of Lacertini (ii) assessing the monophyly of genera which were previously questioned, and (iii) discussing differences between estimates from this and previous studies based on different markers, and phylogenetic methods.

Results uncovered (i) a new phylogenetic clade formed by the monotypic genera *Archaeolacerta*, *Zootoca*, *Teira* and *Scelarcis*; and (ii) support for the monophyly of the *Algyroides* clade, with two sister species pairs represented by western (*A. marchi* and *A. fitzingeri*) and eastern (*A. nigropunctatus* and *A. moreoticus*) lineages. In both cases the members of these groups show peculiar morphology and very different geographical distributions, suggesting that they are relictual groups that were once diverse and widespread. They probably originated about 11-13 million years ago during early events of speciation in the tribe, and the split between their members is estimated to be only slightly older. This scenario may explain why mitochondrial markers (possibly saturated at higher divergence levels) or slower nuclear markers (likely lacking enough phylogenetic signal) used in previous studies failed to recover these relationships.

Finally, the phylogenetic position of most remaining genera was unresolved, corroborating the hypothesis of a hard polytomy in the Lacertini phylogeny due to a fast radiation. This is in agreement with all previous studies but in sharp contrast with a recent squamate megaphylogeny. We show that the supermatrix approach may provide high support for incorrect nodes that are not supported either by original sequence data or by new data from this study. This finding suggests caution when using megaphylogenies to integrate inter-generic relationships in comparative ecological and evolutionary studies.

KEYWORDS

Lacertini radiation; hard polytomy; species tree; concatenation; nucDNA; multilocus phylogeny

INTRODUCTION

The squamate reptile family Lacertidae is a clade of small-bodied lizards distributed in the Palaearctic and Africa. It comprises two sub-families, the Gallotiinae (2 genera, 13 species) and the Lacertinae (41 genera, 308 species), with the latter divided in two tribes, Eremiadini (22 genera, 184 species) and Lacertini (19 genera, 124 species) (Arnold et al., 2007; Uetz and Hošek, 2015). As the most common lizard family in Europe, lacertids have been widely used as model species to answer questions on ecology and evolutionary biology, such as testing hypotheses on functional ecology (e.g. Vanhooydonck and Van Damme, 1999; Herrel et al., 2008; Baeckens et al., 2015), natural selection (e.g. Salvi et al., 2009; Heulin et al., 2011) or biogeography (e.g. Harris et al., 2002; Carranza et al., 2004; Poulakakis et al., 2005; Salvi et al., 2013). All such diverse assessments require an understanding of the evolutionary history of the group, so that comparisons can be drawn within a phylogenetic framework.

Over the last decades several morphological, bio-chemical and molecular studies have been conducted in order to infer the phylogeny of Lacertidae (Harris et al., 1998; Fu, 1998, 2000; Arnold et al., 2007; Mayer and Pavlicev, 2007; Hipsley et al., 2009; Pavlicev and Mayer, 2009; Cox et al., 2010). While the phylogenetic relationships within Gallotiinae and Eremiadini are relatively well known (e.g. Mayer and Pavlicev, 2007; Cox et al., 2010), the phylogeny of the tribe Lacertini is still mainly unresolved, with conflicting hypotheses and little corroboration between studies, particularly in the internal nodes. Indeed, although a few relationships within the tribe have been estimated with confidence and consistently across the previous studies, such as the case of the sister taxa relationships between the monotypic genera *Scelarcis* and *Teira* or between the genera of green lizards *Lacerta* and *Timon*, the phylogenetic position of the majority of taxa remains unknown. Moreover, the monophyly of the genus *Algyroides* was recently questioned (Pavlicev and Mayer, 2009). Since the lack of phylogenetic resolution shown by early studies may be due to insufficient data, Mayer and Pavlicev (2007) and Pavlicev and Mayer (2009) performed phylogenetic analyses including nuclear sequence data and an increasing taxon sampling. Their results yielded no improvements in the basal resolution of the phylogenetic tree and therefore discarded the hypothesis of a soft polytomy due to a methodological artefact. However, a possible alternative explanation for the lack of improvements in these last phylogenetic assessments may be that the nuclear data used in these two later studies consisted in two extremely slow-evolving genes (*c-mos* and *rag1*), possibly holding low information content to recover speciation nodes within Lacertini. On the other hand, a recent study from Pyron et al. (2013) with a

wide focus on relationships between 4161 Squamata taxa, appears to have successfully solved the internal branching within Lacertini recovering high statistical support from internal to tip nodes. In this study, the authors used mainly the same two slow evolving nuclear markers employed by Mayer and Pavlicev and mitochondrial information from previous studies and applied a non-parametric Shimodaira-Hasegawa-Like implementation of the approximate likelihood-ratio test (SHLaLRT) (Anisimova and Gascuel, 2006). Consequently, the current state of knowledge on Lacertini evolutionary history has two contrasting phylogenetic hypotheses drawn from concatenated dataset using mostly the same DNA sequences from mitochondrial and slow evolving nuclear markers.

All previous Lacertini phylogenies were based on the analysis of concatenated sequences from multiple genes. Such concatenation approach can prove problematic due to discordances between gene histories and the true evolutionary relationships among species, or in other words, between the gene trees and the species tree. While several processes can account for the discrepancy between gene trees and species trees (Maddison, 1997), recent studies demonstrate that the common approach of concatenating sequences from multiple genes can result in a well-supported but incorrect tree (Kubatko and Degnan, 2007). Bias caused by the concatenation approach can be produced, for instance, by the overuse of genetically linked and more variable mitochondrial genes, which regularly drives the tree, hiding the information of less variable, usually nuclear, genes. Another major, yet frequently unconsidered, challenge is allele selection in the concatenation process. This substantially influences the phylogenetic results, as heterozygous alleles may have gene tree coalescences deeper than their species divergence, causing gene tree variations according to the chosen allele (Weisrock et al., 2012). Moreover, incongruence across gene tree topologies is an issue of concatenation: if topologies are not significantly different, species trees can be estimated through a concatenation approach. On the other hand, theoretical work has shown that the coalescent process can produce substantial variation in single-gene histories. When single-gene trees are significantly different and incongruent, as it seems the case for Lacertini, the concatenation approach leads to statistically inconsistent estimation of phylogenies (Degnan and Rosenberg, 2006; Kubatko and Degnan, 2007; McVay and Carstens, 2013). In all these cases, bootstraps can provide strong support for an incorrect phylogeny (Kubatko and Degnan, 2007). New methodologies of species tree estimation based on multilocus data from multiple individuals per species allow the reconciliation of a set of gene trees embedded in a shared species phylogeny. Thus, the species tree methods offer a promising tool to assess the reliability of previous

phylogenies based on mainly mitochondrial dataset and to dissect the very different phylogenetic estimates of Lacertini based on the concatenation approach.

In this study we generate a comprehensive DNA sequence dataset for Lacertini, including all the tribe's genera, by sequencing multiple specimens per species, with additional taxa relative to previous studies, and including, for the first time, five fast evolving nuclear molecular markers to complement mitochondrial sequence data. In addition to the common approach of concatenating sequences from multiple genes, we implement a species tree approach to infer the phylogeny of Lacertini. Our main aim is to explore whether the addition of DNA sequences from fast-evolving nuclear genes, combined with a multi-species coalescent approach can resolve or improve the inference of basal relationships of the tribe Lacertini, as well as provide more resolution on the relationships between genera and support for genera monophyly. We also compare the species tree and the trees derived from the concatenation approach based on mitochondrial and nuclear genes from this study and previous ones. By doing this, we investigate the phylogenetic resolution of mitochondrial and nuclear markers, as well as comparing the phylogenetic inferences made by different phylogenetic methods.

MATERIAL AND METHODS

SAMPLING

A total of 78 specimens from all the 19 genera of Lacertini were employed in the phylogenetic analyses. We used an average of two specimens per species, with a minimum of one and a maximum of five specimens. Ten additional samples, two for each of the species *Gallotia atlantica*, *G. stehlini*, *Psammodromus algirus* and *P. hispanicus* from the sub-family Gallotiinae, and *Atlantolacerta andreanskyi* from the tribe Eremiadini were used as outgroups following previous studies (e.g. Arnold et al., 2007; Harris et al., 1998). All samples were obtained from the collections of Centro de Investigação em Biodiversidade e Recursos Genéticos, Universidade do Porto (CIBIO-InBIO) and the Institute of Evolutionary Biology - CSIC-UPF (IBE). Information regarding the sample codes, species, sampling locality and GenBank accession numbers is given in Table 2.1.

DNA EXTRACTION, AMPLIFICATION AND SEQUENCING

Total genomic DNA was extracted from alcohol-preserved tail muscle following standard high-salt protocols (Sambrook et al., 1989). For a reduced number of samples for which

the saline extraction failed we used the Qiagen DNeasy® Blood & Tissue extraction kit, following the manufacture's protocol.

Fragments from the mitochondrial DNA genes NADH Dehydrogenase 4 plus the flanking tRNAs Histidine and Serine (*nd4*), and of the ribosomal 12SrRNA gene (*12S*), and the nuclear genes Acetylcholinergic Receptor M4 (*acm4*), Melanocortin 1 Receptor (*mc1r*), Phosducin (*pdc*), intron 7 of *β*-fibrinogen (*βfib*) and intron 61 of Reelin (*reln*) were amplified through standard Polymerase Chain Reaction (PCR). We selected these mitochondrial and nuclear markers because they have been shown to be highly variable in previous studies on Lacertinae (e.g. Pinho et al., 2008; Salvi et al., 2010, 2014; Barata et al., 2012). Primers and PCR protocols used for the amplification of the molecular markers are reported in Table 2.2. Purification and sequencing of PCR products were carried out by a commercial sequencing company (Macrogen Europe: www.macrogen.com), using the same primers employed for amplification.

PHYLOGENETIC ANALYSES

Sequences were aligned using the MUSCLE algorithm (Edgar, 2004) in Geneious (Biomatters Ltd.) with default settings. Ambiguous and poorly aligned positions were removed by Gblocks v.0.91b using default settings (Castresana, 2000).

Haplotype reconstruction for nuclear gene fragments was performed in PHASE v. 2.1 (Stephens et al., 2001; Stephens and Scheet, 2005). Input files were created in SEQPHASE (Flot, 2010; available at http://seqphase.mpg.de/seqphase/). Haplotypes defined from heterozygous insertion-deletions were manually phased and were incorporated as known phases to improve haplotype determination following Flot et al. (2006). PHASE was run three times to assure consistency, with a phase probability threshold of 0.7 and the remaining settings by default. Recombination detection was performed in RDP v.3.44 (Martin et al., 2010) using five different algorithms, RDP (Martin and Rybicki, 2000), GENECONV (Padidam et al., 1999), MaxChi (Smith, 1992), BootScan (Martin et al., 2005) and SiScan (Gibbs et al., 2000) with default options and applying the auto-masking tool to remove the outgroup and very divergent or very similar sequences, in order to increase statistical power (Martin et al., 2010).

Phylogenetic relationships among the Lacertidae species were inferred by Maximum Likelihood (ML), Bayesian Inference (BI) and the Bayesian species tree approach based on the multi-locus coalescent. For the ML and BI analyses, unphased sequence data were concatenated in three different matrices: mitochondrial DNA

(mtDNA), nuclear DNA (nucDNA) and mitochondrial-nuclear DNA data. Within each matrix the data was partitioned by gene fragment (seven mt-nucDNA partitions). ML analyses were performed in RAxML GUI v.1.1.3 (Silvestro and Michalak, 2012), a graphical front-end for RaxML v.7.4.2 (Stamatakis, 2006). ML searches included 10 random addition replicates and 1000 nonparametric bootstrap replicates, applying the general time-reversible model with gamma model of rate heterogeneity (GTRGAMMA) for each of the three concatenated datasets. BI analyses were performed in BEAST v.1.8.0 (Drummond et al., 2012) for each concatenated dataset. The best model of nucleotide substitution for each gene among 40 different models was assessed in jModelTest v.2.1.3 (Posada, 2008) under the corrected Akaike Information Criterion (AICc) (Table 2.3). We built the input file with evolutionary models, tree priors and Markov Chain Monte Carlo (MCMC) options using the BEAUTi utility included in the BEAST package. Models and prior specifications applied were as follows (otherwise by default): because the K80 model is not available on BEAST, we implemented the next best model available, HKY; the tree model of all gene partitions was linked, while nucleotide substitution and clock models were unlinked; Relaxed Uncorrelated Lognormal Clock set for all genes, Yule process of speciation as tree prior, random starting tree, alpha Uniform (0, 10), ucld.mean Uniform, and operator kappa (2.0). The use of the Yule process of speciation prior requires only one sequence per species, whereas our concatenated alignments contain multiple samples per species. Therefore, to investigate the sensitivity of our estimates to the choice of tree prior, we performed an additional run for each dataset, applying the same settings as above but using only one representative sequence for each species. BEAST was run three times, with 100 million generations, sampling every 10000 generations. We used Tracer v 1.5 (Rambaut and Drummond, 2007) to check the runs for convergence (burn-in = 10%) and to ensure that all effective sample sizes parameters (ESS) were higher than 200, as recommended in the manual. Runs were combined with LogCombiner and afterwards TreeAnnotator (both included in the BEAST package) was used to summarize the trees in a consensus tree representing the posterior distribution.

The species tree was inferred using the *BEAST extension of the BEAST software. *BEAST co-estimates a species tree along with the gene trees and effective population sizes of the species in a single Bayesian Markov Chain Monte Carlo analysis. For this analysis we used the phased alignments of the nuclear genes and their relative models of nucleotide evolution calculated in jModelTest, under the AICc (Table 2.3). Because the models JC and K80 are not available on BEAST, we use the next best option: HKY. Nucleotide substitution, clock and tree models were unlinked, with the

exception of the tree model of the mitochondrial genes *12S* and *nd4* because these genes are genetically linked. The remaining settings were the same as in the BEAST mt-nucDNA concatenated analysis. We used the available estimated rate of evolution of *12S* of lacertid lizards (Carranza and Arnold, 2012) to estimate cladogenetic events within Lacertini. Mean substitution rates and their standard errors for the same *12S* gene regions used in the present study were extracted from a fully-calibrated phylogeny (nine calibration points) including the lacertid lizard Canary Islands radiation of *Gallotia* sp. (Cox et al., 2010) and the Balearic islands *Podarcis pityusensis* and *P. lilfordi* (Brown et al., 2008). For a full account on the specific calibration points and methods used to infer the substitution rate of Lacertid lizard used in the present study please see Carranza and Arnold (2012). Absolute divergence times were estimated in *BEAST by setting a normal distribution prior for the ucld.mean parameter of the *12S* gene fragment with the following parameters: initial: 0.00553, mean: 0.00553, stdev: 0.00128. *BEAST was run five times with 400 million generations, sampling every 40000 generations. Runs were performed in the CIPRES Science Gateway V. 3.3 (Miller et al., 2010, at http://www.phylo.org/). Convergence and ESS of the runs were verified in Tracer v 1.5. Runs were combined with LogCombiner and the maximum clade credibility tree was calculated in TreeAnnotator. All trees were visualized in FIGTREE v1.4 (available at http://tree.bio.ed.ac.uk/software/figtree/).

TOPOLOGY TESTS

In order to compare our phylogenetic hypothesis with previous phylogenies, we performed topological tests between our ML tree based on the concatenated mt-nucDNA dataset and ML trees obtained by previous studies. First, we inspected supported nodes recovered in previous studies that conflicted with our results and then we enforced these nodes in our tree topology. In order to assess the relative contribution on topological comparisons of nodes with different levels of support we generated three constrained topologies enforcing all nodes obtained in previous studies with bootstrap support values equal or higher than (i) 95 or (ii) 90 or (iii) 85. The trees with topological constrains were generated in Mesquite version 3.03 (Maddison and Maddison, 2003). Constrained clades are presented in Table 2.5. Per-site log likelihood values were estimated in RAxMLGUI v.1.1.3. The constrained trees were compared with our best ML tree using the Shimodaira-Hasegawa (SH) and the approximately unbiased (AU) tests (Shimodaira and Hasegawa, 1999 and Shimodaira, 2002, respectively), as implemented in CONSEL (Shimodaira and Hasegawa, 2001) to determine if any of the alternatives could be rejected at the 0.05 level.

Table 2.1. Sample codes, species, sampling locality and sequences' GenBank accession numbers for the 88 samples used in this study.

Code	Species	Locality	*12S*	*nd4*	*acm4*	*βfib*	*mc1r*	*pdc*	*reln*
330	*Algyroides fitzingeri*	Cagliari, Sardinia, Italy	KX080559	KX081002	KX080921	KX080640	KX080711	KX080793	KX080858
701	*Algyroides fitzingeri*	Restonica, Corsica, France	KX080560	KX081003	KX080922	KX080641	KX080712	KX080794	KX080859
4029	*Algyroides fitzingeri*	Mt. Albo, Sardinia, Italy	KX080561	KX081004	KX080923	KX080642	KX080713	KX080795	KX080860
1768	*Algyroides marchi*	La Hueta's waterfall, Spain	KX080563	KX081006	KX080925	KX080644	KX080715	KX080797	KX080862
1859	*Algyroides marchi*	El Toril, Spain	KX080564	KX081007	KX080926	KX080645	KX080716	KX080798	KX080863
1889	*Algyroides marchi*	Puente de las Herrerías, Spain	KX080562	KX081005	KX080924	KX080643	KX080714	KX080796	KX080861
458b	*Algyroides moreoticus*	Kalivia, Greece	KX080555	KX081000	KX080917	KX080637	KX080707		KX080855
4324	*Algyroides moreoticus*	Roitika Patras, Greece	KX080558		KX080920		KX080710	KX080792	KX080857
4325	*Algyroides moreoticus*	Kalavryta, Greece	KX080556		KX080918	KX080638	KX080708	KX080790	
4332	*Algyroides moreoticus*	Zarouchla, Greece	KX080557	KX081001	KX080919	KX080639	KX080709	KX080791	KX080856
416	*Algyroides nigropunctatus*	Metsovo, Greece	KX080552	KX080997	KX080914	KX080634	KX080704	KX080787	
3237	*Algyroides nigropunctatus*	Vitsa, Greece	KX080550	KX080995	KX080913		KX080702	KX080786	
3246	*Algyroides nigropunctatus*	Voidomatis, Greece	KX080551	KX080996		KX080633	KX080703		
15438	*Algyroides nigropunctatus*	Vanganel, Slovenia	KX080553	KX080998	KX080915	KX080635	KX080705	KX080788	
15441	*Algyroides nigropunctatus*	Vanganel, Slovenia	KX080554	KX080999	KX080916	KX080636	KX080706	KX080789	
S10390	*Anatololacerta danfordi*	Çamıyayla, Turkey	KX080617	KX081055	KX080981	KX080690			
12033	*Apathya cappadocica*	Göksun, Turkey	KX080619	KX081057	KX080983		KX080772	KX080843	KX080900
S10388	*Apathya cappadocica*	Eastern Turkey	KX080618	KX081056	KX080982	KX080912	KX080771	KX080842	KX080899
RE1	*Archaeolacerta bedriagae*	Restonica, Corsica, France	KX080585	KX081026	KX080947	KX080659	KX080737	KX080814	
RE2	*Archaeolacerta bedriagae*	Restonica, Corsica, France	KX080586	KX081027	KX080948	KX080660	KX080738	KX080815	KX080880
5015	*Atlantolacerta andreanskyi*	Tizin Tichka, Morocco	JX462057.1	JX462200.1	JX461988.1	KX080693	JX461804.1	JX461634.1	
5058	*Atlantolacerta andreanskyi*	Tizin Tichka, Morocco	JX462054.1	JX462196.1	JX462000.1	KX080694	JX461816.1	JX461644.1	
S10353	*Dalmatolacerta oxycephala*	Bosnia and Herzegovina	KX080610	KX081049	KX080973	KX080684	KX080763	KX080836	

S10354	*Dalmatolacerta oxycephala*	Bosnia and Herzegovina	KX080609	KX081048	KX080972	KX080683	KX080762	KX080835	
7802	*Darevskia derjugini*	Abastumani, Georgia	KX080583		KX080945	KX080657	KX080735	KX080813	KX080878
7803	*Darevskia derjugini*	Abastumani, Georgia	KX080584		KX080946	KX080658	KX080736		KX080879
4985	*Darevskia raddei*	Pia, Georgia	KX080582	KX081025	KX080944	KX080656	KX080734	KX080812	
10126	*Darevskia raddei*	Ganzasar, Nagorno-Karabakh Republic	KX080581	KX081024	KX080943	KX080655	KX080733	KX080811	KX080877
3	*Dinarolacerta montenegrina*	Đebeza, Prokletije Mountains, Montenegro		KX081012	KX080930	KX080909	KX080721	KX080803	KX080868
18	*Dinarolacerta montenegrina*	Đebeza, Prokletije Mountains, Montenegro	KX080566	KX081009	KX080927	KX080646	KX080718	KX080800	KX080865
19	*Dinarolacerta montenegrina*	Đebeza, Prokletije Mountains, Montenegro	KX080567	KX081010	KX080928	KX080907	KX080719	KX080801	KX080866
20	*Dinarolacerta montenegrina*	Đebeza, Prokletije Mountains, Montenegro	KX080565	KX081008		KX080906	KX080717	KX080799	KX080864
22	*Dinarolacerta montenegrina*	Đebeza, Prokletije Mountains, Montenegro	KX080568	KX081011	KX080929	KX080908	KX080720	KX080802	KX080867
9	*Dinarolacerta mosorensis*	Međuvršje, Montenegro	KX080570	KX081014	KX080932	KX080647	KX080723	KX080804	KX080870
13	*Dinarolacerta mosorensis*	Međuvršje, Montenegro	KX080571		KX080933	KX080648	KX080724	KX080805	
15	*Dinarolacerta mosorensis*	Virak, Montenegro	KX080569	KX081013	KX080931		KX080722		KX080869
AM1	*Dinarolacerta mosorensis*	Lovćen Mountains, Montenegro	KX080572	KX081015	KX080934	KX080649			
1244	*Gallotia atlantica*	Nazaret, Lanzarote, Spain	KX080625	KX081062	KX080988	KX080695	KX080778	KX080847	KX080902
1341	*Gallotia atlantica*	Yaiza, Lanzarote, Spain	KX080626	KX081063	KX080989	KX080696	KX080779	KX080848	
1350	*Gallotia stehlini*	San Andrés, Gran Canaria, Spain	KX080627	KX081064	KX080990	KX080697	KX080780	KX080849	
1412	*Gallotia stehlini*	Aldea Blanca, Gran Canaria, Spain	KX080628	KX081065	KX080991	KX080698	KX080781	KX080850	KX080903
456b	*Hellenolacerta graeca*	Agia Kyriaki, Greece	KX080612		KX080975	KX080686	KX080765	KX080838	KX080896
456c	*Hellenolacerta graeca*	Agia Kyriaki, Greece	KX080613		KX080976	KX080687	KX080766	KX080839	KX080897
S10387	*Hellenolacerta graeca*	Greece	KX080611	KX081050	KX080974	KX080685	KX080764	KX080837	KX080895
62	*Iberolacerta cyreni*	Navacerrada, Spain	KX080578	KX081021	KX080940	KX080651	KX080730	KX080809	KX080874
MON1	*Iberolacerta cyreni*	Rascafría, Spain	KX080577	KX081020	KX080939	KX080652	KX080729		KX080873
4282	*Iberolacerta monticola*	Sosas de Laciana, Spain	KX080579	KX081022	KX080941	KX080653	KX080731		KX080875
4283	*Iberolacerta monticola*	Sosas de Laciana, Spain	KX080580	KX081023	KX080942	KX080654	KX080732	KX080810	KX080876
5143	*Iranolacerta brandtii*	Esfahan, Iran	KX080623	KX081060	KX080986		KX080776		
5146	*Iranolacerta brandtii*	Ardabil, Iran	KX080624	KX081061	KX080987		KX080777	KX080846	KX080901

S10397	*Lacerta agilis*	Studland, United Kingdom	KX080600	KX081039	KX080962	KX080673	KX080752	KX080829	KX080888
S10401	*Lacerta agilis*	Alp, Spain	KX080601	KX081040	KX080963	KX080674	KX080753	KX080830	
15306	*Lacerta bileneata*	Bosco Magnano, Italy	KX080602	KX081041	KX080964	KX080675	KX080754		
15307	*Lacerta bileneata*	Pantana, Italy	KX080603	KX081042	KX080965	KX080676	KX080755		KX080889
15308	*Lacerta bileneata*	Abruzzo, Italy	KX080604	KX081043	KX080966	KX080677	KX080756		KX080890
1912	*Lacerta schreiberi*	Garganta de las Lancha, Spain	KX080598	KX081037	KX080960	KX080671	KX080750	KX080827	KX080887
3866	*Lacerta schreiberi*	Tanes, Spain	KX080599	KX081038	KX080961	KX080672	KX080751	KX080828	
445	*Lacerta trilineata*	Agios Vasilios, Greece		KX081044	KX080968	KX080679	KX080758	KX080831	
446	*Lacerta trilineata*	Agios Vasilios, Greece	KX080606	KX081045	KX080969	KX080680	KX080759	KX080832	KX080892
447	*Lacerta trilineata*	Dorio, Greece	KX080607	KX081046	KX080970	KX080681	KX080760	KX080833	KX080893
451	*Lacerta trilineata*	Koutsouroumpas, Greece	KX080608	KX081047	KX080971	KX080682	KX080761	KX080834	KX080894
S10399	*Lacerta trilineata*	Golbasi, Turkey	KX080605		KX080967	KX080678	KX080757		KX080891
S10398	*Parvilacerta parva*	Çorum, Turkey	KX080616	KX081054	KX080980	KX080911	KX080770	KX080841	
JamJB	*Phoenicolacerta kulzeri*	Barouk, Jordan	KX080622	KX081059	KX080985	KX080692	KX080775	KX080845	
Petra	*Phoenicolacerta kulzeri*	Petra, Jordan	KX080621		KX080984	KX080691	KX080774		
S10389	*Phoenicolacerta kulzeri*	Ainata, Lebanon	KX080620	KX081058			KX080773	KX080844	
509	*Podarcis muralis*	Florence, Italy	KX080575	KX081018	KX080937	KX080650	KX080727	KX080807	KX080872
5937	*Podarcis muralis*	Sierra delle Ciavole, Italy	KX080576	KX081019	KX080938		KX080728	KX080808	
771	*Podarcis sicula*	Vulcano Island, Sicily, Italy	KX080573	KX081016	KX080935		KX080725		KX080871
9103	*Podarcis sicula*	Pizzo, Italy	KX080574	KX081017	KX080936	KX080701	KX080726	KX080806	
2347	*Psammodromus algirus*	Iminifri, Morocco	KX080631	KX081068		KX080699	KX080784	KX080853	
2356	*Psammodromus algirus*	Azrou, Morocco	KX080632	KX081069	KX080994	KX080700	KX080785	KX080854	
1723	*Psammodromus hispanicus*	Jaén, Spain	KX080629	KX081066	KX080992		KX080782	KX080851	KX080904
1850	*Psammodromus hispanicus*	Jaén, Spain	KX080630	KX081067	KX080993		KX080783	KX080852	KX080905
139	*Scelarcis perspicillata*	Sidi Yahya Ousaad, Morocco	KX080591	KX081031	KX080953	KX080664	KX080743	KX080820	
3456	*Scelarcis perspicillata*	Taza, Morocco	KX080592		KX080954	KX080665	KX080744	KX080821	
S9282	*Takydromus sexlineatus*	Gangaw, Myanmar	KX080614	KX081051	KX080977	KX080688	KX080767		KX080897

S9294	*Takydromus sexlineatus*	Nat Ma Taung National Park, Myanmar	KX080615	KX081052	KX080978	KX080689	KX080768		KX080898
S10392	*Takydromus smaragdinus*	Okinawa, Japan		KX081053	KX080979		KX080769	KX080840	
5222	*Teira dugesii*	Santa Maria, Azores, Portugal	KX080595	KX081034	KX080957	KX080668	KX080747	KX080824	
5223	*Teira dugesii*	Santa Maria, Azores, Portugal	KX080596	KX081035	KX080958	KX080669	KX080748	KX080825	
5224	*Teira dugesii*	Santa Maria, Azores, Portugal	KX080597	KX081036	KX080959	KX080670	KX080749	KX080826	
4012	*Timon lepidus*	Vairão, Portugal	KX080589		KX080951	KX080662	KX080741	KX080818	KX080883
4846	*Timon lepidus*	Burgos, Spain	KX080590	KX081030	KX080952	KX080663	KX080742	KX080819	KX080884
14	*Timon tangitanus*	Morocco	KX080587	KX081028	KX080949	KX080661	KX080739	KX080816	KX080881
27	*Timon tangitanus*	Cirque de Jafar, Morocco	KX080588	KX081029	KX080950	KX080910	KX080740	KX080817	KX080882
15305	*Zootoca vivipara*	Russia	KX080594	KX081033	KX080956	KX080667	KX080746	KX080823	KX080886
zoo	*Zootoca vivipara*	Baikal Lake, Russia	KX080593	KX081032	KX080955	KX080666	KX080745	KX080822	KX080885

Table 2.2. Primers and PCR protocols used for the amplification of the molecular markers used in this study.

Gene	Primer	Sequence (5'-3')	Source	PCR conditions (ºC(seconds) x number of cicles)
nd4	ND4	CAC CTA TGA CTA CCA AAA GCT CAT GTA GAA GC	Arévalo *et al.* (1994)	94(180), [94(30), 50(30), 72(60) x 35], 72(600)
	Leu	CAT TAC TTT TAC TTG GAA TTT GCA CCA		
12S	12Sa	CTG GGA TTA GAT ACC CCA CTA T	Kocher *et al.* (1989)	94(180), [94(30), 50(30), 72(45) x 35], 72(600)
	12Sb	GAG GGT GAC GGG GCG GTG TGT		
acm4	TgF	CAA GCCTGA GAG CAA RAA GG	Gamble *et al.* (2008)	92(180), [92(30), 62↓0.5(30), 72(45) x 20], [92(30), 50(30), 72(45) x 15], 72(600
	TgR	ACY TGA CTC CTG GCA ATG CT		
βfib	BF8	CAC CAC CGT CTT CTT TGG AAC ACT G	Pinho *et al.* (2008)	92(180), [92(30), 62↓0.5(30), 72(60) x 20], [92(30), 50(30), 72(60) x 15], 72(600
	BfibR	CAG GGA GAG CTA CTT TTG ATT AGA C		
mc1r	MC1R-F	GGC NGC CAT YGT CAA GAA CCG GAA CC	Pinho *et al.* (2009)	92(180), [92(30), 62↓0.5(30), 72(60) x 25], [92(30), 50(30), 72(60) x 15], 72(600
	MC1R-R	CTC CGR AAG GCR TAG ATG ATG GGG TCC AC		
pdc	PHOF2	AGA TGA GCA TGC AGG AGT ATG A	Bauer *et al.* (2007)	92(180), [92(30), 58(30), 72(60) x 35], 72(600)
	PHOR1	TCC ACA TCC ACA GCA AAA AAC TCC T		
reln	62F	GAG TMA CTG AAA TAA ACT GGG AAA C	Pinho *et al.* (2009)	92(180), [92(30), 57(30), 72(60) x 35], 72(600)
	63R	GCC ATG TAA TYC CAT TAT TTA CAC TG		

RESULTS

A total of 529 sequences were obtained and used in the phylogenetic analyses, among which 518 sequences were newly generated for this study and 11 sequences of the species *Atlantolacerta andreanskyi* and *Timon tangitanus* were retrieved from GenBank. The percentage of missing data was 3% for *12S*, 12.5% for *nd4*, 4.5% for *acm4*, 15% for *βfib*, 2 % for *mc1r*, 19% for *pdc* and 42% for *reln*. The percentage of missing nucleotides is very low in the genes *12S*, *nd4* and *reln*, with 1-3 samples having 0.05% of the total length of the alignment missing, and higher for the genes *acm4*, *bfib* and *mc1r*, with an average of 10% of the length missing in a maximum of 6 specimens. The number of sequences, multiple sequence alignments length, models of sequence evolution and number of variable positions are reported for each gene in Table 2.3. Multiple sequences alignments of protein coding genes (*nd4*, *acm4*, *mc1r* and *pdc*) did not require gap positions and their translation into amino acid sequences contained no stop codons. Alignments of both the intronic regions *βfib* and *reln* showed high sequence length polymorphisms. The recombination tests applied in RDP did not find statistically significant evidence for recombination in any of the nuclear genes. All sequences were deposited in GenBank (Table 2.1).

Table 2.3. Number of sequences for each gene, length of the gene fragments, models of sequence evolution for unphased and phased data as selected by jModelTest according to the AICc and number of variable positions inferred in MEGA 5 for the dataset with and without outgroup.

Gene	Nº Seq.	Length (bp)	Model unphased data	Model phased data	Var. pos. Ingroup	Var. pos. with outgroup
12S	85	362	GTR+I+G		125	136
nd4	77	726	TrN+I+G		417	433
acm4	84	379	HKY+I+G	K80+G	73	90
β-fib	75	327	HKY	JC	143	175
mc1r	86	615	HKY+I+G	HKY+I+G	125	138
pdc	71	444	K80+I+G	K80+G	99	113
reln	51	681	HKY+G	HKY+G	265	318

PHYLOGENETIC RELATIONSHIPS WITHIN LACERTIDAE

Phylogenetic results from the concatenation analyses and species tree present three consistent traits: (i) in the Bayesian trees based on the concatenated datasets and in the species tree, in which the outgroup is not enforced, the subfamilies Gallotiinae and Lacertinae are reciprocally monophyletic sister taxa (Figs. 2.1 – 2.4); (ii) the Lacertini tribe presents a basal polytomy pattern with a lack of support for basal nodes in all trees; (iii) all genera are monophyletic with high support (Bayesian Posterior Probabilities (BPP) ≥ 95; Bootstrap Support (BS) ≥ 70), except *Algyroides*, whose monophyly is recovered in all the trees but with low node support, except in the BI tree based on the concatenated mtDNA topology where *Algyroides* is paraphyletic relative to *Dinarolacerta* (Fig. 2.1, in orange). Supported sister genera relationships recovered by all analyses with high support include *Scelarcis perspicillata* and *Teira dugesii* (Figs. 2.1 – 2.4, in green; BPP > 0.98, BS = 100). The sister genus relationships between *Darevskia* and *Iranolacerta* is recovered in all the trees and is statistically well supported (Fig. 2.1, 2.3 and 2.4, in pink; BPP > 0.94, BS > 83) except in the tree based on the concatenated nucDNA (Fig. 2.2). *Algyroides* and *Dinarolacerta* are also recovered as well supported sister taxa in the tree based on the concatenated mt-nucDNA and the species tree analyses (Figs. 2.3 – 2.4, in orange; BPP > 0.97, BS > 98), are recovered as sister taxa but not supported in the nucDNA tree (Fig. 2.2) and *Dinarolacerta* is nested within *Algyroides* in the BI tree based on the concatenated mtDNA (Fig. 2.1). Some clade relationships are recovered when using only one type of molecular markers: when using mtDNA data we recovered the relationships between *Anatololacerta* and *Parvilacerta* in the BI tree based on the concatenated dataset and species tree (Figs. 2.1, 2.3 and 2.4, in brown; BPP > 0.96) and the green lizard genera *Lacerta* and *Timon* in the trees based on the concatenated datasets (Figs. 2.1 and 2.3, in blue; BPP > 0.98, BS > 74); when using the nucDNA data we recovered a well-supported clade containing the taxa *Scelarcis perspicillata*, *Teira dugesii*, *Archaeolacerta bedriagae* and *Zootoca vivipara* (Figs. 2.2 – 2.4, in green; BPP = 1, BS > 87). The position of all the other genera, *Podarcis*, *Hellenolacerta*, *Phoenicolacerta*, *Takydromus*, *Iberolacerta*, *Dalmatolacerta* and *Apathya* is neither resolved nor consistent across the trees.

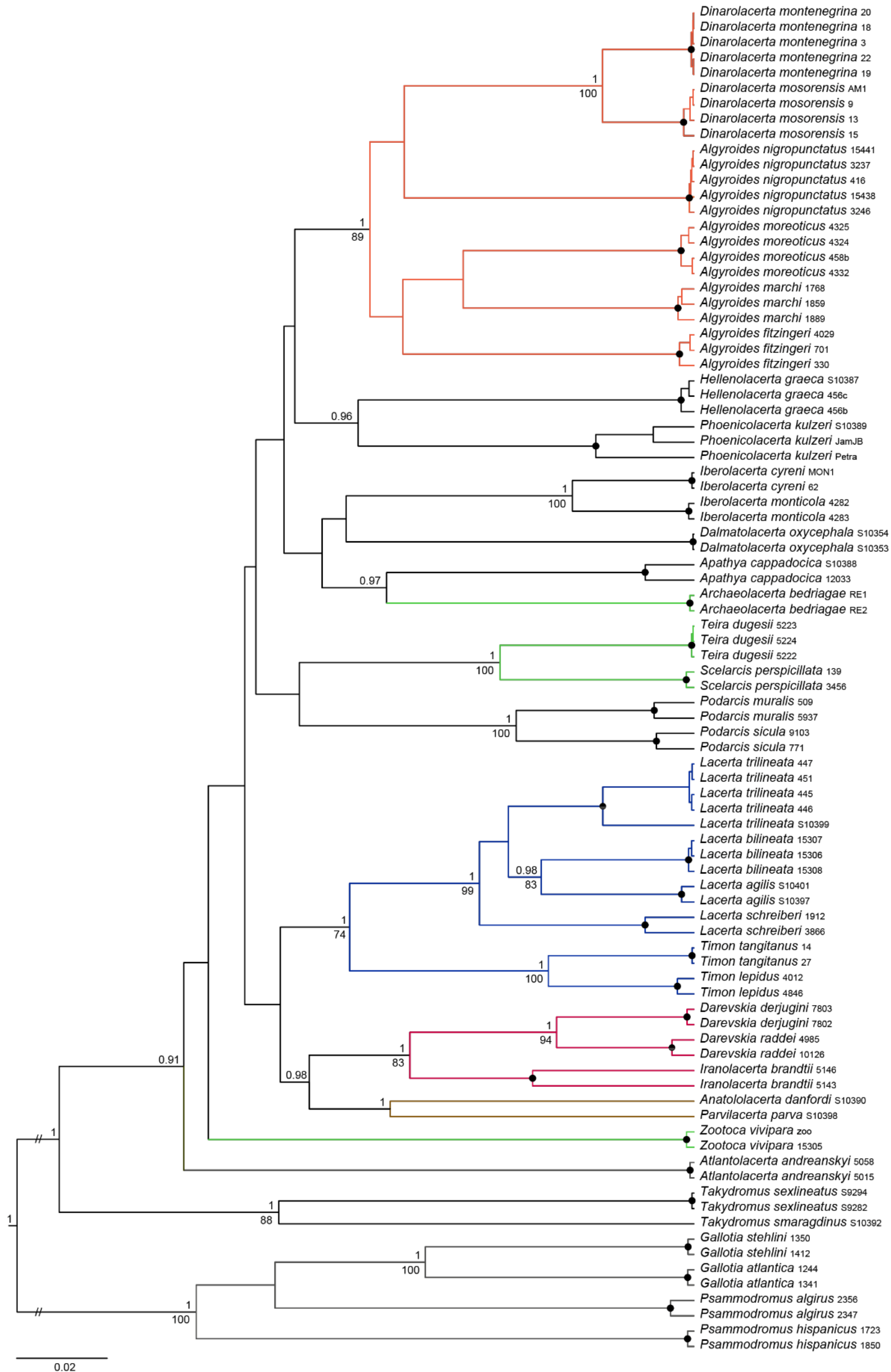


Figure 2.1. Phylogenetic relationships of Lacertini based on Bayesian analyses of concatenated mitochondrial DNA sequences (*12S* and *nd4*). Bayesian posterior probabilities (BPP) ≥ 0.90 are reported above nodes; bootstrap values from Maximum-Likelihood analyses (BS) ≥ 70 are reported below nodes. Within species, black dots represent support = 1 and 100 in both BI and ML analyses; grey part of the dots represent BPP of 0.9 ≥ 0.99 and BS of 70 ≥ 99, respectively.

COMPARISON BETWEEN MITOCHONDRIAL AND NUCLEAR TREES, ML AND BI TREES, AND THE SPECIES TREE

Overall we found higher Bayesian Posterior Probabilities (BPP) support than the Bootstrap Support (BS) when comparing BI and ML trees, irrespective of the dataset used. When comparing results obtained with different phylogenetic methods (BI *vs*. ML approach) or different datasets (mitochondrial *vs.* nuclear data), we found that the position and relationships of some taxa are more sensitive to the markers than to the method used, with some inconsistencies between the results based on mtDNA *vs.* nucDNA data. The Eremiadini species *Atlantolacerta andreanskyi* is sister taxon to the Lacertini tribe in all the analyses including mtDNA + nucDNA data (Figs. 2.3 – 2.4) but not in the trees based on the concatenated nucDNA, where it is positioned within Lacertini (Fig. 2.2 and Fig. S2.2) or in the BI tree based on the concatenated mtDNA, where it is sister taxon to all Lacertini with the exception of *Takydromus* (Fig. 2.1). The genus *Takydromus* is nested within the Lacertini tribe in all the trees, except in the BI tree based on the concatenated mtDNA where it is sister taxon to all the other Lacertini included in the analyses + *Atlantolacerta andreanskyi* (Fig. 2.1; BPP = 0.91). *Archaeolacerta* clustered in a group with *Zootoca*, *Scelarcis* and *Teira* in all the trees containing nucDNA data (Figs.2. 2 – 2.4, in green), in the tree based on the concatenated mtDNA it is either unresolved (ML tree, Fig. S2.1) or sister taxon to *Apathya* (BI tree, Fig. 2.1; BPP: 0.97). A relationship between *Hellenolacerta* and *Phoenicolacerta* is supported only in the BI tree based on the concatenated mtDNA (Fig. 2.1; BPP: 0.96). Regarding *Algyroides*, the monophyly of the genus and sister taxa relationships between the species *A. marchi* and *A. fitzingeri*, and *A. nigropunctatus* and *A. moreoticus* are recovered in all the trees including nucDNA data, the clade formed by *A. marchi* and *A. fitzingeri* with a high statistical support (Figs. 2.2 – 2.4, in orange; BPP > 0.97, BS = 73), whereas in the trees based on mtDNA data only, the genus is either monophyletic (ML trees; Fig. S2.1) or paraphyletic (BI; Fig. 2.1) and a closer relationship between *A. moreoticus* and *A. marchi* is recovered with low statistical support (BPP = 0.85, BS = 59).

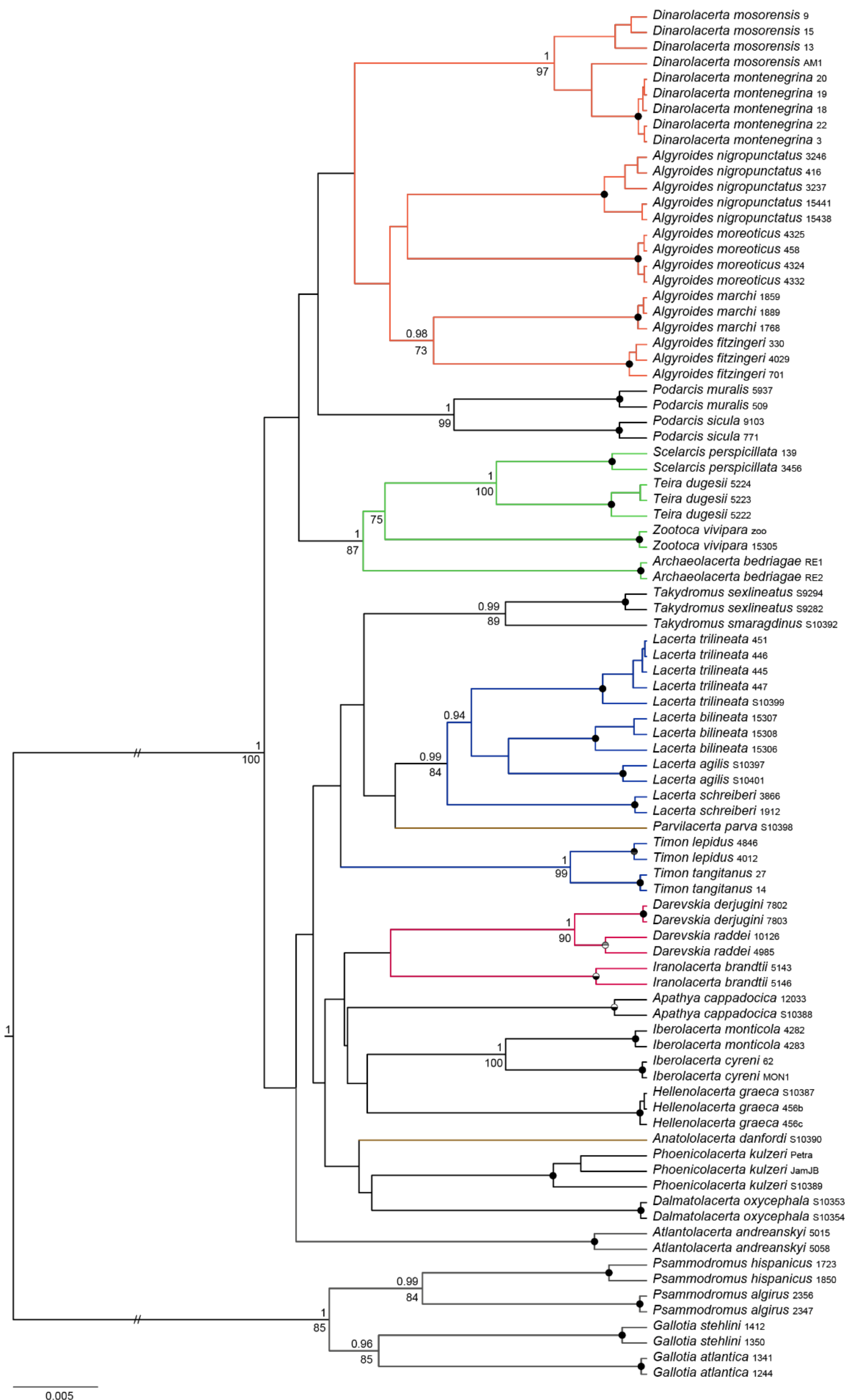


Figure 2.2. Phylogenetic relationships of Lacertini based on Bayesian analyses of concatenated nuclear DNA sequences (*acm4*, *βfib*, *mc1r*, *pdc* and *reln*). Bayesian posterior probabilities ≥ 0.90 are reported above nodes; bootstrap values from Maximum-Likelihood analyses ≥ 0.70 are reported below nodes. Within species, black dots represent support = 1 and 100 in both BI and ML analyses; grey part of the dots represent BPP of 0.9 ≥ 0.99 and BS of 70 ≥ 99, respectively and white part represents no support.

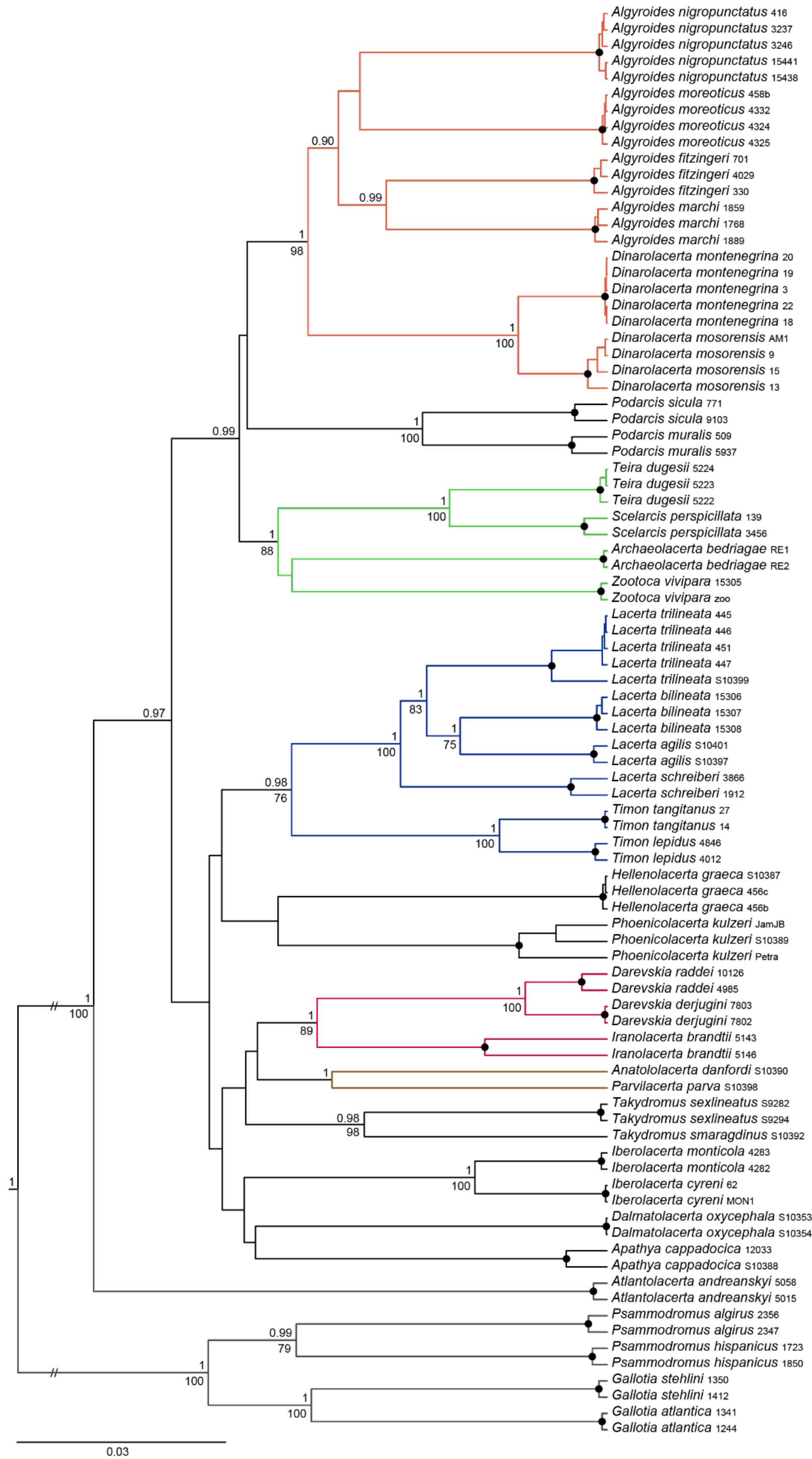


Figure 2.3. Phylogenetic relationships of Lacertini based on concatenated mitochondrial (*12S* and *nd4*) and nuclear (*acm4*, *βfib*, *mc1r*, *pdc* and *reln*) DNA sequences. Bayesian posterior probabilities ≥ 0.90 are reported above nodes; bootstrap values from Maximum-Likelihood analyses ≥ 0.70 are reported below nodes. Within species, black dots represent support = 1 and 100 in both BI and ML analyses.

MOLECULAR DATING

Molecular dating results are shown in Table 2.4, along with the 95% highest probability density (HPD) intervals. The divergence between Gallotiinae and Lacertinae is estimated at around 30 million years ago (Mya) (HPD: 11.77-51.28; node a), with divergence between *Gallotia* and *Psammodromus* about 15 Mya (HPD: 8.92-22.38; node q). Divergence between the tribes Lacertini and Eremiadini is estimated at around 17 Mya (HPD: 10.68-25.7; node b). Within Lacertini, the majority of basal splits are placed in a short time span of about 2.5 million years during the Middle Miocene (15-12.5 Mya). The time to most recent common ancestors (TMRCAs) of the Lacertini genera are estimated in the late Miocene (11-5 Mya; Fig. 2.4, nodes d, f, g, h, i, j, k, l, m, n, o, p).

Table 2.4. Age, in million years (Mya) and node 95% highest posterior density (HPD) intervals at for the major supported nodes in the species tree. Letters represent the nodes in the species tree (Fig. 2.4).

Node	Split/Clade	Age (Mya)	height 95% HPD (Mya)
a	Lacertinae - Gallotinae	33.22	11.77 - 51.28
b	Lacertini - Eremiadini	17.85	10.68 - 25.7
c	Lacertini	15.03	9.42 - 21.58
d	*Algyroides* - *Dinarolacerta*	11.29	6.98 - 16.48
e	*A. fitzingeri* - *A. marchi*	8.59	4.95 - 12.65
f	*Dinarolacerta* genus	2.9	1.35 - 4.76
g	*Podarcis* genus	6.85	3.79 - 10.42
h	*Iberolacerta* genus	4.32	2.09 - 6.85
i	*Archaeolacerta, Zootoca, Teira, Scelarcis*	12.85	7.95 - 18.66
j	*Teira* - *Scelarcis*	5.81	3.17 - 9.02
k	*Lacerta* genus	7.88	4.74 - 11.65
l	*Timon* genus	3.51	1.58 - 5.76
m	*Takydromus* genus	8.45	4.75 - 12.75
n	*Darevskia* - *Iranolacerta*	10.25	5.93 - 15.34
o	*Darevskia* genus	3.18	1.35 - 5.27
p	*Parvilacerta* - *Anatololacerta*	10.91	6.15 - 16.08
q	*Gallotia* - *Psammodromus*	14.92	8.92 - 22.38
r	*Gallotia* genus	10.8	5.74 - 16.55
s	*Psammodromus* genus	11.67	6.56 - 17.64

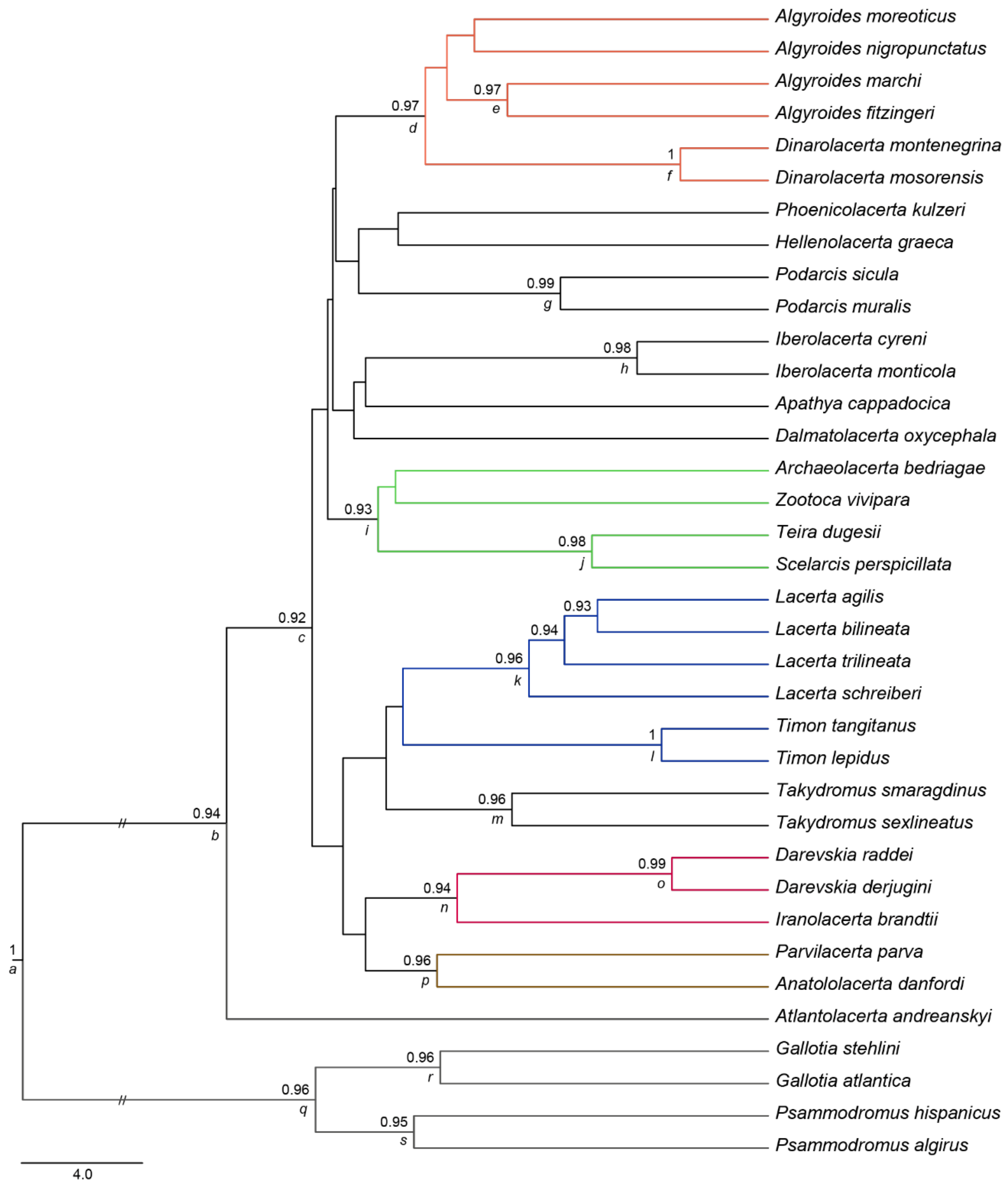


Figure 2.4. Species tree of Lacertini inferred from mitochondrial (*12S* and *nd4*) and nuclear (*acm4*, *βfib*, *mc1r*, *pdc* and *reln*) DNA sequences using the multispecies coalescent model in *BEAST software. The posterior probabilities ≥ 0.90 are shown above nodes. Node ages and 95% highest posterior density intervals (HPD) values for supported nodes (indicated by the letters a-s) are presented in Table 2.4.

TOPOLOGY TESTS

Supported nodes in ML trees recovered by previous studies were consistent with our results (Arnold et al., 2007; Fu, 2000, 1998; Harris et al., 1998; Hipsley et al., 2009; Mayer and Pavlicev, 2007; Pavlicev and Mayer, 2009), except the ML tree by Pyron et al. (2013) that shows six supported nodes (SHLaLRT values ≥ 85) that are not recovered in our ML tree. All the topological hypothesis constrained according to the three levels of support (support ≥ 85: six nodes; support ≥ 90: five nodes; and support ≥ 95: two nodes), were rejected by the SH and AU tests. Results are presented in Table 2.5.

Table 2.5. Results of topological tests using three sets of constrains based on relationships recovered in previous studies with a node support ≥ 85 or ≥ 90 or ≥ 95. The enforced relationships and p-value results of Shimodaira-Hasegawa (SH) and Approximately Unbiased (AU) tests are reported.

Node Support Level	Enforced relationships	SH	AU
95	1. (*Takydromus*, *Zootoca*) 2. (*Dalmatolacerta*, *Hellenolacerta*)	0.018	0.002
90	1. (*Takydromus*, *Zootoca*) 2. (*Dalmatolacerta*, *Hellenolacerta*) 3. ((*Timon*, *Lacerta*) (*Podarcis* (*Teira*, *Scelarcis*))) 4. (*Archaeolacerta*, *Apathya*) 5. (*Algyroides marchi*, *A. fitzingeri*) *Dinarolacerta*)	3e-004	6e-008
85	1. (*Takydromus*, *Zootoca*) 2. (*Dalmatolacerta*, *Hellenolacerta*) 3. ((*Timon*, *Lacerta*) (*Podarcis* (*Teira*, *Scelarcis*))) 4. (*Archaeolacerta*, *Apathya*) 5. (*Algyroides marchi*, *A. fitzingeri*) *Dinarolacerta*) 6. (*Dalmatolacerta*, *Hellenolacerta*, *Archaeolacerta*, *Apathya*, *Iberolacerta*, *Parvilacerta*, *Anatololacerta*, *Algyroides*, *Iranolacerta*, *Darevskia*)	3e-004	8e-006

DISCUSSION

The addition of faster evolving nuclear molecular markers and the use of multi-locus coalescent approaches to the infer the phylogeny of Lacertini enabled the detection of new relationships between genera and provided insights into previously open questions concerning genera monophyly and the rapid radiation of the tribe.

CORROBORATIONS AND ADVANCES IN THE LACERTINI PHYLOGENY

Taxonomy of Lacertidae as described by Arnold et al., (2007) is consistent with our study, with the subfamily Gallotiinae (*Gallotia* and *Psammodromus*) being sister taxon to the subfamily Lacertinae. Within the latter, the Eremiadini tribe, here represented by *Atlantolacerta*, is sister taxon to Lacertini. In the trees based on the concatenated nucDNA *Atlantolacerta* is placed within Lacertini (Fig. 2.2 and Fig. S2.2). This result may be caused either by the lack of a proper taxon sampling within Eremiadini or by the inadequacy of nuclear molecular data. A short time span between the split of the two Lacertinae tribes and the onset of radiations within each tribe could be out of the scope of the nuclear genes used, but relationships corroborate the taxonomy when using the mitochondrial markers either alone or in combination with nuclear data (Figs. 2.1, 2.3 and 2.4).

A completely new and very interesting phylogenetic relationship was detected in all the trees containing nuclear data between the genera *Archaeolacerta*, *Zootoca* and the sister taxa *Teira* and *Scelarcis* that formed a clade. The position of these taxa has been highly unstable across all the previous phylogenetic studies. *Archaeolacerta*, for instance, has been placed with *Algyroides* (Harris et al., 1998; Carranza et al., 2004), *Darevskia* (Pavlicev and Mayer, 2009), *Zootoca* (Fu, 2000; Hipsley et al., 2009), *Scelarcis* (Salvi et al., 2011) and *Apathya* (Pyron et al., 2013), often with low statistical support. In our results, the clade (*Archaeolacerta*, *Zootoca*, *Teira* and *Scelarcis*) is highly supported, although the position of *Archaeolacerta* and *Zootoca* is unresolved. The geographic distribution of these four genera is allopatric: *Archaeolacerta* is endemic to Corsica and Sardinia, which were separated from the Iberian plate around 30-27 Mya, although land connections with Europe and North Africa existed in the Messinian Salinity Crisis at 5.96 - 5.33 Mya (Duggen et al., 2003); *Zootoca* has the widest distribution of all lacertids, covering most of Eurasia north of the Mediterranean peninsulas; *Teira* has the westernmost distribution in the Madeira archipelago and *Scelarcis* is endemic to northwest Africa. In addition to unrelated geographic distribution, these genera are also

morphologically very different, all presenting unique morphological characters within Lacertini and many features found only in a minority of other Lacertini (Arnold et al., 2007). Therefore, the peculiar morphology of the members of this group, which is today represented by four effectively monotypic genera whose geographical distribution show little commonality, indicate that it is a relictual group that was once diverse and widespread. An ancient relationship between the members of this group during early speciation events in the tribe may explain why only the faster evolving nuclear markers used in this study provided enough phylogenetic signal for this relationship. The fact that this group had never been recovered in previous phylogenetic studies could be the result of different scenarios. For instance, the extinction of original lineages from this ancestral group or the possible loss of signal in the mtDNA due to saturation at deep nodes. Other possible cause is the use of slow nuclear markers in previous studies, which might have missed the discrimination between this old split and others slightly older. This finding emphasises the importance of using fast nuclear molecular markers in the phylogenetic inference of fast radiations as they may shed light on some basal polytomies when the clustering of internal nodes occur in a short time span, such as in the case of Lacertini.

The close relationship between the genera *Teira* and *Scelarcis* found in all our trees (Figs. 2.1 – 2.4) is consistent with previous studies. *Scelarcis* was once included in the genus *Teira* (Mayer and Bischoff, 1996), and it has been argued by Pavlicev and Mayer (2009) that they should be reunited again under the genus *Teira*. However, these genera exhibit unique morphological features and, considering the high intraspecific differentiation found within *Teira dugesii* and *Scelarcis perspicillata*, they may represent reciprocally monophyletic species complexes (Brehm et al., 2003; Perera et al., 2007). Moreover, from a taxonomic point of view the repetitive actions of splitting and lumping these two genera may produce taxonomic instability rather than simplifying it, and therefore we suggest to keep the taxonomy proposed by Arnold et al. (2007).

Concerning the monophyly of the Lacertini genera, a previous study by Pavlicev and Mayer (2009) raised the possibility that *Algyroides* could be paraphyletic, as in their results, the monophyletic *Dinarolacerta* clade is nested within the *Algyroides* clade. Our results confirm that these two genera are closely related and form a clade (Figs. 2.1 – 2.4). Moreover, the nucDNA data used in this study support the monophyly of *Algyroides,* which is recovered in all the trees based on nuclear data (Figs. 2.2 – 2.4, Fig. S2.3; but also in the ML mtDNA tree, see Fig. S2.1), whereas the paraphyly of this genus is recovered only in the BI mtDNA tree (Fig. 2.1). Since overall we have no support for paraphyly from molecular data, and considering that the four *Algyroides* species share unique morphological characters that distinguish them from all other lacertids (Arnold,

1973; Arnold et al., 2007) it is highly probable that this genus is monophyletic. Intrageneric relationships of *Algyroides* consist of two sister species pairs, the western clade of *A. marchi* and *A. fitzingeri* from southeast Spain and Corsica-Sardinia, and the eastern clade of *A. nigropunctatus* and *A. moreoticus* from the Balkan Peninsula and Peloponnese. The split between these two clades is represented by very long branches in all the trees, with a short internode between them and the clade including *Dinarolacerta* species. This pattern suggests a scenario where the split between the two lineages including two extant pairs of *Algyroides* sister taxa occurred soon after the cladogenesis between *Algyroides* and *Dinarolacerta*, likely followed by extensive extinction within *Algyroides* lineages which are today represented by four species with a relictual distribution. This would explain the well supported relationships between sister taxa within *Algyroides* and *Dinarolacerta* and the blurred relationships between these genera especially when using mitochondrial (Harris et al., 1999; this study) and slow evolving nuclear markers (Pavlicev and Mayer, 2009; Pyron et al., 2013).

Several sister taxa relationships recovered by our results were previously described, such as the case of the green lizards *Lacerta* and *Timon* (Arnold, 1973; Harris et al., 1998; Fu, 2000; Carranza et al., 2004; Arnold et al., 2007; Pyron et al., 2013). The species from these two genera are morphologically different from all other Lacertini, sharing a significantly bigger size and numerous non-molecular features that do not usually appear in the small body-sized lizards from the rest of the tribe (Arnold et al., 2007).

The sister taxa relationships recovered between the genera *Anatololacerta* and *Parvilacerta* and between *Darevskia* and *Iranolacerta* have already been described before (Harris et al., 1998; Carranza et al., 2004; Arnold et al., 2007; Mayer and Pavlicev, 2007; Hipsley et al., 2009; Pavlicev and Mayer, 2009; Pyron et al., 2013). Species in each of these genera pair occupy the same geographic regions, the former species occur in Anatolia and Middle-East, the latter in the Caucasus and Middle-East, suggesting that they diverged from a common ancestor living somewhere near their shared geographical area.

PHYLOGENETIC HYPOTHESES ON THE EVOLUTIONARY HISTORY OF THE LACERTINI

The phylogenetic position of all other genera of the tribe (*Apathya*, *Dalmatolacerta*, *Iberolacerta*, *Hellenolacerta*, *Podarcis*, *Phoenicolacerta* and *Takydromus*) is not resolved in this study, despite the addition of information from fast evolving nuclear DNA and the

application of coalescent-based phylogenetic methods. Therefore, our results support the hypothesis of a hard polytomy within the evolutionary tree of Lacertini (Pavlicev and Mayer, 2009). The basal polytomy observed in Lacertini would be indicative of a fast radiation, which, according to our molecular dating estimates, place the internal node divergence in a relatively short time span of about 2.5 million years in the Middle Miocene (from 15 to 12.5 Mya). A similar age for the radiation event has been described before by Pavlicev and Mayer (2009). The fast radiation hypothesis agrees with most of the previous molecular studies (Harris et al., 1998; Fu, 1998, 2000; Arnold et al., 2007; Mayer and Pavlicev, 2007) and was further corroborated by Pavlicev and Mayer (2009), but is in sharp contrast with the results from the supermatrix approach applied by Pyron et al. (2013), where the internal branching of the Lacertini (sub)tree is almost completely statistically supported. Results of topological tests comparing our ML tree with that of Pyron et al. (2013) indicate that differences between these trees are statistically significant even for relationships very highly supported in the latter study. While differences between our results and Pyron et al. (2013) can be explained by the use of different molecular data and phylogenetic methods, this cannot explain the differences between Pyron et al. (2013) and the previous studies. Indeed, Pyron et al. (2013) used mostly the data generated in previous studies and implemented the same concatenation approach. On the other hand, since Pyron and colleagues were focused on the species-level relationships between squamate reptiles rather than on Lacertini, they used a very large-scale taxon sampling including 4161 species of lizards and snakes and a non-squamate outgroup taxa, *Sphenodon punctatus*. Estimating a tree of this size required high-speed approximations of tree topology searches, substitution models parameter estimates, as well as to assess node support for which they relied on a non-parametric SHLaLRT approach, since bootstrap analysis was computationally intractable. Such a large-scale taxonomic focus also required the use of a large squamate sequence alignment and a non-squamate outgroup which are certainly appropriate to infer and root relationships between the main squamate lineages but maybe not optimal to assess relationships within Lacertini. These considerations suggest that the supermatrix approach may provide high support for relationships within tip-clades which are actually not supported and inconsistent with those phylogenetic studies, with a narrower taxonomic focus, from where the data used in the supermatrix originated.

Finally, while providing additional support that inferring basal relationships within Lacertini is challenging, this study also highlights how adding a few fast evolving nuclear markers helps to shed some light on many ancient relationships within the tribe. In this context, the application of Next Generation Sequencing approaches makes it possible to

generate information from thousands loci across the whole genome at a reasonable cost (McCormack et al., 2013), thus representing a promising research direction to further investigate early cladogenesis within Lacertini.

CONCLUSION

This study corroborates the difficulties in the recovery of the evolutionary history of Lacertini lizards, with strong evidence that these difficulties reflect a fast radiation event. Implementing nuclear data in the analyses allowed the recovery of novel phylogenetic relationships that solved some basal polytomies in previous studies, as well as support for the monophyly of *Algyroides*, and an overall increase in the node statistical support. Adding many informative nuclear DNA markers to the phylogenetic analyses proved more helpful to the recovery of the evolutionary history of Lacertini than applying different phylogenetic methods. This exemplifies the benefits of the use of fast evolving nuclear DNA to enhance recovery of ancient relationships in groups that experienced fast radiation and extensive extinctions within old lineages.

New data from fast evolving nuclear markers and the multi-species coalescent approach, implemented for the first time in this study, allowed comparisons to be made between two contrasting phylogenetic hypotheses for Lacertini drawn from previous studies focusing on Lacertidae or on a large-scale phylogeny of squamate reptiles. Through topological comparison of supported relationships, we found that the taxon-wide concatenated supermatrix approach provided high support for nodes that are not supported either by analyses of the original sequences data or by new data from this study. These findings have far reaching implication for comparative studies relying on megaphylogenies from supermatrices (Roquet et al., 2013). Indeed, while new large-scale phylogenies built compiling molecular data from previous studies in a supermatrix may be a valuable resource for comparative macroecological and macroevolutionary studies with a focus on wide taxonomic groups or on higher-level relationships, caution is needed when using megaphylogenies as a guide for integrating tip-clades - such as inter-generic - relationships into ecological studies. In the case of lacertids, relying on the phylogenetic estimates produced in the original studies in which the data were generated may be a better choice, especially when these studies are based on a comprehensive taxon and marker sampling. It remains to be investigated if this is a generality or if the case of Lacertini is a rare example in which the megaphylogeny approach appears to fail.

ACKNOWLEDGMENTS

We thank CIBIO colleagues for providing DNA samples, in particular to Ana Perera for providing DNA sequences of *Timon tangitanus* and *Scelarcis perspicillata*; the California Academy of Sciences, San Francisco for providing tissue samples of *Takydromus* specimens CAS219910 and CAS219939; the Asociación Herpetológica Española (AHE) for *Algyroides* photographs. This study was partially funded by the Project "Genomics and Evolutionary Biology" cofinanced by North Portugal Regional Operational Programme 2007/2013 (ON.2–O Novo Norte), under the National Strategic Reference Framework (NSRF), through the European Regional Development Fund (ERDF). JM, DJH, and DS are supported by the Fundação para a Ciência e a Tecnologia (FCT, Portugal): JM, doctoral grant SFRH/BD/81528/2011; DS, post-doctoral grant SFRH/BPD/105274/2014. SC is supported by Grant CGL2012-36970 from the Ministerio de Economía y Competitividad, Spain (co-funded by Fondos FEDER – EU).

REFERENCES

Anisimova, M., Gascuel, O., 2006. Approximate likelihood-ratio test for branches: a fast, accurate, and powerful alternative. Syst. Biol. 55, 539–552. doi:10.1080/10635150600755453

Arévalo, E., Davis, S.K., Sites, J.W.J., 1994. Mitochondrial DNA sequence divergence and phylogenetic relationships among eight chromosome races of the *Sceloporus grammicus* complex (Phrynosomatidae) in Central Mexico. Syst. Bioloy 43, 387–418. doi:10.1093/sysbio/43.3.387

Arnold, E.N., 1973. Relationships of the Palaearctic lizards assigned to the genera *Lacerta*, *Algyroides* and *Psammodromus* (Reptilia: Lacertidae). Bull. Br. Museum 25, 289–366.

Arnold, E.N., Arribas, O., Carranza, S., 2007. Systematics of the Palaearctic and Oriental lizard tribe Lacertini (Squamata: Lacertidae: Lacertinae), with descriptions of eight new genera. Magnolia Press. Zootaxa 1430, 1 – 86.

Baeckens, S., Edwards, S., Huyghe, K., Van Damme, R., 2015. Chemical signalling in lizards: an interspecific comparison of femoral pore numbers in Lacertidae. Biol. J. Linn. Soc. 114, 44–57. doi:10.1111/bij.12414

Barata, M., Carranza, S., Harris, D.J., 2012. Extreme genetic diversity in the lizard *Atlantolacerta andreanskyi* (Werner, 1929): A montane cryptic species complex. BMC Evol. Biol. 12, 167. doi:10.1186/1471-2148-12-167

Bauer, A.M., de Silva, A., Greenbaum, E., Jackman, T., 2007. A new species of day gecko from high elevation in Sri Lanka, with a preliminary phylogeny of Sri Lankan *Cnemaspis* (Reptilia, Squamata, Gekkonidae). Mitteilungen aus dem Museum für Naturkd. Berlin – Zool. R. 83, 22–32. doi:10.1002/mmnz.200600022

Brehm, A., Jesus, J., Spínola, H., Alves, C., Vicente, L., Harris, D.J., 2003. Phylogeography of the Madeiran endemic lizard *Lacerta dugesii* inferred from mtDNA sequences. Mol. Phylogenet. Evol. 26, 222–230. doi:10.1016/S1055-7903(02)00310-X

Brown, R.P., Terrasa, B., Pérez-Mellado, V., Castro, J. A, Hoskisson, P. A, Picornell, A, Ramon, M.M., 2008. Bayesian estimation of post-Messinian divergence times in Balearic Island lizards. Mol. Phylogenet. Evol. 48, 350–8. doi:10.1016/j.ympev.2008.04.013

Carranza, S., Arnold, E.N., 2012. A review of the geckos of the genus *Hemidactylus* (Squamata: Gekkonidae) from Oman based on morphology, mitochondrial and nuclear data, with descriptions of eight new species. Magnolia Press. Zootaxa 95, 1–95.

Carranza, S., Arnold, E.N., Amat, F., 2004. DNA phylogeny of *Lacerta* (*Iberolacerta*) and other lacertine lizards (Reptilia: Lacertidae): Did competition cause longterm mountain restriction? Syst. Biodivers. 2, 57–77. doi:10.1017/S1477200004001355

Castresana, J., 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol. Biol. Evol. 17, 540–52.

Cox, S.C., Carranza, S., Brown, R.P., 2010. Divergence times and colonization of the Canary Islands by *Gallotia* lizards. Mol. Phylogenet. Evol. 56, 747–57. doi:10.1016/j.ympev.2010.03.020

Degnan, J.H., Rosenberg, N.A., 2006. Discordance of species trees with their most likely gene trees. PLoS Genet. 2, e68. doi:10.1371/journal.pgen.0020068

Drummond, A.J., Suchard, M.A., Xie, D., Rambaut, A., 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. Mol. Biol. Evol. 29, 1969–73. doi:10.1093/molbev/mss075

Duggen, S., Hoernie, K., van den Bogaard, P., Rupke, L., Morgan, J.F., 2003. Deep roots of the Messinian salinity crisis. Nature 422, 602–606. doi:10.1038/nature01551.1.

Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32, 1792–7. doi:10.1093/nar/gkh340

Flot, J.-F., 2010. SEQPHASE: a web tool for interconverting PHASE input/output files and FASTA sequence alignments. Mol. Ecol. Resour. 10, 162–166. doi:10.1111/j.1755-0998.2009.02732.x

Flot, J.-F., Tillier, A., Samadi, S., Tillier, S., 2006. Phase determination from direct sequencing of length-variable DNA regions. Mol. Ecol. Notes 6, 627–630. doi:10.1111/j.1471-8286.2006.01355.x

Fu, J., 1998. Toward the phylogeny of the family Lacertidae: implications from mitochondrial DNA 12S and 16S gene sequences (Reptilia: Squamata). Mol. Phylogenet. Evol. 9, 118–30. doi:10.1006/mpev.1997.0456

Fu, J., 2000. Toward the phylogeny of the family Lacertidae - Why 4708 base pairs of mtDNA sequences cannot draw the picture. Biol. J. Linn. Soc. 71, 203–217. doi:10.1006/bijl.1999.0431

Gamble, T., Bauer, A.M., Greenbaum, E., Jackman, T.R., 2008. Evidence for Gondwanan vicariance in an ancient clade of gecko lizards. J. Biogeogr. 35, 88–104. doi:10.1111/j.1365-2699.2007.01770.x

Gibbs, M.J., Armstrong, J.S., Gibbs, A.J., 2000. Sister-Scanning: a Monte Carlo procedure for assessing signals in recombinant sequences. Bioinformatics 16, 573–582. doi:10.1093/bioinformatics/16.7.573

Harris, D.J., Arnold, E.N., Thomas, R.H., 1998. Relationships of lacertid lizards (Reptilia: Lacertidae) estimated from mitochondrial DNA sequences and morphology. Proc. Biol. Sci. 265, 1939–48. doi:10.1098/rspb.1998.0524

Harris, D.J., Arnold, E.N., Thomas, R.H., 1999. A phylogeny of the European lizard genus *Algyroides* (Reptilia: Lacertidae) based on DNA sequences , with comments on the evolution of the group. J. Zool. London 249, 49–60.

Harris, D.J., Carranza, S., Arnold, E.N., Pinho, C., Ferrand, N., 2002. Complex biogeographical distribution of genetic variation within *Podarcis* wall lizards across the Strait of Gibraltar. J. Biogeogr. 29, 1257–1262. doi:10.1046/j.1365-2699.2002.00744.x

Herrel, A., Huyghe, K., Vanhooydonck, B., Backeljau, T., Breugelmans, K., Grbac, I., Van Damme, R., Irschick, D.J., 2008. Rapid large-scale evolutionary divergence in morphology and performance associated with exploitation of a different dietary resource. Proc. Natl. Acad. Sci. 105, 4792–4795. doi:10.1073/pnas.0711998105

Heulin, B., Surget-Groba, Y., Sinervo, B., Miles, D., Guiller, A., 2011. Dynamics of haplogroup frequencies and survival rates in a contact zone of two mtDNA lineages of the lizard *Lacerta vivipara*. Ecography (Cop.). 34, 436–447. doi:10.1111/j.1600-0587.2010.06540.x

Hipsley, C.A., Himmelmann, L., Metzler, D., Müller, J., 2009. Integration of Bayesian molecular clock methods and fossil-based soft bounds reveals early Cenozoic origin of African lacertid lizards. BMC Evol. Biol. 9, 151. doi:10.1186/1471-2148-9-151

Kocher, T.D., Thomas, W.K., Meyer, A., Edwards, S.V., Pääbo, S., Villablanca, F.X., Wilson, A.C., 1989. Dynamics of mitochondrial DNA evolution in animals: amplification and sequencing with conserved primers. Proc. Natl. Acad. Sci. U.S.A. 86, 6196–6200.

Kubatko, L.S., Degnan, J.H., 2007. Inconsistency of phylogenetic estimates from concatenated data under coalescence. Syst. Biol. 56, 17–24. doi:10.1080/10635150601146041

Maddison, W.P., 1997. Gene Trees in Species Trees. Syst. Biol. 46, 523–536. doi:10.2307/2413694

Maddison, W.P., Maddison, D.R., 2003. Mesquite: a modular system for evolutionary analysis.

Martin, D., Rybicki, E., 2000. RDP: detection of recombination amongst aligned sequences. Bioinformatics 16, 562–563. doi:10.1093/bioinformatics/16.6.562

Martin, D.P., Lemey, P., Lott, M., Moulton, V., Posada, D., Lefeuvre, P., 2010. RDP3: a flexible and fast computer program for analyzing recombination. Bioinformatics 26, 2462–2463. doi:10.1093/bioinformatics/btq467

Martin, D.P., Posada, D., Crandall, K.A., Williamson, C., 2005. A Modified Bootscan Algorithm for Automated Identification of Recombinant Sequences and Recombination Breakpoints. AIDS Res. Hum. Retroviruses 21, 98–102. doi:10.1089/aid.2005.21.98

Mayer, W., Bischoff, W., 1996. Beiträge zur taxonomischen Revision der Gattung *Lacerta* (Reptilia: Lacertidae) Teil 1: *Zootoca*, *Omanosaura*, *Timon* und *Teira* als eigenstandige Gattungen. Salamandra 32, 163–170.

Mayer, W., Pavlicev, M., 2007. The phylogeny of the family Lacertidae (Reptilia) based on nuclear DNA sequences: convergent adaptations to arid habitats within the subfamily Eremiainae. Mol. Phylogenet. Evol. 44, 1155–63. doi:10.1016/j.ympev.2007.05.015

McCormack, J.E., Hird, S.M., Zellmer, A.J., Carstens, B.C., Brumfield, R.T., 2013. Applications of next-generation sequencing to phylogeography and phylogenetics. Mol. Phylogenet. Evol. 66, 526–38. doi:10.1016/j.ympev.2011.12.007

McVay, J.D., Carstens, B.C., 2013. Phylogenetic model choice: justifying a species tree or concatenation analysis. J. Phylogenetics Evol. Biol. 01, 1–8. doi:10.4172/2329-9002.1000114

Miller, M. A., Pfeiffer, W., Schwartz, T., 2010. Creating the CIPRES Science Gateway for inference of large phylogenetic trees, in: 2010 Gateway Computing Environments Workshop (GCE). IEEE, pp. 1–8. doi:10.1109/GCE.2010.5676129

Padidam, M., Sawyer, S., Fauquet, C.M., 1999. Possible emergence of new geminiviruses by frequent recombination. Virology 265, 218–25. doi:10.1006/viro.1999.0056

Pavlicev, M., Mayer, W., 2009. Fast radiation of the subfamily Lacertinae (Reptilia: Lacertidae): history or methodical artefact? Mol. Phylogenet. Evol. 52, 727–34. doi:10.1016/j.ympev.2009.04.020

Perera, A., Vasconcelos, R., Harris, D.J., Brown, R.P., Carretero, M.A., Pérez-Mellado, V., 2007. Complex patterns of morphological and mtDNA variation in *Lacerta perspicillata* (Reptilia; Lacertidae). Biol. J. Linn. Soc. 90, 479–490. doi:10.1111/j.1095-8312.2007.00736.x

Pinho, C., Harris, D.J., Ferrand, N., 2008. Non-equilibrium estimates of gene flow inferred from nuclear genealogies suggest that Iberian and North African wall lizards (*Podarcis* spp.) are an assemblage of incipient species. BMC Evol. Biol. 8, 63. doi:10.1186/1471-2148-8-63

Pinho, C., Rocha, S., Carvalho, B.M., Lopes, S., Mourão, S., Vallinoto, M., Brunes, T.O., Haddad, C.F.B., Gonçalves, H., Sequeira, F., Ferrand, N., 2009. New primers for the amplification and sequencing of nuclear loci in a taxonomically wide set of reptiles and amphibians. Conserv. Genet. Resour. 2, 181–185. doi:10.1007/s12686-009-9126-4

Posada, D., 2008. jModelTest: phylogenetic model averaging. Mol. Biol. Evol. 25, 1253–6. doi:10.1093/molbev/msn083

Poulakakis, N., Lymberakis, P., Valakos, E., Pafilis, P., Zouros, E., Mylonas, M., 2005. Phylogeography of Balkan wall lizard (*Podarcis taurica*) and its relatives inferred from mitochondrial DNA sequences. Mol. Ecol. 14, 2433–43. doi:10.1111/j.1365-294X.2005.02588.x

Pyron, R., Burbrink, F.T., Wiens, J.J., 2013. A phylogeny and revised classification of Squamata, including 4161 species of lizards and snakes. BMC Evol. Biol. 13, 93. doi:10.1186/1471-2148-13-93

Rambaut, A., Drummond, A., 2007. Tracer v1.4, Available from http://beast.bio.ed.ac.uk/Tracer.

Roquet, C., Thuiller, W., Lavergne, S., 2013. Building megaphylogenies for macroecology: taking up the challenge. Ecography (Cop.). 36, 13–26. doi:10.1111/j.1600-0587.2012.07773.x

Salvi, D., Bombi, P., Vignoli, L., 2011. Phylogenetic position of the southern rock lizard *Australolacerta australis* within the Lacertidae radiation. African J. Herpetol. 60, 60–69. doi:10.1080/21564574.2010.547609

Salvi, D., Capula, M., Bombi, P., Bologna, M.A., 2009. Genetic variation and its evolutionary implications in a Mediterranean island endemic lizard. Biol. J. Linn. Soc. 98, 661–676. doi:10.1111/j.1095-8312.2009.01313.x

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

Salvi, D., Harris, D.J., Kaliontzopoulou, A., Carretero, M.A., Pinho, C., 2013. Persistence across Pleistocene ice ages in Mediterranean and extra-Mediterranean refugia: phylogeographic insights from the common wall lizard. BMC Evol. Biol. 13, 147. doi:10.1186/1471-2148-13-147

Salvi, D., Schembri, P.J., Sciberras, A., Harris, D.J., 2014. Evolutionary history of the Maltese wall lizard *Podarcis filfolensis*: insights on the "Expansion-Contraction" model of Pleistocene biogeography. Mol. Ecol. 23, 1167–1187. doi:10.1111/mec.12668

Sambrook, J., Fritsch, E.F., Maniatis T., 1989. Molecular Cloning: A Laboratory Manual.

Shimodaira, H., 2002. An approximately unbiased test of phylogenetic tree selection. Syst. Biol. 51, 492–508. doi:10.1080/10635150290069913

Shimodaira, H., Hasegawa, M., 1999. Multiple comparisons of log-likelihoods with applications to phylogenetic inference. Mol. Biol. Evol. 16, 1114–1116.

Shimodaira, H., Hasegawa, M., 2001. CONSEL: for assessing the confidence of phylogenetic tree selection. Bioinformatics 17, 1246–1247. doi:10.1093/bioinformatics/17.12.1246

Silvestro, D., Michalak, I., 2012. raxmlGUI: a graphical front-end for RAxML. Org. Divers. Evol. 12, 335–337. doi:10.1007/s13127-011-0056-0

Smith, J.M., 1992. Analyzing the mosaic structure of genes. J. Mol. Evol. 34, 126–129.

Stamatakis, A., 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22, 2688–90. doi:10.1093/bioinformatics/btl446

Stephens, M., Scheet, P., 2005. Accounting for decay of linkage disequilibrium in haplotype inference and missing-data imputation. Am. J. Hum. Genet. 76, 449–62. doi:10.1086/428594

Stephens, M., Smith, N.J., Donnelly, P., 2001. A new statistical method for haplotype reconstruction from population data. Am. J. Hum. Genet. 68, 978–89. doi:10.1086/319501

Uetz, P., Freed, F., Jirí Hošek (eds). The Reptile Database. Available from http://www.reptile-database.org. (Accessed August 13, 2015)

Vanhooydonck, B., Van Damme, R., 1999. Evolutionary relationships between body shape and habitat use in lacertid lizards. Evol. Ecol. Res. 1, 785–803.

Weisrock, D.W., Smith, S.D., Chan, L.M., Biebouw, K., Kappeler, P.M., Yoder, A.D., 2012. Concatenation and concordance in the reconstruction of mouse lemur phylogeny: An empirical demonstration of the effect of allele sampling in phylogenetics. Mol. Biol. Evol. 29, 1615–1630. doi:10.1093/molbev/mss008

Supporting Information

Additional Supporting Information can be found in the Appendices section.

Figure S2.1. Phylogenetic relationships of Lacertini based on Maximum Likelihood analyses of the concatenated tree based on mitochondrial DNA.

Figure S2.2. Phylogenetic relationships of Lacertini based on Maximum Likelihood analyses of the concatenated nuclear DNA sequences.

Figure S2.3. Phylogenetic relationships of Lacertini based on Maximum Likelihood analyses of concatenated mitochondrial and nuclear DNA sequences.

Figure S2.4. Distribution and phylogenetic relationships of the Lacertini genera.

CHAPTER 3

Multilocus phylogeny of *Zamenis* ratsnakes

Article II. Evolution, biogeography and systematics of the Western Palaearctic ratsnakes *Zamenis*, with the designation of *Zamenis scalaris* comb. nov.

Joana Mendes[1,2,3*], Daniele Salvi[1,4,*], Salvador Carranza[2], D. James Harris[1,3]

In preparation

* These authors contributed equally to this work

[1] CIBIO Research Centre in Biodiversity and Genetic Resources, InBIO, Universidade do Porto, Campus Agrário de Vairão, Rua Padre Armando Quintas, N° 7, 4485-661 Vairão, Vila do Conde, Portugal

[2] Institute of Evolutionay Biology (CSIC-Universitat Pompeu Fabra), Passeig Marítim de la Barceloneta 37-49, 08003 Barcelona, Spain.

[3] Departamento de Biologia, Faculdade de Ciências, Universidade do Porto, Rua do Campo Alegre FC4 4169-007 Porto, Portugal

[4] Department of Health, Life and Environmental Sciences, University of L'Aquila, 67100 Coppito, L'Aquila, Italy

ABSTRACT

The phylogenetic relationships between Western Palaearctic *Zamenis* and *Rhinechis* ratsnakes have been unclear, with recent estimates based on the supermatrix approach questioning their monophyly and providing contradictory results. In this study, we generated a comprehensive molecular dataset for *Zamenis* and closely related ratsnakes to assess their phylogenetic and systematic relationships and infer their spatial and temporal modes of diversification. We obtained a fully-resolved and well-supported phylogeny, which is consistent across markers, taxon-sets and phylogenetic methods. The close phylogenetic relationships between *Rhinechis* and *Zamenis* is well-established. However, the phyletic order between the early branches within this clade, *R. scalaris* and *Z. hohenackeri*, remains poorly supported. The Persian ratsnake *Z. persicus* is sister to the Mediterranean species *Z. situla*, *Z. longissimus* and *Z. lineatus*, and among them *Z. situla* branching off earlier than Aesculapian snakes. Topological tests based on our data and evidence from a recent phylogenomic studies strongly rejected previous phylogenetic estimates based on the supermatrix approach, and demonstrate that these megaphylogenies, with hundreds of taxa and high levels of missing data, have recovered wrong relationships and spurious nodal support. Biogeographic and molecular dating analyses suggest an origin of the ancestor of *Rhinechis* and *Zamenis* in the Aegean region with early cladogenesis during Late Miocene associated with the Aegean arch formation, and support a scenario of East-to-West diversification. Finally, while we have little morphological and phylogenetic evidence for the distinctiveness between *Rhinechis* and *Zamenis*, a classification lumping them in a single genus would better reflect their evolutionary relationships; based on the priority rule, *Rhinechis scalaris* is moved into the genus *Zamenis* and designated as *Zamenis scalaris* **comb. nov**.

KEYWORDS

Ratsnakes, *Zamenis*, species tree, multilocus phylogeny, fast radiation, supermatrix

INTRODUCTION

Ratsnakes represent an evolutionary lineage of the subfamily Colubrinae Oppel, 1811 within Colubridae Oppel, 1811, sometimes considered as a distinct subfamily, Coronellini Jan, 1863 (Utiger, Schatti, & Helfenberger, 2005). Some 90 species and 20 genera of ratsnakes are currently recognized (Uetz, P., Freed, P., & Jirí Hošek (eds), The Reptile Database, http://www.reptile-database.org/, accessed August 2017), mainly distributed in the Holarctic region but also in the Oriental and norther portions of the Neotropical regions. Over the last decades, molecular studies have contributed to a drastic reconsideration of ratsnakes' systematics and evolution. Until the end of the last century, most species distributed in the Northern Hemisphere were included in a single genus *Elaphe* auct. based on shared morphological features (see Schulz, 1996). Extensive research based on DNA sequence data restricted the genus *Elaphe* Fitzinger, 1833 to a dozen of Palaearctic species and provided strong phylogenetic evidence for a monophyletic origin of New World ratsnakes (Lampropeltini), which were assigned to distinct genera (e.g. Burbrink & Lawson, 2007; Rodríguez-Robles & De Jesús-Escobar, 1999; Utiger et al., 2002, 2005).

Western Palaearctic ratsnakes belong to four distinct lineages (Helfenberger, 2001; Lenk, Joger, & Wink, 2001; Schulz, 1996; Utiger et al., 2002). These include representatives of the genera *Coronella* Laurenti, 1768, *Elaphe* and two genera more recently designated. Helfenberger (2001), based on anatomical and allozyme data, resurrected the monotypic genus *Rhinechis* Michahelles, 1833 for the Western Mediterranean species *R. scalaris* Schinz, 1822. Utiger et al. (2002), based on mitochondrial DNA (mtDNA) sequence data, established the monophyletic genus *Zamenis* Wagler, 1830 including the Mediterranean species *Z. longissimus* Laurenti, 1768, *Z. lineatus* Camerano, 1891 and *Z. situla* Linnaeus, 1758, and the species *Z. persicus* Werner, 1913 and *Z. hohenackeri* Strauch, 1873 from the Middle East in Southwestern Asia (Fig. 3.1).

Phylogenetic estimates of the relationships between *Zamenis* and *Rhinechis* have been contradictory in different studies (Fig. 3.2). First molecular studies based mainly on mtDNA showed *R. scalaris* either branching off early in a solitary phylogenetic lineage within the European ratsnakes (Lenk et al., 2001; Utiger et al., 2002) or sister to the *Zamenis* clade (Burbrink & Lawson, 2007; Utiger et al., 2002), although these phylogenetic relationships received low statistical support. In contrast, recent studies based on the supermatrix approach, with hundreds or thousands of Colubridae or Squamata taxa, recovered *R. scalaris* deeply nested within *Zamenis* in a highly supported clade with

resolved intra-clade relationships (Figueroa, McKelvy, Grismer, Bell, & Lailvaux, 2016; Pyron et al., 2011; Pyron, Burbrink, & Wiens, 2013; Zheng & Wiens, 2016; see Fig. 3.2). Moreover, in the phylogeny obtained by Figueroa et al. (2016), also the Indo-Chinese ratsnake *Ptyas korros* Schlegel, 1837 was nested within *Zamenis*, as sister species to *Z. longissimus*, in a highly supported clade. These megaphylogenies have questioned the monophyly of *Zamenis* and its distinction from *Rhinechis*. Furthermore, the relationships between *Zamenis* species are unstable across these studies, with each work presenting a different topology (except the trees from Pyron et al. (2013) and Zheng & Wiens (2016), because these used the same dataset).

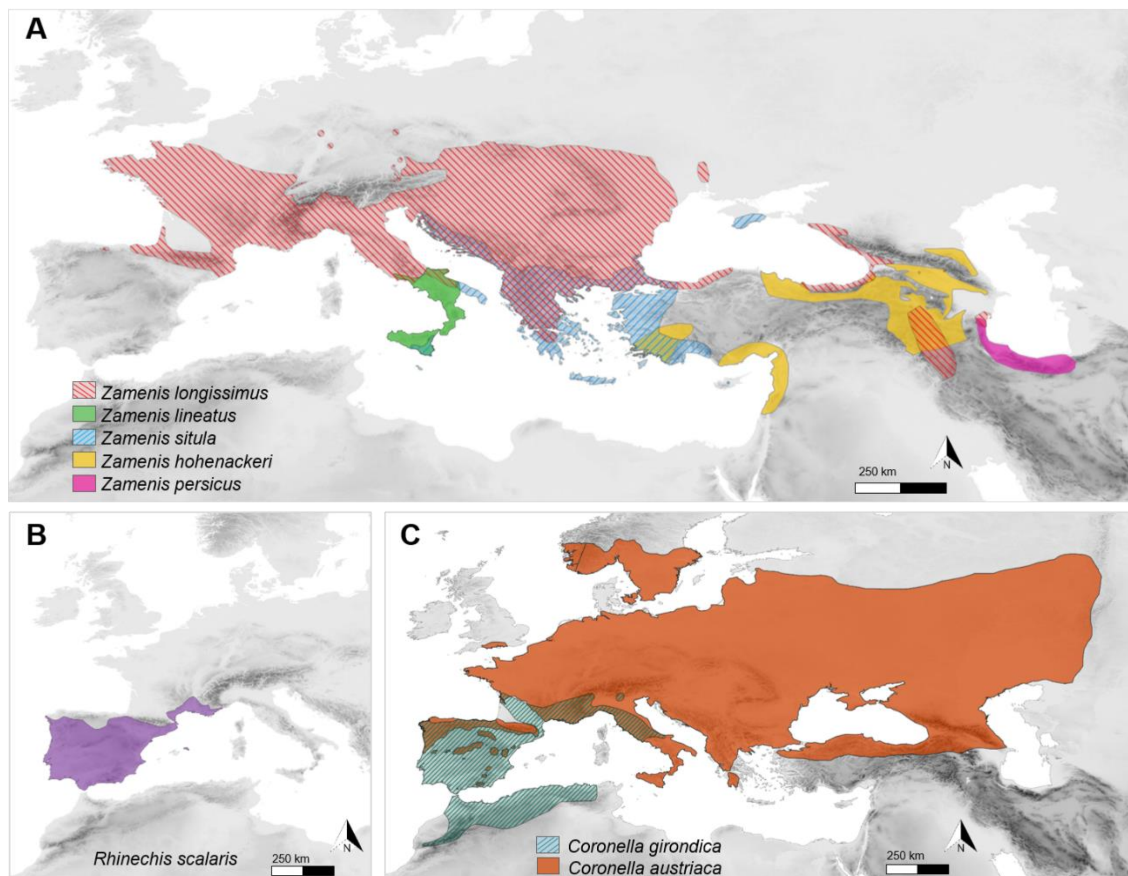


Figure 3.1. Distribution of the ratsnakes genera *Zamenis*, *Rhinechis* and *Coronella* in the Palaearctic: a) distribution of *Zamenis* species*; b) distribution of *Rhinechis scalaris*; c) distribution of *Coronella* species (*: in the area of range overlap between *Z. lineatus* and *Z. longissimus* in southern Italy, Salvi et al. (2017) found introgressed individuals instead of pure forms of either species). Distribution of *Zamenis*, *Rhinechis* and *Coronella* girondica retrieved from the IUCN Red List of Threatened Species, and of *C. austriaca* adapted from Santos et al. (2015).

Rapid evolution of Palaearctic ratsnakes, and within *Zamenis,* might be an additional explanation, besides methodological artefacts, for the phylogenetic instability of this clade in previous molecular studies. Rapid radiations generate a phylogenetic pattern

with short deep branches combined with long terminal branches, which is notoriously difficult to resolve in phylogenetic analysis, especially when this is based on fast evolving characters such as mtDNA sequences (Cummins & McInerney, 2011; Hoelzer & Meinick, 1994). Indeed, previous studies showed deep phylogenetic divisions, with associated low statistical support, between Palaearctic ratsnakes (Lenk et al., 2001; Utiger et al., 2002) and suggest that most cladogenetic events – such as those leading to *Rhinechis* and *Zamenis* - took place in a short time frame (Burbrink & Lawson, 2007). Rapid diversification within *Zamenis* may have been associated with the invasion of the Western Palearctic from the east as suggested by the east-west pattern of phyletic diversification, with eastern species *Z. hohenackeri* and *Z. persicus* branching off earlier than Mediterranean species *Z. situla*, *Z. longissimus* and *Z. lineatus* (Utiger et al., 2002). However, this biogeographical hypothesis has not yet been formally tested.

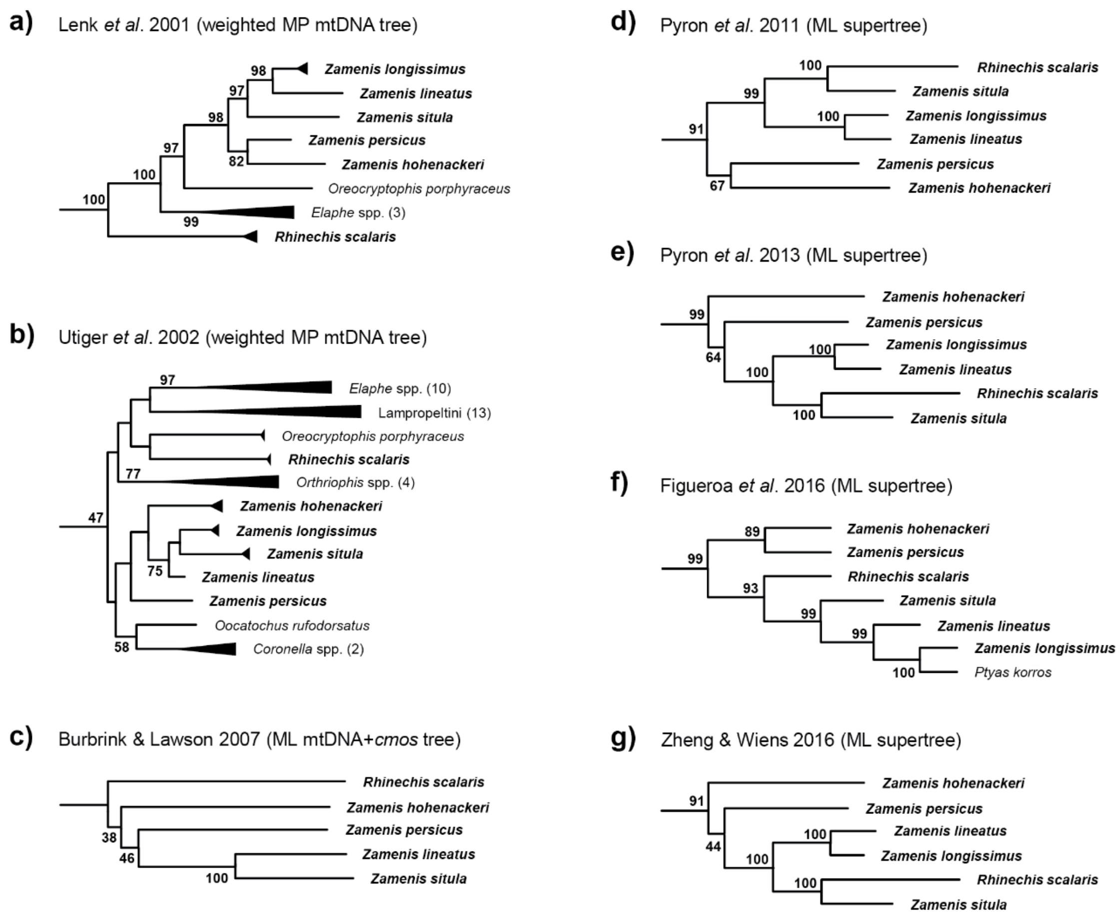


Figure 3.2. Summary of the previous phylogenetic hypotheses on *Zamenis* and *Rhinechis* based on molecular data (mainly mitochondrial data) and different taxon-sets: ratsnakes taxon-sets (a-c), Serpentes taxon-sets (d, f), and Squamata taxon-sets (e, g).

In this study, we generated a comprehensive molecular phylogenetic framework for *Zamenis* and closely related ratsnakes based on seven gene fragments and multiple individuals from each species. We tested the robustness of this phylogeny to ingroup and outgroup choice, marker choice (mitochondrial *vs* nuclear) and missing data, phylogenetic methods (Maximum Likelihood *vs*. Bayesian Inference) and approach (gene trees *vs*. species trees). Molecular dating and biogeographic analysis were used to further explore the temporal and spatial pattern of phyletic diversification. The main aims of this study are to infer the evolutionary and biogeographic history of *Zamenis* and *Rhinechis* ratsnakes and to assess their phylogenetic and systematic relationships resolving the current controversial estimates based on mitochondrial and supermatrix datasets.

MATERIAL AND METHODS

SAMPLE AND SEQUENCE DATA COLLECTION

We collected, either from live individuals or museum specimens, muscle tissues from three samples of each *Zamenis* species, of *R. scalaris* and of *Coronella austriaca* Laurenti, 1768 and *C. girondica* Daudin, 1803 and two samples of the species *Hierophis viridiflavus* Lacépède, 1789 and *Hemorrhois algirus* Jan, 1863. Sequences obtained from these specimens were complemented with GenBank data including, whenever possible, samples from a close geographical location to our specimens. Additional sequences were downloaded for Old and New World ratsnake species of the genera *Elaphe*, *Lampropeltis* Fitzinger, 1843, *Pantherophis* Fitzinger, 1843, *Pituophis* Holbrook, 1842 and *Oocatochus* Helfenberger, 2001 to be used in the phylogenetic analyses for ingroup and outgroup testing. Information regarding the original specimens used in the analyses, the sampling localities and the GenBank accession numbers of new and published sequences is given in Table 3.1.

Total genomic DNA was extracted from alcohol-preserved muscle following the standard high-salt protocol (Sambrook, Fritsch, & Maniatis T., 1989). We generated DNA sequences from seven gene fragments, including two mitochondrial genes: cytochrome-*b* (*cytb*) and NADH Dehydrogenase 4 plus flanking tRNAs Serine, Histidine and Leucine (*nd4*); and five nuclear genes: oocyte maturation factor mos (*cmos*), dynein axonemal heavy chain 3 (*dnah3*), prolactin receptor (*prlr*), spectrin beta, non-erythrocytic 1 intron 1 (*sptbn1*) and vimentin intron 5 (*vim*). Target sequences were amplified through polymerase chain reaction (PCR) in 25µL volume, containing 1X PCR buffer (50mm Tris–HCl, 50mm NaCl, pH 8.5); 3mM $MgCl_2$; 0.6mM each dNTP, 2U of

GoTaq DNA polymerase (Promega), 0.4μM each primer and approximately 50ng of genomic DNA. Amplification conditions consisted of a preliminary denaturation step at 94°C for 5 minutes, followed by 35 cycles of 30 seconds denaturation at 94°C, 40 seconds at varying annealing temperature according to the gene (48.5ºC for *cytb*; 50ºC for *nd4*; 51.5ºC for *dnah3*, *vim* and *sptbn1*; 54ºC for *prlr* and 55ºC for *cmos*) and 60 seconds of extension at 72°C. A final extension was carried out at 72°C for 10 minutes [see Salvi et al. (2017) for further details]. The primers used for amplification of each gene fragment are reported in Table 3.2. Purification and sequencing of PCR products were carried out by a commercial sequencing company (GENEWIZ: www.genewiz.com), using the same primers employed for amplification.

Table 3.1. Code, location coordinates and GenBank accession numbers of the colubrid specimens used in this study ([1]: samples excluded in the reduced dataset analyses; *: sequences downloaded from GenBank, XX: accession numbers requested to GenBank).

Species	Sample	Sampling locality	Geographic Coordinates	*nd4*	*cytb*	*cmos*	*dnah3*	*prlr*	*sptbn1*	*vim*
Zamenis longissimus	z106	Bressanone, Italy	46.779 / 11.644	KY495534*	HQ392561*	KY495516*	-	-	KY495573*	XX
Zamenis longissimus	z85	Penne, Italy	42.458 / 13.929	KY495542*	HQ392562*	KY495520*	XX	XX	KY495577*	XX
Zamenis longissimus	z93	Varco Sabino, Italy	42.241 / 13.021	KY495545*	HQ392564*	KY495524*	-	XX	KY495581*	XX
Zamenis lineatus	z204	Pietrapertosa, Italy	40.519 / 16.063	KY495560*	HQ392567*	KY495529*	XX	-	KY495588*	XX
Zamenis lineatus	z68	Torre Salsa, Italy	37.337 / 13.424	KY495570*	KF639746*	KY495526*	-	-	KY495591*	XX
Zamenis lineatus	z99	Potame, Italy	39.189 / 16.199	KY495565*	KF639747*	KY495531*	-	XX	KY495590*	XX
Zamenis persicus	SARI	Sari, Iran	36.565 / 53.058	XX	XX	XX	XX	XX	XX	XX
Zamenis persicus	GB1[1]	-	-	DQ902297*	HQ392568*	DQ902075*	-	-	-	KM870881*
Zamenis persicus	ZPF	Neka, Iran	36.647 / 53.298	XX	XX	XX	XX	-	XX	XX
Zamenis situla	SF01	Putignano, Italy	40.836 / 17.080	XX	KF639748*	XX	XX	XX	XX	XX
Zamenis situla	GB2[1]	-	-	DQ902303*	DQ902125*	DQ902083*	-	-	-	-
Zamenis situla	SIT1	Aristi, Greece	39.941 / 20.676	XX	JX315468*	XX	XX	XX	XX	XX
Zamenis hohenackeri	1591	Borçka, Turkey	41.333 / 41.640	XX	XX	XX	-	XX	XX	XX
Zamenis hohenackeri	GB3	-	-	DQ902320*	DQ902137*	DQ902098*	-	-	KM870820*	KM870880*
Zamenis hohenackeri	1589	Ağri, Turkey	39.484 / 43.987	XX	XX	XX	-	XX	XX	XX
Rhinechis scalaris	DB22	Guadalupe, Spain	39.514 / -5.347	XX	XX	XX	XX	XX	XX	XX
Rhinechis scalaris	DB38	Resende, Portugal	41.103 / -7.964	XX	XX	XX	XX	XX	XX	XX
Rhinechis scalaris	DB41[1]	Fermoselle, Spain	41.318 / -6.398	XX	XX	XX	-	-	-	FJ627909*
Coronella austriaca	DB4002	Espinho, Portugal	41.027 / -8.645	XX	EU022663*	XX	-	XX	XX	XX
Coronella austriaca	GB5	-	-	AY487065*	AY486930*	AY486954*	-	-	FJ627921*	-
Coronella austriaca	DB4003[1]	Espinho, Portugal	41.027 / -8.645	-	EU022668*	XX	-	-	-	XX
Coronella girondica	DB2686	Alvão, Portugal	41.350 / -7.783	AY487066*	JQ837595*	XX	-	XX	XX	XX
Coronella girondica	DB2689	Sopeira, Portugal	42.316 / 0.733	XX	JQ837597*	XX	XX	XX	XX	XX

Coronella girondica	DB1725[1]	Okaimeden, Morocco	31.208 / -7.860	-	XX	AF471113*	-	-	-	-
Hemorrhois algirus	DB1525	Michelifene, Morocco	32.262 / -5.146	AY487037*	XX	XX	XX	XX	XX	XX
Hemorrhois algirus	DB1562	Imellalen, Morocco	32.221 / -4.676	XX	XX	AY486935*	XX	XX	-	XX
Hierophis viridiflavus	EL304	Italy		XX	XX	XX	XX	XX	-	XX
Hierophis viridiflavus	GB4[1]	-	-	LN552062*	LN551980*	AY376803*	-	LN551943*	-	-
Hierophis viridiflavus	DB2291	Lesina, Italy	41.883 / 15.433	XX	LN551964*	XX	XX	XX	-	XX
Elaphe quatuorlineata	GB10	-	-	KF728020*	AY487067*	AY486955*	-	-	JX648617*	KM870871*
Elaphe quatuorlineata	GB11	-	-	KF728019*	-	-	-	-	-	JX648636*
Elaphe carinata	GB12	-	-	KF669244*	JN799414*	JN799416*	-	-	KF669162*	KF669217*
Elaphe carinata	GB13	-	-	KF669243*	JN799413*	JN799415*	-	-	KF669161*	KF669216*
Elaphe taeniura	GB14	--	-	KF669248*	DQ902305*	DQ902087*	-	-	KF669166*	KF669221*
Elaphe taeniura	GB15	-	-	KF669247*	AH015912*	EF076705*	-	-	KF669165*	KF669220*
Oocatochus rufodorsatus	GB16	-	-	DQ902123*	DQ902301*	DQ902081*	-	-	KM870834*	-
Oocatochus rufodorsatus	GB17	-	-	JQ798793*	NC022146*	AF435015*	-	-	-	-
Pituophis catenifer	GB18	-	-	KX835885*	JF308311*	FJ627790*	-	-	FJ627939*	FJ627902*
Pituophis deppei	GB19	-	-	KX694867*	AF138766*	KX694814*	-	-	FJ627924*	FJ627897*
Pituophis deppei	GB20	-	-	FJ627818*	AF138765*	FJ627801*	-	-	-	-
Lampropeltis triangulum	GB21	-	-	KF216143*	FJ627850*	FJ627798*	-	-	FJ627938*	FJ627888*
Lampropeltis triangulum	GB22	-	-	KF216244*	-	-	-	KF215308*	KF215163*	KF215552*
Lampropeltis ruthveni	GB23	-	-	AF337065*	AY739642*	FJ627803*	-	KF215243*	KF215130*	FJ627882*
Lampropeltis ruthveni	GB24	-	-	AF337064*	AY739641*	-	-	KF215244*	KF215129*	-
Lampropeltis calligaster	GB25	-	-	KF216289*	DQ902311*	DQ902091*	-	KF215239*	KF215126*	KF215529*
Lampropeltis calligaster	GB26	-	-	KX881114*	AY739644*	-	-	KF215240*	KF215135*	KF215534*
Pantherophis obsoletus	GB27	-	-	DQ538344*	KM655218*	FJ627805*	-	-	FJ627930*	FJ627886*
Pantherophis obsoletus	GB28	-	-	DQ538343*	DQ902296*	AF471140*	-	-	KM655046*	-
Ptyas korros	GB6	-	-	AY487062*	KX694869*	KX694817*	-	-	KF669171*	KF669227*
Ptyas korros	GB7	-	-	-	AY486929*	AY486953*	-	-	KF669172*	KF669226*
Ptyas mucosa	GB8	-	-	KR814726*	KR814695*	KR814679*	-	-	KM870839*	KM870878*
Ptyas mucosa	GB9	-	-	NC030041*	LC105628*	AF471151*	-	-	-	-

PHYLOGENETIC DATASETS

DNA sequences were edited in Geneious (Kearse et al., 2012) and heterozygous positions in the nuclear genes were coded according to the IUPAC ambiguity codes. All sequences were aligned using the MUSCLE algorithm (Edgar, 2004) implemented in Geneious with default parameters. Insertions or deletions (indels) in the intronic fragments *sptbn1* and *vim* were manually phased. Haplotype reconstruction of the nuclear genes was performed in PHASE 2.1 (Stephens, Smith, & Donnelly, 2001; Stephens & Scheet, 2005) using the input files obtained in SeqPHASE (Flot, 2010; available at http://seqphase.mpg.de/seqphase/). We ran PHASE three times for each nuclear gene to ensure consistency, applying a probability threshold of 0.7, 100 iterations and the remaining settings as default. The possible occurrence of recombination events was assessed using the Pairwise Homoplasy Index (phi) test (Bruen, Philippe, & Bryant, 2006) implemented in SplitsTree4 (Huson & Bryant, 2006). We used MEGA 6 (Tamura, Stecher, Peterson, Filipski, & Kumar, 2013) to estimate the number of variable and parsimony informative sites in all genes.

Table 3.2. Name, sequence and reference of the primers used in this study

Gene	Primer name	Primer sequence (5'-3')	Reference
nd4	ND4	CAC CTA TGA CTA CCA AAA GCT CAT GTA GAA GC	(Arevalo, Davis, & Sites, 1994)
	LEU	CAT TAC TTT TAC TTG GAA TTT GCA CCA	
cytb	GluDG	TGA CTT GAA RAA CCA YCG TTG	(Palumbi et al., 1991)
	cytb2	CCC TCA GAA TGA TAT TTG TCC TCA	
cmos	S77	CAT GGA CTG GGA TCA GTT ATG	(Lawson, Slowinski, Crother, & Burbrink, 2005)
	S78	CCT TGG GTG TGA TTT TCT CAC CT	
dnah3	DNAH3_f1	GGT AAA ATG ATA GAA GAY TAC TG	(Townsend, Alegre, Kelley, Wiens, & Reeder, 2008)
	DNAH3_r6	CTK GAG TTR GAH ACA ATK ATG CCA T	
prlr	PRLR_f1	GAC ARY GAR GAC CAG CAA CTR ATG CC	(Townsend et al., 2008)
	PRLR_r3	GAC YTT GTG RAC TTC YAC RTA ATC CAT	
vim	Vim_Ex5_F2	AAC AAT GAT GCC CTG CGC CA	(Pyron & Burbrink, 2009)
	Vim_Ex6_R2	CAA TAT CAA GAG CCA TCT TTA CAT T	
sptbn1	SPTBN1-F	TTG GTC GAT GCC AGT TGT A	(Chen et al., 2014)
	SPTBN1-R	CAG GGT TTG TAA CCT KTC CA	

Phylogenetic inference was performed by (i) single-locus network analyses, based on full length sequences of individual nuclear gene alignments and the concatenated mtDNA (*cytb*+*nd4*) alignment; (ii) multi-locus Maximum Likelihood (ML) and Bayesian Inference (BI) methods, based on a concatenated alignment of unphased nuclear genes (nucDNA) and with the mitochondrial and unphased nuclear genes combined (mt-nucDNA); and (iii) the multi-locus coalescent species tree approach based on single gene datasets for mtDNA and phased nucDNA. *Hierophis viridiflavus* and *Hemorrhois algirus* were used as outgroups in ML analyses following Burbrink & Lawson (2007).

In order to test for the effect of missing data in the phylogenetic inference, we performed the phylogenetic analysis with two distinct datasets: (i) with all samples ('complete dataset', with 20.7% of missing sequences), and (ii) excluding samples that missed data for more than two gene fragments ('reduced dataset', with 9.8% of missing sequences) (Table 3.1).

In order to test for the effect of taxon set in the phylogenetic inference, and to compare our results with previous studies, we performed multi-locus phylogenetic analyses (ML and BI analyses based on concatenated loci and species tree) with two different taxon-sets: (i) the *Zamenis* taxon-set, including species of the genera *Zamenis*, *Rhinechis*, *Coronella* and the outgroups *Hierophis viridiflavus* and *Hemorrhois algirus*; and (ii) the *Zamenis* + ratsnakes taxon-set adding GenBank data from New and Old World species of the genera *Elaphe*, *Lampropeltis*, *Pantherophis*, *Pituophis* and *Oocatochus* (Table 3.1). The two taxon-sets were used to verify that relationships between Western Palaearctic ratsnakes were consistent either excluding (*Zamenis* taxon-set) or including (*Zamenis* + ratsnakes taxon-set) taxa from Asia and the New World.

PHYLOGENETIC ANALYSES AND DIVERGENCE TIME ESTIMATION

Phylogenetic networks were inferred with the Neighbor-Net algorithm (Bryant & Moulton, 2004) implemented in SplitsTree4 (Huson & Bryant, 2006) and with the Median-Joining algorithm (Bandelt, Forster, & Rohl, 1999) implemented in the software Network 5.0.0.0. (available at http://www.fluxus-engineering.com/sharenet.htm), using in both cases default parameters.

Partition schemes and models of nucleotide substitution for ML and BI phylogenetic analyses were defined by PartitionFinder 2.1.0 (Lanfear, Frandsen, Wright, Senfeld, & Calcott, 2016) with the following search parameters: linked branch length,

beast models, BIC model selection, and user scheme with data blocks by genes and codons positions (1st, 2nd, 3rd or 1st+2nd, 3rd) for the protein coding genes (*cytb*, *nd4*, *cmos*, *dnah3* and *prlr*) and single data block for the intronic fragments *sptbn1* and *vim*. We obtained identical models for gene and codon positions and thus we used the gene partitions.

Maximum likelihood (ML) analyses were performed in raxmlGUI 1.3 (Silvestro & Michalak, 2012), a graphical front-end for RAxML 7.4.2 (Stamatakis, 2006). Searches included 10 random addition replicates and 1000 nonparametric bootstrap replicates, applying the general time-reversible model with a gamma model of rate heterogeneity (GTRGAMMA), with individual gene partitions.

Bayesian inference (BI) analyses were performed in BEAST 1.8.0 (Drummond, Suchard, Xie, & Rambaut, 2012). The input file was built in the BEAUTi utility with models and prior specifications applied were as follows (otherwise by default): each gene was used as a partition; nucleotide substitution and relaxed uncorrelated lognormal clock models were unlinked across partitions, except for the clock models of the mtDNA genes *cytb* and *nd4*, which were linked; the tree model was linked across all partitions; models of nucleotide substitution for each gene partition as selected by PartitionFinder (reported in Table 3.3); Yule process of speciation as tree prior, random starting tree, alpha uniform prior (0, 10), ucld.mean gamma prior (1, 1); kappa operator (2.0). The xml file was manually edited to account for variability in heterozygous positions by defining "ambiguities=true" for the nuclear partitions. BEAST was run three times with 100 million generations each, sampling every 10000 generations. Since the Yule process of speciation prior assumes that each terminal represents a distinct species (i.e. requires only one single sequence per species), whereas our dataset contains multiple samples per species, in order to inspect the sensitivity of our estimates to the choice of tree prior, we performed an additional run applying the same settings described above but using only one specimen for each species and we obtained identical estimates (Fig. S3.1).

Time calibrated species trees were estimated with *BEAST, an extension of the BEAST software implementing the multi-species coalescent model (Heled & Drummond, 2010). The ratsnakes fossil record is scarce and unevenly distributed in the Palaearctic and offer little help for calibrating the ratsnakes tree. Most importantly, the phylogenetic position of extinct ratsnakes is very hard to define because they show a mixture of characters found in diverse extant species and some unique features. In order to assign them to the crown group of some of the extant genera a detailed phylogenetic analysis including all fossils and extant ratsnakes would be necessary (Massimo Delfino,

personal communication). For example, the fossils of the extinct species *Elaphe praelongissima* (Venczel, 1994) have been tentatively interpreted as remains of the ancestor of *Z. longissimus* or *Z. situla* and used in previous dating estimates (Lenk et al., 2001), but many features of *Elaphe praelongissima* are shared with Old World ratsnakes which are distantly related to *Zamenis* such as *Elaphe dione* Pallas, 1773. The same applies for *Elaphe algorensis* (Szyndlar, 1985), which has been related to the ancestor of *R. scalaris* (Lenk et al., 2001). However, *Elaphe algorensis* displays most morphological features attributed to both *Elaphe* and *Coluber* Linnaeus, 1758 (Szyndlar, 1985), which makes the assignment of *Elaphe algorensis* as an ancestor to *R. scalaris* unreliable. Therefore, due to the lack of internal calibration points in *Zamenis* and the general difficulty in the attribution of fossils to stem groups of extant ratsnakes genera, we recurred to literature for rates of *cytb* and *nd4* evolution in colubrids. The review of available literature indicated that rates of evolution estimated for colubrids fall under two classes: fast (~1.3 % substitutions per million years, s/my) and medium (~0.8% s/my) rates. Therefore, we performed time estimation analyses with two different calibration rates. In the calibration I (Cal I) we used a fast rate of evolution of 1.34% s/my (with $CI_{95\%}$ = 0.99–1.70%) as estimated by Daza et al. (2009) and similar to rates applied in previous papers on Serpentes (e.g. Carranza, Arnold, & Pleguezuelos, 2006; Lenk et al., 2001; Mezzasalma et al., 2015). In calibration II (Cal II) we used a medium rate of evolution of 0.815% s/my as implemented by Bryson, de Oca, & Velasco (2008) and similar to other studies on Serpentes based on New World snake fossils (Myers et al., 2013; Ruane, Bryson, Pyron, & Burbrink, 2014). The standard deviation for this rate was not available in Bryson et al. (2008), thus we decided to apply a range from 0.7 to 0.9%, which also includes estimated rates of evolution close to 0.8% which were applied previously (Guicking & Lawson, 2006; Nagy, Lawson, Joger, & Wink, 2004; Pyron & Burbrink, 2009; Ruane, Torres-Carvajal, & Burbrink, 2015). We implemented these rate priors in *BEAST by defining a lognormal distribution for the mitochondrial ucld.mean parameter, with mean 0.0134 and standard deviation 0.0035 for Cal I and mean 0.00815 and standard deviation 0.0010 for Cal II. Model and prior specification were as follows (otherwise by default): the nucleotide substitution models were unlinked across loci and implemented for each gene as selected by PartitionFinder (Table 3.3), relaxed uncorrelated lognormal clock models and tree models were unlinked, with the exception of the clock and tree models of the mitochondrial genes *cytb* and *nd4* which were linked; Yule process of speciation as tree prior, random starting tree; alpha uniform (0, 10); ucld.mean of nuclear genes Gamma (1, 1); operator kappa (2.0). *BEAST was run three times with 500 million generations, sampling every 50000 generations.

All BEAST and *BEAST runs were performed in the CIPRES Science Gateway 3.3. (available at http://www.phylo.org/). We used Tracer 1.6 (Rambaut, Suchard, & Drummond, 2014) to check the runs for convergence (burn-in = 10%) and to ensure that all effective sample size (ESS) parameters were higher than 200, as recommended in the software's manual. Tree files were combined with LogCombiner and TreeAnnotator (both included in the BEAST package) was applied to calculate the maximum clade credibility tree (MCC) summarizing the posterior distribution of tree topologies and branch lengths. All trees were visualized with FigTree 1.4 (available at http://tree.bio.ed.ac.uk/software/figtree/).

TOPOLOGY TESTS

We performed topological tests in order to compare our phylogenetic tree with estimates from previous studies, which inferred the following phylogenetic relationships with the maximum statistical support: (i) *Rhinechis* nested within *Zamenis* and sister to *Z. situla*, Test A (Pyron et al., 2011; Pyron et al., 2013; Zheng & Wiens, 2016); (ii) *Rhinechis* and *Ptyas korros* nested within *Zamenis* and *Ptyas korros* sister to *Z. longissimus*, Test B (Figueroa et al., 2016). For the Test B, we downloaded sequences from *P. korros* and *P. mucosa* Linnaeus, 1758 available for the same genes applied in this study (Table 3.1). These sequences were added to our *Zamenis* taxon-set and used to infer a ML mt-nucDNA tree, which was compared with the ML tree enforcing the topology obtained by Figueroa et al. (2016).

Additionally, given the low support for the relationships between *R. scalaris*, *Z. hohenackeri* and *Z. persicus*, and between them and the remaining *Zamenis* species, we also tested for significant differences between our best tree and four alternative topological hypotheses with varying phyletic order between *R. scalaris*, *Z. hohenackeri* and *Z. persicus* (Tests C-F; Table 3.4).

The topological constrains (Table 3.4) were built in Mesquite 3.2 (Maddison & Maddison, 2003) and the per-site log likelihoods were estimated in RAxMLGUI 1.3. The constrained trees were compared with our best ML tree using the Shimodaira-Hasegawa (SH) and the approximately unbiased (AU) tests (Shimodaira & Hasegawa, 1999; Shimodaira, 2002; respectively), as implemented in Consel 0.2 (Shimodaira & Hasegawa, 2001) to determine if the constrained topology could be rejected at the 0.05 level.

BIOGEOGRAPHICAL RECONSTRUCTION

Biogeographical analyses were performed with the programs *rase* (Quintero, Keil, Jetz, & Crawford, 2015) and RASP (Yu, Harris, Blair, & He, 2015). In order to perform the biogeographical analyses, we built a species tree including only *Zamenis* and *Rhinechis* species (i.e. without outgroups), applying the same procedure and settings as described above. Results from biogeographical analyses were identical irrespective of the calibration used to build the species tree, given that species trees recovered with either Cal I or II had identical topologies and branch lengths and *rase* time-slices “cut” the tree at the same points in both time-trees.

The *rase* method allows using of the whole distribution range of species (instead of single localities) without an a-priori definition of areas to infer the geographical location of the ancestors in a Bayesian framework. The geographical distributions of these species were extracted from the IUCN Red List of Threatened Species (IUCN, 2009; Fig. 3.1); we assumed no correlation between dispersal rates in longitude and latitude. We ran *rase* for 10,000 iterations, discarded the first 1000 as burn-in, logged every 10th iteration and obtained the posterior distributions of ancestral nodes and migration rates (σ^2). We plotted the trace to evaluate the MCMC results, estimate the mean and posterior densities for each of the estimated parameters and confirm that the algorithm converged to the posterior distribution with the *coda* package (Plummer, Best, Cowles, & Vines, 2006). We applied evenly spaced time slices to infer the location of the ancestral nodes. The *rase* and *coda* packages were ran in R (R Core Team, 2017) using the interface RStudio 1.0.153.

RASP analyses were performed in RASP 3.0 (Yu et al., 2015) using the statistical dispersal-vicariance analysis (S-DIVA) (Yu, Harris, & He, 2010) and the Bayesian binary MCMC (BBM) (Yu, Harris, & Xingjin, 2011). The distribution range of *Zamenis* and *Rhinechis* was divided into four main ranges: Southwestern Asia (Caspian, Caucasus and Anatolia), the Balkan Peninsula, the Italian Peninsula, and Western Europe, and each species was attributed to one or more areas where it is present (see Fig. 3.1 for species distribution). Besides the four-area analyses, we performed additional runs with three or five areas and different areas assemblages and we obtained consistent results. We used all the post-burnin trees and the MCC tree resultant from the species tree run as input. S-DIVA was run using 1000 trees randomly sampled from the input trees and the BBM analyses were conducted with the Jukes-Cantor model, site variation set to equal and with two simultaneous runs with 5 million generations, sampling each 100th generation.

RESULTS

We generated a total of 105 new sequences which were deposited in GenBank (Table 3.1). Length of individual genes and the number of variable sites are presented in Table 3.3. Sequence alignments of the protein coding genes *cytb*, *nd4*, *cmos* and *dnah3* did not require gap positions; the alignment of *prlr* required an 18 base insertion. The translation into amino acids of these genes did not contain stop codons. The phi test did not find statistically significant evidence for recombination in any gene fragment ($p > 0.05$).

Table 3.3. Length of gene alignments (base pairs, bp), models of nucleotide substitution and number of variable positions for the datasets used in this study: (i) *Zamenis* (complete dataset), and (ii) *Zamenis*+ratsnakes (*Zam*+ratsnakes) taxon-sets. (*: models calculated on phased or unphased nuclear DNA sequence alignments were identical).

Gene	Lenght (bp)	Model of nucleotide substitution*		Variable Positions	
		***Zamenis* taxon-set**	***Zam*+ratsnakes taxon-set**	***Zamenis* taxon-set**	***Zam*+ratsnakes taxon-set**
nd4	844	HKY+G	HKY+G+I	285	359
cytb	282	HKY+G	HKY+G	98	111
cmos	542	HKY	HKY	25	35
dnah3	664	TrN+I	TrN+I	44	44
prlr	495	HKY	HKY	64	73
sptbn1	831	HKY	HKY	61	118
vim	610	HKY	HKY+G	62	113

PHYLOGENETIC RELATIONSHIPS AND TIME OF DIVERGENCE ESTIMATES

Haplotype network reconstructions for both nuclear and mitochondrial loci recovered equivalent relationships between Neighbor-Net and Median-Joining analyses (Figs. 3.3 & S3.2, respectively). The least variable genes were the nuclear *cmos* and *prlr*, that showed the lower number of haplotypes and mutational steps between the ingroup species. The most variable loci were mtDNA and the nuclear *sptbn1* and *vim. Coronella* species are well distinct from *Rhinechis* and *Zamenis* in the mtDNA and fast-evolving nuclear genes *sptbn1* and *vim*, whereas these genera are not well-sorted at slower-evolving nuclear genes. *Rhinechis* is always closely related with, or nested within,

Zamenis species. *Zamenis* species are closely related among them but form a paraphyletic assemblage in the haplotype networks of most of the loci analysed (mtDNA, *cmos*, *prlr*, *sptbn1* and *vim*). There was no haplotype sharing between species, except in the nuclear gene *vim*, where *Z. situla* and *Z. lineatus* shared a single haplotype.

Phylogenetic relationships recovered by multi-locus analyses were consistent between methods (ML and BI; Figs. 3.4 and 3.5), datasets (concatenated nucDNA and mt-nucDNA datasets; Figs. 3.4 and S3.3) and approaches (based on the concatenation and the species tree approaches; Figs. 3.4 and 3.6) and with any of the taxon-sets used (the *Zamenis* taxon-set and the *Zamenis* + ratsnakes; Figs. 3.4, 3.5, 3.6, S3.6 and S3.7). Moreover, results from the phylogenetic analyses based on the *Zamenis* taxon-set either with the complete dataset or the reduced dataset were identical at supported nodes, suggesting that adding taxa with higher proportion of missing data did not affect phylogenetic estimates (Figs. 3.4 and S3.3 – S3.5).

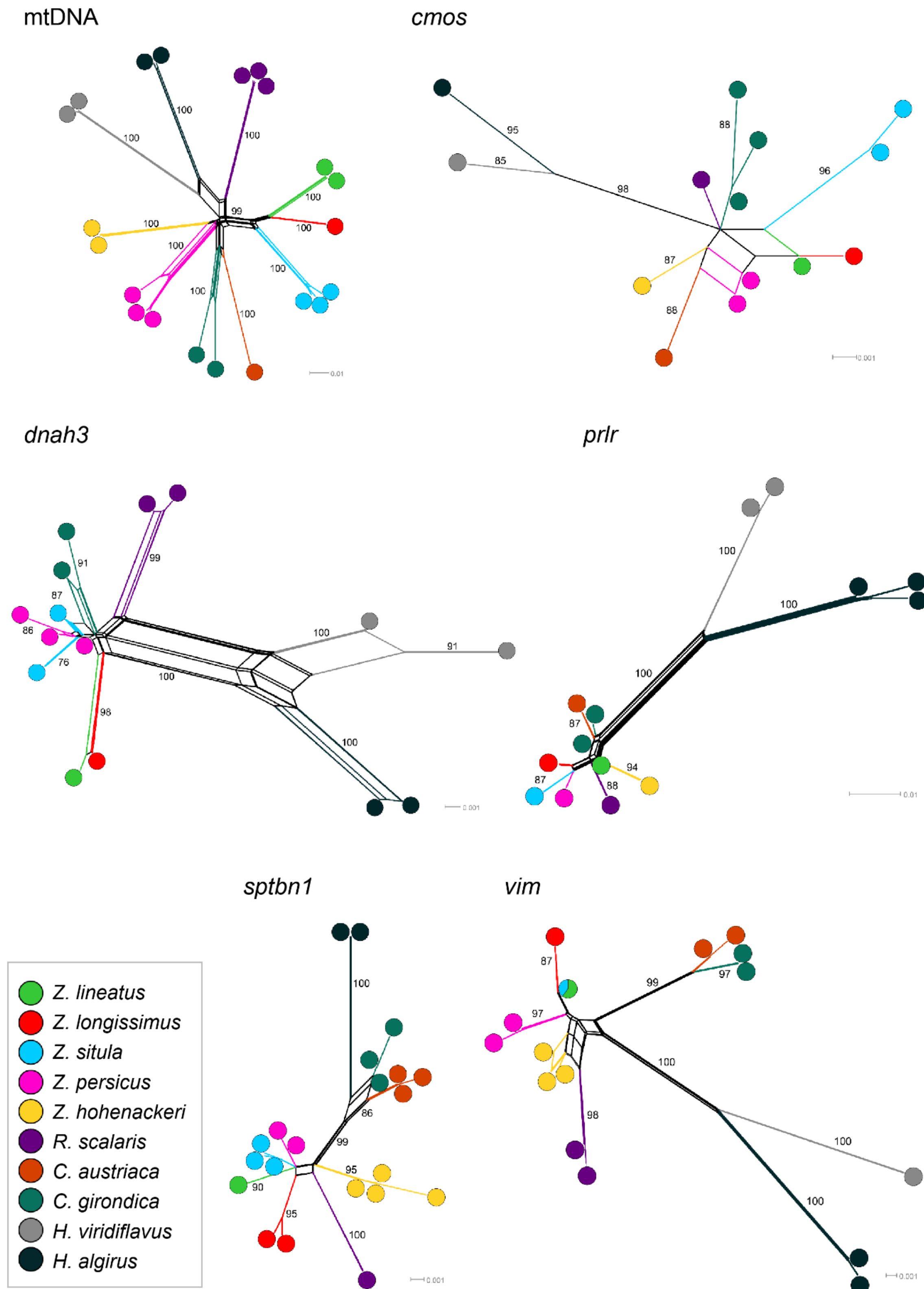


Figure 3.3. Neighbor-Net haplotype networks for *Zamenis*, *Rhinechis*, *Coronella*, *Hierophis* and *Hemorrhois* species inferred from single locus DNA sequences data: mitochondrial (concatenated *cytb* and *nd4*) and individual nuclear loci (*cmos*, *dnah3*, *prlr*, *sptbn1* and *vim*). Circles represent different haplotypes. Values next to lines represent bootstrap support.

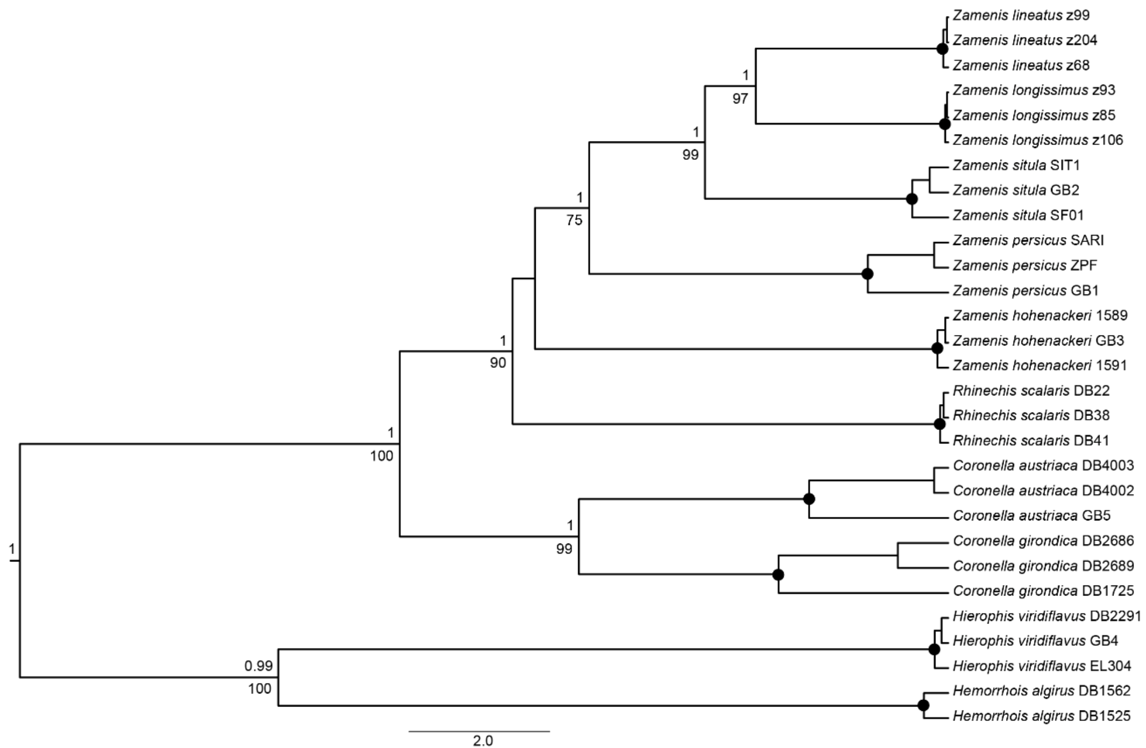


Figure 3.4. Bayesian tree depicting the phylogenetic relationships between *Zamenis*, *Rhinechis* and *Coronella* ratsnakes (referred in the text as *Zamenis* taxon-set) based on concatenated mitochondrial (*cytb*, *nd4*) and nuclear (*cmos*, *dnah3*, *prlr*, *sptbn1* and *vim*) DNA sequences. Values above nodes represent Bayesian posterior probabilities (BPP) ≥ 0.9; values below nodes represent maximum likelihood bootstrap support (BS) values ≥ 70. Node support for intra-specific clades is represented by black circles: BPP ≥ 0.98 (upper half) and BS ≥ 95 (bottom half).

The genus *Coronella* is monophyletic (Bayesian Posterior Probabilities, BPP=1, Bootstrap Support, BS>90; Figs. 3.4, 3.5 and 3.6) and sister to the clade formed by *Rhinechis* and *Zamenis* species. The latter is well supported in all the phylogenetic analyses (BPP>0.98, BS>90) and show a topology with *R. scalaris* sister to all *Zamenis* species. However, the *Zamenis* clade is not supported in any phylogenetic analysis (BPP<0.90, BS<70), with all the trees showing a short internode between the *R. scalaris* and Z*. hohenackeri* lineages. Relationships among the remaining *Zamenis* species are well resolved, with *Z. persicus* sister to the clade formed by *Z. situla*, *Z. longissimus* and *Z. lineatus* (BPP>0.95, BS>75) and with the latter two always recovered as sister species with high statistical support (BPP>0.92, BS>92).

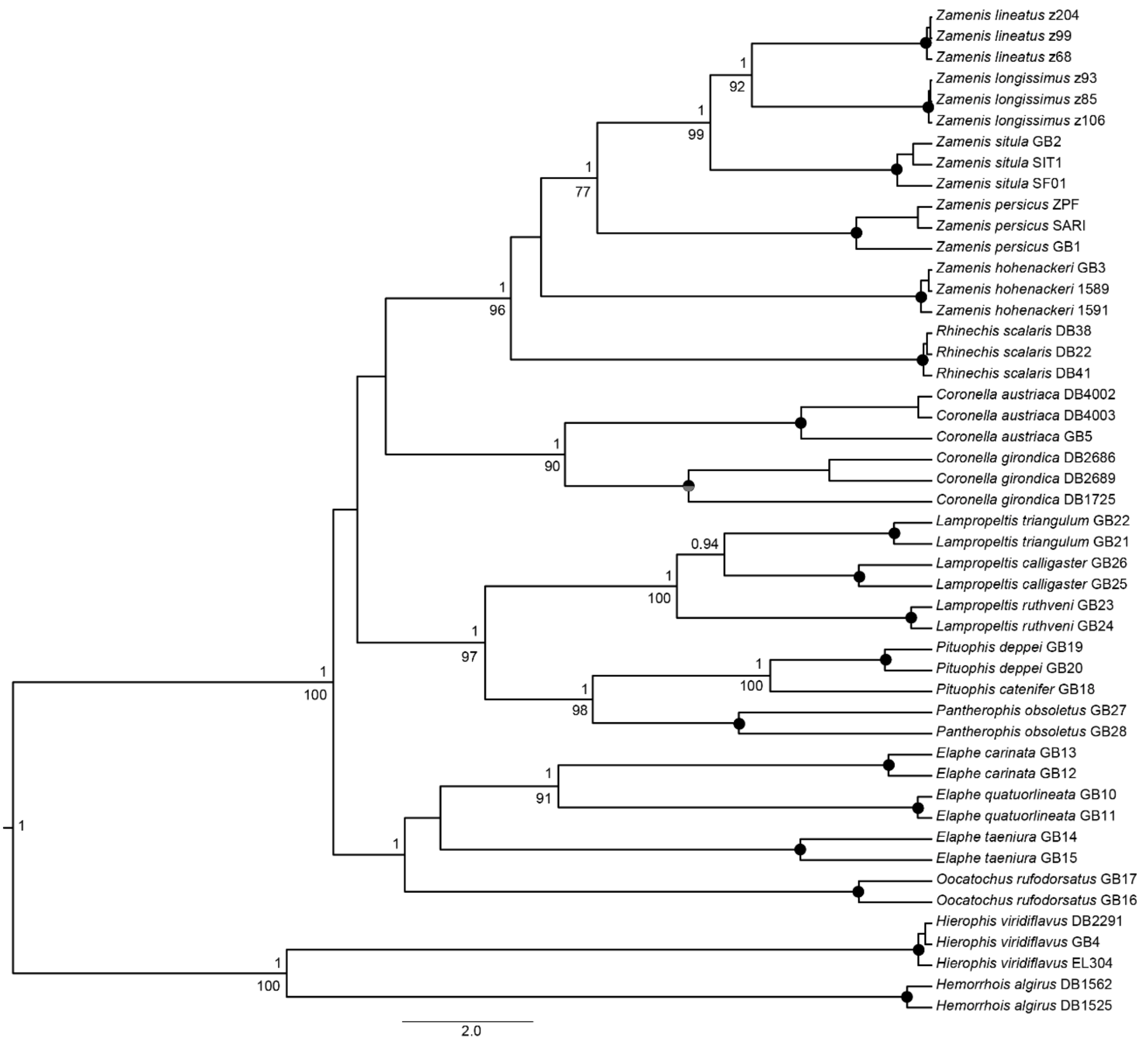


Figure 3.5. Bayesian tree depicting the phylogenetic relationships between *Zamenis*, *Rhinechis*, *Coronella* and representative of five other ratsnakes' genera (referred in the text as *Zamenis* + ratsnakes taxon-set), based on concatenated mitochondrial (*cytb*, *nd4*) and nuclear (*cmos*, *dnah3*, *prlr*, *sptbn1* and *vim*) DNA sequences. Values above nodes represent Bayesian posterior probabilities (BPP) ≥ 0.9; values below nodes represent maximum likelihood bootstrap support (BS) values ≥ 70. Node support for intra-specific clades is represented by black circles: BPP ≥ 0.98 (upper half) and BS ≥ 95 (bottom half) and grey circles: 0.9 ≤ BPP < 0.98 (up) and 70 ≤ BS < 95 (bottom).

Divergence time estimates based on Cal I and Cal II are reported below the species tree (Figs. 3.6 and S3.7) and in Tables 3.5 and S3.1 (along with 95% High Posterior Density intervals, 95%HPD). The Time to the Most Recent Common Ancestor (TMRCA) between *Rhinechis* and *Zamenis* is estimated in the Middle Miocene according to Cal II (~13 Million years ago, Ma) or in the Late Miocene according to Cal I (~7 Ma). The split of Z. *hohenackeri* is placed about a million year later according to both calibrations. The following cladogenetic events leading to the remaining *Zamenis* species are estimated during the Late Miocene (~9-5 Ma) or in the Pliocene (~5-3 Ma) according to Cal II and I, respectively.

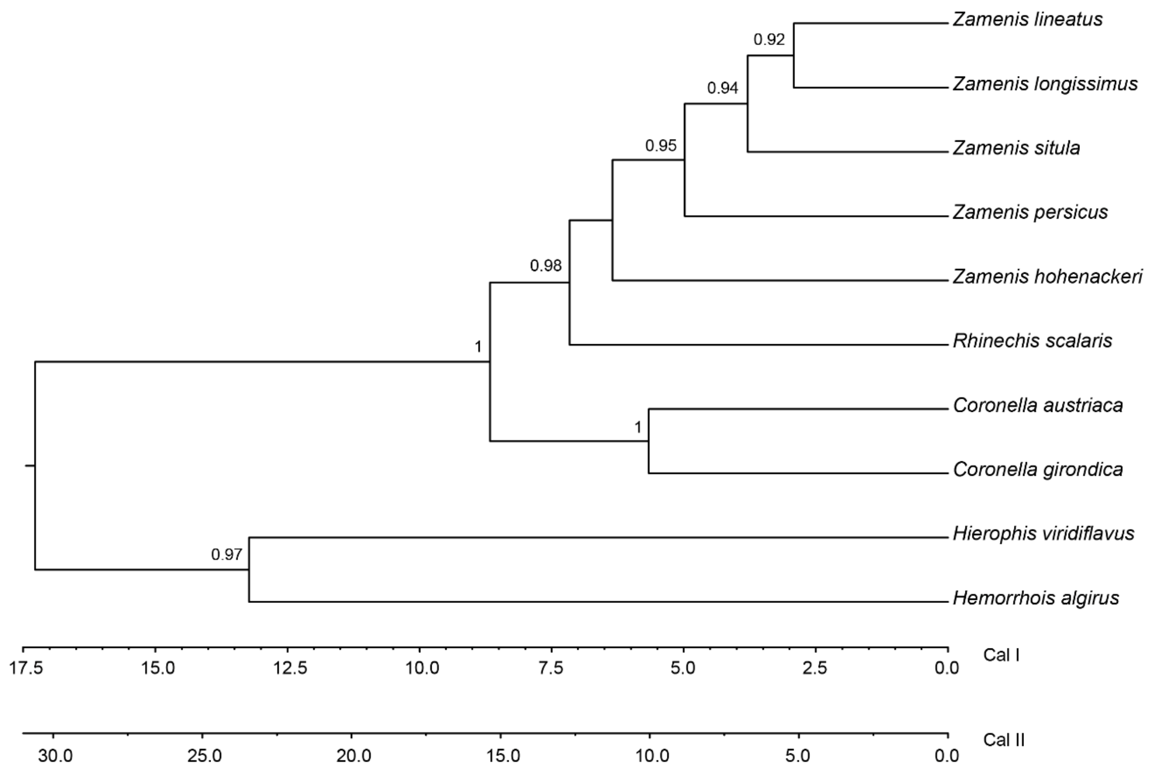


Figure 3.6. Species tree of *Zamenis*, *Rhinechis* and *Coronella* ratsnakes (referred in the text as *Zamenis* taxon-set), inferred from mitochondrial (*cytb* and *nd4*) and nuclear (*cmos*, *dnah3*, *prlr*, *sptbn1* and *vim*) DNA sequences using the multispecies coalescent model in *BEAST. Values above nodes represent Bayesian posterior probabilities (BPP) ≥ 0.9. The time axis for the calibrations I (Cal I) and II (Cal II) are represented below the tree; the age of the nodes and the associated 95% Highest Posterior Density intervals are presented in Table 3.5.

Table 3.4. Results of topological tests. Test A and B used topological constrains based on highly supported relationships recovered by previous studies (see Figure 3.2): Test A – *Rhinechis* nested within *Zamenis* and sister to *Z. situla* as recovered by Pyron et al. (2011, 2013) and Zheng & Wiens (2016); Test B – *Rhinechis* and *Ptyas korros* nested within *Zamenis* and *Ptyas korros* sister to *Z. longissimus* as recovered by Figueroa et al. (2016). Tests C to F used topological constrains based on different phyletic order between *R. scalaris*, *Z. hohenackeri* and *Z. persicus* in order to compare our best tree with alternative topological hypotheses of relationships between *Rhinechis* and *Zamenis*. The constrained relationships and *p*-value results of Approximately Unbiased (AU) and Shimodaira-Hasegawa (SH) tests are reported; significant *p*-values are in bold (*Z. lin*: *Zamenis lineatus*, *Z. lon*: *Zamenis longissimus*, *Z. sit*: *Zamenis situla*, *Z. hoh*: *Zamenis hohenackeri*, *Z. per*: *Zamenis persicus*, *R. sca*: *Rhinechis scalaris*).

TEST	CONSTRAINT	REFERENCE	AU	SH
A	((((*Z. lin*, *Z. lon*) (*Z. sit*, *R. sca*)) *Z. per*) *Z. hoh*)	Pyron et al 2011, 2013; Zheng & Wiens 2016	**0.025**	**0.041**
B	((((*Ptyas korros*, *Z. lon*) *Z. lin*) *Z. sit*) *R. sca*) (*Z. per*, *Z. hoh*)	Figueroa et al. 2016	**7e-0.005**	**0**
C	(((((*Z. lin*, *Z. lon*) *Z. sit*) *Z. per*) *R. sca*) *Z. hoh*)		0.117	0.161
D	(((((*Z. lin*, *Z. lon*) *Z. sit*) *R. sca*) *Z. per*) *Z. hoh*)		0.197	0.207
E	(((((*Z. lin*, *Z. lon*) *Z. sit*) *Z. hoh*) *R. sca*) *Z. per*)		0.06	0.097
F	(((((*Z. lin*, *Z. lon*) *Z. sit*) *R. sca*) *Z. hoh*) *Z. per*)		0.217	0.228

TOPOLOGY TESTS

The phylogenetic hypotheses from previous studies of a sister relationship between *R. scalaris* and *Z. situla* (Test A), and of a sister relationship between *Z. longissimus* and *Ptyas korros* with *R. scalaris* nested within *Zamenis* (Test B), were rejected by the SH and AU tests (Table 3.4). On the contrary, varying the phyletic order between *R. scalaris*, *Z. hohenackeri* and *Z. persicus* resulted in topologies which were not statistically rejected by the SH and AU tests (Tests C – F; Table 3.4).

Table 3.5. Divergence time estimates, in million years, based on rates from calibration I (Cal I) and calibration I (Cal II) with the *Zamenis* taxon-set; 95% HPD intervals are provided in brackets. *Z. lin*: *Zamenis lineatus*, *Z. lon*: *Zamenis longissimus*, *Z. sit*: *Zamenis situla*, *Z. hoh*: *Zamenis hohenackeri*, *Z. per*: *Zamenis persicus*, *R. sca*: *Rhinechis scalaris*, *C. aus*: *Coronella austriaca*, *C. gir*: *Coronella girondica*, *He. alg*: *Hemorrhois algirus*, *Hi. vir*: *Hierophis viridiflavus*.

	Age and 95% HPD	
Node	**Cal I**	**Cal II**
(*Z. lin*, *Z. lon*)	2.92 (1.44 – 4.38)	5.26 (3.02 – 7.53)
(*Z. sit*, Z. *lin*, *Z. lon*)	3.79 (2.22 – 5.65)	6.83 (4.54 – 9.31)
(*Z. per*, *Z. sit*, Z. *lin*, *Z. lon*)	4.98 (2.80 – 7.60)	8.93 (5.58 – 12.45)
Zamenis spp.	6.35 (3.85 – 9.10)	11.35 (8.17 – 15.07)
(*R. sca*, *Zamenis* spp.)	7.16 (4.44 – 10.23)	12.85 (9.35 – 16.71)
(*Coronella* spp., (*R. sca*, *Zamenis* spp.))	8.67 (5.26 – 12.24)	15.49 (11.23 – 20.13)
Coronella spp.	5.66 (3.19 – 8.30)	10.11 (6.59 – 13.80)
(*He. alg*, *Hi. vir*)	13.23 (7.02 – 21.07)	23.36 (14.13 – 34.48)
Root	17.28 (10.01 – 25.52)	30.61 (20.54 – 41.94)

BIOGEOGRAPHICAL RECONSTRUCTION

The results from the biogeographical reconstruction with *rase* showed a remarkable longitudinal migration throughout the evolutionary history of *Zamenis* and *Rhinechis*, whereas the latitudinal migration was relatively low (Fig. 3.7A). Indeed, the estimated longitudinal dispersal rate, σ_x^2 was high (mean: 47.8604 degrees^2/Myr, 95%HPD: 29.5124 – 77.4254) and the latitudinal dispersal rate, σ_y^2 was comparatively low (mean: 0.4667 degrees^2/Myr, 95%HPD: 0.3199 – 0.6882). According to the *rase* results, the location of the ancestor of *Rhinechis* and *Zamenis* was distributed in an area presently comprehending the Aegean region in the southern Balkan Peninsula (Fig. 3.7a), at latitudinal mean (latM) 39.3020 and longitudinal mean (lonM) 22.2972 (95% HPD values of latitude and longitude for the ancestral locations are presented in Table S3.2). The ancestor of *R. scalaris* underwent a westward migration (Fig. 3.7b, c; purple), while the ancestor of *Zamenis* migrated eastwards (Fig. 3.7b). The split between the ancestors of *Z. hohenackeri* (Fig. 3.7c; yellow) and of the remaining *Zamenis* species (Fig. 3.7c; dark green) occurred in the current Anatolia (latM: 39.0857, lonM: 28.5675) alike the

subsequent split between *Z. persicus* (Fig. 3.7e, pink) and the ancestor of *Z. situla, Z. longissimus* and *Z. lineatus* (latM: 38.6477, lonM: 33.3818). The divergence of *Z. situla* (Fig. 3.7g, blue) occurred in present western Anatolia (latM: 39.6099, lonM: 25.2771) and the divergence between *Z. longissimus* (Fig. 3.7g, red) and *Z. lineatus* (Fig. 3.7g, green) occurred in an area presently including the southwestern Balkan Peninsula and the westernmost part of Anatolia (latM: 39.5779, lonM: 23.1189). By the time of the divergence between *Z. lineatus* and *Z. longissimus* the species *Z. hohenackeri* (Fig. 3.7g, yellow) and *Z. persicus* (Fig. 3.7g, pink) would had migrated towards the eastern Anatolia and Caucasus regions (latM: 38.9985, lonM: 35.1886, and latM: 37.6885, lonM: 41.6369, respectively), while *R. scalaris* (Fig. 3.7g, purple) had migrated west towards the Iberian Peninsula (latM: 39.9817, lonM: 2.8080).

The results from the RASP analyses suggest an eastern origin of the *Zamenis* clade (Fig. 3.7B). The basal nodes within the clade leading to the split of the *Z. hohenackeri* and *Z. persicus* lineages were estimated either in southwestern Asia, as inferred the BMM method (>90%), or shared by the southwestern Asia, Balkans and the Italian Peninsula, as inferred the by S-DIVA method (98%). The ancestral area of *Z. situla*, *Z. longissimus* and *Z. lineatus* was either shared by the Balkans and the Italian Peninsula or these areas plus southwestern Asia according to S-DIVA and BBM. The ancestral area of *Z. lineatus* and *Z. longissimus* was in the Italian Peninsula according to S-DIVA (100%), whereas it is highly uncertain in the BBM inference. Finally, it remains undetermined the ancestral area for the root node *Zamenis*+*Rhinechis*, with S-DIVA inferring a shared origin between all areas (100%) while BBM estimated with equal probability either the easternmost area (southwestern Asia) or the westernmost area (Western Europe) as probable ancestral areas.

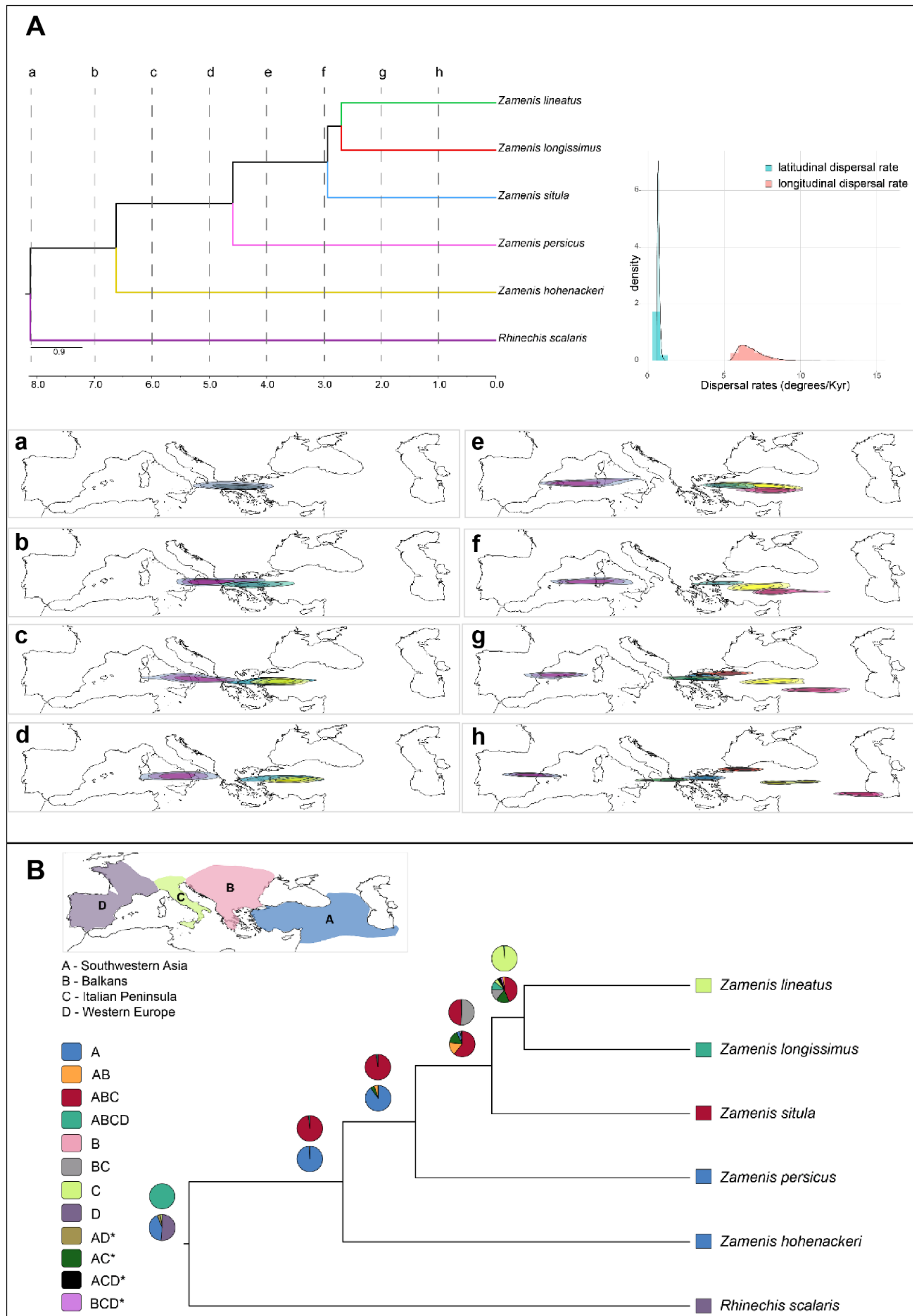


Figure 3.7. Biogeographical reconstruction of the evolutionary history of *Zamenis* and *Rhinechis* according to *rase* and RASP analyses based on the species tree applying the time calibration I. A) biogeographical reconstruction according to the *rase* analysis at eight evenly spaced time slices from 8 to 1 Ma. The 5%, 10%, and 15% highest posterior density is plotted for each extant branch. Top-right: posterior distributions of the dispersal rate parameters (σ^2) for latitude, in blue, and longitude, in pink, represented in degrees/Kyrs. B) Biogeographical reconstruction according to the RASP analyses

based on four areas (A, B, C, D as represented in the top-left box): coloured pie charts in correspondence of the nodes represent results from ancestral reconstruction conducted with S-DIVA (top) and BBM (bottom); * represent areas estimated by BBM alone.

DISCUSSION

In this study we obtained a fully-resolved and well-supported phylogeny of Western Palaearctic ratsnakes, which is consistent across markers, taxon-sets, phylogenetic methods and incorporates coalescent models accounting for gene-tree/species-tree conflicts (Figs. 3.4, 3.5, 3.6, S3.3 – S3.7). Smooth snakes of the genus *Coronella* form a sister clade to *Rhinechis* and *Zamenis* ratsnakes (Figs. 3.4, 3.5 and 3.6). The close phylogenetic relationships between *Rhinechis* and *Zamenis* is well-established. However, the phyletic order between the early branches within this clade, *R. scalaris* and Z. *hohenackeri*, remains poorly supported. Relationships between remaining *Zamenis* species are fully resolved with the Persian ratsnake *Z. persicus* sister to the Mediterranean species *Z. situla*, *Z. longissimus* and *Z. lineatus*, and among them *Z. situla* branching off earlier than Aesculapian snakes *Z. longissimus* and *Z. lineatus*.

In the following sections we discuss how these results allowed resolving previous uncertainties and controversies on *Rhinechis* and *Zamenis* phylogenetics and systematics, and also how the incorporation in this phylogeny of clade age and biogeographical information allowed inferring their spatial and temporal modes of diversification.

RESOLVING PHYLOGENETIC CONTROVERSIES ON *ZAMENIS* AND *RHINECHIS*

The phylogenetic relationships between *Zamenis* and *Rhinechis* inferred in this study are in line with those estimated by early studies mainly based on mtDNA data (Burbrink & Lawson, 2007; Lenk et al., 2001; Utiger et al., 2002; Fig. 3.2a, b, c), with a few discrepancies which are easy to conciliate. Our topology is fully congruent with the tree of Burbrink & Lawson (2007) based on data from mtDNA and the nuclear *cmos* gene, although in this latter study the nodal support was low and the *Zamenis* taxon-set was not complete. The close relationships between *Rhinechis* and *Zamenis* is not retrieved in the Maximum Parsimony (MP) trees by Lenk et al. (2001) and by Utiger et al. (2002) (Fig. 3.2). The different results from these latter studies are probably due to the use of small mitochondrial fragments in combination with the MP method, which is not ideal for molecular data analyses. Indeed, when using the Minimum Evolution method Utiger et al. (2002) found *Rhinechis* sister to *Zamenis* (not shown in Fig. 3.2). The phylogeny by

Lenk et al. (2001) recovered a supported *Zamenis* clade sister to *Oreocryptophis porphyraceus* Cantor, 1839 (formerly *Elaphe porphyracea*) and *R. scalaris* early branching as sister to all in-group taxa analysed (*Elaphe* and *Zamenis*). However, these results are clearly due to limited taxon sampling and the unfortunate use of *Lampropeltis triangulum* as outgroup of Palaearctic ratsnakes. The addition of New and Old World ratsnakes to the phylogenetic analyses clearly showed that *Lampropeltis triangulum* and other Lampropeltini are nested within the clade including Palaearctic ratsnakes (Figs. 3.5, S3.6, and S3.7), as described in further studies (Burbrink & Lawson, 2007; Chen, Lemmon, Lemmon, Pyron, & Burbrink, 2017). The inspection of sequence divergence in the short mitochondrial fragment dataset of Lenk et al. (2001) show similar values between ingroup and outgroup taxa. Therefore, the use of such a close outgroup violates the assumption that in-group taxa are more closely related to each other than any are to the outgroup, resulting in the attraction of one or more taxa towards the root. As to the high nodal support recovered for the tree by Lenk et al. (2001), and specifically for the *Zamenis* clade, these authors used MP bootstrapping using character weights derived from the successive weighting. This procedure inflates bootstrap values as it is likely that the support values will tend to increase when incongruent characters are down-weighted and characters supporting that specific topology are over-weighted (Cummins & McInerney, 2011). By adding a proper outgroup (*Hemorrhois algirus* and *Hierophis viridiflavus* instead of *Lampropeltis triangulum*) to the dataset of Lenk et al. (2001), and repeating the bootstrap analysis using unweighted matrix, we recovered a tree which is congruent with results from this study and other previous studies: i.e. *Rhinechis* is sister to *Zamenis* and *Oreocryptophis porphyraceus* to *Elaphe*, and the clade *Zamenis* is not supported.

More difficult to conciliate are the discrepancies on the phylogenetic relationships obtained in our study and early studies discussed above (Burbrink & Lawson, 2007; Lenk et al., 2001; Utiger et al., 2002; Figs. 3.2a, b, c, 3.4, 3.5 and 3.6) with those recovered by the analyses of supermatrices with hundreds of squamate taxa (Figueroa et al., 2016; Pyron et al., 2011; Pyron et al., 2013; Zheng & Wiens, 2016; Fig. 3.2d, e, f). All these latter mega-phylogenies recovered *Rhinechis* deeply nested within *Zamenis* with maximum statistical support (99-100), while showing conflicting results on the relationships between *Zamenis* species (Fig. 3.2d, e, f, g). Moreover, the phylogeny by Figueroa et al. (2016) shows a paraphyletic *Zamenis* clade, which included *Rhinechis* but also the Indo-Chinese ratsnake *Ptyas korros* (Fig. 3.2f). These phylogenetic hypotheses are strongly rejected by topological tests based on our data (Table 3.4). On the other hand, it is unlikely that incongruences between early phylogenetic estimates (Fig.

3.2a, b, c) and recent mega-phylogenies based on supermatrices (Fig. 3.2d, e, f, g) can be explained by differences in molecular data between studies. Indeed, although supermatrices are built with dozens of mitochondrial and nuclear markers – up to 52 genes in Zheng & Wiens (2016) – in each dataset the representation of these genes in *Zamenis* and *Rhinechis* is limited to sequences of the same four-five mtDNA genes (differently distributed across studies) and the slow evolving nuclear gene *cmos* generated by those early molecular studies. A compelling evidence that the nodes estimated by the supermatrix approach are incorrect comes from a recent phylogenomic study on ratsnakes that was published at the time of completion of our study (Chen et al., 2017). The phylogenomic tree of ratsnakes by Chen and colleagues based on over 300 loci shows identical phylogenetic relationships between *Zamenis* and *Rhinechis* as inferred in our study, with *Rhinechis* and *Z. hohenackeri* as early branches of the clade and sister to the remaining four *Zamenis* species, whose relationships are well resolved and highly supported (see Figs. 3.4, 3.5, 3.6 and Chen et al., 2017). The bad performance of the supermatrix approach in estimating relationships within terminal groups was documented also for lacertid lizards by Mendes, Harris, Carranza, & Salvi (2016). This is probably due to the fact that the analysis of supermatrices with thousands of terminals require high-speed approximations of tree topology searches, substitution models and nodal support as well as outgroups that are excessively distant from the tips of the tree (Mendes et al., 2016). Overall these findings indicate that while the supermatrix approach with hundreds of taxa and high levels of missing data is prone to recover wrong relationships and spurious support, a set of fast evolving nuclear markers such as the one used in this study is enough to recover a correct species tree.

EVOLUTIONARY HISTORY AND BIOGEOGRAPHY OF *ZAMENIS* AND *RHINECHIS* RATSNAKES

The ancestral area inferred by biogeographical analyses suggest an eastern origin of *Zamenis* ratsnakes, with the Anatolia and Balkan peninsulas as ancestral areas of the basal nodes and western areas inferred for terminal nodes (Fig. 3.7). Such a scenario of east to west diversification is consistent with the phyletic pattern of diversification of *Zamenis*, with early diverging species having an eastern distribution (*Z. hohenackeri* and *Z. persicus*) and species with a western distribution splitting more recently (*Z. situla*, *Z. longissimus* and *Z. lineatus*). An eastern origin of *Rhinechis* and *Zamenis*, and of the Palaearctic ratsnakes in general, followed by dispersal towards Western Europe in the case of *Rhinechis* and *Zamenis*, had been suggested by early molecular studies on

ratsnakes (Burbrink & Lawson, 2007; Helfenberger & Schulz, 2013; Lenk et al., 2001; Utiger et al., 2002).

The biogeographical reconstruction from *rase* attributes the origin of the ancestor of *Rhinechis* and *Zamenis* to the Aegean region (Fig. 3.7A). The Aegean area has a complex geological history since the late Tertiary (summarized in Poulakakis et al., 2015). From a paleogeographic point of view, one of the major geologic events affecting this area was the formation of the mid-Aegean trench in the middle Miocene, from 12 to 9 Ma, which caused the west-east split of the former Aegean landmass (Creutzburg, 1963; Dermitzakis & Papanikolaou, 1981). The formation of the mid-Aegean trench acted as a vicariant agent for the separation of lineages present in the area at the time, including vertebrates and invertebrates genera (see Poulakakis et al., 2015 for a review on phylogeographical studies on the Aegean region). Likewise, in a scenario of Aegean origin of *Zamenis* and *Rhinechis*, the divergence between these genera could be the result of the vicariance mediated by the formation of the mid-Aegean trench and the east-west division of the region. This is in agreement with estimated time of divergence between these genera (about 7 or 13 Ma according to Cal I or Cal II respectively; Fig. 3.6).

According to the biogeographical reconstruction, these early cladogenetic events were followed by the eastward dispersal of the ancestors of *Z. hohenackeri* and *Z. persicus* and by the westward dispersal of the ancestors of *Rhinechis* and remaining *Zamenis* species during Late Miocene (Fig. 3.7A). Land connections were available at the time for these dispersal events, both towards western Europe and towards Anatolia, due to the existing land connections between this region and the eastern Aegean (Çağatay et al., 2006; Elmas, 2003; Melinte-Dobrinescu et al., 2009). The Anatolian and Caucasus regions have been affected by the orogenic activity leading to the formation of the Anatolian mountain chains (including the Anatolian Diagonal, Taurus and Black Sea Mountains) as well as the Central Anatolian Plateau and the Central Anatolian lake system, and the uplift of the Greater and Lesser Caucasus. The formation of these mountain chains, in combination with past habitat changes produced by climatic oscillations between dryer and wetter conditions, have created significant barriers to gene flow (Eronen et al., 2009; Oberprieler, 2005; Popov et al., 2006; Rögl, 1998). Therefore, it is plausible that the climatic and geological changes occurring in this area have been responsible for the divergence of *Z. hohenackeri* and *Z. persicus*, as suggested for many genera of vertebrates in previous studies (e.g. Kapli et al., 2013; Kornilios et al., 2011; Skourtanioti et al., 2016; Weisrock, Macey, Ugurtas, Larson, & Papenfuss, 2001) and invertebrates (e.g. Micó, Sanmartín, & Galante, 2009).

The divergence of the western *Zamenis* species occurred, according to the ancestral reconstruction analyses, around the western Anatolia and southern Balkans for *Z. situla* and in the Italian Peninsula between *Z. longissimus* and *Z. lineatus* (Fig. 3.7). The location of these splits implicates a dispersal event of the ancestor of these three species from Anatolia/Caucasus to the Balkans and the Western Europe. Such dispersal events have been described for plants (Lo Presti & Oberprieler, 2009), warblers (Blondel, Catzeflis, & Perret, 1996), bush crickets (Chobanov, Kaya, Grzywacz, Warchalowska-Sliwa, & Ciplak, 2017; Kaya & Çiplak, 2017) and lizards (Ahmadzadeh et al., 2013). The connection between Anatolia and the Balkans has been intermittent during the last millions of years, particularly at the Bosphorus – Marmara Sea – Dardanelles. Land bridges between these landmasses have been established during the Messinian Salinity Crisis, and repeatedly during the Quaternary (Çağatay et al., 2006; Elmas, 2003; Gökaşan et al., 1997; Melinte-Dobrinescu et al., 2009). The divergence time estimation for the split of *Z. situla* supports a dispersal of the ancestor of *Z. situla*, *Z. longissimus* and *Z. lineatus* towards west either during the Pliocene (according to Cal I) or during the late Miocene (according to Cal II), followed by the split between *Z. lineatus* and *Z. longissimus* in the Italian Peninsula (Figs. 3.6 and 3.7; Table 3.4).

The climatic changes associated to the glacial/interglacial cycles during the Pleistocene have played an important role in shaping the current distribution of extant *Rhinechis* and *Zamenis* species and of their genetic diversity, as indicated by recent studies. Over 35 fossil records of *Rhinechis* found in the Iberian Peninsula, support the establishment of this species in its current distribution since the Pliocene (from 2.9 Ma until the Upper Pleistocene; Database of Vertebrates: fossil fishes, amphibians, reptiles and birds, available at: http://www.wahre-staerke.com/). The low genetic diversity observed in *R. scalaris* across its range likely reflects a drastic demographic and range contraction during the last glacial period, followed by a recent expansion across the Iberian Peninsula (Nulchis, Biaggini, Carretero, & Harris, 2008; Silva-Rocha, Salvi, Sillero, Mateo, & Carretero, 2015). Range contraction and divergence during the last glacial periods were suggested also for *Z. hohenackeri* *Z. longissimus* and *Z. situla*. The former presents a fragmented distribution with three genetic lineages probably associated with distinct glacial refugia: one from the Caucasus and northeastern Anatolia, other from southern Turkey and the last from the southern Amanos and Lebanon Mountains (Jandzik, Avci, & Gvoždík, 2013). According to Musilová, Zavadil, Marková, & Kotlík (2010), the current distribution of *Z. longissimus* is the result of the expansion of two main lineages from their respective western and eastern refugia after the last glacial maximum. Holocene fossils in Denmark, Germany and Poland suggest a

wider range of the species in northern areas where the species is no longer present (e.g. Böhme, 2000; Gomille, 2002; Helfenberger & Schulz, 2013; Ljungar, 1995; Szyndlar, 1984). Similarly, the range of *Z. situla* extended further north than presently, with fossils attributed to the lower Pleistocene in Austria and the Czech Republic, where the species became extinct likely during glacial periods. Molecular data including specimens of *Z. situla* from continental Greece, Aegean islands and Turkey revealed the existence of two distinct clades that diverged in the Pleistocene: one from Crete, Thera and the Peloponnese and the other from Turkey, northern continental Greece and the eastern Aegean island Samos (Kyriazi et al., 2013). It remains unclear whether the distribution of *Z. situla* in the Aegean area is a result of over-sea dispersal from mainland Greece or if it was mediated by humans (Kyriazi et al., 2013; Poulakakis et al., 2015). Phenotypic and genetic data on *Z. lineatus* indicate reduced diversity and documented a wide area of introgression with *Z. longissimus* at the eastern contact zone between these species (Salvi et al., 2017). The pattern of introgression seems asymmetric, with a prevalent introgression of *Z. lineatus* traits into the *Z. longissimus* background. Whether such a pattern resulted from a scenario of demographic and geographic range expansion on one species into the range of the other is a question under study (Salvi et al., in prep).

SYSTEMATICS OF *ZAMENIS* AND *RHINECHIS*: ONE OR TWO GENERA?

From a phenotypic point of view *R. scalaris* shares several morphological features with *Zamenis* species (Schulz, 1996). The head scalation pattern is particularly similar between *R. scalaris* and *Z. situla*/*Z. longissimus* (Schulz, 1996), except as regards the rostral shield. While in *Zamenis* species the rostral shield is broader than deeper, like in other members of the genus *Elaphe*, in *R. scalaris* the rostral shield is deeper than broader and forms an acute angle posteriorly (pointing behind), which represents a unique trait in ratsnakes (Boulenger, 1894). To our knowledge this is the only external feature that differentiates the genera *Rhinechis* and *Zamenis* (Boulenger, 1894; Venchi & Sindaco, 2006).

In contrast, Helfenberger (2001) pointed out the phylogenetic distinctiveness of *R. scalaris* based on the analysis of the anatomical, osteological and allozyme variation of Palaearctic ratsnakes. However, the characters analysed in this study show little phylogenetic value and high level of homoplasy, and the uniqueness of *R. scalaris* is not apparent in any of these datasets. Each derived visceral feature listed for *R. scalaris* (the cranial reduction of the hypapophyses, a cranially shifted heart, long right and rudimentary left lung, a tracheal opening which ends far caudally and large overlap of lung and liver) is shared with one or more species of any of the genera analysed

Oocatochus, *Coronella*, *Elaphe* and *Coelognathus* Fitzinger, 1843. Phenograms based on soft anatomical characters (position and length of visceral organs) and vertebral characters revealed conflicting patterns across characters and sexes (diffuse homoplasy, sensu Lee, 2000), whereas one third of the analysed taxa lacks private alleles at the allozyme loci analysed suggesting the occurrence of homoplasic alleles (distinct alleles with equal electrophoretic mobility). Many phenograms based on visceral and vertebrae features show similarity between *R. scalaris* and *Z. hohenackeri* or *Z. situla* (e.g. Figs. 14-17 in Helfenberger, 2001), while that one based on allozyme data shows a close relationship between the genera *Rhinechis*, *Coronella* and *Zamenis* (Fig. 19 in Helfenberger, 2001), which is consistent with molecular phylogenetic studies (Burbrink & Lawson, 2007; Chen et al., 2017; this study). However, it is problematic drawing 'phylogenetic' inferences from the analysis of characters such those analysed by Helfenberger (2001) because of the lack of a statistical or genealogical framework to understand the evolutionary relationships between character states or alleles.

Phylogenetic reconstruction based on molecular data provide compelling evidence for a close relationship between *Rhinechis* and *Zamenis*. These taxa form a well-supported monophyletic clade in all previous phylogenetic assessments based on different molecular markers, taxon-set, phylogenetic methods (Burbrink & Lawson, 2007; Pyron et al., 2011; Pyron et al., 2013; Zheng & Wiens, 2016, and also Lenk et al., 2001, or see section were Lenk et al., 2001 is discussed). On the other hand, while in many phylogenies *R. scalaris* is sister to the clade including all *Zamenis* species, there is no statistical support for the monophyly of *Zamenis*. It is unlikely that this phylogenetic uncertainty stem from a lack of sufficient data in previous studies and in this study. Indeed, even a recent phylogenomic study employing as much as 304 nuclear loci failed to recover a statistically supported monophyletic clade for *Zamenis* and to resolve the relationships between *R. scalaris* and *Z. hohenackeri* and the remaining *Zamenis* species (Chen et al., 2017). We argue that this phylogenetic uncertainty reflects an evolutionary scenario in which early cladogenetic events within the *Rhinechis*-*Zamenis* clade took place at the same time or in a short time frame so that it is hard disentangling which (if any) lineage branched off earlier between *R. scalaris, Z. hohenackeri* and possibly *Z. persicus* (Figs. 3.4 – 3.6, Table 3.4, see also Chen et al., 2017). As a consequence, while the monophyletic clade of *Zamenis* received little support in all phylogenetic analyses, the clade including *Rhinechis* and *Zamenis* received maximum support.

In conclusion, while we have little morphological and phylogenetic evidence for the distinctiveness of *Rhinechis*, a classification combining the genera *Rhinechis* and

Zamenis into a single unit would better reflect their evolutionary relationships. Such a classification is based on a straightforward definition of monophyletic clade, and allows avoiding the use of monotypic genera, which are questionable especially when compelling evidence for merging them with other genera is available. Additionally, *R. scalaris* shows no signs of phylogeographic structure across its distribution range, further confirming the monotypic status of *Rhinechis* (Nulchis et al., 2008). The genus *Zamenis* Wagler, 1830, has precedence over *Rhinechis* Michahelles, 1833. Therefore, based on the priority rule, the species *R. scalaris* (Schinz, 1822) is moved into the genus *Zamenis* and designated as *Zamenis scalaris* **comb. nov.**

ACKNOWLEDGMENTS

We are indebted with Philippe Geniez (CEFE, Centre d'Écologie Fonctionnelle et Évolutive, Montpellier, France) and Faraham Ahmadzadeh (Shahid Beheshti University, Tehran, Iran) for donating tissue samples of Middle-East species and with Notker Helfenberger for constructive discussion. We thank Àlex Cortada for the help with the biogeographical reconstruction analyses. DS is currently supported by the program 'Rita Levi Montalcini' for the recruitment of young researchers at the University of L'Aquila. JM and DJH are supported by the Fundação para a Ciência e a Tecnologia (FCT, Portugal): JM, doctoral grant SFRH/BD/81528/2011; DJH IF contract 01627/2014. SC is supported by Grant CGL2015-70390 from the Ministerio de Economía y Competitividad, Spain (co-funded by Fondos FEDER – EU).

REFERENCES

Ahmadzadeh, F., Flecks, M., Rödder, D., Böhme, W., Ilgaz, Ç., Harris, D. J., … 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. http://doi.org/10.1111/bij.12129

Arevalo, E., Davis, S. K., & Sites, J. W. J. (1994). Mitochondrial DNA sequence divergence and phylogenetic relationships among eight chromosome races of the *Sceloporus grammicus* complex (Phrynosomatidae) in Central Mexico. *Systematic Bioloy*, *43*, 387–418. http://doi.org/10.1093/sysbio/43.3.387

Bandelt, H.-J., Forster, P., & Rohl, A. (1999). Median-Joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution*, *16*, 37–48. http://doi.org/10.1093/oxfordjournals.molbev.a026036

Blondel, J., Catzeflis, F., & Perret, P. (1996). Molecular phylogeny and the historical biogeography of the warblers of the genus *Sylvia* (Aves). *Journal of Evolutionary Biology*, *9*, 871–891. http://doi.org/10.1046/j.1420-9101.1996.9060871.x

Böhme, G. (2000). Fossile Amphibien und Reptilien im Quartär Thüringens. *Veröffentlichungen - Naturkundemuseum Erfurt*, *19*, 79–97.

Boulenger, G. A. (1894). *Catalogue of the Snakes in the British Museum (Natural History). Volume II., Containing the Conclusion of the Colubridæ Aglyphæ*. London. British Museum of Natural History, Ed.

Bruen, T. C., Philippe, H., & Bryant, D. (2006). A simple and robust statistical test for detecting the presence of recombination. *Genetics*, *172*, 2665–2681. http://doi.org/10.1534/genetics.105.048975

Bryant, D., & Moulton, V. (2004). Neighbor-Net: An agglomerative method for the construction of phylogenetic networks. *Molecular Biology and Evolution*, *21*, 255–265. http://doi.org/10.1093/molbev/msh018

Bryson, R. W., de Oca, A. N.-M., & Velasco, J. R. (2008). Phylogenetic position of *Porthidium hespere* (Viperidae: Crotalinae) and phylogeography of arid-adapted Hognosed Pitvipers based on mitochondrial DNA. *Copeia*, *2008*, 172–178. http://doi.org/10.1643/CH-07-043

Burbrink, F. T., & Lawson, R. (2007). How and when did Old World ratsnakes disperse into the New World? *Molecular Phylogenetics and Evolution*, *43*, 173–189. http://doi.org/10.1016/j.ympev.2006.09.009

Çağatay, M. N., Görür, N., Flecker, R., Sakinç, M., Tünoğlu, C., Ellam, R., … Dikbaş, A. (2006). Paratethyan-Mediterranean connectivity in the Sea of Marmara region (NW Turkey) during the Messinian. *Sedimentary Geology*, *188–189*, 171–187. http://doi.org/10.1016/j.sedgeo.2006.03.004

Carranza, S., Arnold, E. N., & Pleguezuelos, J. M. (2006). Phylogeny, biogeography, and evolution of two Mediterranean snakes, *Malpolon monspessulanus* and *Hemorrhois hippocrepis* (Squamata, Colubridae), using mtDNA sequences. *Molecular Phylogenetics and Evolution*, *40*, 532–546. http://doi.org/10.1016/j.ympev.2006.03.028

Chen, X., Jiang, K., Guo, P., Huang, S., Rao, D., Ding, L., … Burbrink, F. T. (2014). Assessing species boundaries and the phylogenetic position of the rare Szechwan ratsnake, *Euprepiophis perlaceus* (Serpentes: Colubridae), using coalescent-based methods. *Molecular Phylogenetics and Evolution*, *70*, 130–136. http://doi.org/10.1016/j.ympev.2013.09.003

Chen, X., Lemmon, A. R., Lemmon, E. M., Pyron, R. A., & Burbrink, F. T. (2017). Using phylogenomics to understand the link between biogeographic origins and regional diversification in ratsnakes. *Molecular Phylogenetics and Evolution*, *111*, 206–218. http://doi.org/10.1016/j.ympev.2017.03.017

Chobanov, D. P., Kaya, S., Grzywacz, B., Warchalowska-Sliwa, E., & Ciplak, B. (2017). The Anatolio-Balkan phylogeographic fault: a snapshot from the genus *Isophya* (Orthoptera, Tettigoniidae). *Zoologica Scripta*, *46*, 165–179. http://doi.org/10.1111/zsc.12194

Creutzburg, N. (1963). Paleogeographic evolution of Crete from Miocene till our days. *Cretan Annals*, *15–16*, 336–342.

Cummins, C. A., & McInerney, J. O. (2011). A method for inferring the rate of evolution of homologous characters that can potentially improve phylogenetic inference, resolve deep divergence and correct systematic biases. *Systematic Biology*, *60*, 833–844. http://doi.org/10.1093/sysbio/syr064

Daza, J. M., Smith, E. N., Páez, V. P., & Parkinson, C. L. (2009). Complex evolution in the Neotropics: The origin and diversification of the widespread genus *Leptodeira* (Serpentes: Colubridae). *Molecular Phylogenetics and Evolution*, *53*, 653–667. http://doi.org/10.1016/j.ympev.2009.07.022

Dermitzakis, D. M., & Papanikolaou, D. J. (1981). Paleogeography and geodynamics of the Aegean region during the Neogene. *Annales Geologique Des Pays Helleenique*, *30*, 245–289.

Drummond, A. J., Suchard, M. A., Xie, D., & Rambaut, A. (2012). Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution*, *29*, 1969–1973. http://doi.org/10.1093/molbev/mss075

Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, *32*, 1792–1797. http://doi.org/10.1093/nar/gkh340

Elmas, A. (2003). Late Cenozoic tectonics and stratigraphy and northwestern Anatolia: The effects of the North Anatolian Fault to the region. *International Journal of Earth Sciences*, *92*, 380–396. http://doi.org/10.1007/s00531-003-0322-2

Eronen, J. T., Ataabadia, M. M., Micheelsb, A., Karme, A., Bernor, R. L., & Fortelius, M. (2009). Distribution history and climatic controls of the Late Miocene Pikermian chronofauna. *Proceedings of the National Academy of Sciences of the United States of America*, *106*, 11867–11871. http://doi.org/DOI 10.1073/pnas.0902598106

Figueroa, A., McKelvy, A. D., Grismer, L. L., Bell, C. D., & Lailvaux, S. P. (2016). A Species-level phylogeny of extant snakes with description of a new colubrid subfamily and genus. *PLOS ONE*, *11*, e0161070. http://doi.org/10.1371/journal.pone.0161070

Flot, J.-F. (2010). SEQPHASE: a web tool for interconverting PHASE input/output files and FASTA sequence alignments. *Molecular Ecology Resources*, *10*, 162–166. http://doi.org/10.1111/j.1755-0998.2009.02732.x

Gökaşan, E., Demirbag̨, E., Oktay, F. Y., Ecevitog̨lu, B., Şimşek, M., & Yüce, H. (1997). On the origin of the Bosphorus. *Marine Geology*, *140*, 183–199. http://doi.org/10.1016/S0025-3227(97)00022-4

Gomille, A. (2002). *Die Äskulapnatter Elaphe longissima – Verbreitung und Lebensweise in Mitteleuropa*. Frankfurt am Main: Edition Chimaira.

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

Heled, J., & Drummond, A. J. (2010). Bayesian inference of species trees from multilocus data. *Molecular Biology and Evolution*, *27*, 570–580. http://doi.org/10.1093/molbev/msp274

Helfenberger, N. (2001). Phylogenetic relationships of Old World ratsnakes based on visceral organ topography, osteology, and allozyme variation. *Russian Journal of Herpetology*, *8*, 1–62.

Helfenberger, N., & Schulz, K.-D. (2013). A checklist of the extinct taxa allied to the Old World ratsnakes, with remarks on the phylogeny and radiation of *Elaphe*, *Rhinechis* and *Zamenis*. In K.-D. Schultz (Ed.), *Old World ratsnakes - A collection of papers.* (pp. 423–427). Berg SG: Bushmaster Publications.

Hoelzer, G. A., & Meinick, D. J. (1994). Patterns of speciation and limits to phylogenetic resolution. *Trends in Ecology & Evolution*, *9*, 104–107.

Huson, D. H., & Bryant, D. (2006). Application ofphylogenetic networks in evolutionary studies. *Molecular Biology and Evolution*, *23*, 254–267. http://doi.org/10.1093/molbev/msj030

Jandzik, D., Avci, A., & Gvoždík, V. (2013). Incongruence between taxonomy and genetics: Three divergent lineages within two subspecies of the rare Transcaucasian rat snake (*Zamenis hohenackeri*). *Amphibia Reptilia*, *34*, 579–584. http://doi.org/10.1163/15685381-00002911

Kapli, P., Botoni, D., Ilgaz, C., Kumlutaş, Y., Avcı, A., Rastegar-Pouyani, N., … Poulakakis, N. (2013). Molecular phylogeny and historical biogeography of the Anatolian lizard *Apathya* (Squamata, Lacertidae). *Molecular Phylogenetics and Evolution*, *66*, 992–1001. http://doi.org/10.1016/j.ympev.2012.12.002

Kaya, S., & Çiplak, B. (2017). Phylogeography and taxonomy of the *Psorodonotus caucasicus* (Orthoptera, Tettigoniidae) group: independent double invasion of the Balkans from the Caucasus. *Systematic Entomology*, *42*, 118–133. http://doi.org/10.1111/syen.12197

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., … Drummond, A. (2012). Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, *28*, 1647–1649. http://doi.org/10.1093/bioinformatics/bts199

Kornilios, P., Ilgaz, Ç., Kumlutaş, Y., Giokas, S., Fraguedakis-Tsolis, S., & Chondropoulos, B. (2011). The role of Anatolian refugia in herpetofaunal diversity: an mtDNA analysis of *Typhlops vermicularis* Merrem, 1820 (Squamata, Typhlopidae). *Amphibia-Reptilia*, *32*, 351–363. http://doi.org/10.1163/017353711X579858

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 Biogeography*, *40*, 1143–1155. http://doi.org/10.1111/jbi.12057

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

Lawson, R., Slowinski, J. B., Crother, B. I., & Burbrink, F. T. (2005). Phylogeny of the Colubroidea (Serpentes): New evidence from mitochondrial and nuclear genes. *Molecular Phylogenetics and Evolution*, *37*, 581–601. http://doi.org/10.1016/j.ympev.2005.07.016

Lee, M. S. (2000). Soft anatomy, diffuse homoplasy, and the relationships of lizards and snakes. *Zoologica Scripta*, *29*, 101–130. http://doi.org/10.1046/j.1463-6409.2000.00035.x

Lenk, P., Joger, U., & Wink, M. (2001). Phylogenetic relationships among European rat snakes of the genus *Elaphe* Fitzinger based on mitochondrial DNA sequence comparisons. *Amphibia-Reptilia*, *22*, 329-339.

Ljungar, L. (1995). First subfossil find of the Aesculapian snake, *Elaphe longissima* (Laur.) from a Mesolithic settlement in Denmark. *Amphibia-Reptilia*, *16*, 93–94.

Lo Presti, R. M., & Oberprieler, C. (2009). Evolutionary history, biogeography and eco-climatological differentiation of the genus *Anthemis* L. (Compositae, Anthemideae) in the circum-Mediterranean area. *Journal of Biogeography*, *36*, 1313–1332. http://doi.org/10.1111/j.1365-2699.2009.02121.x

Maddison, W. P., & Maddison, D. R. (2003). Mesquite: a modular system for evolutionary analysis. http://mesquiteproject.org

Melinte-Dobrinescu, M. C., Suc, J.-P., Clauzon, G., Popescu, S.-M., Armijo, R., Meyer, B., … Çakir, Z. (2009). The Messinian Salinity Crisis in the Dardanelles region: Chronostratigraphic constraints. *Palaeogeography, Palaeoclimatology, Palaeoecology*, *278*, 24–39. http://doi.org/10.1016/j.palaeo.2009.04.009

Mendes, J., Harris, D. J., Carranza, S., & Salvi, D. (2016). Evaluating the phylogenetic signal limit from mitogenomes, slow evolving nuclear genes, and the concatenation approach. New insights into the Lacertini radiation using fast evolving nuclear genes and species trees. *Molecular Phylogenetics and Evolution*, *100*, 254–267. http://doi.org/10.1016/j.ympev.2016.04.016

Mezzasalma, M., Dall'Asta, A., Loy, A., Cheylan, M., Lymberakis, P., Zuffi, M. A. L., … Guarino, F. M. (2015). A sisters' story: Comparative phylogeography and taxonomy of *Hierophis viridiflavus* and *H. gemonensis* (Serpentes, Colubridae). *Zoologica Scripta*, *44*, 495–508. http://doi.org/10.1111/zsc.12115

Micó, E., Sanmartín, I., & Galante, E. (2009). Mediterranean diversification of the grass-feeding *Anisopliina* beetles (Scarabaeidae, Rutelinae, Anomalini) as inferred by

bootstrap-averaged dispersal-vicariance analysis. *Journal of Biogeography*, *36*, 546–560. http://doi.org/10.1111/j.1365-2699.2008.02010.x

Musilová, R., Zavadil, V., Marková, S., & Kotlík, P. (2010). Relics of the Europe's warm past: Phylogeography of the Aesculapian snake. *Molecular Phylogenetics and Evolution*, *57*, 1245–1252. http://doi.org/10.1016/j.ympev.2010.09.017

Myers, E. A., Rodríguez-Robles, J. A., Denardo, D. F., Staub, R. E., Stropoli, A., Ruane, S., & Burbrink, F. T. (2013). Multilocus phylogeographic assessment of the California Mountain Kingsnake (*Lampropeltis zonata*) suggests alternative patterns of diversification for the California Floristic Province. *Molecular Ecology*, *22*, 5418–5429. http://doi.org/10.1111/mec.12478

Nagy, Z. T., Lawson, R., Joger, U., & Wink, M. (2004). Molecular systematics of racers, whipsnakes and relatives (Reptilia: Colubridae) using mitochondrial and nuclear markers. *Journal of Zoological Systematics and Evolutionary Research*, *42*, 223–233. http://doi.org/10.1111/j.1439-0469.2004.00249.x

Nulchis, V., Biaggini, M., Carretero, M. A., & Harris, D. J. (2008). Unexpectedly low mitochondrial DNA variation within the ladder snake *Rhinechis scalaris*. *North-Western Journal of Zoology*, *4*, 119–124.

Oberprieler, C. (2005). Temporal and spatial diversification of Circum-Mediterranean Compositae–Anthemideae. *Taxon*, *54*, 951–966.

Palumbi, S. R., Martin, A., Romano, S., McMillan, S., Stice, W. O., & Grabowski, L. (1991). *The simple fool's guide to PCR* (1st ed.). Honolulu: Department of Zoology, University of Hawaii.

Plummer, M., Best, N., Cowles, K., & Vines, K. (2006). CODA: convergence diagnosis and output analysis for MCMC. *R News*, *6*, 7–11. http://doi.org/10.1159/000323281

Popov, S. V., Shcherba, I. G., Ilyina, L. B., Nevesskaya, L. A., Paramonova, N. P., Khondkarian, S. O., & Magyar, I. (2006). Late Miocene to Pliocene palaeogeography of the Paratethys and its relation to the Mediterranean. *Palaeogeography, Palaeoclimatology, Palaeoecology*, *238*, 91–106. http://doi.org/10.1016/j.palaeo.2006.03.020

Poulakakis, N., Kapli, P., Lymberakis, P., Trichas, A., Vardinoyiannis, K., Sfenthourakis, S., & Mylonas, M. (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. http://doi.org/10.1111/jzs.12071

Pyron, R. A., & Burbrink, F. T. (2009). Neogene diversification and taxonomic stability in the snake tribe Lampropeltini (Serpentes: Colubridae). *Molecular Phylogenetics and Evolution*, *52*, 524–529. http://doi.org/10.1016/j.ympev.2009.02.008

Pyron, R. A., Burbrink, F. T., Colli, G. R., de Oca, A. N. M., Vitt, L. J., Kuczynski, C. A., & Wiens, J. J. (2011). The phylogeny of advanced snakes (Colubroidea), with discovery of a new subfamily and comparison of support methods for likelihood trees. *Molecular Phylogenetics and Evolution*, *58*, 329–342. http://doi.org/10.1016/j.ympev.2010.11.006

Pyron, R., Burbrink, F. T., & Wiens, J. J. (2013). A phylogeny and revised classification of Squamata, including 4161 species of lizards and snakes. *BMC Evolutionary Biology*, *13*, 93. http://doi.org/10.1186/1471-2148-13-93

Quintero, I., Keil, P., Jetz, W., & Crawford, F. W. (2015). Historical biogeography using species geographical ranges. *Systematic Biology*, *64*, 1059–1073. http://doi.org/10.1093/sysbio/syv057

Rambaut, A., Suchard, M. A., & Drummond, A. J. (2014). Tracer v1.6, Available from http://tree.bio.ed.ac.uk/software/tracer/.

Rodríguez-Robles, J. A., & De Jesús-Escobar, J. M. (1999). Molecular systematics of New World lampropeltine snakes (Colubridae): implications for biogeography and evolution of food habits. *Biological Journal of the Linnean Society*, *68*, 355–385. http://doi.org/10.1006/bijl.1999.0320

Rögl, F. (1998). Palaeogeographic considerations for Mediterranean and Paratethys seaways (Oligocene to Miocene). *Annalen Des Naturhistorischen Museums in Wien*, *99*, 279–310.

Ruane, S., Bryson, R. W., Pyron, R. A., & Burbrink, F. T. (2014). Coalescent species delimitation in Milksnakes (Genus *Lampropeltis*) and impacts on phylogenetic comparative analyses. *Systematic Biology*, *63*, 231–250. http://doi.org/10.1093/sysbio/syt099

Ruane, S., Torres-Carvajal, O., & Burbrink, F. T. (2015). Independent demographic responses to climate change among temperate and tropical milksnakes (Colubridae: Genus *Lampropeltis*). *PLoS ONE*, *10*, 1–17. http://doi.org/10.1371/journal.pone.0128543

Salvi, D., Lucente, D., Mendes, J., Liuzzi, C., Harris, D. J., & Bologna, M. A. (2017). Diversity and distribution of the Italian Aesculapian snake *Zamenis lineatus*: A phylogeographic assessment with implications for conservation. *Journal of*

Zoological Systematics and Evolutionary Research, *55*, 222–237. http://doi.org/10.1111/jzs.12167

Sambrook, J., Fritsch, E. F., & Maniatis T. (1989). *Molecular Cloning: A Laboratory Manual*. New York, Cold Spring Harbor Press.

Santos, X., Roca, J., Pleguezuelos, J. M., Donaire, D., & Carranza, S. (2008). Biogeography and evolution of the Smooth snake *Coronella austriaca* (Serpentes: Colubridae) in the Iberian Peninsula: evidence for Messinian refuges and Pleistocenic range expansions. *Amphibia-Reptilia*, *29*, 35-47.

Schulz, K.-D. (1996). *A monograph of the colubrid snakes of the genus* Elaphe *Fitzinger*. Würselen, Koeltz Scientific Books.

Shimodaira, H. (2002). An approximately unbiased test of phylogenetic tree selection. *Systematic Biology*, *51*, 492–508. http://doi.org/10.1080/10635150290069913

Shimodaira, H., & Hasegawa, M. (1999). Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Molecular Biology and Evolution*, *16*, 1114–1116.

Shimodaira, H., & Hasegawa, M. (2001). CONSEL: for assessing the confidence of phylogenetic tree selection. *Bioinformatics*, *17*, 1246–1247. http://doi.org/10.1093/bioinformatics/17.12.1246

Silva-Rocha, I., Salvi, D., Sillero, N., Mateo, J. A., & Carretero, M. A. (2015). Snakes on the Balearic Islands: An Invasion tale with implications for native biodiversity conservation. *PLOS ONE*, *10*, e0121026. http://doi.org/10.1371/journal.pone.0121026

Silvestro, D., & Michalak, I. (2012). raxmlGUI: a graphical front-end for RAxML. *Organisms Diversity & Evolution*, *12*, 335–337. http://doi.org/10.1007/s13127-011-0056-0

Skourtanioti, E., Kapli, P., Ilgaz, Ç., Kumlutaş, Y., Avcı, A., Ahmadzadeh, F., … Poulakakis, N. (2016). A reinvestigation of phylogeny and divergence times of the *Ablepharus kitaibelii* species complex (Sauria, Scincidae) based on mtDNA and nuDNA genes. *Molecular Phylogenetics and Evolution*, *103*, 199–214. http://doi.org/10.1016/j.ympev.2016.07.005

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

Stephens, M., & Scheet, P. (2005). Accounting for decay of linkage disequilibrium in haplotype inference and missing-data imputation. *American Journal of Human Genetics*, *76*, 449–462. http://doi.org/10.1086/428594

Stephens, M., Smith, N. J., & Donnelly, P. (2001). A new statistical method for haplotype reconstruction from population data. *American Journal of Human Genetics*, *68*, 978–989. http://doi.org/10.1086/319501

Szyndlar, Z. (1984). Fossil snakes from Poland. *Acta Zoologica Cracoviensia*, *28*, 1–156.

Szyndlar, Z. (1985). Ophidian fauna (Reptilia, Serpentes) from the uppermost Miocene of Algora (Spain). *Estudios Geológicos*, *41*, 447–465.

Tamura, K., Stecher, G., Peterson, D., Filipski, A., & Kumar, S. (2013). MEGA6: Molecular evolutionary genetics analysis version 6.0. *Molecular Biology and Evolution*, *30*, 2725–2729. http://doi.org/10.1093/molbev/mst197

Team, Rs. (2017). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. https://www.r-project.org/

Townsend, T. M., Alegre, R. E., Kelley, S. T., Wiens, J. J., & Reeder, T. W. (2008). Rapid development of multiple nuclear loci for phylogenetic analysis using genomic resources: An example from squamate reptiles. *Molecular Phylogenetics and Evolution*, *47*, 129–142. http://doi.org/10.1016/j.ympev.2008.01.008

Utiger, U., Helfenberger, N., Schätti, B., Schmidt, C., Ruf, M., & Ziswiler, V. (2002). Molecular systematics and phylogeny of Old World and New World ratsnakes, *Elaphe* Auct., and related genera (Reptilia, Squamata, Colubridae). *Russian Journal of Herpetology*, *9*, 105–124.

Utiger, U., Schatti, B., & Helfenberger, N. (2005). The oriental colubrine genus *Coelognathus* Fitzinger, 1843 and classification of old and new world racers and ratsnakes (Reptilia, Squamata, Colubridae, Colubrinae). *Russian Journal of Herpetology*, *12*, 39–60.

Venchi, A., & Sindaco, R. (2006). Annotated checklist of the reptiles of the Mediterranean countries, with keys to species identification. Part 2 - Snakes (Reptilia, Serpentes). *Annali Del Museo Civico Di Storia Naturale "G.Doria," XCVIII*, 259–364.

Venczel, M. (1994). Late Miocene snakes from Polgárdi (Hungary). *Acta Zoologica. Cracoviensia*, *37*, 1–29.

Weisrock, D. W., Macey, J. R., Ugurtas, I. H., Larson, A., & Papenfuss, T. J. (2001). Molecular phylogenetics and historical biogeography among salamandrids of the

"true" salamander clade: rapid branching of numerous highly divergent lineages in *Mertensiella luschani* associated with the rise of Anatolia. *Molecular Phylogenetics and Evolution*, *18*, 434–448. http://doi.org/10.1006/mpev.2000.0905

Yu, Y., Harris, A. J., Blair, C., & He, X. (2015). RASP (Reconstruct Ancestral State in Phylogenies): a tool for historical biogeography. *Molecular Phylogenetics and Evolution*, *87*, 46–49. http://doi.org/10.1016/j.ympev.2015.03.008

Yu, Y., Harris, A. J., & He, X. (2010). S-DIVA (Statistical Dispersal-Vicariance Analysis): A tool for inferring biogeographic histories. *Molecular Phylogenetics and Evolution*, *56*, 848–850. http://doi.org/10.1016/j.ympev.2010.04.011

Yu, Y., Harris, A. J., & Xingjin, H. (2011). RASP version2.1b. Available at http://mnh.scu.edu.cn/soft/blog/RASP.

Zheng, Y., & Wiens, J. J. (2016). Combining phylogenomic and supermatrix approaches, and a time-calibrated phylogeny for squamate reptiles (lizards and snakes) based on 52 genes and 4162 species. *Molecular Phylogenetics and Evolution*, *94*, 537–547. http://doi.org/10.1016/j.ympev.2015.10.009

SUPPORTING INFORMATION

Additional Supporting Information can be found in the Appendices section.

Table S3.1. Divergence time estimates and 95% HPD interval based on rates from calibrations I and II with the *Zamenis* + ratsnakes taxon-set.

Table S3.2. Geographical location of the ancestors and lineages of *Rhinechis* and *Zamenis* for each time slice.

Figure S3.1. Bayesian tree topology with one individual per species and the Yule prior of speciation.

Figure S3.2. Median-joining haplotype networks for *Zamenis*, *Rhinechis*, *Coronella* and outgroups inferred with mitochondrial and phased nuclear DNA sequences.

Figure S3.3. Bayesian tree depicting phylogenetic relationships of *Zamenis*, *Rhinechis* and *Coronella* based on concatenated nuclear DNA sequences.

Figure S3.4. Bayesian tree depicting phylogenetic relationships of *Zamenis*, *Rhinechis* and *Coronella* based on concatenated nuclear DNA sequences using the reduced dataset.

Figure S3.5. Bayesian tree depicting phylogenetic relationships of *Zamenis*, *Rhinechis*, *Coronella* based on concatenated mitochondrial and nuclear DNA sequences using the reduced dataset.

Figure S3.6. Bayesian tree depicting phylogenetic relationships of *Zamenis*, *Rhinechis*, *Coronella* and representatives of five other ratsnakes' genera based on concatenated nuclear DNA sequences.

Figure S3.7. Species tree of *Zamenis*, *Rhinechis*, *Coronella* and representatives of five other ratsnakes' genera based on DNA sequences from mitochondrial and nuclear loci.

Chapter 4

Multilocus phylogeny of *Psammodromus* lizards

Article III. Biogeographic crossroad across the Pillars of Hercules: evolutionary history of *Psammodromus* lizards in space and time

Joana Mendes[1,2,3], D. James Harris[1,2], Salvador Carranza[3], Daniele Salvi[1,4]

Journal of Biogeography, 2017, 44: 2877-2890;

DOI: 10.1111/jbi.13100

[1] CIBIO Research Centre in Biodiversity and Genetic Resources, InBIO, Universidade do Porto, Campus Agrário de Vairão, Rua Padre Armando Quintas, N° 7, 4485-661 Vairão, Vila do Conde, Portugal

[2] Departamento de Biologia, Faculdade de Ciências, Universidade do Porto, Rua do Campo Alegre FC4 4169-007 Porto, Portugal

[3] Institute of Evolutionay Biology (CSIC-Universitat Pompeu Fabra), Passeig Marítim de la Barceloneta 37-49, 08003 Barcelona, Spain.

[4] Department of Health, Life and Environmental Sciences, University of L'Aquila, 67100 Coppito, L'Aquila, Italy

ABSTRACT

Aim To infer the biogeographical and evolutionary history of the Western Mediterranean *Psammodromus* lizards with the aim of assessing the role of vicariance and dispersal on the cladogenetic events within the palaeogeological dynamics of the Strait of Gibraltar.

Location North Africa and Western Europe

Methods We built a dataset including all six species of *Psammodromus* using mitochondrial (*12S*, *cytb*, *nd4*) and nuclear (*acm4*, *mc1r*, *pomc*) gene fragments. Species tree and concatenation methods were used to infer phylogenetic relationships and divergence times. Phylogenies were used for biogeographical inference using S-DIVA, DEC and BBM.

Results *Psammodromus* probably originated in Iberia, with *P. algirus* diverging early. The ancestor of the African *P. blanci* and *P. microdactylus* dispersed to Africa through the Betic-Rif massif, approximately 10 Ma. The cladogenetic events within Africa and Iberia were probably due to vicariance mediated by habitat and climatic changes at the end of the Miocene (*P. blanci* and *P. microdactylus*) and during the Pliocene (*P. occidentalis*, *P. hispanicus* and *P. edwardsianus*). *Psammodromus algirus* shows three lineages, two in Iberia and one in Africa, the latter originated following a transmarine dispersal during the Middle Pleistocene (1.5 Ma).

Main Conclusions Over-sea dispersal has played a major role in intercontinental exchange and divergence in *Psammodromus*, with two dispersal events towards Africa that occurred 10 Ma and 1.5 Ma originating the African lineages. This study, combined with previous literature, provides compelling evidence that major biotic exchanges took place across the Strait of Gibraltar well before or long after the land connection during the Messinian Salinity Crisis (MSC, 5.9 – 5.33 Ma). These findings suggest caution in the application of the relatively short event of Atlantic flooding at the end of the MSC as cause for divergence in molecular clock calibrations, which is a popular approach in literature.

KEYWORDS

Dispersal, Iberian Peninsula, Lacertidae, Messinian Salinity Crisis, phylogeography, North Africa, species tree, Strait of Gibraltar, transmarine colonization, vicariance

INTRODUCTION

Biotic exchange between continents through vicariance and dispersal has classically been the focus of many biogeographic studies. The main faunal exchanges of North Africa have been with the Arabian Peninsula and the Levant through the Sinai in the east and with south-western Europe across the Strait of Gibraltar in the west. While dispersal has played a major role in the stable land bridge at the eastern crossroad (e.g. Carranza, Arnold, Geniez, Roca, & Mateo, 2008; Metallinou et al., 2015; Tamar et al., 2016) the history of intermittent connections at the Strait of Gibraltar has created opportunities for both vicariance and dispersal to contribute to the biotic exchange between North Africa and Western Europe (e.g. Carranza et al., 2008; García-Vázquez et al., 2016; Jaramillo-Correa et al., 2010).

The Pillars of Hercules, as the Strait of Gibraltar has been known since ancient time, is an excellent setting for biogeographic inference, due to its well-known complex geological history (Fig. 4.1). Briefly, in the middle Oligocene, the Hercynian belt detached from Iberia and broke into smaller blocks, including the Betic-Rif, which drifted southwards until reaching its final position in the Strait of Gibraltar, 10 Ma. Here, from the Middle Miocene, tectonic uplift caused progressive land emergence and reduction of seaways (Braga, Martin, & Quesada, 2003; Duggen, Hoernie, van den Bogaard, Rupke, & Morgan, 2003; Rosenbaum, Lister, & Duboz, 2002a), until the connection between the Mediterranean and the Atlantic was maintained by the Betic, Guadalhorce, and the Rif corridors. The Betic corridor closed approximately 10 Ma and re-opened 8-9 Ma. As the sea level gradually decreased, this corridor closed definitively (Krijgsman et al., 2000), followed by the Guadalhorce (Pérez-Asensio, Aguirre, Schmiedl, & Civis, 2012) and the Rif corridors, creating a land bridge between Africa and Iberia (Krijgsman & Langereis, 2000), setting in motion the Messinian Salinity Crisis (MSC), from 5.96 to 5.33 Ma. The MSC ended abruptly with the formation of the Strait of Gibraltar, which remained open until the present (Duggen et al., 2003), although with varying extent. During the Pliocene the sea level oscillations shortened the distance between Africa and Iberia from the actual 14 km to 4-5 km, with newly-formed islands available as stepping-stones for transcontinental colonizations of terrestrial species (Zazo, 1999).

As a consequence of this complex geological history, multiple dispersal and vicariance events took place across the Strait of Gibraltar, profoundly shaping genetic and demographic patterns of a wide number of taxa. Before the formation of the Strait of Gibraltar, the Betic-Rif Massif acted as an agent for dispersal and vicariance for taxa differentiation. Both the first closure of the Betic corridor, 10 Ma, and the MSC, 5.33 Ma,

led to the dispersal of taxa in the area of the sub-Betic region and between Iberia and North Africa, respectively, whereas the re-opening of the Betic-corridor and the Strait of Gibraltar triggered the divergence of populations on both sides of the Strait (e.g. Martínez-Solano et al., 2004; Carranza et al., 2008; Velo-Antón et al., 2012). Transmarine colonizations long after the MSC had finished have also been reported (Carranza et al., 2008; Carranza, Arnold, & Pleguezuelos, 2006; Carranza, Arnold, Wade, & Fahd, 2004; Carranza, Harris, Arnold, Batista, & Gonzalez de la Vega, 2006; Harris, Batista, Carretero, & Ferrand, 2004).

Given the intermittent land connections between Iberia and North Africa during the Miocene and varying extent of sea channels due to marine sea level oscillations during the Pleistocene, biogeographic inference across the Strait of Gibraltar has been documented as if it were a puzzle through time. The legacy of ancient biogeographic events is nowadays visible in distribution patterns at high taxonomic levels, generally at the genus level (Carranza et al., 2008; García-Vázquez et al., 2016; Sampaio, Harris, Perera, & Salvi, 2014), whereas more recent processes underlie the phylogeographic patterns of intraspecific groups (Carranza, Harris, et al., 2006; Jaramillo-Correa et al., 2010). While studies at the genus level are scarce, in most studies the genetic lineage distribution on both sides of the Strait is intraspecific, thus providing information for biogeographical events restricted to a relatively recent time-frame. A comprehensive vision of the biogeographical events across the Strait of Gibraltar might be enhanced by using as model a single genus, comprising taxa demonstrating both intra and interspecific genetic distributions on both sides of the Strait, such as the sand racer lizard genus *Psammodromus*.

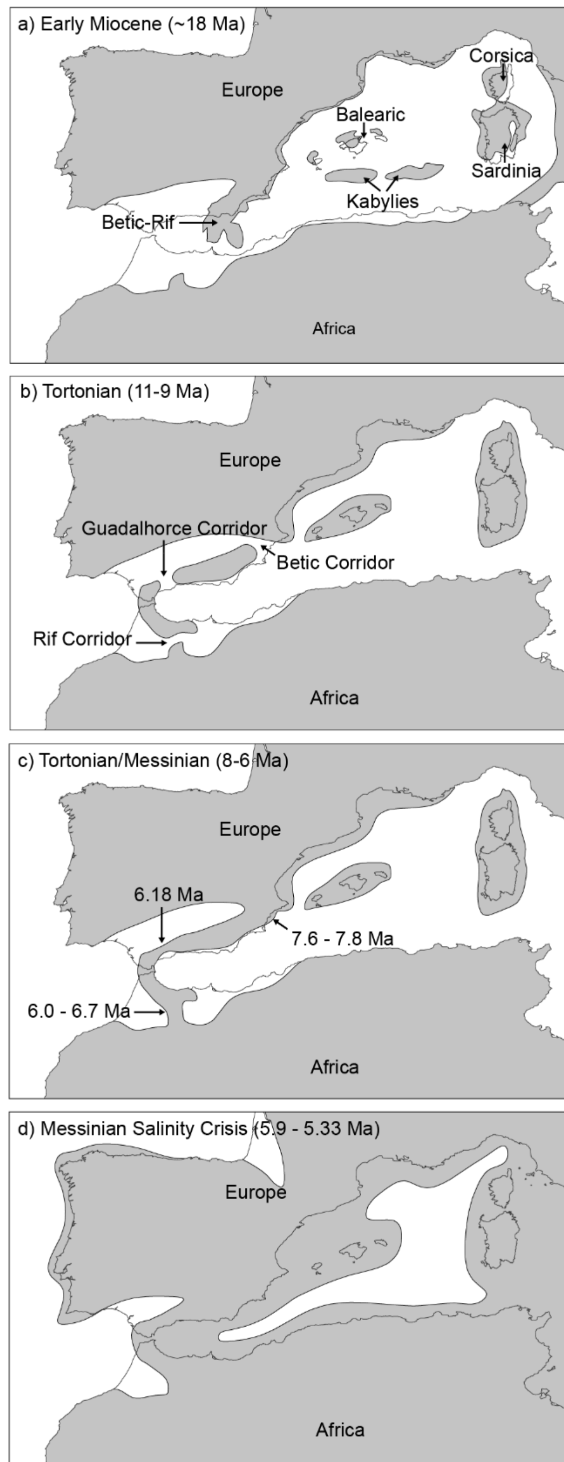


Figure 4.1. Simplified palaeogeological history of the Strait of Gibraltar area from the Middle Miocene (18 Ma) until the Messinian Salinity Crisis (adapted from Rosenbaum et al. 2002a and Paulo et al., 2008).

Psammodromus, together with the Canary Islands endemic *Gallotia*, form the Gallotiinae subfamily, an ancient lineage of the Lacertidae lizard family (Arnold, Arribas, & Carranza, 2007). Six species are distributed across the Strait of Gibraltar: two North African endemic species, *P. blanci* (Lataste, 1880) and *P. microdactylus* (Boettger,

1881); three species in the Iberian Peninsula and Languedoc region, *P. edwarsianus* Dugès, 1829, *P. occidentalis* Fitze, Gonzalez-Jimena, San-Jose, San Mauro & Zardoya, 2012 and *P. hispanicus* Fitzinger, 1826; and *P. algirus* (Linnaeus, 1758) distributed across the Iberian Peninsula and North Africa. Previous studies have proposed biogeographical hypotheses for the evolution of *Psammodromus* species. In the Iberian Peninsula, divergence due to geographic barriers and environmental differences following the closure of the Betic corridor may have led to the evolution of the three extant species (Fitze et al., 2011). Within *P. algirus*, mtDNA data indicate the existence of two main lineages: an eastern Iberian lineage and a western lineage including specimens from western Iberia and North Africa (Busack & Lawson, 2006; Carranza, Harris, et al., 2006; Verdú-Ricoy, Carranza, Salvador, Busack, & Díaz, 2010). These results could indicate that *P. algirus* first diverged in Iberia and later colonized North Africa. Molecular dating places the colonization of Africa by *P. algirus* approximately 1.9 Ma, long after the end of the MSC (Carranza, Harris, et al., 2006). The same study also puts the divergence between the African *P. blanci* and the Iberian clade long before the MSC, at approximately 20 Ma, thereby suggesting two transmarine colonizations to explain the distribution of *Psammodromus* lineages.

In this study, we generate a comprehensive molecular dataset for all the species of *Psammodromus*, with novel taxa and nuclear markers relative to previous studies. We apply both concatenation and coalescent species tree approaches, and estimate divergence times of the main lineages and their ancestral geographic distribution. Our main aim is to frame the temporal and spatial patterns of the evolutionary history of *Psammodromus* within the biogeographic dynamics of the Strait of Gibraltar in order to contribute to the understanding of the faunal exchange across the Pillars of Hercules.

MATERIAL AND METHODS

SAMPLING

Twenty-seven *Psammodromus* specimens were employed in the phylogenetic analyses, including all six species currently described (Fig. 4.2). Lizards were captured and handled under permit of relevant authorities. Six samples of *Gallotia caesaris caesaris*, *G. c. gomerae* and *G. atlantica* were used as outgroups. Information regarding the sample codes and sampling localities is given in Table 4.1.

DNA EXTRACTION, AMPLIFICATION AND SEQUENCING

Genomic DNA was extracted from alcohol-preserved tail muscle following standard high-salt protocols (Sambrook, Fritsch, & Maniatis, 1989). We amplified fragments of three mitochondrial (mtDNA) genes, NADH Dehydrogenase 4 plus flanking tRNAs (*nd4*), ribosomal 12S rRNA gene (*12S*) and cytochrome *b* (*cytb*), and of three nuclear (nucDNA) genes, Acetylcholinergic Receptor M4 (*acm4*), Melanocortin 1 Receptor (*mc1r*) and Proopiomelanocortin (*pomc*). These markers have been shown to be highly variable in lacertid lizards (e.g. Mendes, Harris, Carranza, & Salvi, 2016; Salvi, Schembri, Sciberras, & Harris, 2014). Primers and conditions of polymerase chain reactions (PCR) are reported in Table 4.2.

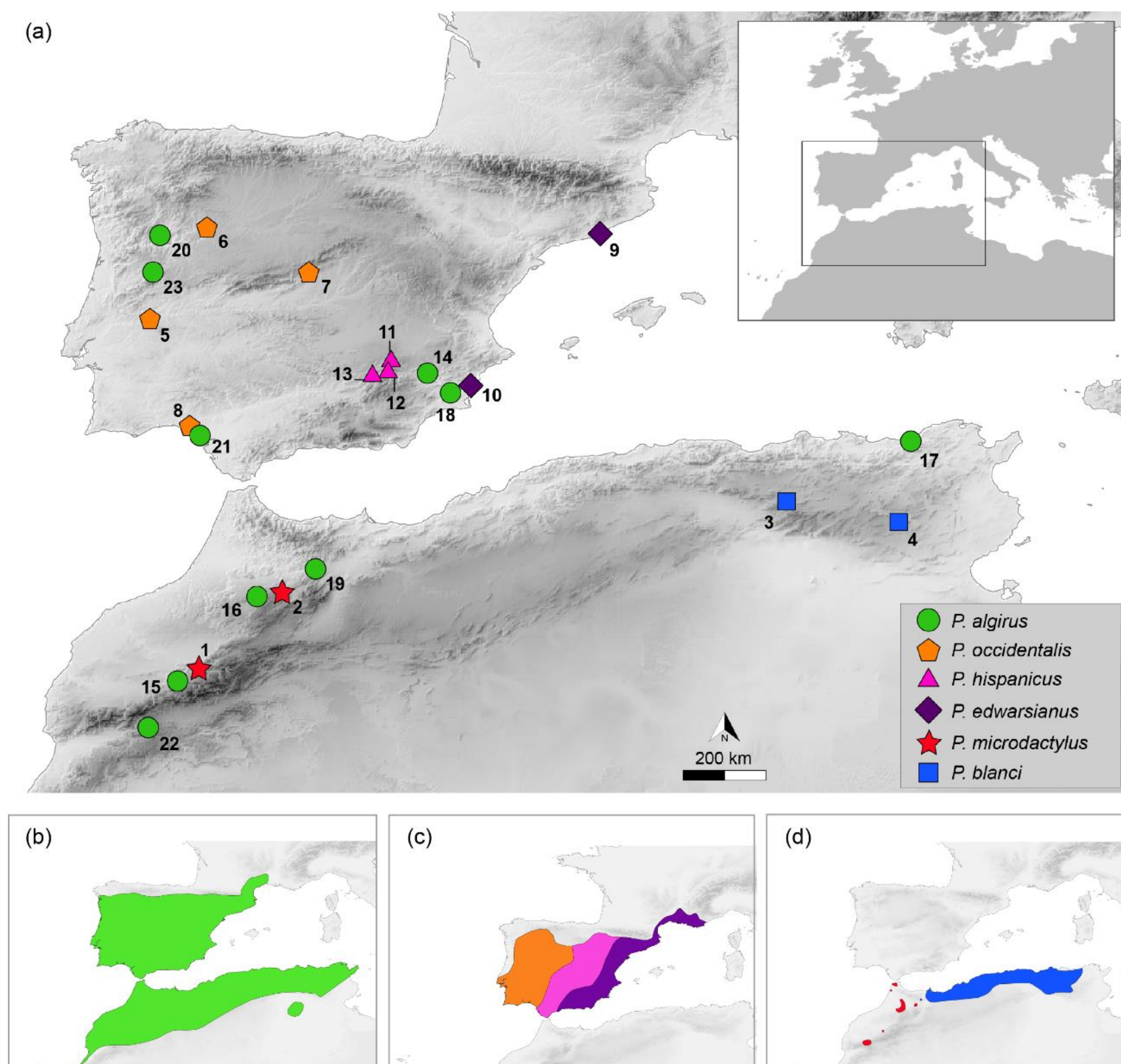


Figure 4.2. Map showing the location in the Iberian Peninsula and North Africa of the *Psammodromus* specimens used in this study (a). Detailed information on sampling localities and specimens is reported in Table 4.1. Approximate distribution of *Psammodromus* species displayed in smaller maps: b) *P. algirus*, c) *P. occidentalis*, *P. hispanicus* and *P. edwarsianus*, d) *P. microdactylus* and *P. blanci*.

PHYLOGENETIC INFERENCE

Nucleotide sequences were aligned using the MUSCLE algorithm (Edgar, 2004) in GENEIOUS (Biomatters Ltd.). Haplotype reconstruction for nuclear markers was performed in PHASE 2.1 (Stephens, Smith, & Donnelly, 2001). PHASE was run three times to assure consistency of results, with a phase probability threshold of 0.7. The occurrence of recombination events in nucDNA was evaluated using the Pairwise Homoplasy Index (phi) test implemented in SPLITSTREE4 4.13.1 (Huson & Bryant, 2006).

Phylogenetic relationships among the *Psammodromus* species were inferred by Maximum Likelihood (ML), Bayesian Inference (BI) and the Bayesian species tree approach based on the multi-species coalescent model (Heled & Drummond, 2010).

For the ML and BI analyses, the mitochondrial and (unphased) nuclear DNA sequence data (mt-nucDNA) were concatenated. The ML analyses were performed in RAXML 7.4.2 (Stamatakis, 2006). ML searches included 10 random addition replicates and 1000 nonparametric bootstrap replicates, applying the general time-reversible models with a gamma model of rate heterogeneity (GTRGAMMA) for each of the six gene partitions. Bayesian analyses were performed in BEAST 1.8.0 (Drummond, Suchard, Xie, & Rambaut, 2012) The best model of nucleotide substitution for each gene was assessed in JMODELTEST 2.1.7 (Posada, 2008) under the Bayesian information criterion (Table 4.3). Models and prior specifications applied in BEAUTi were as follows (otherwise by default): tree model of all gene partitions linked, nucleotide substitution and clock models unlinked; model of nucleotide substitution for *12S* set as HKY, because preliminary runs with the GTR model presented low effective sample size (ESS) values. A relaxed uncorrelated lognormal clock was set for all genes, Yule process of speciation as tree prior, random starting tree, alpha Uniform (0, 10), ucld.mean uniform and operator kappa (2.0). BEAST was run three times, with 50 million generations, sampling every 5000 generations. The use of the Yule process of speciation prior requires only one sequence per species, whereas our concatenated dataset includes multiple samples per species. Therefore, to inspect the sensitivity of our estimates to the choice of tree prior, we confirmed that identical results were obtained in an additional run applying the same settings as above but using only one representative sequence for each species (Fig. S4.1).

Table 4.1. Species, code, map location code represented in Fig. 4.2, locality, coordinates and GenBank accession number of the specimens used in this study.

Species	Code	Map Code	Locality	Geographic Coordinates	GenBank accession number					
					12S	***cytb***	***nd4***	***acm4***	***mc1r***	***pomc***
P. microdactylus	25232	1	Azilal, Morocco	31.92 N, -6.66 E	MF684874	MF684919	MF684964	MF684898	MF684938	MF684987
P. microdactylus	25250	1	Azilal, Morocco	31.92 N, -6.66 E	MF684875	MF684920	MF684965	MF684899	MF684939	
P. microdactylus	25384	2	Ait Khalifa, Sefrou, Morocco	33.69 N, -4.84 E	MF684876	MF684921	MF684966		MF684940	
P. blanci	11219	3	Belezma National Park, Algeria	35.58 N, 6.08 E	MF684877	MF684922	MF684967		MF684941	
P. blanci	11220	3	Belezma National Park, Algeria	35.58 N, 6.08 E	MF684878	MF684923	MF684968		MF684942	
P. blanci	T31-47a	4	Bou Chebka, Tunisia	35.12 N, 8.49 E	MF684879	MF684924	MF684969	MF684900	MF684943	MF684988
P. blanci	T31-47b	4	Bou Chebka, Tunisia	35.12 N, 8.49 E	MF684880	MF684925	MF684970			
P. occidentalis	14	5	Castelo de Vide, Portugal	39.43 N, -7.58 E				MF684901	MF684944	MF684989
P. occidentalis	4213	6	Miranda do Douro, Portugal	41.40 N, -6.37 E				MF684902	MF684945	
P. occidentalis	25453	7	Colmenar del Arroyo, Spain	40.43 N, -4.18 E	MF684881		MF684971	MF684903	MF684946	MF684990
P. occidentalis	25454	8	Doñana National Park, Spain	37.01 N, -6.57 E	MF684882	MF684926	MF684972	MF684904	MF684947	MF684991
P. edwarsianus	1869	9	El Prat de Llobregat, Spain	41.29 N, 2.10 E				MF684905	MF684948	MF684992
P. edwarsianus	14359	10	La Mata y Torrevieja Natural Park, Spain	38.02 N, -0.70 E	MF684883		MF684973	MF684906	MF684949	MF684993
P. hispanicus	1723	11	La Hueta, Spain	38.36 N, -2.51 E	KX080629		KX081066	KX080992	KX080782	
P. hispanicus	1784	11	La Hueta, Spain	38.34 N, -2.61 E	MF684884		MF684974	MF684907	MF684950	MF684994
P. hispanicus	1850	12	Segura de la Sierra, Spain	38.29 N, -2.59 E	KX080630		KX081067	KX080993	KX080783	
P. hispanicus	1856	13	Cortijos de Tortas, Spain	38.54 N, -2.42 E	MF684885		MF684975		MF684951	MF684995
P. algirus	1291	14	Embalse de Camarillas, Spain	38.34 N, -1.65 E	MF684886		MF684976		MF684952	MF684996
P. algirus	2347	15	Imi n'Ifri, Morocco	31.72 N, -6.97 E	KX080631		KX081068		KX080784	
P. algirus	2356	16	Azrou, Morocco	33.54 N, -5.32 E	KX080632		KX081069	KX080994	KX080785	MF684997
P. algirus	3784	17	Tabarka, Tunisia	36.95 N, 8.76 E	MF684887	MF684927	MF684977	MF684908	MF684953	MF684998

P. algirus	3854	18	Múrcia, Spain	37.93 N, -1.10 E	MF684888	MF684928	MF684978	MF684909	MF684954	MF684999
P. algirus	4235	19	Taza, Morocco	34.13 N, -4.03 E	MF684889	MF684929	MF684979	MF684910	MF684955	MF685000
P. algirus	16794	20	São Lourenço, Portugal	41.29 N, -7.37 E	MF684890	MF684930	MF684980	MF684911	MF684956	MF685001
P. algirus	21567	21	Matalascañas, Spain	37.00 N, -6.56 E	MF684891	MF684931	MF684981	MF684912	MF684957	MF685002
P. algirus	23902	22	Jbel Siroua, Morocco	30.74 N, -7.61 E	MF684892	MF684932	MF684982	MF684913	MF684958	
P. algirus	24987	23	Gouveia, Portugal	40.51 N, -7.53 E	MF684893	MF684933		MF684914	MF684959	MF685003
Gallotia atlantica	1244	-	Lanzarote, Canary Islands, Spain	29.05 N, -13.56 E	KX080625		KX081062	KX080988	KX080778	MF685004
Gallotia atlantica	1341	-	Lanzarote, Canary Islands, Spain	28.95 N, -13.77 E	KX080626		KX081063	KX080989	KX080779	MF685005
Gallotia c. gomerae	19355	-	La Gomera, Canary Islands, Spain	28.12 N, -17.16 E	MF684894	MF684934	MF684983	MF684915	MF684960	MF685006
Gallotia c. gomerae	19375	-	La Gomera, Canary Islands, Spain	28.10 N, -17.13 E	MF684895	MF684935	MF684984	MF684916	MF684961	MF685007
Gallotia c. caesaris	19413	-	El Hierro, Canary Islands, Spain	27.81 N, -17.91 E	MF684896	MF684936	MF684985	MF684917	MF684962	MF685008
Gallotia c. caesaris	19470	-	El Hierro, Canary Islands, Spain	27.73 N, -18.12 E	MF684897	MF684937	MF684986	MF684918	MF684963	MF685009

Table 4.2. Name, sequence and reference of the primers, and the amplification conditions for the genes used in this study.

Gene	Primer Name	Primer Sequence (5'-3')	Source	PCR conditions (ºC(seconds) x number of cycles)
12S	12Sa	CTG GGA TTA GAT ACC CCA CTA T	Kocher et al., 1989	94(180), [94(30), 50(30), 72(45) x 35], 72(600)
	12Sb	GAG GGT GAC GGG GCG GTG TGT		
cytb	GluDG	TGA CTT GAA RAA CCA YCG TTG	Palumbi et al., 1991	94(300), [94(30), 49(40), 72(50) x 35], 72(600)
	Cytb2	CCC TCA GAA TGA TAT TTG TCC TCA		
nd4	ND4	CAC CTA TGA CTA CCA AAA GCT CAT GTA GAA GC	Arevalo et al., 1994	94(180), [94(30), 50(30), 72(60) x 35], 72(600)
	LEU	CAT TAC TTT TAC TTG GAA TTT GCA CCA		
acm4	TgF	CAA GCCTGA GAG CAA RAA GG	Gamble et al., 2008	92(180), [92(30), 62↓0.5(30), 72(45) x 20], [92(30), 50(30), 72(45) x 15], 72(600)
	TgR	ACY TGA CTC CTG GCA ATG CT		
mc1r	MC1R-F	GGC NGC CAT YGT CAA GAA CCG GAA CC	Pinho et al., 2009	92(180), [92(30), 62↓0.5(30), 72(60) x 25], [92(30), 50(30), 72(60) x 15], 72(600)
	MC1R-R	CTC CGR AAG GCR TAG ATG ATG GGG TCC AC		
pomc	POMC_DRV_F1	ATA TGT CAT GAS CCA YTT YCG CTG GAA	Vieites et al., 2007	94(180), [94(30), 50(35), 72(60) x 35], 72(600)
	POMC_DRV_R1	GGC RTT YTT GAA WAG AGT CAT TAG WGG		

The species tree was inferred using the *BEAST extension of the BEAST software. We used the mtDNA sequences and the phased alignments of the nuclear genes and their relative models of nucleotide evolution (Table 4.3). The settings applied in *BEAST for the species tree reconstruction were similar to the concatenated BEAST analyses, except the tree model of the mitochondrial genes *12S*, *cytb* and *nd4* was linked. *BEAST was run three times with 400 million generations, sampling every 40,000 generations.

We used TRACER 1.6 to check all BEAST runs for convergence (burn-in = 10%) and that all ESS parameters were higher than 200. All runs were combined with LOGCOMBINER and TREEANNOTATOR was used to calculate the Maximum Clade Credibility (MCC) tree summarizing the posterior distribution of tree topologies and branch lengths. Trees were visualised in FIGTREE 1.4 (available at http://tree.bio.ed.ac.uk/software/figtree/). BEAST analyses were performed on the CIPRES Science Gateway 3.3 (Miller et al., 2010).

MOLECULAR DATING

In order to infer the time of cladogenetic events within Gallotiinae we used the estimated rates of evolution of *12S* and *cytb* of lacertid lizards (Carranza & Arnold, 2012). We implemented a normal distribution prior for the ucld.mean parameters of the *12S* (mean: 0.00553, stdev: 0.00128) and *cytb* (mean: 0.0164, stdev: 0.00317) gene fragments in *BEAST. To infer divergence time within species we also applied the same priors in BEAST on the concatenated mt-nucDNA dataset (all gene partitions linked). We cross-validated the substitution rates by performing (i) an estimate of *12S* and *cytb* rates based on the procedure and the dataset of Carranza & Arnold (2012) but excluding the MSC vicariance between *Podarcis* species as calibration point (since the use of the MSC in molecular clock calibrations is under debate, see Hewitt (2011) and also the Discussion section) and (ii) a time tree estimate applying on our dataset the priors on *Gallotia* nodes used by Carranza & Arnold (2012): a uniform prior from 14.5 to 0 Ma for the root of *Gallotia* and a prior constraining the split between *Gallotia c. caesaris* and *Gallotia c. gomerae* soon after the emergence of the El Hierro Island 1.12 Ma (Normal prior: mean=1, SD=0.1, upper limit truncated to 1.12). Rates and time estimates resulting from all molecular dating analyses were almost identical (Fig. 3 & S4.2). Additionally, the *12S* rate used in this study is in agreement with the *12S* rate estimated in other lacertid species (Salvi, Pinho, & Harris, 2017).

ANCESTRAL AREA RECONSTRUCTION

Biogeographic reconstructions were performed in RASP 3.0 (Yu, Harris, Blair, & He, 2015) using the statistical dispersal-vicariance analysis (S-DIVA) (Yu, Harris, & He, 2010), dispersal-extinction cladogenesis (DEC) (Ree & Smith, 2008) and Bayesian binary MCMC (BBM) (Yu, Harris, & Xingjin, 2011) methods. We used all the post-burnin trees and the MCC tree resultant from the concatenated BEAST run as input. To evaluate the effect of changing the number of areas in the reconstruction of the ancestral geographic distribution, we performed two distinct sets of analyses: with two areas: North Africa (A) and Iberian Peninsula (B) (Fig. 4.3), and with three areas: North Africa (A), Western Iberia (B) and Eastern Iberia (C) (Fig. S4.3).

S-DIVA was conducted using 1000 trees randomly sampled from the input trees. DEC analyses were run either with the default connectivity value (equal to 1.0) between ancestral ranges along the time trees or implementing the known palaeogeographic model of the Strait of Gibraltar. We stratified the phylogeny into five different time slices each one associated to a specific connectivity value (i.e. dispersal cost) reflecting the changing continental configuration over time in the Gibraltar region. In particular, we implemented the following time slices and associated connectivity values: 0-5.33 Ma/connectivity set as 0.5 (opening of the Strait of Gibraltar), 5.33-7.0 Ma/ connectivity 1 (land connection between Africa and Iberia during the MSC and closing of the northern sea corridors), 7.0-8.5 Ma/ connectivity 0.75 (sea corridors opened between Iberia and Africa with the Betic-Rif blocks as stepping stones), 8.5-10 Ma/ connectivity 1 (closure of the Betic corridor, land bridge between Iberia and Betic-Rif) and 10-13 Ma/ connectivity 0.75 (Betic-Rif was between Iberia and Africa, as possible stepping stones). Results were identical between the simple and the stratified models (likely because of the low number of cladogenetic events in each time slice, see Ree & Sanmartín (2009)), so only results from the default connectivity model are provided here. BBM analyses were run with the Jukes-Cantor model, site variation set to equal and maximum number of areas set to two and three, respectively. We performed two simultaneous runs with 5 million generations, sampling each 100 generation, without outgroup species.

RESULTS

MOLECULAR DATA, *PSAMMODROMUS* PHYLOGENY AND DIVERGENCE TIME

A total of 159 new sequences were obtained (GenBank accession numbers in Table 4.1). In order to concatenate only sequences from the same individual we avoided including published sequences from GenBank. The length of the gene fragments and the number of polymorphic sites are reported in Table 4.3. The translation of these genes into amino acid sequences did not contain stop codons. The phi test did not find statistically significant evidence for recombination in the nuclear gene fragments (p>0.05).

Table 4.3. Length of gene fragments in base pairs, models of nucleotide substitution for unphased and phased molecular data and number of variable positions within *Psammodromus* genus and in *Psammodromus* plus outgroup *Gallotia* data set.

Gene	Length (bp)	Model unphased data	Model phased data	Variable Positions	Variable Positions with Outgroup
12S	368	GTR+I+G		79	89
cytb	392	HKY+G		120	138
nd4	825	TrN+I+G		294	335
acm4	357	HKY	HKY+I	24	29
mc1r	610	HKY+I	HKY+I+G	36	52
pomc	428	TrN+I	TrN+I	50	72

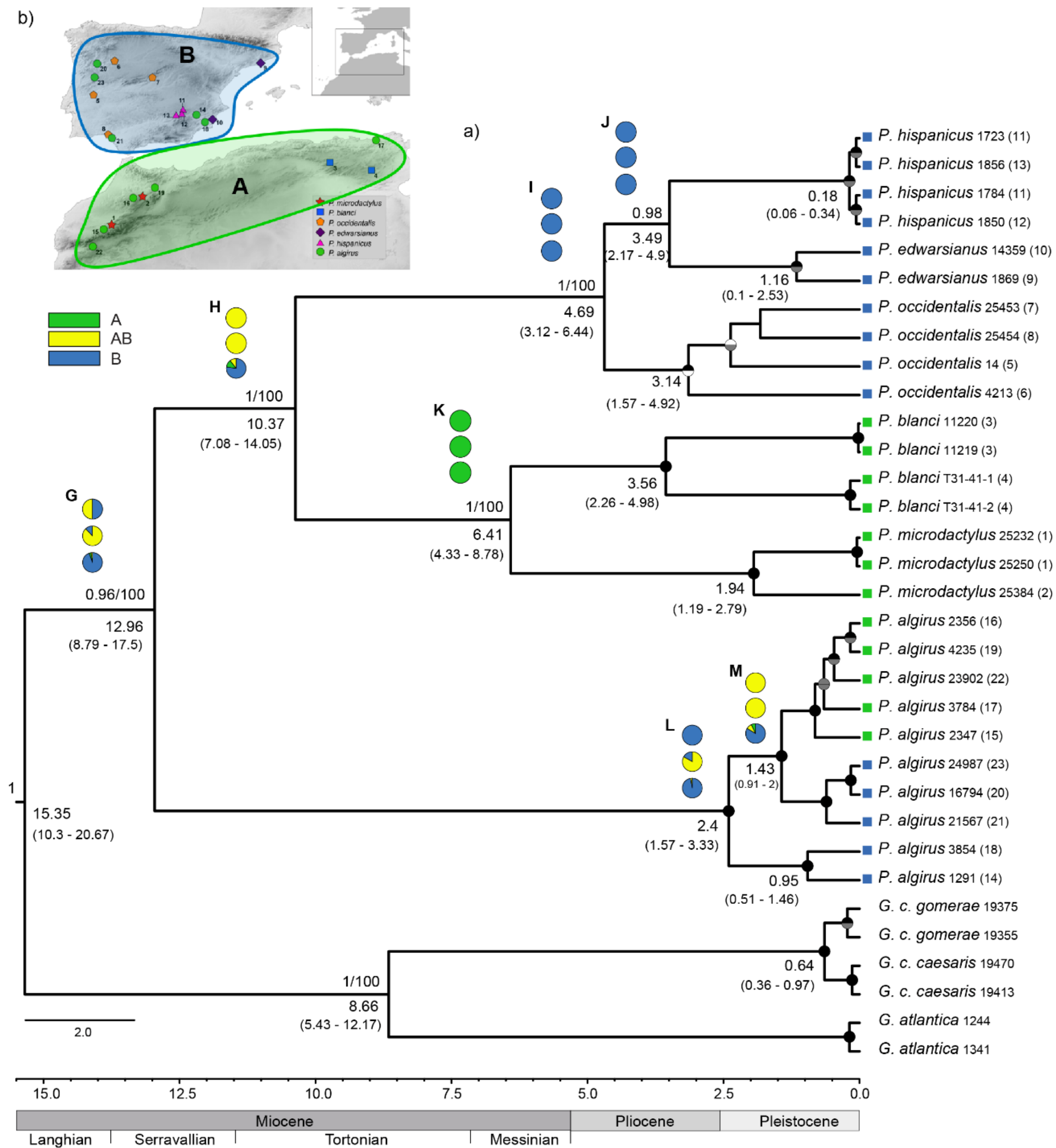


Figure 4.3. a) Bayesian tree depicting the phylogenetic relationships between *Psammodromus* lizards based on concatenated mitochondrial (*12S*, *cytb*, *nd4*) and nuclear (*acm4*, *mc1r*, *pomc*) DNA sequence. Numbers beside species names indicate specimen codes (Table 4.1) and numbers between brackets indicate locality codes as shown in Fig. 3.2 and Fig. 4.3b Values above nodes represent Bayesian posterior probabilities (BPP) ≥0.9 (left) and Maximum Likelihood bootstrap (BS) values ≥70 (right). Node support for intra-specific clades is represented by black circles: BPP≥0.98 (upper half) and BS≥95 (bottom half); grey circles: 0.9≤BPP<0.98 (up) and 70≤BS < 95 (bottom) and empty circles for nodes with lower support. Values below nodes represent the estimated age of the node with associated 95% Highest Posterior Density interval (in parentheses) according to molecular dating analyses. Coloured squares represent the geographic origin of each tip sample in either Iberia or Africa according to the inset (b). Coloured pie charts in correspondence of nodes (which are named from G to M).

Phylogenetic relationships between *Psammodromus* species recovered by species tree and concatenated dataset were identical (Figs. 4.3a & 4.4). Maximum Likelihood and BI analyses based on the concatenated dataset resulted in identical relationships between species with overall high bootstrap (BS) and Bayesian posterior probabilities (BPP) values (Fig. 4.3a). The divergence time estimates were similar between concatenated (Fig. 4.3a) and species trees (Fig. 4.4), therefore from here on we will refer only to time estimates of the concatenated tree (Fig. 4.3a). The genera *Gallotia* and *Psammodromus* are reciprocally monophyletic with divergence between them estimated at 15 Ma (95% Highest Posterior Densities interval (95%HPD): 10.3-20.67). All *Psammodromus* species are monophyletic (Figs. 4.3a & 4.4; BPP≥0.9, BS≥70). *Psammodromus algirus* was the first species to diverge within the *Psammodromus* clade (Figs. 4.3a & 4.4; BPP=0.96, BS=100), approximately 13 Ma (95%HPD: 8.79-17.5), and includes three lineages: the Eastern Iberia lineage, which diverged circa 2.5 Ma (95%HPD: 1.57-3.33) from the Western Iberia lineage and the North African lineage. The latter two diverged about 1.4 Ma (95%HPD: 0.91-2.0) (Figs. 4.2 & 4.3). The North African endemic *P. blanci* and *P. microdactylus* are sister species (Figs. 4.3a & 4.4; BPP=1, BS=100) and diverged from the Iberian clade (*P. occidentalis*, *P. hispanicus* and *P. edwarsianus*) approximately 10 Ma (95%HPD: 7.08-14.05) and between each other 6 Ma (95%HPD: 4.33-8.78) (Figs. 4.3a & 4.4). *Psammodromus blanci* shows a deep differentiation between two lineages in the Aurès Mountains, estimated at 3.5 Ma (95%HPD: 2.26-4.98), (Figs. 4.2 & 4.3a). *Psammodromus microdactylus* presents a comparatively lower, but still notable, differentiation between the two populations sampled (Figs. 4.2 & 4.3a). Diversification within the Iberian clade started approximately 4.7 Ma (95%HPD: 3.12-6.44) with the split of *P. occidentalis* followed by the split between the sister species *P. hispanicus* and *P. edwarsianus* at 3.5 Ma (95%HPD: 2.17-4.9) (Figs. 4.3a & 4.4).

ANCESTRAL AREA RECONSTRUCTION

All biogeographic analyses, based on either two areas (Africa and Iberia) or thee areas (Africa, Western Iberia and Eastern Iberia), inferred the same biogeographical scenario for the nodes K, I and J: the ancestor of the species *P. blanci* and *P. microdactylus* has an African origin (node K), and the ancestors of *P. occidentalis*, *P. hispanicus* and *P. edwarsianus* (nodes I and J) have an Iberian origin (Fig. 4.3 & Fig. S4.3). The ancestral areas inferred for the basal nodes within the genus (G and H) are either in Iberia (BBM: 95% in G, 77% in H; S-DIVA: 50% in G) or shared by Iberia and Africa (S-DIVA: 50% in G, 100% in H; DEC 87%: in G, 100% in H) in the two areas analyses (Fig. 4.3). A similar

inference is presented in node M for the most recent lineages within *P. algirus*, with BBM (80%) supporting with higher probability an Iberian ancestry, while S-DIVA and DEC support an Iberian and African origin. Finally, the origin of *P. algirus*, node L, is likely to be Iberian, as indicated by S-DIVA and BBM (Fig. 4.3).

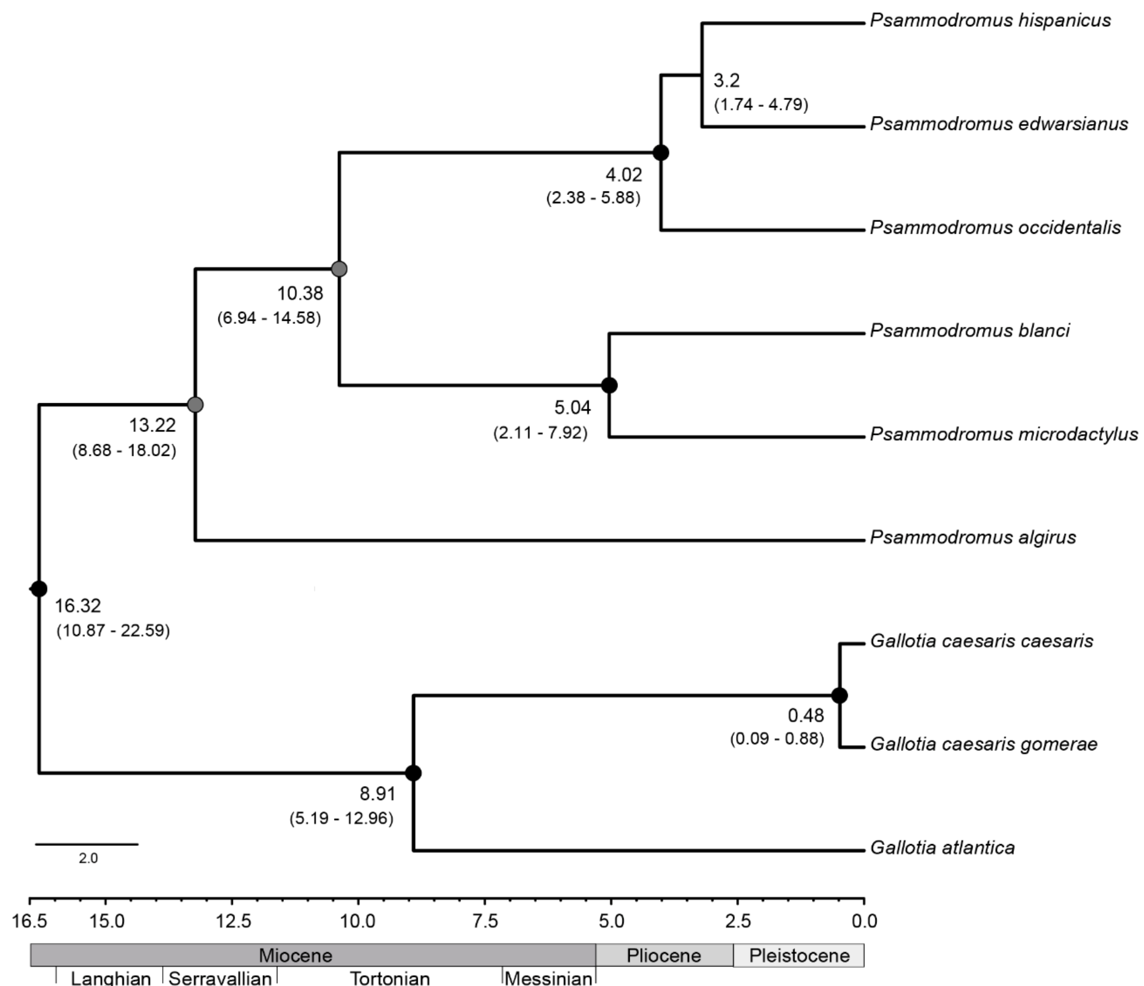


Figure 4.4. Species tree of *Psammodromus* lizards inferred from mitochondrial (*12S*, *cytb*, *nd4*) and nuclear (*acm4*, *mc1r*, *pomc*) DNA sequences using the multispecies coalescent model in *BEAST. Black circles represent Bayesian posterior probabilities (BPP) =1 and grey circles represent 0.98≤BPP<1. Values below nodes represent the age of the node and the associated 95% Highest Posterior Density interval (in parentheses).

Regarding the nodes G, H, L, and M the results from RASP analyses based on three areas are congruent with two areas analyses but show increased inference uncertainty. For example, S-DIVA and DEC infer a similar pattern in the two and three areas analyses for the nodes G, H, L and M, with the exception of S-DIVA in node L suggesting an Iberian origin in two areas and uncertain origin in three areas (Fig. 4.3 &

Fig. S4.3). Increased inference uncertainty in the three areas compared to the two areas analyses is also apparent in the BBM results with higher probabilities for an Iberian origin of nodes G, H and M in the analyses based on two areas and comparable probabilities associated to either an Iberian or an African origin in the three areas analysis.

DISCUSSION

The use of a multilocus dataset including all species allowed estimation of a fully resolved phylogeny of the genus *Psammodromus*. Relationships between species received high statistical support (Fig. 4.3a & 4.4) and are congruent with relationships previously inferred using mostly mtDNA data from incomplete taxon sets (Busack & Lawson, 2006; Carranza, Harris, et al., 2006; Fitze et al., 2011; Fitze, Gonzalez-Jimena, San-Jose, Mauro, & Zardoya, 2012; Verdú-Ricoy et al., 2010). *Psammodromus algirus* represents an early branch within the genus and is sister to two clades endemic to Africa and Iberia, respectively. In the Iberian clade, the species *P. occidentalis* is sister to the sister species *P. hispanicus* and *P. edwarsianus*. Unlike previous studies, our phylogeny includes *P. microdactylus* which is sister to the other African species *P. blanci*. Both species of this African clade show two deeply divergent lineages separated by a relatively small geographic distance (Fig. 4.2 & 4.3).

Here we integrate the information of the main cladogenetic events within *Psammodromus*, and their estimated time and ancestral geographic distribution, within the context of the palaeogeographic evolution of the Strait of Gibraltar to trace the biogeographic history of *Psammodromus* across North Africa and Western Europe. The biogeographic inference from the Bayesian binary MCMC (BBM) is in agreement with the fossil and palaeogeographical evidence and with the timing and pattern of cladogenetic events inferred for *Psammodromus* from molecular data (Fig. 4.1 & 4.3). Biogeographical inferences obtained with the S-DIVA and DEC methods show higher uncertainty and wider ancestral ranges at deep nodes. Such a pattern has been observed before (Liu et al., 2016; Perea, Cobo-Simon, & Doadrio, 2016) and is likely due to the underlying assumptions of these methods (Buerki et al., 2011). Both DIVA and DEC models are based on the assumption of 'vicariance-mediated allopatry', in which dispersal is modelled as wide ancestral range divided by vicariance (or also peripheral isolate speciation in DEC) at speciation nodes (Sanmartín, 2007), so speciation following early dispersal events in the phylogeny imply wide ancestral ranges at deep nodes (see e.g. node H in Fig. 4.3). Additionally, in DEC analyses widespread terminals subtended

by a long branch, such as *P. algirus* in our case, may have the effect of increasing the uncertainty at deeper nodes (Buerki et al., 2011; Ree, Moore, Webb, & Donoghue, 2005).

HISTORICAL BIOGEOGRAPHY OF *PSAMMODROMUS*

The origin of the genus *Psammodromus* was likely in Western Europe according to previous knowledge. The European origin of lacertids, and of the Gallotiinae in particular, is supported by (i) fossil evidence (Augé, 2005; Čerňanský, Klembara, & Smith, 2016; Fig. S4.4, Table S4.1), (ii) the mainly European distribution of the first genera to branch off in the lacertid phylogeny (Arnold et al., 2007) and (iii) a high species diversity in Europe compared to Africa (Arnold et al., 2007; Pavlicev & Mayer, 2009).

The origin of the *Psammodromus* radiation is estimated to have started in the Middle Miocene (Fitze et al., 2011, present study) or even in the Early Miocene (Carranza, Harris, et al., 2006). At that time, the westernmost Mediterranean was characterised by intense geological activity, with the southward drift of the Betic-Rif block across the Gibraltar region. The exact position of the Betic-Rif after the breaking of the Hercynian belt until it reached its final position in the Strait of Gibraltar remains debatable. While some authors advocate that the Betic-Rif was attached to Iberia until a marine transgression separated them, 16-14 Ma (Lonergan & White, 1997), others suggest that, separated from Iberia, it formed a continuous block with the Balearic Islands and Kabylies until 15 Ma (Rosenbaum et al., 2002a; Rosenbaum, Lister, & Duboz, 2002b) or with the Balearic Islands alone, until 10 Ma (Fontboté et al., 1990). Given the absence of remains referable to *Psammodromus* within the well-studied fossil record of the Balearic Islands (Bailon, Boistel, Bover, & Alcover, 2014; Pinya & Carretero, 2011 and references therein) it is likely that the genus would have arrived to Africa by dispersal from the paleo-Iberia through the Betic-Rif, after the detachment of the latter from the other blocks. The alternative hypotheses that the ancestor of *P. blanci* and *P. microdactylus* would have reached Africa by vicariance following the detachment of either the Hercynian belt (Betic-Rif+Balearics+ Kabylies+Corsica-Sardinia+Calabria) from Iberia (30 Ma) or the Betic-Rif block from the other Hercynian blocks (16-14 Ma) is less plausible not only for the lack of *Psammodromus* fossils in these blocks but also considering the estimated time of divergence between the Iberian and the African clades, although the latter cannot be definitively rejected. According to our estimates, the cladogenetic event that gave rise to the ancestor of *P. blanci* and *P. microdactylus* occurred during the Middle to Late Miocene, or even earlier in the Early to Middle Miocene (Carranza, Harris, et al., 2006). Assuming no undocumented extinctions, this

time estimate provides an approximate upper limit for the colonization of Africa. Whether the cladogenesis between the ancestors of the Iberian clade (*P. occidentalis*, *P. edwarsianus* and *P. hispanicus*) and of the African clade (*P. blanci* and *P. microdactylus*) was due to a dispersal event between these areas, or if the divergence of the ancestor of the African clade slightly predated its dispersal, the colonization of Africa took place during a period when the Betic-Rif massif was fragmented and separated from mainland by sea corridors to the north and south (Fig. 4.1b). The northern Betic sea corridor closed approximately 10 Ma, originating a temporary land bridge between Iberian and the sub-Betic massif that disappeared when this sea corridor re-opened 9-8 Ma. Therefore the arrival into North Africa of the ancestor of *P. blanci* and *P. microdactylus* was likely achieved through stepping stones across the Betic-Rif (Fig. 4.5a).

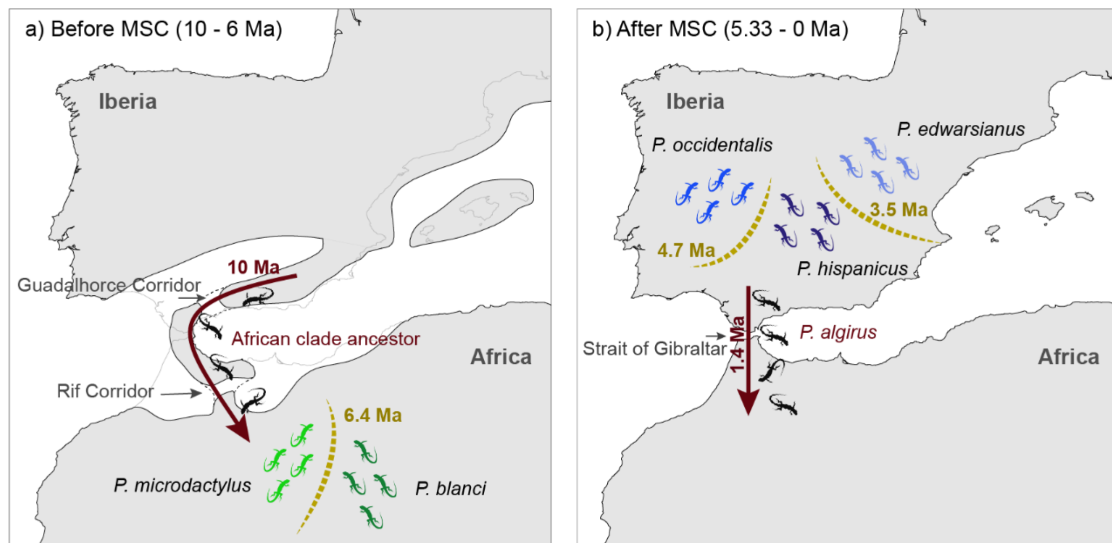


Figure 4.5. Dispersal and vicariance events in *Psammodromus* in Iberia and North Africa before (a) and after (b) the end of the Messinian Salinity Crisis, 5.33 Ma. Dark red represents the direction, age and taxa in the dispersal events. Dashed yellow represent age and vicariance events in Africa and Iberia.

The ancestor of *P. blanci* and *P. microdactylus* underwent considerable diversification after it reached North Africa. The divergence between the two species is dated at the end of the Miocene (Fig. 4.5a). Climatic conditions, shifting towards aridity in North Africa during the Miocene and the Pliocene, may have played a role in their divergence, as has been described for other reptile species during this period (Amer & Kumazawa, 2005; Gonçalves et al., 2012). We identified two main evolutionary lineages within each species. In *P. blanci*, the divergence between the two lineages is as deep as between the species *P. hispanicus* and *P. edwarsianus* (3.5 Ma). Such a level of divergence between two sampling localities closely located in the Eastern Maghreb (Fig.

4.2) is suggestive of undescribed diversity, especially considering that the range of *P. blanci* extends to Eastern Morocco. The high genetic differentiation within *P. microdactylus* and its fragmented distribution are indicative of a relictual pattern possibly resultant of widespread extinction mediated by climatic or ecological changes. Additional studies are necessary to fully understand the diversity and the evolutionary history of these species, with a possible novel taxonomical assessment.

A major role for vicariance processes behind the diversification of the Iberian clade has been previously suggested (Fitze et al. 2011) and is supported by our results. Under this biogeographical scenario, the final closure of the Betic corridor (7.6-7.8 Ma) caused the formation of the Guadalquivir basin, which acted as a vicariant factor driving the divergence between *P. occidentalis* and the remaining Iberian species following an east-west differentiation (Fig. S4.1). This geographical pattern has been described for multiple Iberian species (e.g. Harris, Carranza, Arnold, Pinho, & Ferrand, 2002; Harris & Sá-Sousa, 2002; Martínez-Solano et al., 2004). Later, during the MSC, geological and climatic changes associated with the uplift of the Spanish Central System and the re-configuration of the main river drainages in Iberia would have ultimately led to the differentiation between *P. edwarsianus* and *P. hispanicus* (Fitze et al., 2011). The main difference between our results and those from Fitze et al. (2011) is that the inferred time is more recent in our study. This may be attributed to the use of different molecular data and rates of evolution. However, the uncertainty of such estimates overlap for the majority of the associated confidence intervals, while supporting the same pattern of evolution of these three species.

Finally, the pattern of diversification in the most widespread species of the genus, *P. algirus*, shows some striking parallels with the wall lizard *Podarcis vaucheri* (Kaliontzopoulou, Pinho, Harris, & Carretero, 2011). The fact that *P. algirus*, like *Podarcis vaucheri*, has an African lineage nested within two paraphyletic Iberian lineages is an indication that both species diverged in Iberia and then reached North Africa. According to our dated phylogeny, the divergence of the African lineage of *P. algirus* took place approximately 1.4 Ma, and therefore this species colonized North Africa through transmarine dispersal across the Strait of Gibraltar (Fig. 4.5b). This colonization was likely associated with a sea level low-stand during a glacial stage, when the distance between Iberia and North Africa was reduced. Such a scenario has been described by Carranza et al. (2006b) and is concordant with our ancestral area reconstructions (Fig. 4.3, nodes L and M). During the last five million years, several species have independently crossed in both directions the Strait of Gibraltar in different occasions after its opening (see e.g. Carranza et al., 2008, 2004; Carranza, Arnold, et al., 2006;

Carranza, Harris, et al., 2006; Gaubert et al., 2011; Harris, Batista, Lymberakis, & Carretero, 2004; Harris et al., 2004).

The historical biogeography of *Psammodromus* indicates that the Strait of Gibraltar has been a permeable biogeographic barrier for these lizards since their early evolution during the Middle Miocene. While the MSC (5.9-5.33 Ma) is often seen as a major period for biotic exchanges between Africa and Europe, due to their wide land connection, this study shows that dispersal over narrow sea channels before and after the MSC (Fig. 4.5) possibly had a major role in shaping the biogeographic patterns of *Psammodromus* across the Pillars of Hercules.

CONCLUDING REMARKS ON BIOTIC EXCHANGE BETWEEN WESTERN EUROPE AND NORTH AFRICA

In this study, our dated phylogeny and different ancestral area reconstructions provided a robust statistical framework for biogeographic inference, while the well-known geological history of this region provided an independent temporal (and spatial) framework for such an inference. This may compensate the intrinsic uncertainty associated with time estimates based on molecular clocks (Pulquério & Nichols, 2007) and the paucity of fossil data associated to the internal nodes of our phylogeny.

The historical biogeography of *Psammodromus* seems to have been predominantly a history of dispersals across the Strait of Gibraltar. Both interspecific and intraspecific diversification in this genus has been associated with dispersal from Iberia to North Africa through transmarine colonizations on different occasions (Fig. 4.5). On the other hand, vicariance most likely played a role in the diversification of lineages within Iberia and North Africa (Fig. 4.5), but more phylogeographic data, hopefully combined with fossil information, are needed to develop a stronger understanding of diversification patterns of the African taxa.

The cladogenetic events between Iberian and African lineages within *Psammodromus* are dated well before or long after the vicariant event at the end of the MSC (5.33 Ma). Indeed, none of the Bayesian analyses of divergence time (BEAST or *BEAST), which incorporated a wide range of molecular rates, provided 95% posterior intervals on these nodes that overlapped with the MSC. Therefore in *Psammodromus* range evolution and cladogenesis are clearly associated to over-seas dispersals from Iberia toward Africa rather than with the vicariance at the end of the MSC. This pattern is not restricted to *Psammodromus* and is well documented in recent literature. Indeed,

for many terrestrial organisms such as worm lizards and salamanders the high divergence of populations across the Strait of Gibraltar is associated to early crossing of the Strait since the Middle Miocene, likely through the Betic-Rif stepping stones, whereas in many other - such as shrews, larks, snakes, chameleons, skinks, salamanders, frogs and mongooses - shallow divergence underlie dispersal during the Late Pleistocene, possibly in correspondence of sea level low-stand in glacial maxima (reviewed in Hewitt, 2011; see also Gaubert et al., 2011; Santos, Rato, Carranza, Carretero, & Pleguezuelos, 2012). This emerging body of evidence has far reaching implications for molecular biogeography studies beyond the Strait of Gibraltar. Indeed, to date numerous studies have extensively assumed the end of the MSC as the main vicariant cause of species divergence across the Strait and many sea channels across the Mediterranean (e.g. Brown et al., 2008; Chueca, Gómez-Moliner, Forés, & Madeira, 2017; Prüser & Mossakowski, 1998) and used this timing point in phylogenies to calibrate rates of molecular divergence and speciation. However, this and other recent studies (references listed in Hewitt 2011) demonstrate that the assumption of such a 'vicariance-mediated allopatry' model may not be justified in many cases and a 'dispersal-mediated allopatry' model of speciation fits most of the available data significantly better. Since the latter model implies either much older or more recent divergence compared with the vicariance model used so far, this realization warrant a re-assessment of many available molecular rates inferred around the MSC paradigm.

ACKNOWLEDGEMENTS

We thank CIBIO and IBE (CSIC/UPF) colleagues for their help during samples collection. Lizards were captured and handled under permit of the Haut Commisariat aux Eaux and Forets of Morocco (HCEFLCD/DLCDPN/DPRN/DFF N°14/2010). JM and DJH are supported by the Fundação para a Ciência e a Tecnologia (FCT, Portugal): JM, doctoral grant SFRH/BD/81528/2011; DJH IF contract 01627/2014. SC is supported by Grant CGL2015-70390 from the Ministerio de Economía y Competitividad, Spain (co-funded by Fondos FEDER – EU). DS is currently supported by the program 'Rita Levi Montalcini' for the recruitment of young researchers at the University of L'Aquila.

REFERENCES

Amer, S. A. M., & Kumazawa, Y. (2005). Mitochondrial DNA sequences of the Afro-Arabian spiny-tailed lizards (genus *Uromastyx*; family Agamidae): phylogenetic analyses and evolution of gene arrangements. *Biological Journal of the Linnean Society*, *85*, 247–260.

Arevalo, E., Davis, S. K., & Sites, J. W. J. (1994). Mitochondrial DNA sequence divergence and phylogenetic relationships among eight chromosome races of the *Sceloporus grammicus* complex (Phrynosomatidae) in Central Mexico. *Systematic Bioloy*, *43*, 387–418. doi:10.1093/sysbio/43.3.387

Arnold, E. N., Arribas, O., & Carranza, S. (2007). Systematics of the Palaearctic and Oriental lizard tribe Lacertini (Squamata: Lacertidae: Lacertinae), with descriptions of eight new genera. Magnolia Press. *Zootaxa*, *1430*, 1–86.

Augé, M. L. (2005). *Évolution des lézards du Paléogène en Europe. Mémoires du Muséum national d'histoire naturelle, Tome 192*. Paris: Muséum National d'Histoire Naturelle.

Bailon, S., Boistel, R., Bover, P., & Alcover, J. A. (2014). *Maioricalacerta rafelinensi*s, gen. et sp. nov. (Squamata, Lacertidae), from the early Pliocene of Mallorca (Balearic Islands, western Mediterranean Sea). *Journal of Vertebrate Paleontology*, *34*, 318–326. doi:10.1080/02724634.2013.799481

Braga, J. C., Martin, J. M., & Quesada, C. (2003). Patterns and average rates of late Neogene - Recent uplift of the Betic Cordillera, SE Spain. *Geomorphology*, *50*, 3–26. doi:10.1016/S0169-555X(02)00205-2

Brown, R. P., Terrasa, B., Pérez-Mellado, V., Castro, J. A., Hoskisson, P. A., Picornell, A., & Ramon, M. M. (2008). Bayesian estimation of post-Messinian divergence times in Balearic Island lizards. *Molecular Phylogenetics and Evolution*, *48*, 350–358. doi:10.1016/j.ympev.2008.04.013

Buerki, S., Forest, F., Alvarez, N., Nylander, J. A. A., Arrigo, N., & Sanmartín, I. (2011). An evaluation of new parsimony-based versus parametric inference methods in biogeography: a case study using the globally distributed plant family Sapindaceae. *Journal of Biogeography*, *38,* 531–550. doi:10.1111/j.1365-2699.2010.02432.x

Busack, S., & Lawson, R. (2006). Historical biogeography, mitochondrial DNA, and allozymes of *Psammodromus algirus* (Lacertidae): a preliminary hypothesis.

Amphibia-Reptilia, *27*, 181–193. doi:10.1163/156853806777239968

Carranza, S., & Arnold, E. N. (2012). A review of the geckos of the genus *Hemidactylus* (Squamata: Gekkonidae) from Oman based on morphology, mitochondrial and nuclear data, with descriptions of eight new species. Magnolia Press. *Zootaxa*, *95*, 1–95.

Carranza, S., Arnold, E. N., Geniez, P., Roca, J., & Mateo, J. A. (2008). Radiation, multiple dispersal and parallelism in the skinks, *Chalcides* and *Sphenops* (Squamata: Scincidae), with comments on *Scincus* and *Scincopus* and the age of the Sahara Desert. *Molecular Phylogenetics and Evolution*, *46*, 1071–1094. doi:10.1016/j.ympev.2007.11.018

Carranza, S., Arnold, E. N., & Pleguezuelos, J. M. (2006). Phylogeny, biogeography, and evolution of two Mediterranean snakes, *Malpolon monspessulanus* and *Hemorrhois hippocrepis* (Squamata, Colubridae), using mtDNA sequences. *Molecular Phylogenetics and Evolution*, *40*, 532–546. doi:10.1016/j.ympev.2006.03.028

Carranza, S., Arnold, E. N., Wade, E., & Fahd, S. (2004). Phylogeography of the false smooth snakes, *Macroprotodon* (Serpentes, Colubridae): mitochondrial DNA sequences show European populations arrived recently from Northwest Africa. *Molecular Phylogenetics and Evolution*, *33*, 523–532. doi:10.1016/j.ympev.2004.07.009

Carranza, S., Harris, D. J., Arnold, E. N., Batista, V., & Gonzalez de la Vega, J. P. (2006). Phylogeography of the lacertid lizard, *Psammodromus algirus*, in Iberia and across the Strait of Gibraltar. *Journal of Biogeography*, *33*, 1279–1288. doi:10.1111/j.1365-2699.2006.01491.x

Čerňanský, A., Klembara, J., & Smith, K. T. (2016). Fossil lizard from central Europe resolves the origin of large body size and herbivory in giant Canary Island lacertids. *Zoological Journal of the Linnean Society*, *176*, 861–877. doi:10.1111/zoj.12340

Chueca, L. J., Gómez-Moliner, B. J., Forés, M., & Madeira, M. J. (2017). Biogeography and radiation of the land snail genus *Xerocrassa* (Geomitridae) in the Balearic Islands. *Journal of Biogeography*, *44*, 760–772. doi:10.1111/jbi.12923

Drummond, A. J., Suchard, M. A., Xie, D., & Rambaut, A. (2012). Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution*, *29*, 1969–1973. doi:10.1093/molbev/mss075

Duggen, S., Hoernie, K., van den Bogaard, P., Rupke, L., & Morgan, J. F. (2003). Deep

roots of the Messinian salinity crisis. *Nature*, *422*, 602–606. doi:10.1038/nature01551.1.

Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, *32*, 1792–1797. doi:10.1093/nar/gkh340

Fitze, P. S., Gonzalez-Jimena, V., San-Jose, L. M., Mauro, D. S., & Zardoya, R. (2012). A new species of sand racer, *Psammodromus* (Squamata: Lacertidae), from the Western Iberian Peninsula. *Zootaxa*, *52*, 41–52.

Fitze, P. S., Gonzalez-Jimena, V., San-Jose, L. M., San Mauro, D., Aragón, P., Suarez, T., & Zardoya, R. (2011). Integrative analyses of speciation and divergence in *Psammodromus hispanicus* (Squamata: Lacertidae). *BMC Evolutionary Biology*, *11*, 347. doi:10.1186/1471-2148-11-347

Fontboté, J., Guimerà, J., Roca, E., Sàbat, F., Santanach, P., & Fernández-Ortigosa, F. (1990). The Cenozoic geodynamic evolution of the Valencia Trough (Western Mediterranean). *Revista de La Sociedad Geológica de Espana*, *3*, 7–18.

Gamble, T., Bauer, A. M., Greenbaum, E., & Jackman, T. R. (2008). Evidence for Gondwanan vicariance in an ancient clade of gecko lizards. *Journal of Biogeography*, *35*, 88–104. doi:10.1111/j.1365-2699.2007.01770.x

García-Vázquez, D., Bilton, D. T., Alonso, R., Benetti, C. J., Garrido, J., Valladares, L. F., & Ribera, I. (2016). Reconstructing ancient Mediterranean crossroads in *Deronectes* diving beetles. *Journal of Biogeography*, *43*, 1533–1545. doi:10.1111/jbi.12740

Gaubert, P., Machordom, A., Morales, A., López-Bao, J. V., Veron, G., Amin, M., Barros, T., Basuony, M., Djagoun, C. A. M. S., San, E. D. L., Fonseca, C., Geffen, E., Ozkurt, S. O., Cruaud, C., Couloux, A. & Palomares, F. (2011). Comparative phylogeography of two African carnivorans presumably introduced into Europe: disentangling natural versus human-mediated dispersal across the Strait of Gibraltar. *Journal of Biogeography*, *38*, 341–358. doi:10.1111/j.1365-2699.2010.02406.x

Gonçalves, D. V., Brito, J. C., Crochet, P.-A., Geniez, P., Padial, J. M., & Harris, D. J. (2012). Phylogeny of North African *Agama* lizards (Reptilia: Agamidae) and the role of the Sahara desert in vertebrate speciation. *Molecular Phylogenetics and Evolution*, *64*, 582–591. doi:10.1016/j.ympev.2012.05.007

Harris, D. J., Batista, V., Carretero, M. A., & Ferrand, N. (2004). Genetic variation in

Tarentola mauritanica (Reptilia: Gekkonidae) across the Strait of Gibraltar derived from mitochondrial and nuclear DNA sequences. *Amphibia-Reptilia*, *25*, 451–459.

Harris, D. J., Batista, V., Lymberakis, P., & Carretero, M. A. (2004). Complex estimates of evolutionary relationships in *Tarentola mauritanica* (Reptilia: Gekkonidae) derived from mitochondrial DNA sequences. *Molecular Phylogenetics and Evolution*, *30*, 855–859. doi:10.1016/S1055-7903(03)00260-4

Harris, D. J., Carranza, S., Arnold, E. N., Pinho, C., & Ferrand, N. (2002). Complex biogeographical distribution of genetic variation within *Podarcis* wall lizards across the Strait of Gibraltar. *Journal of Biogeography*, *29*, 1257–1262. doi:10.1046/j.1365-2699.2002.00744.x

Harris, D. J., & Sá-Sousa, P. (2002). Molecular pylogenetics of Iberian Wall Lizards (*Podarcis*): is *Podarcis hispanica* a species complex? *Molecular Phylogenetics and Evolution*, *23*, 75–81. doi:10.1006/mpev.2001.1079

Heled, J., & Drummond, A. J. (2010). Bayesian inference of species trees from multilocus data. *Molecular Biology and Evolution*, *27*, 570–580. doi:10.1093/molbev/msp274

Hewitt, G. M. (2011). Mediterranean Peninsulas: The Evolution of Hotspots. In F. E. Zachos & J. C. Habel (Eds.), *Biodiversity Hotspots: Distribution and Protection of Conservation Priority Areas* (pp. 123–147). Berlin, Heidelberg: Springer Berlin Heidelberg. doi:10.1007/978-3-642-20992-5_7

Huson, D. H., & Bryant, D. (2006). Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution*, *23*, 254–267. doi:10.1093/molbev/msj030

Jaramillo-Correa, J. P., Grivet, D., Terrab, A., Kurt, Y., De-Lucas, A. I., Wahid, N., Vendramin, G. G. & González-Martínez, S. C. (2010). The Strait of Gibraltar as a major biogeographic barrier in Mediterranean conifers: A comparative phylogeographic survey. *Molecular Ecology*, *19*, 5452–5468. doi:10.1111/j.1365-294X.2010.04912.x

Kaliontzopoulou, A., Pinho, C., Harris, D. J., & Carretero, M. A. (2011). When cryptic diversity blurs the picture: a cautionary tale from Iberian and North African *Podarcis* wall lizards. *Biological Journal of the Linnean Society*, *103*, 779–800. doi:10.1111/j.1095-8312.2011.01703.x

Kocher, T. D., Thomas, W. K., Meyer, A., Edwards, S. V., Pääbo, S., Villablanca, F. X., & Wilson, A. C. (1989). Dynamics of mitochondrial DNA evolution in animals: amplification and sequencing with conserved primers. *Proceedings of the National*

Academy of Sciences of the United States of America, *86*, 6196–6200.

Krijgsman, W., Garces, M., Agusti, J., Raffi, I., Taberner, C., & Zachariasse, W. J. (2000). The "Tortonian salinity crisis" of the eastern Betics (Spain). *Earth and Planetary Science Letters*, *181*, 497–511. doi:10.1016/S0012-821X(00)00224-7

Krijgsman, W., & Langereis, C. G. (2000). Magnetostratigraphy of the Zobzit and Koudiat Zarga sections (Taza-Guercif basin, Morocco): Implications for the evolution of the Rifian Corridor. *Marine and Petroleum Geology*, *17*, 359–371. doi:10.1016/S0264-8172(99)00029-X

Liu, X.-Q., Ickert-Bond, S. M., Nie, Z.-L., Zhou, Z., Chen, L.-Q., & Wen, J. (2016). Phylogeny of the *Ampelocissus–Vitis* clade in Vitaceae supports the New World origin of the grape genus. *Molecular Phylogenetics and Evolution*, *95*, 217–228. doi:10.1016/j.ympev.2015.10.013

Lonergan, L., & White, N. (1997). Origin of the Betic-Rif mountain belt. *Tectonics*, *16*, 504–522.

Martínez-Solano, I., Gonçalves, H. A., Arntzen, J. W., & García-París, M. (2004). Phylogenetic relationships and biogeography of midwife toads (Discoglossidae: *Alytes*). *Journal of Biogeography*, *31*, 603–618. doi:10.1046/j.1365-2699.2003.01033.x

Mendes, J., Harris, D. J., Carranza, S., & Salvi, D. (2016). Evaluating the phylogenetic signal limit from mitogenomes, slow evolving nuclear genes, and the concatenation approach. New insights into the Lacertini radiation using fast evolving nuclear genes and species trees. *Molecular Phylogenetics and Evolution*, *100*, 254–267. doi:10.1016/j.ympev.2016.04.016

Metallinou, M., Červenka, J., Crochet, P.-A., Kratochvíl, L., Wilms, T., Geniez, P., Shobrak, M. Y., Brito, J. C. & Carranza, S. (2015). Species on the rocks: Systematics and biogeography of the rock-dwelling *Ptyodactylus* geckos (Squamata: Phyllodactylidae) in North Africa and Arabia. *Molecular Phylogenetics and Evolution*, *85*, 208–220. doi:10.1016/j.ympev.2015.02.010

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*, New Orleans. doi:10.1109/GCE.2010.5676129

Palumbi, S. R., Martin, A., Romano, S., McMillan, S., Stice, W. O., & Grabowski, L. (1991). *The simple fool's guide to PCR* (1st ed.). Honolulu: Department of Zoology,

University of Hawaii.

Pavlicev, M., & Mayer, W. (2009). Fast radiation of the subfamily Lacertinae (Reptilia: Lacertidae): history or methodical artefact? *Molecular Phylogenetics and Evolution*, *52*, 727–34. doi:10.1016/j.ympev.2009.04.020

Perea, S., Cobo-Simon, M., & Doadrio, I. (2016). Cenozoic tectonic and climatic events in southern Iberian Peninsula: Implications for the evolutionary history of freshwater fish of the genus *Squalius* (Actinopterygii, Cyprinidae). *Molecular Phylogenetics and Evolution*, *97*, 155–169. doi:10.1016/j.ympev.2016.01.007

Pérez-Asensio, J. N., Aguirre, J., Schmiedl, G., & Civis, J. (2012). Impact of restriction of the Atlantic-Mediterranean gateway on the Mediterranean Outflow Water and eastern Atlantic circulation during the Messinian. *Paleoceanography*, *27*, PA3222. doi:10.1029/2012PA002309

Pinho, C., Rocha, S., Carvalho, B. M., Lopes, S., Mourão, S., Vallinoto, M., Brunes, T., Haddad, C. F. B., Gonçalves, H., Sequeira, F. & Ferrand, N. (2009). New primers for the amplification and sequencing of nuclear loci in a taxonomically wide set of reptiles and amphibians. *Conservation Genetics Resources*, *2*, 181–185. doi:10.1007/s12686-009-9126-4

Pinya, S., & Carretero, M. A. (2011). The Balearic herpetofauna: a species update and a review on the evidence. *Acta Herpetologica*, *6*, 59–80.

Posada, D. (2008). jModelTest: phylogenetic model averaging. *Molecular Biology and Evolution*, *25*, 1253–1256. doi:10.1093/molbev/msn083

Prüser, F., & Mossakowski, D. (1998). Low substitution rates in mitochondrial DNA in Mediterranean carabid beetles. *Insect Molecular Biology*, *7*, 121–128. doi:10.1046/j.1365-2583.1998.72056.x/abs

Pulquério, M. J. F., & Nichols, R. A. (2007). Dates from the molecular clock: how wrong can we be? *Trends in Ecology & Evolution*, *22*, 180–184. doi:10.1016/j.tree.2006.11.013

Ree, R. H., Moore, B. R., Webb, C. O., & Donoghue, M. J. (2005). A Likelihood framework for inferring the evolution of geographic range on phylogenetic trees. *Evolution*, *59*, 2299–2311. doi:10.1111/j.0014-3820.2005.tb00940.x

Ree, R. H., & Sanmartín, I. (2009). Prospects and challenges for parametric models in historical biogeographical inference. *Journal of Biogeography*, *36*, 1211–1220.

doi:10.1111/j.1365-2699.2008.02068.x

Ree, R. H., & Smith, S. (2008). Maximum likelihood inference of geographic range evolution by dispersal, local extinction, and cladogenesis. *Systematic Biology*, *57*, 4–14. doi:10.1080/10635150701883881

Rosenbaum, G., Lister, G. S., & Duboz, C. (2002a). Reconstruction of the tectonic evolution of the western Mediterranean since the Oligocene. *Journal of the Virtual Explorer*, *8*, 107–126. doi:10.3809/jvirtex.2002.00053

Rosenbaum, G., Lister, G. S., & Duboz, C. (2002b). Relative motions of Africa, Iberia and Europe during Alpine orogeny. *Tectonophysics*, *359*, 117–129. doi:10.1016/S0040-1951(02)00442-0

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

Salvi, D., Schembri, P. J., Sciberras, A., & Harris, D. J. (2014). Evolutionary history of the Maltese wall lizard *Podarcis filfolensis*: insights on the “Expansion-Contraction” model of Pleistocene biogeography. *Molecular Ecology*, *23*, 1167–1187. doi:10.1111/mec.12668

Sambrook, J., Fritsch, E. F., & Maniatis T. (1989). *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Press, New York.

Sampaio, F. L., Harris, D. J., Perera, A., & Salvi, D. (2014). Phylogenetic and diversity patterns of *Blanus* worm lizards (Squamata: Amphisbaenia): Insights from mitochondrial and nuclear gene genealogies and species tree. *Journal of Zoological Systematics and Evolutionary Research*, *53*, 45–54. doi:10.1111/jzs.12075

Sanmartín, I. (2007). Event-based biogeography: Integrating patterns, processes, and time. In M. C. Ebach & R. Tangney (Eds.), *Biogeography in a changing world* (pp. 135–159). Boca Raton, Florida: CRC Press.

Santos, X., Rato, C., Carranza, S., Carretero, M. A., & Pleguezuelos, J. M. (2012). Complex phylogeography in the Southern Smooth Snake (*Coronella girondica*) supported by mtDNA sequences. *Journal of Zoological Systematics and Evolutionary Research*, *50*, 210–219. doi:10.1111/j.1439-0469.2012.00663.x

Stamatakis, A. (2006). RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics (Oxford,*

England), *22*, 2688–2690. doi:10.1093/bioinformatics/btl446

Stephens, M., Smith, N. J., & Donnelly, P. (2001). A new statistical method for haplotype reconstruction from population data. *American Journal of Human Genetics*, *68*, 978–989. doi:10.1086/319501

Tamar, K., Scholz, S., Crochet, P.-A., Geniez, P., Meiri, S., Schmitz, A., Wilms, T. & Carranza, S. (2016). Evolution around the Red Sea: Systematics and biogeography of the agamid genus *Pseudotrapelus* (Squamata: Agamidae) from North Africa and Arabia. *Molecular Phylogenetics and Evolution*, *97*, 55–68. doi:10.1016/j.ympev.2015.12.021

Velo-Antón, G., Godinho, R., Harris, D. J., Santos, X., Martínez-Freiria, F., Fahd, S., Larbes, S., Pleguezuelos, J. M. & Brito, J. C. (2012). Deep evolutionary lineages in a Western Mediterranean snake (*Vipera latastei*/*monticola* group) and high genetic structuring in Southern Iberian populations. *Molecular Phylogenetics and Evolution*, *65*, 965–973. doi:10.1016/j.ympev.2012.08.016

Verdú-Ricoy, J., Carranza, S., Salvador, A., Busack, S., & Díaz, J. (2010). Phylogeography of *Psammodromus algirus* (Lacertidae) revisited: systematic implications. *Amphibia-Reptilia*, *31*, 576–582. doi:10.1163/017353710X521555

Vieites, D. R., Min, M.-S., & Wake, D. B. (2007). Rapid diversification and dispersal during periods of global warming by plethodontid salamanders. *Proceedings of the National Academy of Sciences*, *104*, 19903–19907. doi:10.1073/pnas.0705056104

Yu, Y., Harris, A. J., Blair, C., & He, X. (2015). RASP (Reconstruct Ancestral State in Phylogenies): a tool for historical biogeography. *Molecular Phylogenetics and Evolution*, *87*, 46–49. doi:10.1016/j.ympev.2015.03.008

Yu, Y., Harris, A. J., & He, X. (2010). S-DIVA (Statistical Dispersal-Vicariance Analysis): A tool for inferring biogeographic histories. *Molecular Phylogenetics and Evolution*, *56*, 848–850. doi:10.1016/j.ympev.2010.04.011

Yu, Y., Harris, A. J., & Xingjin, H. (2011). RASP version2.1b. Available at http://mnh.scu.edu.cn/soft/blog/RASP.

Zazo, C. (1999). Interglacial sea levels. *Quaternary International*, *55*, 101–113. doi:10.1016/S1040-6182(98)00031-7

SUPPORTING INFORMATION

Additional Supporting Information can be found in the Appendices section

Figure S4.1 Bayesian tree with one individual per species

Figure S4.2 Cross-validation of the evolution rates

Figure S4.3 Three areas ancestral reconstruction tree

Figure S4.3 Map and Table of the described fossils of *Psammodromus* and *Gallotia* clade

CHAPTER 5

Multilocus phylogeny of *Omanosaura* lizards

Article IV. Hidden in the Arabian mountains: multilocus phylogeny reveals cryptic diversity in the endemic *Omanosaura* lizards

Joana Mendes[1,2,3*], Daniele Salvi[1,4], D. James Harris[1,2], Johannes Els[5], Salvador Carranza[3]

Accepted for publication by the *Journal of Zoological Systematics and Evolutionary Research* on December 25, 2017

DOI: 10.1111/jzs.12210

[1] CIBIO Research Centre in Biodiversity and Genetic Resources, InBIO, Universidade do Porto, Campus de Vairão, Rua Padre Armando Quintas, N° 7, 4485-661 Vairão, Vila do Conde, Portugal

[2] Departamento de Biologia, Faculdade de Ciências, Universidade do Porto, Rua do Campo Alegre FC4 4169-007 Porto, Portugal

[3] Institute of Evolutionay Biology (CSIC-Universitat Pompeu Fabra), Passeig Marítim de la Barceloneta 37-49, 08003 Barcelona, Spain.

[4] Department of Health, Life and Environmental Sciences, University of L'Aquila, 67100 Coppito, L'Aquila, Italy

[5] Breeding Centre for Endangered Arabian Wildlife, Environment and Protected Areas Authority, PO Box 29922, Sharjah, UAE

ABSTRACT

An increase of studies in the Hajar Mountains from the south-eastern Arabian Peninsula has revealed a high richness of endemic evolutionary lineages with many cryptic taxa. *Omanosaura* is the only lacertid lizard genus endemic to the Hajar Mountains, with the two species *O. cyanura* and *O. jayakari* distributed throughout this mountain range. The phylogenetic relationships and genetic diversity between and within these species have been poorly studied. In this study, we collected mitochondrial (*12S*, *cytb* and *nd4*) and nuclear (*cmos* and *mc1r*) sequences for 25 specimens of *Omanosaura*, including 15 individuals of *O. jayakari* and 10 of *O. cyanura*. We performed phylogenetic analyses based on network reconstruction, Maximum Likelihood and Bayesian inference to infer the relationships and intraspecific genetic diversity of these species. We estimated the time of divergence between the two species in the Miocene, around 8.5 million years ago. *Omanosaura jayakari* shows little genetic diversity, while *O. cyanura* presents two differentiated lineages. These are reciprocally monophyletic at mitochondrial and nuclear genes and present a high genetic distance between them. These two lineages are associated with the geographical features of the Hajar Mountains, with one lineage distributed in the northernmost part of the Hajar Mountains and the other in the rest of the western Hajars, the Jebel Akhdar and the eastern Hajars. This geographical relationship has been recovered previously in other reptile taxa, and is generally associated with high levels of local genetic diversity. Our results suggest the existence of cryptic diversity within *O. cyanura*, and support a general biogeographic pattern of high diversity and endemism in the northern Hajar Mountains that certainly deserves additional research in the future.

KEYWORDS

Lacertidae, Hajar Mountains, biodiversity, species tree

INTRODUCTION

Knowledge concerning biodiversity patterns around the globe is imbalanced. The existence of whole regions for which information is either absent or very scarce, can be associated with difficult or impossible access to these areas, whether connected with remote or isolated geographical features (Ficetola, Bonardi, Sindaco, & Padoa-Schioppa, 2013), or with socio-political instability (Brito et al., 2014). The lack of data normally culminates in a general underestimation of the local biodiversity levels. Case examples of regions with recently discovered hidden genetic variability are the arid mountains in North Africa and the Arabian Peninsula (e.g. Carranza, Simó-Riudalbas, Jayasinghe, Wilms, & Els, 2016; Garcia-Porta, Simó-Riudalbas, Robinson, & Carranza, 2016; Metallinou et al., 2015; Rato, Harris, Carranza, Machado, & Perera, 2016; Rosado, Rato, Salvi, & Harris, 2017).

The Atlas Mountains in the Maghreb, for instance, belong to the Mediterranean basin hotspot but, contrary to the European margin, still hide a considerable amount of genetic diversity. Over the last years, an increase of expeditions to the Maghreb, together with the application of new tools in molecular analyses, led to the discovery of cryptic diversity in this area (Barata, Carranza, & Harris, 2012; Metallinou et al., 2015; Rato et al., 2016; Rosado et al., 2017; Tamar, Geniez, Brito, & Crochet, 2017). A similar scenario can be found in the Hajar Mountains of south-eastern Arabia. This impressive mountain range, with the highest peaks reaching 3000 m a.s.l., extends from the Musandam Peninsula in the north and runs almost parallel to the Gulf of Oman through the eastern United Arab Emirates (UAE) and northern Oman for 650 km (Edgell, 2006; Figure 5.1). Despite being close to the sea, the general low precipitation and high evaporation levels make it an arid mountain desert (Edgell, 2006; Mandaville, 1977). However, the deep gorges running between the peaks (wadis) have water and vegetation, at least intermittently, and are the places where the majority of fauna and flora can be found. The geological setting, together with a local microclimate and the presence of wadis make the Hajar Mountains an important refuge for endemic and relict species (Mandaville, 1977), including, among others, endemic species of plants (MacLaren, 2016) and of reptiles, such as two species of the lizard genus *Omanosaura* Lutz, Bischoff & Mayer, 1986, several species of the geckos of the genus *Asaccus* Dixon & Anderson, 1973 (Carranza et al., 2016; Simó-Riudalbas, Tarroso, Papenfuss, Al-Sariri, & Carranza, 2017), *Hemidactylus* Oken, 1817 (Carranza & Arnold, 2012), *Pristurus* Rüppell, 1835 (Arnold, 2009; Badiane et al., 2014; Garcia-Porta et al., 2016), *Ptyodactylus* Godfuss, 1820 (Metallinou et al., 2015; Simó-Riudalbas, Metallinou, et al., 2017) and

Trachydactylus Haas and Battersby, 1959 (de Pous et al., 2016), an agamid of the genus *Pseudotrapelus* Fitzinger, 1843 (Tamar, Scholz, et al., 2016) and a viper of the genus *Echis* Merrem, 1820 (Babocsay, 2004; Robinson, Carranza, & Arnold, 2009). With an increase of scientific attention given to the Hajar Mountains, the number of microendemic species or lineages in this area has increased remarkably over the last few years, particularly among reptiles (Carranza et al., 2016; Garcia-Porta et al., 2016; Simó-Riudalbas, Metallinou, et al., 2017) an indication of very high levels of cryptic diversity in a still understudied biodiversity hotspot.

Omanosaura is the only endemic genus of the family Lacertidae in the Arabian Peninsula and the only endemic reptile genus of the Hajar Mountains. The genus has two species, *O. cyanura* (Arnold, 1972) and *O. jayakari* (Boulenger, 1887), with similar distribution ranges, including many areas of sympatry (Carranza et al., 2017; Gardner, 2013; Sindaco & Jeremcenko, 2008; Sindaco, Venchi, & Grieco, 2011). These species are morphologically very distinct, although both have very long tails of up to 2.75 times their mean body sizes (snout vent length, SVL). *Omanosaura cyanura* is a small-bodied lizard, with SVL of about 60 mm, and presents a flattened head and body, with dorsal colouration that can be brown or blue, with a long electric blue tail, the feature to which it owes its name. Little is known about this lizard, which is normally spotted in wadis with vegetation and water, from sea level up to 2400 m a.s.l (Carranza et al., 2017; Gardner, 2013). On the other hand, *O. jayakari* is the largest lacertid lizard in the Hajar Mountains, with SVL up to 200 mm. This robust species is an active hunter and can predate on other lizards, including juveniles of *O. cyanura* (Gardner, 2013). First described as *Lacerta cyanura* and *L. jayakari*, these species were placed in a separate sub-genus, *Omanosaura*, by Lutz et al. (1986), based on the genetic distance between *O. jayakari* and the rest of the former members of the genus *Lacerta*. The inclusion of *O. cyanura* in the same sub-genus was based on morphological grounds (Arnold, 1972; Arnold, 1973) and only later, the relatively low immunological distance to *O. jayakari* corroborated their close relationship (Mayer & Benyr, 1994). *Omanosaura* was elevated to the generic level by Mayer & Bischoff (1996). Despite the interest from an evolutionary point of view, sequences in GenBank are only available for one specimen of *O. cyanura* (three mitochondrial genes -Harris, Arnold, & Thomas (1998)) and two specimens of *O. jayakari*, with one specimen sequenced for three mitochondrial genes (Harris et al., 1998) and a different specimen sequenced for two slow-evolving nuclear genes (Mayer & Pavlicev, 2007). Most of these studies had the main purpose of establishing the phylogenetic relationships among genera of Lacertidae or within the tribe Eremiadini, to which *Omanosaura* belongs (Arnold, Arribas, & Carranza, 2007). The results show that

O. jayakari and *O. cyanura* form a well-supported clade (Harris et al., 1998) of unresolved phylogenetic position within the Eremiadini. However, the genetic diversity and phylogeography of the two *Omanosaura* species remain essentially unknown.

In this study, we analyse for the first time multiple samples of *O. cyanura* and *O. jayakari* from across their distribution ranges in the Hajar Mountains. We performed phylogenetic analyses based on a multilocus approach, including both mitochondrial and nuclear genes and estimated the divergence time between and within species. The main aim of this study is to infer the phylogeographic relationships and intraspecific genetic diversity within *Omanosaura* and to increase our knowledge concerning biogeographic patterns and diversity in this understudied region.

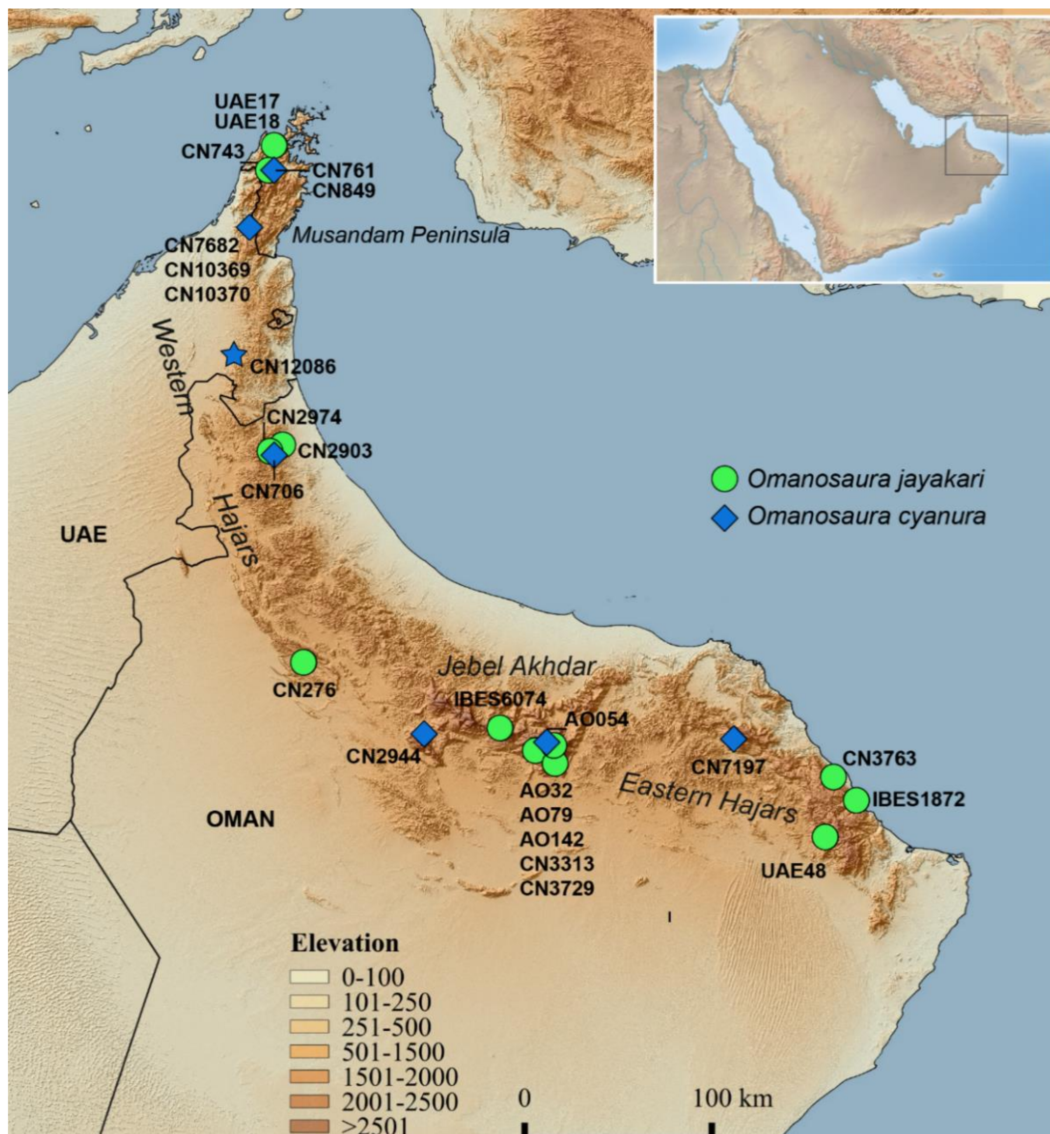


Figure 5.1. Sampling localities in the Hajar Mountains of the individuals of *Omanosaura cyanura* and *O. jayakari* included in this study. The blue star represents the type locality of *O. cyanura*.

MATERIAL AND METHODS

TAXON SAMPLING

We included in the phylogenetic analyses 15 samples of *O. jayakari* and 10 of *O. cyanura*, covering the species distribution range in the Hajar Mountains (Figure 5.1). All *Omanosaura* samples and vouchers belong to the collection of the Institute of Evolutionary Biology (IBE-CSIC). Two samples of the closely related Eremiadini species *Acanthodactylus blanfordii* Boulenger, 1918, *Acanthodactylus schmidti* Haas, 1957 and *Mesalina guttulata* Lichtenstein, 1823 were included in the analyses as outgroups (Arnold et al., 2007; Greenbaum, Villanueva, Kusamba, Aristote, & Branch, 2011; Mayer & Pavlicev, 2007). Sample codes, geographic coordinates and GenBank accession numbers are reported in Table 5.1.

DNA EXTRACTION, AMPLIFICATION AND SEQUENCE ANALYSES

Total genomic DNA was obtained from ethanol-preserved tissue samples following the standard high-salt protocol (Sambrook, Fritsch, & Maniatis T., 1989). We amplified three mitochondrial gene fragments –*ribosomal 12S rRNA* (*12S*), *cytochrome b* (*cytb*) and *NADH dehydrogenase 4* with flanking tRNAs Serine, Histidine and Leucine (*nd4*) – and two nuclear genes – *oocyte maturation factor mos* (*cmos*) and *melanocortin 1 receptor* (*mc1r*). These markers have been successfully used in many intra- and inter- specific studies on lacertid lizards (Mendes, Harris, Carranza, & Salvi, 2016; Salvi, Harris, Bombi, Carretero, & Bologna, 2010; Salvi, Schembri, Sciberras, & Harris, 2014; Tamar, Carranza et al., 2016). Amplification was performed through Polymerase Chain Reaction (PCR). Primers, PCR conditions and references are listed in the Supporting Information (Table S5.1; see also Mendes et al. (2016) and Salvi et al. (2017)). Purification of PCR products and sequencing were carried out by Macrogen (www.macrogen.com) with the same primers used for amplification.

Nucleotide sequences were manually checked and edited in GENEIOUS 4.8.5 (Kearse et al., 2012). Heterozygous positions in the nuclear genes *cmos* and *mc1r* were coded according to the IUPAC ambiguity codes. All protein-coding gene fragments (*cytb*, *nd4, cmos* and *mc1r*) were translated into amino acids sequences and no stop codons were observed, suggesting that all the sequences were functional. Multiple DNA sequences were aligned using the MUSCLE algorithm (Edgar, 2004) implemented in GENEIOUS. For each nuclear gene, the possible occurrence of recombination events was assessed using the Pairwise Homoplasy Index (phi) test (Bruen, Philippe, & Bryant,

2006) implemented in SPLITSTREE 4.14.4 (Huson & Bryant, 2006). We used MEGA 6.0 (Tamura, Stecher, Peterson, Filipski, & Kumar, 2013) to calculate pairwise genetic distances (*p*-distance) of mitochondrial haplotypes and the number of variable and parsimony informative sites in all genes.

PHYLOGENETIC ANALYSES AND DIVERGENCE TIME ESTIMATION

In order to reconstruct the phylogenetic relationships of *Omanosaura* species we performed haplotype network reconstructions based on single loci; Maximum Likelihood (ML) and Bayesian Inference (BI) analyses based on the concatenated mitochondrial (mtDNA), and mitochondrial and nuclear (mt-nucDNA) datatsets; and species tree based on the multispecies coalescent approach (Heled & Drummond, 2010). The best fitting models of sequence evolution for the gene partition and phased nuclear sequences were inferred with jMODELTEST 2.1.7 (Posada, 2008) under the corrected Akaike Information Criterion.

Haplotype networks were performed for the concatenated mtDNA genes and for the phased nucDNA genes *cmos* and *mc1r* using only full length sequences. The haplotype reconstruction for *cmos* and *mc1r* was performed in PHASE 2.1 (Stephens et al., 2001; Stephens and Scheet, 2005). The input files were obtained in SEQPHASE (Flot, 2010; available at http://seqphase.mpg.de/seqphase/). PHASE was run three times with 100 iterations, phase probability of 70% and remaining settings as by default. Haplotype networks were inferred with Median-Joining algorithm in the software NETWORK 5.0.0.0. (available at http://www.fluxus-engineering.com/sharenet.htm), with default parameters.

ML analyses were performed with RAxML 7.4.2 (Stamatakis, 2006), using the graphical front-end RaxML GUI 1.3 (Silvestro & Michalak, 2012). ML searches were conducted with 10 random addition replicates using the GTR+G model of evolution with parameters estimated independently for each gene partition. To assess the node support we performed 1000 nonparametric bootstrap replicates.

Table 5.1. Code and geographic coordinates of the *O. jayakari* and *O. cyanura* lineages and of outgroup species included in this study. GenBank accession numbers of *12S*, *cytb*, *nd4*, *cmos*, and *mc1r* are MG672294-318, MG672340-364, MG672389-410, MG672319-339 and MG672365-388, respectively. †Sequences downloaded from GenBank; ‡ specimens with vouchers.

Species (lineage)	Code	Geographic Coordinates (Latitude, Longitude)	*12S*	*cytb*	*nd4*	*c-mos*	*mc1r*
Omanosaura jayakari	UAE17	26.18 N, 56.26 E	MG672297	MG672343	MG672392	MG672322	MG672368
Omanosaura jayakari	UAE18	26.18 N, 56.26 E	MG672298	MG672344		MG672323	MG672369
Omanosaura jayakari	CN743	26.05 N, 56.23 E	MG672302	MG672348	MG672396	MG672327	MG672373
Omanosaura jayakari	CN2974	24.61 N, 56.24 E	MG672304	MG672350	MG672398	MG672329	MG672375
Omanosaura jayakari	CN2903	24.64 N, 56.30 E	MG672305	MG672351	MG672399		MG672376
Omanosaura jayakari	CN276	23.52 N, 56.41 E	MG672303	MG672349	MG672397	MG672328	MG672374
Omanosaura jayakari	IBES6074	23.18 N, 57.42 E	MG672308	MG672354	MG672402	MG672331	MG672379
Omanosaura jayakari	AO32	23.07 N, 57.60 E	MG672294	MG672340	MG672389	MG672319	MG672365
Omanosaura jayakari	AO79	23.09 N, 57.69 E	MG672295	MG672341	MG672390	MG672320	MG672366
Omanosaura jayakari	AO142	23.09 N, 57.69 E	MG672296	MG672342	MG672391	MG672321	MG672367
Omanosaura jayakari	CN3313	23.07 N, 57.63 E	MG672306	MG672352	MG672400		MG672377
Omanosaura jayakari	CN3729	23.00 N, 57.70 E	MG672307	MG672353	MG672401	MG672330	MG672378
Omanosaura jayakari	CN3763	22.93 N, 59.13 E	MG672301	MG672347	MG672395	MG672326	MG672372
Omanosaura jayakari	IBES1872	22.81 N, 59.25 E	MG672300	MG672346	MG672394	MG672325	MG672371
Omanosaura jayakari	UAE48	22.62 N, 59.09 E	MG672299	MG672345	MG672393	MG672324	MG672370
Omanosaura cyanura (North)	CN761‡	26.05 N, 56.23 E	MG672311	MG672357	MG672404	MG672334	MG672382
Omanosaura cyanura (North)	CN849‡	26.05 N, 56.23 E	MG672310	MG672356	MG672403	MG672333	MG672381
Omanosaura cyanura (North)	CN7682	25.76 N, 56.13 E	MG672315	MG672361	MG672408		MG672386

Omanosaura cyanura (North)	CN10369	25.76 N, 56.13 E	MG672314	MG672360	MG672407	MG672337	MG672385
Omanosaura cyanura (North)	CN10370	25.76 N, 56.13 E	MG672313	MG672359	MG672406	MG672336	MG672384
Omanosaura cyanura (North)	CN12086	25.10 N, 56.07 E	MG672318	MG672364	MG672410	MG672339	
Omanosaura cyanura (South)	CN706	24.61 N, 56.24 E	MG672316	MG672362	MG672409	MG672338	MG672387
Omanosaura cyanura (South)	CN2944‡	23.15 N, 57.03 E	MG672309	MG672355		MG672332	MG672380
Omanosaura cyanura (South)	AO054	23.11 N, 57.66E	MG672317	MG672363			MG672388
Omanosaura cyanura (South)	CN7197‡	23.13 N, 58.62 E	MG672312	MG672358	MG672405	MG672335	MG672383
Mesalina guttulata	M17		†KX296984	†KX297187		†KX297735	†KX297361
Mesalina guttulata	M18		†KX296985	†KX297188		†KX297736	†KX297362
Acanthodactylus blanfordii	A59		†KX296907	†KX297018		†KX297577	†KX297306
Acanthodactylus blanfordii	A290		†KX296906	†KX297017		†KX297575	†KX297305
Acanthodactylus schmidti	A51		†KX296911	†KX297020		†KX297583	†KX297308
Acanthodactylus schmidti	A228		†KX296910	†KX297019		†KX297576	†KX297307
Gallotia caesaris gomerae	DB19355		†MF684894	†MF684934	†MF684983	†AY152005	†MF684960
Gallotia caesaris gomerae	DB19375		†MF684895	†MF684935	†MF684984	†AF435101	†MF684961
Gallotia caesaris caesaris	DB19413		†MF684896	†MF684936	†MF684985	†AY152006	†MF684962
Gallotia caesaris caesaris	DB19470		†MF684897	†MF684937	†MF684986		†MF684963
Atlantolacerta andreanskyi	5015		†JX462057		†JX462200	†JX485204	†JX461803
Atlantolacerta andreanskyi	5058		†JX462054		†JX462196	†JX485206	†JX461816

Currently, no fossils of *Omanosaura* are known, preventing the use of internal calibration points to directly estimate the divergence time within the genus. Therefore, we used two strategies to estimate cladogenetic events within *Omanosaura* based on previous estimates on lacertid lizards by Carranza & Arnold (2012): (i) we applied the rate of evolution of *12S* and *cytb* estimated by Carranza & Arnold (2012) based on seven biogeographic calibration points; and (ii) we calibrated our tree by applying directly the node prior for the split between *Gallotia caesaris caesaris* and *G. c. gomerae* used in Carranza & Arnold (2012) (Table 5.1). This allowed consistency of time estimates using priors on rates *versus* nodes (i.e. secondary *versus* primary calibrations) and including or excluding the biogeographic calibration centred on the Messinian Salinity Crisis event, which has been recently criticized (Hewitt, 2011; Mendes, Harris, Carranza, & Salvi, 2017). The BI and the divergence time estimation were performed in BEAST 1.8.0. (Drummond & Rambaut, 2007). We implemented the relaxed uncorrelated lognormal clock model for all genes because it overcomes the hard assumptions of the strict clock. We built the input file with evolutionary models, tree priors and Markov Chain Monte Carlo (MCMC) options using the BEAUTi utility included in the BEAST package. Models and prior specifications applied were as follows (otherwise by default): nucleotide substitution and clock models were unlinked; models of nucleotide substitution as specified in Table 5.2; relaxed uncorrelated lognormal clock for all genes. Yule process of speciation; random starting tree; base substitution Uniform (0, 100); alpha Uniform (0, 10); ucld.mean of *12S* Normal (initial: 0.00553, mean: 0.00553, stdev: 0.00128); ucld.mean of *cytb* Normal (initial: 0.0164, mean: 0.0164, stdev: 0.00317); clock rate of *nd4*, *cmos* and *mc1r* Uniform (0, 0.25); operator kappa (2.0). The xml file was manually modified to set “Ambiguities = true” for the nuclear partitions to allow a full account of nuclear polymorphisms during phylogeny estimation. BEAST was run three times, with 100 million generations, sampling every 10,000 generations. The use of the Yule process of speciation prior requires only one sequence per species, whereas our concatenated alignments contained multiple samples per species. Therefore, to investigate the sensitivity of our estimates to the choice of tree prior, we performed an additional run applying the same settings as above but using only one representative sequence for each species.

In addition, we inferred a phylogeny using the species tree approach implemented in the *BEAST extension of the BEAST software because we found evidence for three evolutionary independent lineages (which were defined as “species” in *BEAST) within *Omanosaura* taxa analysed (*O. jayakari*, *O. cyanura* North lineage and *O. cyanura* South lineage, see Results). We used the mtDNA sequences and the

phased alignments of the nuclear genes and their relative models of nucleotide evolution (Table 5.2). The settings applied in *BEAST for the species tree reconstruction were similar to the concatenated BEAST analyses, except the tree model of the mitochondrial genes *12S*, *cytb* and *nd4* was linked, since these genes are genetically linked. In order to calibrate the species tree, we applied the *12S* and *cytb* substitution rates estimated by Carranza and Arnold (2012), as in the concatenated BEAST analysis. *BEAST was run three times with 200 million generations, sampling every 20,000 generations. All BEAST runs were performed on the CIPRES Science Gateway 3.3 (Miller et al., 2010, at http://www.phylo.org/). Results were analysed in TRACER 1.6 (Rambaut & Drummond, 2007), applying a 10 % bur-in, to check for convergence and to ensure that all ESS parameters were higher than 200, as recommended in the software's manual. LOGCOMBINER and TREEANNOTATOR (both included in the BEAST package) were used to calculate the Maximum Clade Credibility (MCC) tree summarizing the posterior distribution of tree topologies and branch lengths. All trees were visualized in FIGTREE 1.4 (available at http://tree.bio.ed.ac.uk/software/figtree/).

RESULTS

A total of 117 new sequences of *Omanosaura* were generated and deposited in GenBank; the accession numbers are reported in Table 5.1. The concatenated mt-nucDNA alignment of *Omanosaura* was 2,648 bp long. The phi test did not find statistically significant evidence for recombination in the nuclear gene fragments ($p>0.05$). The length of each gene fragment is reported in Table 5.2, along with the number of variable positions between and within *Omanosaura* species.

Table 5.2. Length, models of nucleotide substitution, number of intra and interspecific variable positions and parsimony informative sites for each gene fragment used in this study.

Gene	Length		Model		Var. pos. *O. jayakari*	Var. pos. *O. cyanura*	Var. pos. *Omanosaura*	Pars. Inf
	Alignment	Amplicon	Unphased	Phased				
12S	387	435	GTR+I		1	27	47	41
cytb	425	460	GTR+I		8	76	103	93
nd4	851	900	GTR+G		12	120	185	144
cmos	353	390	HKY	TN93	1	1	3	2
mc1r	629	690	HKY+I	HKY+I	3	7	12	8

PHYLOGENETIC RELATIONSHIPS WITHIN OMANOSAURA

The phylogenetic relationships inferred by the ML, BI and species tree analyses are similar and present overall high levels of bootstrap support (BS) and Bayesian Posterior Probabilities (BPP) (Figures 5.2, 5.3 and Supplementary information Figures S5.1 and S5.2). The age estimates based on the substitution rates were similar between the concatenated BI tree (Figure 5.2) and the species tree (Figure 5.3) and are in agreement with the estimates obtained using the split between *Gallotia c. caesaris* and *G. c. gomerae* as calibration (Figure S5.3), thus we will refer only to the estimates of the concatenated BI tree (Figure 5.2). The two *Omanosaura* species, *O. jayakari* and *O. cyanura,* are reciprocally monophyletic (BPP = 1, BS=100; Figures 5.2 and 5.3) and the results from the time calibrated tree (Figures 5.2, 5.3 and S5.3) indicate that they diverged around 8.4 million years ago (Ma) (95% HPD: 5.51-11.15). *Omanosaura jayakari* presents low levels of genetic diversity, with two very recent clades, recovered only in the BI analyses. On the other hand, *O. cyanura* presents two highly divergent lineages (*p*-distance: *12S*: 5%, *cytb*: 12%, *nd4*: 9%; Table 5.3) that are strictly associated with the geographic distribution of the samples (Figures 5.1 and 5.2). One lineage is distributed in the northern Hajar Mountains, *O. cyanura* North; and the second lineage is distributed in the remaining range of the Hajar Mountains, *O. cyanura* South. The divergence time between the two *O. cyanura* lineages is estimated at 4.25 Ma (95% HPD: 2.87-5.82). The northern lineage of *O. cyanura* presents low levels of genetic diversity, with diversification starting recently at 0.24 Ma (95% HPD: 0.10-0.41), while the southern lineage presents higher levels of diversity, with deeper intra-lineage divergence starting around 3.12 Ma (95% HPD: 2.04-4.36).

Table 5.3. Genetic pairwise *p*-distance values (d), range (minimum and maximum values) and standard error (S.E.) between and within *O. jayakari* and *O. cyanura* lineages for the mitochondrial genes *12S*, *cytb* and *nd4*.

	12S		***cytb***		***nd4***	
Between groups	**d (min-max)**	**S.E**	**d (min-max)**	**S.E**	**d (min-max)**	**S.E**
O. jayakari – *O. cyanura* North	0.085 (0.081 – 0.090)	0.014	0.166 (0.158 – 0.170)	0.017	0.155 (0.148 – 0.177)	0.014
O. jayakari – *O. cyanura* South	0.085 (0.081 – 0.092)	0.012	0.153 (0.148 – 0.160)	0.017	0.148 (0.143 – 0.161)	0.011
O. cyanura North – *O. cyanura* South	0.052 (0.047 – 0.060)	0.009	0.124 (0.0 – 0.141)	0.015	0.094 (0.0 – 0.112)	0.009
Within groups	**d (min-max)**	**S.E**	**d (min-max)**	**S.E**	**d (min-max)**	**S.E**
O. jayakari	0.0014 (0.0 – 0.003)	0.0014	0.0028 (0.0 – 0.0070)	0.0009	0.0059 (0.0 – 0.011)	0.0018
O. cyanura North	0.0 (0.0 – 0.0)	0.0	0.0050 (0.0 – 0.0090)	0.0023	0.004 (0.0 – 0.007)	0.0015
O. cyanura South	0.0201 (0.0 – 0.0340)	0.005	0.0635 (0.002 – 0.096)	0.0078	0.0989 (0.0989 – 0.099)	0.0101

The same phylogenetic structure was recovered by the network analyses (Figures 5.4 and 5.5). The mtDNA network recovered the same clades described in the BI analyses (Figures 5.2, 5.4 and S5.2): the *O. jayakari* clade (showing low intra-clade differentiation) and the two lineages of *O. cyanura* (Figure 5.4). The same clades are also recovered by the two nuclear networks, with a sharp pattern of no allele sharing between species and between *O. cyanura* lineages. In particular, the nuclear gene *cmos* presents four unique haplotypes, each separated from neighbouring haplotypes by one mutational step. The two lineages of *O. cyanura* have one private haplotype each and the remaining two haplotypes are found in *O. jayakari* (Figure 5.5). The *mc1r* is more variable and presents 10 unique haplotypes, with a shallow level of diversity within *O. jayakari*, with four neighbouring haplotypes separated by one mutation. The two lineages of *O. cyanura*, North and South, are recovered with three mutational steps between them and three haplotypes each (Figure 5.5).

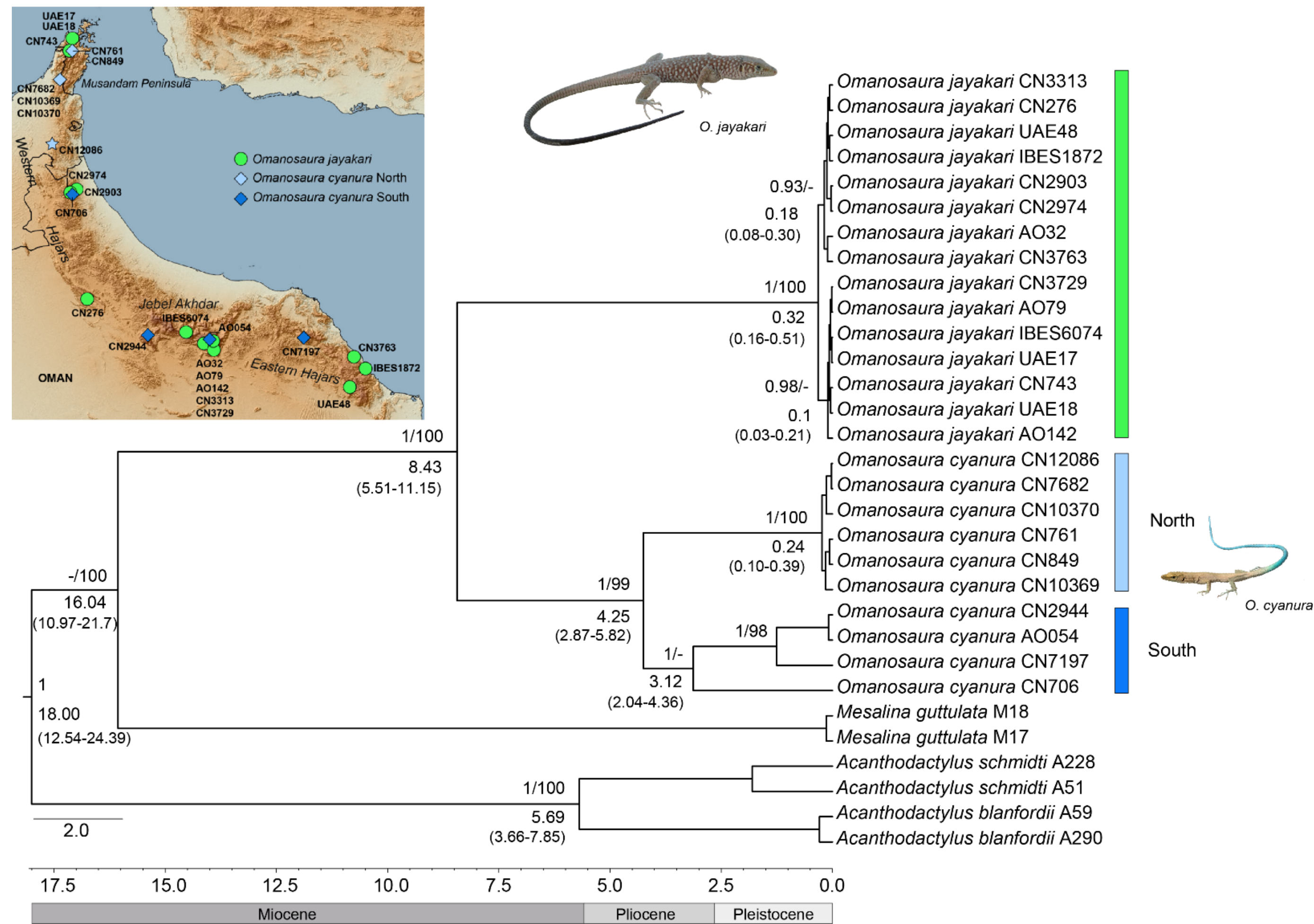


Figure 5.2. Bayesian tree topology of *Omanosaura* inferred using mtDNA (*12S*, *cytb* and *nd4*) and nucDNA (*cmos* and *mc1r*) sequences. Bayesian Posterior Probability (left) and maximum likelihood Bootstrap Support (right) values are represented above nodes. Age estimates are represented below the relevant nodes and include the mean and, between brackets, the HPD 95% confidence interval (in million years). Photos of *O. jayakari* by Salvador Carranza and of *O. cyanura* by Roberto Sindaco.

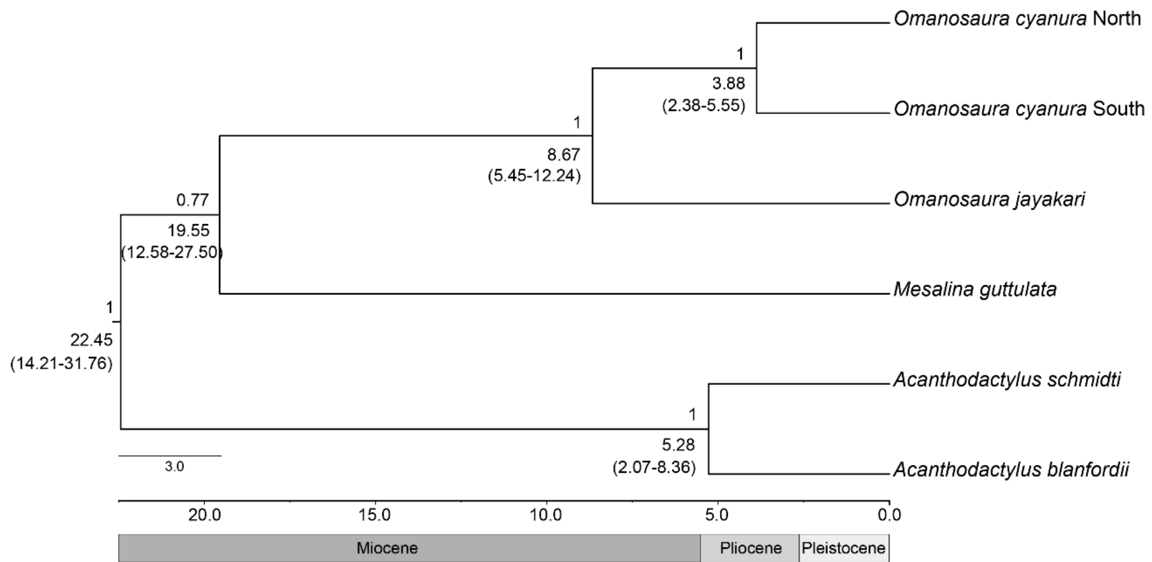


Figure 5.3. Species tree of *Omanosaura* lizards inferred from mitochondrial (*12S*, *cytb* and *nd4*) and nuclear (*cmos* and *mc1r*) sequences using the multispecies coalescent approach. Bayesian Posterior Probabilities are represented above nodes and the age estimates are represented below nodes and include the mean and, between brackets, the HPD 95% confidence interval (in million years).

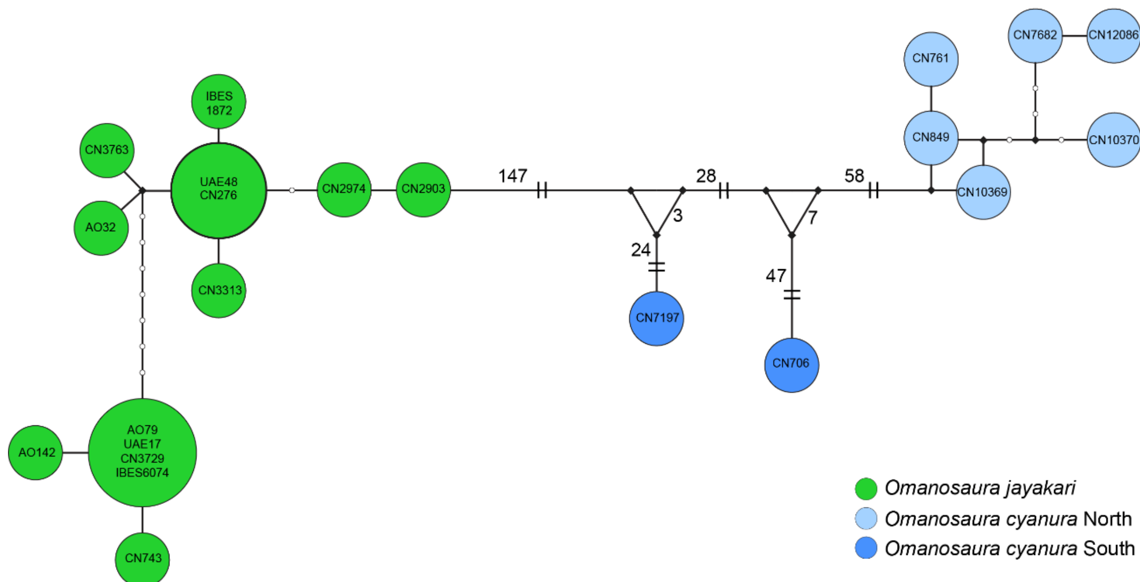


Figure 5.4. Median-joining haplotype networks for *Omanosaura* species inferred from concatenated mtDNA (*12S*, *cytb* and *nd4*). Circles represent different haplotypes and the size is proportional to sample frequency. Small diamonds represent median vectors. Mutations are represented by black lines, separated by white circles, or represented by a number next to the line, if presenting high numbers of mutational steps.

DISCUSSION

Our molecular assessment provides a clear phylogenetic and biogeographic pattern for the *Omanosaura* species and allows the identification of a sharp intraspecific partition of the observed genetic diversity within *O. cyanura*. Overall inter- and intra-specific phylogenetic patterns are congruent between mtDNA and nucDNA, highlighting the importance of including fast evolving nuclear genes to obtain a robust multi-locus phylogeography below the genus level, in this case within *Omanosaura.*

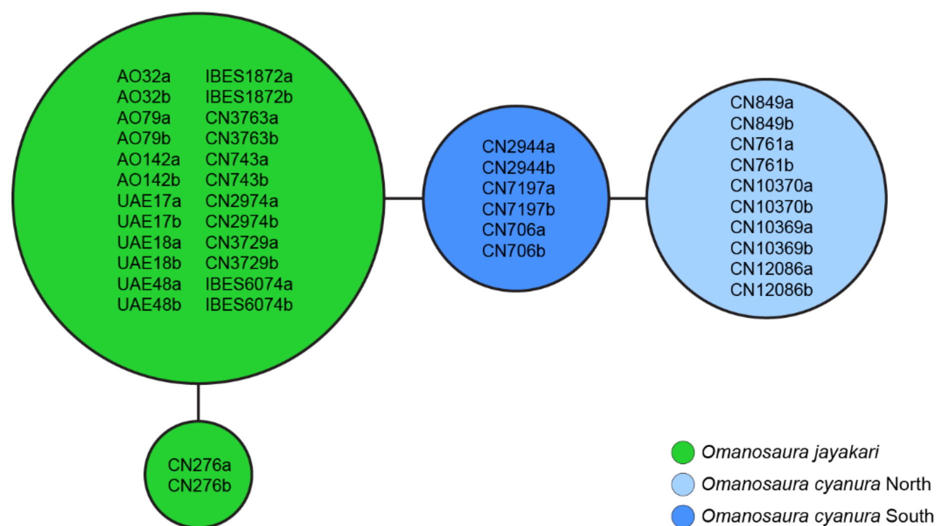
The divergence time between the two *Omanosaura* species estimated in this study, approximately 8 Ma, is much more recent than the estimated 35 Ma from a recent squamate time-tree by Zheng & Wiens (2016). Such difference is not limited to this study and seems to present a general discordance between divergence time estimates obtained in the squamate time-tree and in studies focused on terminal groups of squamates such as lacertids (Kapli et al., 2015; Mendes et al., 2016; Tamar, Carranza, et al., 2016), amphisbaenians (Albert, Zardoya, & Garcia-Paris, 2007; Sampaio, Harris, Perera, & Salvi, 2014), geckos (Carranza & Arnold, 2012; Gamble et al., 2011; Smíd et al., 2013), skinks (Carranza, Arnold, Geniez, Roca, & Mateo, 2008; Pereira & Schrago, 2017) and snakes (Chen, Lemmon, Lemmon, Pyron, & Burbrink, 2017; Daza, Smith, Páez, & Parkinson, 2009). This discordance might be attributed to the differences in the calibration methods – fossils of higher taxa in Zheng & Wiens (2016) *versus* calibration based on rates or recent nodes in other studies. It might also be caused by the methodological approximations in phylogenetic and divergence time estimates required by analyses with thousands of taxa as the supermatrix approach applied by Zheng & Wiens (2016) (see Mendes et al., (2016) for a comparison between phylogenetic estimates based on the supermatrix and the multispecies coalescent approaches in lacertids). The tendency to over estimates of divergence times by Zheng & Wiens (2016) is exemplified by the observation that, according to their study, the divergence within the endemic Canary Islands' *Gallotia* is older than the appearance of the Canary Islands.

Despite sharing the distribution range and being found in sympatry, the two morphologically distinct species of *Omanosaura* present contrasting levels of intraspecific genetic differentiation (Figures 5.2, 5.4 and 5.5, Table 5.3) in both the mtDNA (Figures 5.2, 5.4, and S5.2, Table 5.3) and in the nucDNA genes *cmos* and *mc1r* (Figures 5.2 and 5.5). The much larger and robust *O. jayakari* presents fairly homogeneous low levels of genetic variation across its distribution range, with two very recent clades recovered by the mtDNA data, which diverged by the end of the Middle Pleistocene, around 0.3 Ma. Such low levels of genetic diversity could tentatively be

related to the body size and the ecology of *O. jayakari* (e.g. Gaston & Blackburn, 1996; Meiri, 2008). The relatively large size of these lizards, together with a generalist feeding and habitat traits (Gardner, 2013) may confer them with a higher dispersal ability, hence making them less susceptible to mild historic climatic variation and thus to structure in distinct phylogeographic groups. In contrast, the small and delicately-built *O. cyanura* presents two very well differentiated lineages that diverged in the Pliocene, more than 4.25 Ma. The mtDNA genetic distance between *O. cyanura* lineages is very high (*12S*: 5%, *cytb*: 12%, *nd4*: 9%; Table 5.3). Similar levels of genetic distances have been reported within the Eremiadini species *Acanthodactylus schreiberi* (*12S* and *cytb*; Tamar et al., 2014), which is considered a species complex. Unfortunately, there are not sufficient museum vouchers or photographical material to properly assess morphological differentiation between the two *O. cyanura* lineages, especially from the southern lineage. The inclusion of a genetic sample from the type locality in Wadi Shawkah, UAE (see Figure 5.1; Arnold, 1972) shows that the holotype of *O. cyanura* is likely to belong to the northern lineage. Both paratypes fall within the distribution range of the northern lineage, one (BM1972.710) is from Wadi Qidah (Qadah in GoogleEarth), near Qasab (26.18N 56.22E) and the other (BM1971.1292) is from Wadi Siji, near Masafi (25.31N 56.15E). Although the paratypes could not be genetically assigned to any of the two lineages, their distribution indicates that, similarly to the holotype, they belong to the northern lineage of *O. cyanura*. Therefore, in a future taxonomical assessment of this species, the northern lineage would retain the name *O. cyanura*. However, in order to perform an integrative and comprehensive taxonomic assessment of this species, it will be necessary to collect additional morphological and ecological data across the distribution range of *O. cyanura* in the Hajar Mountains, including other areas within the geographical gap between the two lineages.

The distribution of the two lineages of *O. cyanura* present a clear geographical association with different regions within the Hajar Mountains, with a lineage mostly restricted to the northernmost part of the Hajar Mountains and a southern lineage occurring in the remaining part of the western Hajars, the Jebel Akhdar and the eastern Hajars (Figure 5.2). The northern lineage shows little genetic structure, with a very recent diversification starting around 0.24 Ma, whereas the southern lineage shows higher levels of genetic diversity, with an ancient split (3.12 Ma) between the northernmost sample (CN706) of the southern lineage and the samples located in the Jebel Akhdar and the eastern Hajars (Figure 5.2). The inclusion of more samples in between these sub-lineages of the southern *O. cyanura* lineage, would allow to better locate the geographic break between them.

cmos



mc1r

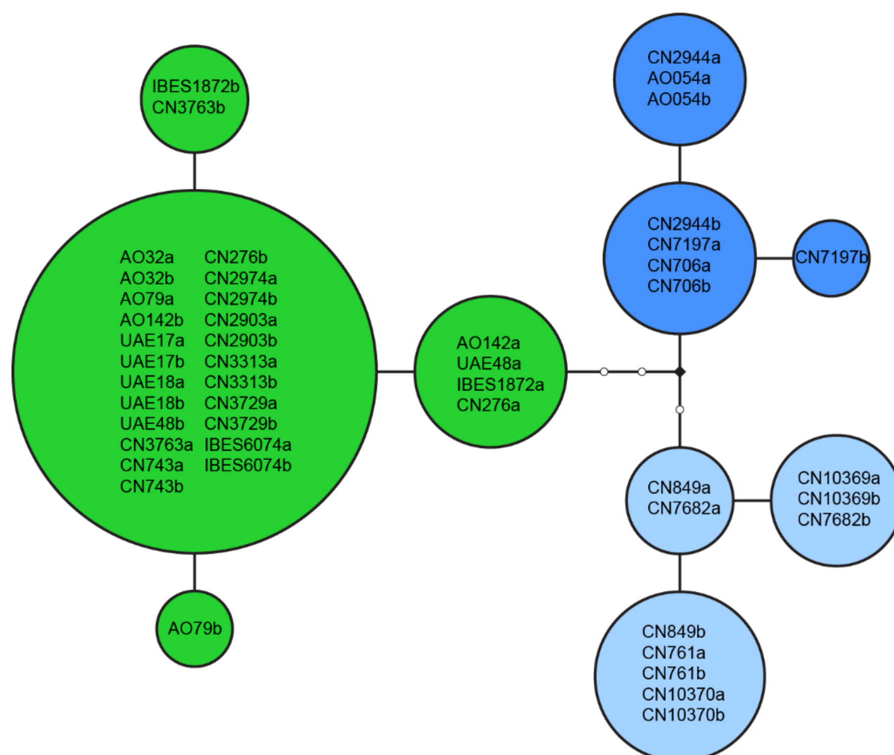


Figure 5.5. Median-joining haplotype networks for *Omanosaura* species inferred from nucDNA *cmos* and *mc1r*. Circles represent different haplotypes and the size is proportional to sample frequency. Alleles are represented by the letters “a” and “b” after the sample code. Small diamonds represent median vectors, and mutations are represented by the black lines, separated by white circles.

The main phylogeographic partition between the different lineages observed within *O. cyanura* strictly matches the biogeographic pattern recovered by previous studies on reptiles of the genera *Asaccus* (Carranza et al., 2016; Simó-Riudalbas et al., 2017), *Trachydactylus* (de Pous et al., 2016) and *Ptyodactylus* (Metallinou et al., 2015; Simó-Riudalbas, Metallinou, et al., 2017). Historically, knowledge of biodiversity patterns was not equally distributed across the Hajar Mountains. The Jebel Akhdar, which includes the highest peaks in the Hajar Mountains, and the eastern Hajars have received more scientific attention over the last decades and have recently been shown to be areas with maximum genetic diversity for the reptile genera *Asaccus* (Simó-Riudalbas, Tarroso, et al., 2017), *Hemidactylus* (Carranza & Arnold, 2012), *Pristurus* Rüppell, 1835 (Badiane et al., 2014) and *Trachydactylus* (de Pous et al., 2016). Comparatively, less research effort has been devoted to the northernmost part of the Hajar Mountains and especially to the Musandam Peninsula until very recently, when the discovery of high genetic diversity and new species of other reptile genera such as *Ptyodactylus* (Metallinou et al., 2015; Simó-Riudalbas, Metallinou, et al., 2017) and *Asaccus* (Carranza et al., 2016) highlighted this area as a hotspot of diversity within the Hajar Mountains. The processes that have shaped the distribution patterns of taxa in the Hajar Mountains remain unclear, although a relation with geological and climatic events could be suggested. The Hajar Mountains have had a complex geological history dating back to 300 Ma, but the uplift into a mountain range started around 30 Ma with the opening of the Gulf of Aden (Bosworth, Huchon, & McClay, 2005; Glennie, 2006; Laughton, 1966), and it probably rose to a high mountain range in the last 4 to 6 Ma, during an intense plate tectonic phase that affected Oman (Glennie, 2006). The diversification of the *Omanosaura* species might have been influenced by the uplift of the mountain range, particularly the divergence between the two *O. cyanura* lineages, which seems to coincide with the most recent part of the final uplift of these mountains, around 4 Ma. Climatic events might also have influenced the diversification of *Omanosaura* lineages, since the high elevation areas of the Hajar Mountains, which harbour the highest levels of genetic diversity of *O. cyanura* among many other taxa – the Musandam Peninsula and the Jebel Akhdar massif – could have acted as refuge and centre of diversification for these reptile groups during periods of climatic changes.

The distribution of the two *Omanosaura* species as currently assessed by the IUCN Red List of Threatened Species (Soorae, Wilms, & Al Rasbi, 2012) is more restricted than our sampling locations and records for *Omanosaura* shown by Gardner (2013) and Carranza et al. (2017) (*O. cyanura* CN706 and *O. jayakari* UAE17, UAE18, CN743 and CN276; Figure 5.1), and reflect record gaps, rather than true absence

(Gardner, 2013). This can be associated with (i) the low number of expeditions to the Hajar Mountains until very recently and the inaccessibility of some of the areas, (ii) the lack of scientific attention given to this genus, and also (iii) with the general elusive nature of *O. cyanura*, which makes it particularly difficult to see and capture. Further studies are needed to complete the diversity inventory of this region and to better understand the association of the current biodiversity patterns with past geologic and climatic events. Finally, the results of this study highlight the importance of biodiversity assessments in mountainous regions characterized by high endemicity but which are difficult to access.

ACKNOWLEDGMENTS

We wish to thank Marc Simó-Riudalbas, Margarita Metallinou, Jiri Smíd, Raquel Vasconcelos, Roberto Sindaco, Sithum Jayasinghe, Philip de Pous, Fèlix Amat, Ali Alghafri, Sultan Khalifa, Hamed Al Furkani, for assisting in sample collection in the field. Special thanks are due to Saleh Al Saadi, Mohammed Al Shariani, Ali Al Kiyumi, Mohammed Abdullah Al Maharmi and the other members of the Nature Conservation Department of the Ministry of Environment and Climate, Sultanate of Oman and to His Highness Sheikh Dr. Sultan bin Mohammed Al Qasimi, Supreme Council Member and Ruler of Sharjah and Ms. Hana Al Suwaidi (Chairperson of the Environment and Protected Areas Authority, Sharjah) for their help and support. Specimens were collected and manipulated with the authorization and under control and permission of the governments of Oman (Ministry of Environment and Climate Affairs, MECA) and the United Arab Emirates (Environment and Protected Areas Authority, Government of Sharjah). Specimens were captured and processed following the guidelines and protocols stated in the collecting permits and agreements obtained from the competent authorities (see references below). Members of the government supervised collecting activities. All efforts were made to minimize animal suffering. All the necessary collecting and export permits for this study in Oman were issued by the Nature Conservation Department of the Ministry of Environment and Climate Affairs, Oman (Refs: 08/2005; 16/2008; 38/2010; 12/2011; 13/2013; 21/2013) and the research in the United Arab Emirates was done under the supervision and permission of the Environment and Protected Areas Authority, Government of Sharjah. JM and DJH are supported by the Fundação para a Ciência e a Tecnologia (FCT, Portugal): JM, doctoral grant SFRH/BD/81528/2011; DJH IF contract 01627/2014. DS is currently supported by the program 'Rita Levi Montalcini' for the recruitment of young researchers at the University

of L'Aquila. SC was supported by the Ministerio de Economía y Competitividad, Spain (co-funded by FEDER) under grant number CGL2015-70390-P; Ministry of Environment and Climate Affairs under Grant number 22412027; Secretaria d'Universitats i Recerca del Departament d'Economia i Coneixement de la Generalitat de Catalunya under grant number 2014-SGR-1532.

REFERENCES

Albert, E. M., Zardoya, R., & Garcia-Paris, M. (2007). Phylogeographical and speciation patterns in subterranean worm lizards of the genus *Blanus* (Amphisbaenia: Blanidae). *Molecular Ecology*, 16, 1519–1531. doi:10.1111/j.1365-294X.2007.03248.x

Arnold, E. N. (1972). Lizards with northern afinities from the mountains of Oman. *Zoologische Mededelingen*, 47, 111–128.

Arnold, E. N. (1973). Relationships of the Palaearctic lizards assigned to the genera *Lacerta*, *Algyroides* and *Psammodromus* (Reptilia: Lacertidae). *Bulletin of the British Museum*, 25, 289–366.

Arnold, E. N. (2009). Relationships, evolution and biogeography of Semaphore geckos, *Pristurus* (Squamata, Sphaerodactylidae) based on morphology. *Zootaxa*, 2060, 1–21.

Arnold, E. N., Arribas, O., & Carranza, S. (2007). Systematics of the Palaearctic and Oriental lizard tribe Lacertini (Squamata: Lacertidae: Lacertinae), with descriptions of eight new genera. Magnolia Press. *Zootaxa*, 1430, 1–86.

Babocsay, G. (2004). A new species of saw-scaled viper of the *Echis coloratus* complex (Ophidia: Viperidae) from Oman, Eastern Arabia. *Systematics and Biodiversity*, 1, 503–514. doi:10.1017/S1477200003001294

Badiane, A., Garcia-Porta, J., Červenka, J., Kratochvíl, L., Sindaco, R., Robinson, M. D., Morales, H., Mazuch, T., Price, T., Amat, F., Shobrak, M. Y., Wilms, T., Simó-Riudalbas, M., Ahmadzadeh, F., Papenfuss, T. J., Cluchier, A., Viglione, J. & Carranza, S (2014). Phylogenetic relationships of Semaphore geckos (Squamata:

Sphaerodactylidae: *Pristurus*) with an assessment of the taxonomy of *Pristurus rupestris*. Zootaxa, 3835, 33-58. doi:10.11646/zootaxa.3835.1.2

Barata, M., Carranza, S., & Harris, D. (2012). Extreme genetic diversity in the lizard *Atlantolacerta andreanskyi* (Werner, 1929): A montane cryptic species complex. *BMC Evolutionary Biology*, 12, 167. doi:10.1186/1471-2148-12-167

Bosworth, W., Huchon, P., & McClay, K. (2005). The Red Sea and Gulf of Aden Basins. *Journal of African Earth Sciences*, 43, 334–378. doi:10.1016/j.jafrearsci.2005.07.020

Brito, J. C., Godinho, R., Martínez-Freiría, F., Pleguezuelos, J. M., Rebelo, H., Santos, X., Vale, C. G., Velo-Antón, G., Boratyński, Z., Carvalho, S. B., Ferreira, S., Gonçalves, D. V., Silva, T. L., Tarroso, P., Campos, J. C., Leite, J. V., Nogueira, J., Álvares, F., Sillero, N., Sow, A. S., Fahd, S., Crochet, P.- A. & Carranza, S. (2014). Unravelling biodiversity, evolution and threats to conservation in the SaharaSahel. *Biological Reviews*, 89, 215–231. doi:10.1111/brv.12049

Bruen, T. C., Philippe, H., & Bryant, D. (2006). A simple and robust statistical test for detecting the presence of recombination. *Genetics*, 172, 2665–2681. doi:10.1534/genetics.105.048975

Carranza, S., & Arnold, E. N. (2012). A review of the geckos of the genus *Hemidactylus* (Squamata: Gekkonidae) from Oman based on morphology, mitochondrial and nuclear data, with descriptions of eight new species. Magnolia Press. *Zootaxa*, 95, 1–95.

Carranza, S., Arnold, E. N., Geniez, P., Roca, J., & Mateo, J. A. (2008). Radiation, multiple dispersal and parallelism in the skinks, *Chalcides* and *Sphenops* (Squamata: Scincidae), with comments on *Scincus* and *Scincopus* and the age of the Sahara Desert. *Molecular Phylogenetics and Evolution*, 46, 1071–1094. doi:10.1016/j.ympev.2007.11.018

Carranza, S., Simó-Riudalbas, M., Jayasinghe, S., Wilms, T., & Els, J. (2016). Microendemicity in the northern Hajar Mountains of Oman and the United Arab Emirates with the description of two new species of geckos of the genus *Asaccus* (Squamata: Phyllodactylidae). *PeerJ*, *4*, e2371. doi:10.7717/peerj.2371

Carranza, S., Xipell, M., Tarroso, P., Gardner, D., Arnold, E. N., Robinson, M. D., Simó-Riudalbas, M., Vasconcelos, R., de Pous, P., Amat, P., Smíd, J., Sindaco, R., Metallinou, M., Els, J., Pleguezuelos, J. M., Machado, L., Donaire, L., Martínez, G.,

Garcia-Porta, J., Mazuch, T., Wilms, T., Gebhart, J., Aznar, J., Gallego, J., Zwanzig, B. M., Fernández-Guiberteau, D., Papenfuss, T., Al Saadi, S., Alghafri, S., Khalifa, S., Al Fargani, H., Bilal, S. B., Al Adhoobi, A. S., Al Omairi, Z. S., Al Shariani, M., Al Kiyumi, A., Al Sariri, T., Al Shukaili, A. S.& Al Akhzami, S. N. (2017). Diversity, distribution and conservation of the terrestrial reptiles of Oman (Sauropsida, Squamata). *PloS One*, in press.

Chen, X., Lemmon, A. R., Lemmon, E. M., Pyron, R. A., & Burbrink, F. T. (2017). Using phylogenomics to understand the link between biogeographic origins and regional diversification in ratsnakes. *Molecular Phylogenetics and Evolution*, 111, 206–218. doi:10.1016/j.ympev.2017.03.017

Daza, J. M., Smith, E. N., Páez, V. P., & Parkinson, C. L. (2009). Complex evolution in the Neotropics: The origin and diversification of the widespread genus *Leptodeira* (Serpentes: Colubridae). *Molecular Phylogenetics and Evolution*, 53, 653–667. doi:10.1016/j.ympev.2009.07.022

de Pous, P., Machado, L., Metallinou, M., Červenka, J., Kratochvíl, L., Paschou, N., Mazuch, T., Šmíd, J., Simó-Riudalbas, M., Sanuy, D. & Carranza, S. (2016). Taxonomy and biogeography of *Bunopus spatalurus* (Reptilia; Gekkonidae) from the Arabian Peninsula. *Journal of Zoological Systematics and Evolutionary Research*, 54, 67–81. doi:10.1111/jzs.12107

Drummond, A. J., & Rambaut, A. (2007). BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology*, 7, 214. doi:10.1186/1471-2148-7-214

Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, 32, 1792–1797. doi:10.1093/nar/gkh340

Edgell, H. S. (2006). *Arabian Deserts: Nature, Origin and Evolution*. Dordrecht: Springer.

Ficetola, G. F., Bonardi, A., Sindaco, R., & Padoa-Schioppa, E. (2013). Estimating patterns of reptile biodiversity in remote regions. *Journal of Biogeography*, 40, 1202–1211. doi:10.1111/jbi.12060

Flot, J.-F. (2010). SEQPHASE: a web tool for interconverting PHASE input/output files and FASTA sequence alignments. *Molecular Ecology Resources*, 10, 162–166. doi:10.1111/j.1755-0998.2009.02732.x

Gamble, T., Bauer, A. M., Colli, G. R., Greenbaum, E., Jackman, T. R., Vitt, L. J., & Simons, A. M. (2011). Coming to America: Multiple origins of New World geckos.

Journal of Evolutionary Biology, 24, 231–244. doi:10.1111/j.1420-9101.2010.02184.x

Garcia-Porta, J., Simó-Riudalbas, M., Robinson, M., & Carranza, S. (2016). Diversification in arid mountains: biogeography and cryptic diversity of *Pristurus rupestris rupestris* in Arabia. *Journal of Biogeography*, 1694–1704. doi:10.1111/jbi.12929

Gardner, A. S. (2013). *The Amphibians and Reptiles of Oman and the UAE*. Frankfurt: Edition Chimaira.

Gaston, K. J., & Blackburn, T. M. (1996). Range size-body size relationships: Evidence of scale dependence. *Oikos*, 75, 479–485. doi:10.2307/3545889

Glennie, K. (2006). *Oman's Geological Heritage*. London: Stacey Internationl Publishers.

Greenbaum, E., Villanueva, C. O., Kusamba, C., Aristote, M. M., & Branch, W. R. (2011). A molecular phylogeny of Equatorial African Lacertidae, with the description of a new genus and species from eastern Democratic Republic of the Congo. *Zoological Journal of the Linnean Society*, 163, 913–942. doi:10.1111/j.1096-3642.2011.00732.x

Harris, D. J., Arnold, E. N., & Thomas, R. H. (1998). Relationships of lacertid lizards (Reptilia: Lacertidae) estimated from mitochondrial DNA sequences and morphology. *Proceedings of the Royal Society B Biological Sciences*, 265, 1939–1948. doi:10.1098/rspb.1998.0524

Heled, J., & Drummond, A. J. (2010). Bayesian inference of species trees from multilocus data. *Molecular Biology and Evolution*, 27, 570–580. doi:10.1093/molbev/msp274

Hewitt, G. M. (2011). Mediterranean Peninsulas: The Evolution of Hotspots. In F. E. Zachos & J. C. Habel (Eds.), *Biodiversity Hotspots: Distribution and Protection of Conservation Priority Areas* (pp. 123–147). Berlin, Heidelberg: Springer Berlin Heidelberg. doi:10.1007/978-3-642-20992-5_7

Huson, D. H., & Bryant, D. (2006). Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution*, 23, 254–267. doi:10.1093/molbev/msj030

Kapli, P., Lymberakis, P., Crochet, P. -A., Geniez, P., Brito, J. C., Almutairi, M., Ahmadzadeh, F., Schmitz, A., Wilms, T., Pouyani, N. R. & Poulakakis, N (2015). Historical biogeography of the lacertid lizard *Mesalina* in North Africa and the Middle East. *Journal of Biogeography*, 42, 267–279. doi:10.1111/jbi.12420

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P. & Drummond, A. (2012). Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, 28, 1647–1649. doi:10.1093/bioinformatics/bts199

Laughton, A. S. (1966). The Gulf of Aden. *Philosophical Transactions of the Royal Society of London Series A, Mathematical and Physical Sciences*, 259, 150–171.

Lutz, D., Bischoff, W., & Mayer, W. (1986). Chemosystematische untersuchungen zur stellung von *Lacerta jayakari* Boulenger, 1887 sowie der gattungen *Gallotia* Boulenger und *Psammodromus* Fitzinger (Sauria; Lacertidae). *Journal of Zoological Systematics and Evolutionary Research*, 24, 144–157.

MacLaren, C. A. (2016). Climate change drives decline of *Juniperus seravschanica* in Oman. *Journal of Arid Environments*, 128, 91–100. doi:10.1016/j.jaridenv.2016.02.001

Mandaville, J. (1977). The scientific results of the Oman flora and fauna survey 1975. Plants. *Journal Oman Studies Special Report*, 1, 229-267.

Mayer, W., & Benyr, G. (1994). Albumin-evolution und phylogenese in der Familie Lacertidae (Reptilia: Sauria). *Annalen Des Naturhistorischen Museums in Wien*, 96, 621–648.

Mayer, W., & Bischoff, W. (1996). Beiträge zur taxonomischen revision der gattung *Lacerta* (Reptilia: Lacertidae) Teil 1: *Zootoca*, *Omanosaura*, *Timon* und *Teira* als eigenstandige Gattungen. *Salamandra*, 32, 163–170.

Mayer, W., & Pavlicev, M. (2007). The phylogeny of the family Lacertidae (Reptilia) based on nuclear DNA sequences: convergent adaptations to arid habitats within the subfamily Eremiainae. *Molecular Phylogenetics and Evolution*, 44, 1155–63. doi:10.1016/j.ympev.2007.05.015

Meiri, S. (2008). Evolution and ecology of lizard body sizes. *Global Ecology and Biogeography*, 17, 724–734. doi:10.1111/j.1466-8238.2008.00414.x

Mendes, J., Harris, D. J., Carranza, S., & Salvi, D. (2016). Evaluating the phylogenetic signal limit from mitogenomes, slow evolving nuclear genes, and the concatenation approach. New insights into the Lacertini radiation using fast evolving nuclear genes and species trees. *Molecular Phylogenetics and Evolution*, 100, 254–267.

doi:10.1016/j.ympev.2016.04.016

Mendes, J., Harris, D. J., Carranza, S., & Salvi, D. (2017). Biogeographical crossroad across the Pillars of Hercules: Evolutionary history of *Psammodromus* lizards in space and time. *Journal of Biogeography*, 44, 2877–2890. doi:10.1111/jbi.13100

Metallinou, M., Červenka, J., Crochet, P.-A., Kratochvíl, L., Wilms, T., Geniez, P., Shobrak, M. Y., Brito, J. C. & Carranza, S. (2015). Species on the rocks: Systematics and biogeography of the rock-dwelling *Ptyodactylus* geckos (Squamata: Phyllodactylidae) in North Africa and Arabia. *Molecular Phylogenetics and Evolution*, 85, 208–220. doi:10.1016/j.ympev.2015.02.010

Miller, M. A., Pfeiffer, W., & Schwartz, T. (2010). Creating the CIPRES Science Gateway for inference of large phylogenetic trees. In *2010 Gateway Computing Environments Workshop (GCE)* (pp. 1–8). doi:10.1109/GCE.2010.5676129

Pereira, A. G., & Schrago, C. G. (2017). Arrival and diversification of mabuyine skinks (Squamata: Scincidae) in the Neotropics based on a fossil-calibrated timetree. *PeerJ*, 5, e3194. doi:10.7717/peerj.3194

Posada, D. (2008). jModelTest: phylogenetic model averaging. *Molecular Biology and Evolution*, 25, 1253–1256. doi:10.1093/molbev/msn083

Pyron, R., Burbrink, F. T., & Wiens, J. J. (2013). A phylogeny and revised classification of Squamata, including 4161 species of lizards and snakes. *BMC Evolutionary Biology*, 13, 93. doi:10.1186/1471-2148-13-93

Rambaut, A., & Drummond, A. (2007). Tracer v1.4, Available from http://beast.bio.ed.ac.uk/Tracer.

Rato, C., Harris, D. J., Carranza, S., Machado, L., & Perera, A. (2016). The taxonomy of the *Tarentola mauritanica* species complex (Gekkota: Phyllodactylidae): Bayesian species delimitation supports six candidate species. *Molecular Phylogenetics and Evolution*, 94, 271–278. doi:10.1016/j.ympev.2015.09.008

Robinson, M., Carranza, S., & Arnold, N. (2009). A preliminary analysis of phylogenetic relationships and biogeography of the dangerously venomous Carpet Vipers, *Echis* (Squamata, Serpentes, Viperidae) based on mitochondrial DNA sequences. *Amphibia-Reptilia*, 30, 273–282. doi:10.1163/156853809788201090

Rosado, D., Rato, C., Salvi, D., & Harris, D. J. (2017). Evolutionary history of the Morocco lizard-fingered geckos of the *Saurodactylus brosseti* complex. *Evolutionary Biology*,

44, 386-400. doi:10.1007/s11692-017-9417-8

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

Salvi, D., Lucente, D., Mendes, J., Liuzzi, C., Harris, D. J., & Bologna, M. A. (2017). Diversity and distribution of the Italian Aesculapian snake *Zamenis lineatus*: A phylogeographic assessment with implications for conservation. *Journal of Zoological Systematics and Evolutionary Research*, 55, 222–237. doi:10.1111/jzs.12167

Salvi, D., Schembri, P. J., Sciberras, A., & Harris, D. J. (2014). Evolutionary history of the Maltese wall lizard *Podarcis filfolensis*: insights on the "Expansion-Contraction" model of Pleistocene biogeography. *Molecular Ecology*, 23, 1167–1187. doi:10.1111/mec.12668

Sambrook, J., Fritsch, E. F., & Maniatis T. (1989). *Molecular Cloning: A Laboratory Manual*. New York: Cold Spring Harbor Press.

Sampaio, F. L., Harris, D. J., Perera, A., & Salvi, D. (2014). Phylogenetic and diversity patterns of *Blanus* worm lizards (Squamata: Amphisbaenia): Insights from mitochondrial and nuclear gene genealogies and species tree. *Journal of Zoological Systematics and Evolutionary Research*, 53, 45-54. doi:10.1111/jzs.12075

Silvestro, D., & Michalak, I. (2012). raxmlGUI: a graphical front-end for RAxML. *Organisms Diversity & Evolution*, 12, 335–337. doi:10.1007/s13127-011-0056-0

Simó-Riudalbas, M., Metallinou, M., de Pous, P., Els, J., Jayasinghe, S., Péntek-Zakar, E., Wilms, T., Al-Saadi, S. & Carranza, S. (2017). Cryptic diversity in *Ptyodactylus* (Reptilia: Gekkonidae) from the northern Hajar Mountains of Oman and the United Arab Emirates uncovered by an integrative taxonomic approach. *PloS One*, 12, e0180397. doi:10.1371/journal.pone.0180397

Simó-Riudalbas, M., Tarroso, P., Papenfuss, T., Al-Sariri, T., & Carranza, S. (2017). Systematics, biogeography and evolution of *Asaccus gallagheri* (Squamata, Phyllodactylidae) with the description of a new endemic species from Oman. *Systematics and Biodiversity*, in press. doi:10.1080/14772000.2017.1403496

Sindaco, R., & Jeremcenko, V. K. (2008). *The reptiles of the Western Palearctic. 1. Annotated checklist and distributional atlas of the turtles, crocodiles,*

amphisbaenians and lizards of Europe, North Africa, Middle East and Central Asia. Sindaco R, Jeremcenko VK. The reptiles of the Western Pal. Italy: Societas Herpetologica Italica.

Sindaco, R., Venchi, A., & Grieco, C. (2011). The southern limit of *Omanosaura* lizards (Squamata: Lacertidae). *Herpetology Notes*, 4, 171–172.

Šmíd, J., Carranza, S., Kratochvíl, L., Gvoždík, V., Nasher, A. K., & Moravec, J. (2013). Out of Arabia: a complex biogeographic history of multiple vicariance and dispersal events in the gecko genus *Hemidactylus* (Reptilia: Gekkonidae). *PloS One*, 8, e64018. doi:10.1371/journal.pone.0064018

Soorae, P., Wilms, T., & Al Rasbi, K. J. M. (2012). *Omanosaura cyanura*. The IUCN Red List of Threatened Species 2012.

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

Stephens, M., & Scheet, P. (2005). Accounting for decay of linkage disequilibrium in haplotype inference and missing-data imputation. *American Journal of Human Genetics*, 76, 449–462. doi:10.1086/428594

Stephens, M., Smith, N. J., & Donnelly, P. (2001). A new statistical method for haplotype reconstruction from population data. *American Journal of Human Genetics*, 68, 978–89. doi:10.1086/319501

Tamar, K., Carranza, S., Sindaco, R., Moravec, J., & Meiri, S. (2014). Systematics and phylogeography of *Acanthodactylus schreiberi* and its relationships with *Acanthodactylus boskianus* (Reptilia: Squamata: Lacertidae). *Zoological Journal of the Linnean Society*, 172, 720–739. doi:10.1111/zoj.12170

Tamar, K., Carranza, S., Sindaco, R., Moravec, J., Trape, J.-F., & Meiri, S. (2016). Out of Africa: Phylogeny and biogeography of the widespread genus *Acanthodactylus* (Reptilia: Lacertidae). *Molecular Phylogenetics and Evolution*, 103, 6–18. doi:10.1016/j.ympev.2016.07.003

Tamar, K., Geniez, P., Brito, J. C., & Crochet, P. A. (2017). Systematic revision of *Acanthodactylus busacki* (Squamata: Lacertidae) with a description of a new species from Morocco. *Zootaxa*, 4276, 357–386. doi:10.11646/zootaxa.4276.3.3

Tamar, K., Scholz, S., Crochet, P.-A., Geniez, P., Meiri, S., Schmitz, A., Wilms, T. & Carranza, S. (2016). Evolution around the Red Sea: Systematics and biogeography

of the agamid genus *Pseudotrapelus* (Squamata: Agamidae) from North Africa and Arabia. *Molecular Phylogenetics and Evolution*, 97, 55–68. doi:10.1016/j.ympev.2015.12.021

Tamura, K., Stecher, G., Peterson, D., Filipski, A., & Kumar, S. (2013). MEGA6: Molecular evolutionary genetics analysis version 6.0. *Molecular Biology and Evolution*, 30, 2725–2729. doi:10.1093/molbev/mst197

Zheng, Y., & Wiens, J. J. (2016). Combining phylogenomic and supermatrix approaches, and a time-calibrated phylogeny for squamate reptiles (lizards and snakes) based on 52 genes and 4162 species. *Molecular Phylogenetics and Evolution*, 94, 537–547. doi:10.1016/j.ympev.2015.10.009

Supporting Information

Additional Supporting Information can be found in the Appendices section.

Table 5.1. Primers and PCR conditions.

Figure S5.1. Phylogenetic relationships of *Omanosaura* inferred by Maximum Likelihood.

Figure S5.2. Phylogenetic relationships of *Omanosaura* inferred by Bayesian Inference and Maximum Likelihood using mtDNA sequences.

Figure S5.3. Phylogenetic relationships and divergence time of *Omanosaura* based on time split between *Gallotia caesaris caesaris* and *G. c. gomerae* as described in Carranza & Arnold (2012).

CHAPTER 6

General Discussion

GENERAL DISCUSSION

The main aim of this thesis was to investigate if and how the addition of novel molecular data, in particular fast evolving nuclear markers, and the use of recently developed phylogenetic methods, particularly the species tree based on the multispecies coalescent approach, could improve the estimate of phylogenetic relationships within selected squamate groups covering different systematic levels, particularly:

i. Test if the application of nuclear genes and species tree approach could solve the basal relationships of the Lacertini tribe and discriminate between contrasting proposed phylogenetic hypotheses;

ii. Test if the application of fast evolving nuclear genes and species tree approach could solve the intrageneric relationships and the monophyly of the colubrid genus *Zamenis* and related ratsnakes;

iii. Frame the phylogeographical history of the Gallotiinae genus *Psammodromus* within the Strait of Gibraltar geodynamics with the application of fast evolving nuclear genes and species tree;

iv. Assess the pattern of phyletic diversification between and within *Omanosaura* species based on mitochondrial and nuclear genes.

The studies developed and presented in this thesis have demonstrated the positive impact of the integrative analyses implementing the coalescent species tree phylogenetic tool and the fast evolving nuclear markers on the phylogenetic inference of these different squamate groups. They have also shed some light on the differences between the inference based on (i) the use of mitochondrial DNA, together with slow nuclear genes, and the use of faster evolving nuclear genes and (ii) the application of the phylogenetic methods of concatenation and species tree. The phylogenetic inferences presented in the different chapters of this thesis are the result of the combined use of datasets with comprehensive taxon sampling, including sequences from both mitochondrial DNA, slow and fast evolving nuclear genes (analysed separately or combined) and are based on the application of the coalescent species tree method, as well as the concatenation approach. Although all of these factors have contributed to the resolution of the proposed phylogenetic questions as a whole, it was possible to assess

the phylogenetic signal of different markers and the resolution power of different approaches used for the phylogenetic inference.

THE IMPORTANCE OF THE CHOICE OF MOLECULAR MARKERS

The role of the mitochondrial and slow and fast evolving nuclear genes in phylogenetic inference is particularly evident in the contrasting results provided by these different molecular markers in Chapter 2, with the Lacertini study case and in Chapter 3, with the colubrid ratsnakes genera *Zamenis* and *Rhinechis*.

In the Lacertini case, the inclusion of five fast evolving nuclear genes in the analyses, comprising two introns and three exons, was more influential for the phylogenetic results than the application of the different phylogenetic methods, including the species tree and the concatenation approach based on Maximum Likelihood and Bayesian Inference. The phylogenetic tree based on mitochondrial DNA (Fig. 2.1) provided little support for the internal branching of the Lacertini tree and did not resolve the position of different clades that have been questioned in previous works, such as the position of the Asian genus *Takydromus* (Harris et al., 1998; Mayer & Pavlicev, 2007; Pavlicev & Mayer, 2009) and the monophyly of the genus *Algyroides* (Pavlicev & Mayer, 2009). On the other hand, the trees based on nuclear DNA (Fig. 2.2) provided new insights into the Lacertini phylogeny, particularly: the genus *Algyroides* was recovered as monophyletic, *Takydromus* is nested within Lacertini and a completely new relationship was recovered between the four monotypic genera *Teira*, *Scelarcis*, *Archaeolacerta* and *Zootoca*, in a strongly supported clade. In previous studies, mainly applying mitochondrial genes, the position of *Archaeolacerta* and *Zootoca* had been highly unstable. This clade and other relationships supported by the nuclear data were corroborated by the analyses of the combined concatenated mitochondrial and nuclear DNA and by the species tree. These results highlight the importance of the choice of molecular markers in phylogenetic analyses, particularly in the presence of a suspected old and fast radiation and with extensive extinctions within old lineages. In the case of Lacertini, the mitochondrial and slow evolving nuclear genes applied in all previous studies (Fu, 1998, 2000; Harris et al., 1998; Mayer & Pavlicev, 2007; Hipsley et al., 2009; Pavlicev & Mayer, 2009; Pyron et al., 2013) partially explain the lack of resolution to recover new clades due to the processes of fast radiation. In this case, the extinction of old lineages or the possible loss of signal in the mitochondrial DNA due to saturation, and the incapacity of slow nuclear genes to discriminate between successive cladogenetic events in a short time interval may explain the low resolution of these markers in the phylogenetic inference. This highlights the importance of the use of fast

nuclear molecular markers in the phylogenetic inference of fast radiations, since they may shed light on basal polytomies when the clustering of internal nodes occurs in a short time span, such as in the case of Lacertini. In the complex evolutionary history of Lacertini, if the addition of only five fast evolving nuclear genes was sufficient to uncover a new clade and provide resolution for the position of some genera, it is possible that adding several hundreds or thousands of these loci through a phylogenomic approach would provide enough resolution to solve the basal cladogenetic events within this lacertid tribe.

In the colubrid ratsnake *Zamenis* study case from Chapter 3, the results from the phylogenetic inference between mitochondrial and nuclear genes allow different conclusions to be drawn. While in the Lacertini tribe the relationships recovered by the nuclear genes are corroborated by the combined mitochondrial and nuclear analyses and by the species tree, in the case of *Zamenis* it is the combination of the mitochondrial and nuclear genes that provides the best estimate of phylogenetic relationships within *Zamenis* and between *Zamenis* and *Rhinechis*. This study case provides a good example of the benefits of the inclusion of both mitochondrial and nuclear markers in phylogenetic inference, since each kind of marker adds resolution at different levels of the phylogenetic tree of *Zamenis* and its allies. While previous phylogenetic studies on *Zamenis*, based on concatenated mitochondrial DNA and the slow evolving nuclear gene *cmos*, presented conflicting results regarding the intrageneric relationships of *Zamenis* and its monophyly (Lenk et al., 2001; Utiger et al., 2002; Burbrink & Lawson, 2007; Pyron et al., 2011, 2013; Figueroa et al., 2016; Zheng & Wiens, 2016), in the study from this thesis, the inclusion of four fast evolving nuclear markers, in addition to the slow evolving nuclear gene *cmos* and two mitochondrial genes, has provided a fully resolved phylogeny.

In the genera *Psammodromus* from Chapter 4 and *Omanosaura* from Chapter 5 the individual analyses from the mitochondrial and nuclear genes resulted in similar relationships within and between the species from these two lacertid genera. This may due to the fact that cladogenetic events within these two lacertid genera have been well spaced in time so that diversification at fast evolving genes, both mitochondrial and nuclear, has followed speciation patterns. The inclusion of both mitochondrial and fast evolving nuclear genes, in both cases, provides a more robust inference and allows to determine if distinct species or lineages acquired reciprocal monophyly (i.e. have complete lineage sorting) at both mitochondrial and nuclear genes.

Overall, and well exemplified in the papers presented in this thesis (Chapters 2 – 5), the phylogenetic inference in general can only benefit from the inclusion of fast evolving nuclear genes in the analyses, particularly if these loci have been shown to be highly variable in the group under study. The presence of nuclear markers can generally overcome the disadvantages of the mitochondrial genes, such as the small size of the mitogenome and the fact that it essentially consists of a single locus and therefore provides a single tree that can be affected by multiple processes, such as selection and introgression, in addition to stochastic sorting (for more detailed information see the section "Mitochondrial and Nuclear data" in the Chapter 1). The use of fast evolving nuclear genes also allows the recovery of cladogenic events in older and fast radiations that are outside the scope of the resolution power of the mitochondrial DNA. On the other hand, more recent events might be better resolved by the mitochondrial DNA and, for this reason, the combination of mitochondrial DNA and fast evolving nuclear genes seems to be, for now, the most appropriate choice of molecular markers for phylogenetic inference.

DIFFERENCES FROM PHYLOGENETIC TOOLS: CONCATENATION AND SPECIES TREE

The species tree approach based on the multispecies coalescent model was applied in all the study cases presented in this thesis (Chapters 2, 3, 4 and 5). In each of these studies, the concatenation or supermatrix method was also applied and directly compared with the results from the species tree. In general, the phylogenetic relationships inferred using species trees and concatenated datasets were similar in all of these squamate groups, with equivalent relationships being recovered by both methods. The results from these particular studies are examples of the conciliation between the most commonly occurring gene tree estimated from multilocus data with the species tree estimated by the coalescent approach. However, the results of this thesis are very different from previous studies of the same squamate groups based on concatenation approaches. In the studies presented here, the congruent results between concatenation and species tree can be credited to the type and number of molecular markers applied in each study, particularly the use of several independent fast evolving nuclear genes in addition to the mitochondrial DNA genes, as explained in the previous section. It is possible to compare between the different phylogenetic methods of the concatenation and species tree approaches in two groups for which previous inference has provided contrasting phylogenetic hypotheses: the Lacertini and *Zamenis*. These groups offer great insights on both the use of different molecular markers and

phylogenetic inference methods due to the complex evolutionary history of lacertids and colubrids, which has made them the subject of a great number of phylogenetic studies.

In the case of the Lacertini tribe (Chapter 2), as explained in the general introduction section, two different phylogenetic hypotheses were proposed from two different studies using the same molecular markers, consisting on mitochondrial DNA and two slow evolving genes (*cmos* and *rag1*) and the concatenation approach. Pavlicev & Mayer (2009) recovered a basal polytomy associated to a fast radiation with low support for most of the cladogenic events, and (Pyron et al., 2013) recovered a Lacertini clade with well resolved internal branching. By applying the species tree approach combined with new nuclear data, it was possible to step out of the concatenation approach used by these two studies and to obtain a better picture on the contrasting phylogenetic hypotheses. On one side, the deeper relationships of Lacertini were not fully resolved in the species tree, also suggesting a fast radiation, yet some other relationships were corroborated and new clades were recovered. On the other side, results from the Lacertini species tree were in clear contrast with those obtained by Pyron et al. (2013). In this case, the difference between the coalescent species tree and the gene trees is highly evident, with the supermatrix approach producing well supported but phylogenetic estimates that are not congruent with those obtained via alternative approaches. These results highlight the role of the species tree with appropriate sampling of molecular markers in the phylogenetic resolution of Lacertini.

A very similar case is represented by the ratsnakes genus *Zamenis* (Chapter 3). Throughout the years, different phylogenetic studies have provided contrasting results concerning two subjects: the monophyly of *Zamenis* regarding *Rhinechis* (now classified as *Zamenis scalaris*) and the relationships between *Zamenis* species. The results from the species tree approach supported a *Zamenis* and *Rhinechis* clade. In this case, the speciation of *Rhinechis* and *Zamenis* oldest species is very close in time, suggesting a relatively fast radiation that precluded the inference on the monophyly of these genera. The relationships within *Zamenis*, however, are well supported in the species tree, in contrast with the hypotheses from previous studies with different molecular markers (mitochondrial DNA and the slow nuclear *cmos* gene) and phylogenetic methods (supermatrices with high levels of missing data). This case also shows the differences obtained by the coalescent species tree and the supermatrix approach of previous studies, highlighting the importance of the species tree, allowing the reconciliation of a set of gene trees embedded in a shared species phylogeny, and of the nuclear markers in the phylogenetic inference. While the early studies with few concatenated genes for

few taxa agree with estimates obtained in the chapters of this thesis, the supermatrix approach, with the high levels of missing data from the use of few genes for many taxa provides supported estimates that are not congruent with those obtained by the methods applied in the chapters of this thesis.

MISSING DATA

When comparing the phylogenetic hypotheses proposed by previous studies for the squamate groups studied in this thesis, particularly the Lacertini tribe and the colubrid genus *Zamenis,* and the results obtained in the Chapters of this thesis, we found a significant impact of the amount of missing data in phylogenetic estimates. The phylogenetic estimates from previous studies were mainly based on datasets with a high percentage of missing data for Lacertini and Colubrinae (in some cases reaching 80% of missing sequences), resulting in a final dataset that heavily relied on mitochondrial data and a few slow evolving nuclear genes available in GenBank. So, for these two squamate groups, however high the number of nuclear genes might be in the final supermatrix (47 nuclear genes in the latest squamate phylogenetic inference by Zheng & Wiens (2016)), the inference of the phylogenetic relationships within the Lacertini and Colubrinae sub-trees were based on sequences of a few concatenated genes that had been generated and used by previous phylogenetic studies on these two groups.

Estimating a tree applying the supermatrix with thousands of taxa requires high-speed approximations of tree topology searches, substitution models parameter estimates, as well as to assess node support, since bootstrap analysis is computationally intractable. This, combined with the high levels of missing data associated with this method suggest that the supermatrix approach may provide high support for relationships within tip-clades which are actually not supported and inconsistent with the phylogenetic studies with a narrower taxonomic focus, from where the data used in the supermatrix originated. This is likely the case of the phylogenetic inference for Lacertini and *Zamenis*, where the high support levels within these groups were a result of the missing sequences for these taxa and the kind of approximations required by the supermatrix approach.

THE ROLE OF NEXT GENERATION SEQUENCING

One of the main finding of this thesis is the demonstration of the benefits of the addition of fast evolving nuclear markers to resolve difficult phylogenetic questions including

phylogenetic instability and polytomies potentially related to fast radiations in selected squamate groups (Leaché et al., 2014; Leaché & Linkem, 2015; Leaché & Oaks, 2017). In our study cases, improvement of the phylogenetic estimates was already apparent by using five nuclear markers. Following this line, we expect that analyses of data from hundreds of loci generated by the application of next generation sequencing (NGS) will further improve the phylogenetic inference of these and other groups.

In recent decades the representation of nuclear genes among the molecular markers used for phylogenetic inference has been slowly increasing. However, in most cases, only a few nuclear markers were used, normally slow evolving genes, and as a complement to mitochondrial data. Even in supermatrices with a relatively high number of nuclear genes, these are underrepresented for several taxa, as exemplified for previous studies on squamate phylogeny including, among others, Lacertini and *Zamenis* (Pyron et al., 2011, 2013; Figueroa et al., 2016; Zheng & Wiens, 2016). Over the years, however, the beneficial use of highly informative nuclear genes on phylogenetics has been documented, as demonstrated in the studies presented in this thesis and other works (e.g. Godinho et al., 2008; Townsend et al., 2011; Fontanella et al., 2012) and the use of nuclear data has become more common.

The use of nuclear genes has reached its full potential with the development of next generation sequencing (NGS) techniques and the possibility of generating data from thousands of loci in a relatively fast and economic way, accompanied by the development of powerful statistical methods able to deal with such kind of data (McCormack et al., 2013). This has led to the phylogenomics era, and several studies have already documented new results obtained with this method in several different groups including animals (Dunn et al., 2008), reptiles (Chiari et al., 2012; Crawford et al., 2012), birds (Jarvis et al., 2014), arthropods (Meusemann et al., 2010), annelids (Weigert et al., 2014) and fungi (Chang et al., 2015).

In squamates, and in the light of the results provided by few fast evolving nuclear genes, the application of phylogenomics reveals itself as promising tool, particularly in phylogenetically challenging groups such as the Lacertini tribe. In Chapter 2, the addition of five nuclear genes has allowed the recovery of new clades and higher support for the position of some genera, while the basal polytomy remained, potentially indicating a fast radiation. Perhaps the basal polytomy associated with the early cladogenesis events in Lacertini may actually be a soft polytomy, resulting from limited amount of data that could be resolved by the addition of several hundreds or thousands of loci.

On the other hand, results from the NGS approach in the Colubrinae sub-family, particularly from the ratsnakes, are already available from a very recent work by Chen et al. (2017). In this study the authors infer the phylogeny of Old World ratsnakes based on 304 nuclear loci (452 698 bp) generated through NGS techniques (anchored hybrid enrichment; Lemmon et al., 2012). Interestingly, the results from this study, regarding the phylogenetic relationships of the genera *Zamenis* and *Rhinechis* are fully concordant with the relationships recovered in the study from this thesis (Chapter 3). This finding shows that using as few as five fast evolving nuclear genes was enough to recover the 'true' phylogenetic relationships between *Zamenis* and *Rhinechis* as provided by a phylogenomic approach using 304 loci. This might be relevant for circumstances where developing a phylogenomic dataset is difficult, at least in studies with a narrow taxonomical focus as we do not know if analyses based on our five nuclear genes would be able to resolve relationships between all Old World ratsnakes as carried out in the phylogenomic approach of Chen et al. (2017).

TAXONOMICAL IMPLICATIONS

The increasing number of phylogenetic studies using squamates and, in particular, lacertids and colubrids has been the responsible for the increasing number of species included in these groups and for a high number of taxonomical revisions over the last years (Table 1.1, Fig. 1.1). Within Lacertini, after the description of eight new genera previously included in the genus *Lacerta*, a decade ago (Arnold et al., 2007), the taxonomy seems to be stable, at least at the genus level. The only taxonomical questions raised in recent studies refer to the monotypic genera *Teira* and *Scelarcis* and the monophyly of the genus *Algyroides*. The genera *Teira* and *Scelarcis* have consistently been recovered as sister genera in all previous studies. *Scelarcis* was once included in the genus *Teira* and Pavlicev & Mayer (2009) suggested that these should be merged. As described in Chapter 2, these genera exhibit unique morphological features and intraspecific diversity indicative of species complexes (Arnold, 1973; Arnold et al., 2007; Brehm et al., 2003; Perera et al., 2007) and this is why it has been argued that they should be kept as separated genera. The monophyly of *Algyroides* was questioned by Pavlicev and Mayer (2009), because *Algyroides* was paraphyletic, with *Dinarolacerta* nested within. The results from the Lacertini study showed that these genera are closely related and the monophyly of *Algyroides* is supported by the nuclear DNA. Moreover, the four *Algyroides* species share unique morphological characters that distinguish them from all other lacertids (Arnold, 1973; Arnold et al., 2007) and are likely to constitute a single monophyletic genus.

The genus *Psammodromus* has been recently taxonomically revised, with the description of the three Iberian species, previously included in *P. hispanicus*, used in Chapter 4 (*P. occidentalis*, *P. edwarsianus* and *P. hispanicus*). The results from Chapter 4 recognise different lineages within *P. algirus* (Fig. 4.3). Other studies have recovered a similar structure and one of them described two species from different lineages of *P. algirus* in the Iberian Peninsula: *P. manuelae* and *P. jeanneae* (Busack et al., 2006). A later reassessment of the phylogenetic structure of *P. algirus* rejected the validity of these species and advised on the need of a proper phylogeny, with sufficient individuals across the entire distribution range, and nuclear and morphological data (Verdú-Ricoy et al., 2010). The results from this chapter also showed two highly divergent lineages within each of the African endemic species *P. blanci* and *P. microdactylus* (Fig. 4.3), which is suggestive of undescribed diversity within both species. Additional data are necessary to fully understand the diversity and the evolutionary history of these species, with a comprehensive taxon sampling and further molecular and morphological assessments.

From all the lacertid groups studied in this thesis, the genus *Omanosaura* is the one that received the less taxonomic attention in the past and that is more likely to be the subject of a taxonomical revision in the future. The study included in this thesis shows cryptic diversity within the species *O. cyanura* with two highly differentiated lineages associated with the northern and the southern portion of the species range (Figs. 5.2, 5.4 and 5.5, Table 5.3) and indicates that a comprehensive taxonomical assessment should be performed for this species. The inclusion of a genetic sample from the type locality in Wadi Shawkah, UAE (Fig. 5.1) shows that the holotype of *O. cyanura* is likely to belong to the northern lineage and therefore, in a future taxonomical assessment of this species, the northern lineage would retain the name *O. cyanura*. In order to perform such taxonomical assessment additional individuals should be collected from unsampled areas to test whether the genetic distance between these two lineages underlie a sharp phylogeographic break or rather a zone if intergradation. Moreover, additional voucher specimens need to be collected for a comprehensive morphological characterisation.

Finally, the results of the study of the colubrid genera *Zamenis* and *Rhinechis* from Chapter 3 supported a classification lumping that better reflects the evolutionary history of these genera and so, according to the priority rule, the species *R. scalaris* was moved into the genus *Zamenis*, and designated as *Zamenis scalaris* **comb. nov.**. A recent study (Figueroa et al., 2016) proposed to synonymise *Rhinechis* and *Zamenis*, although their phylogenetic inference showed that not only *Rhinechis* was nested within *Zamenis* but also the Indo-Chinese snake *Ptyas korros*. In contrast, the phylogenetic

trees from the study presented in Chapter 3 (Figs. 3.4, 3.5 & 3.6) showed little support for the monophyly of *Zamenis* but maximum support for the clade *Zamenis* + *Rhinechis*. The speciation events of the species from these genera are very close in time, which could mask the true phylogenetic relationships between these genera. The recent phylogenomic study by Chen et al., (2017) applying 304 nuclear loci has recovered the same phylogenetic structure as the study from this thesis, showing that our estimate is robust to the increase of the number of molecular markers. In addition, only one or a few morphological characters in the head scalation distinguish *Rhinechis* from *Zamenis*, demonstrating the lack of distinctiveness between these genera.

CONCLUDING REMARKS

The main aim of this thesis was to understand if and how the addition of fast evolving nuclear markers and species tree based on the multispecies coalescent approach could improve the estimate of phylogenetic relationships within selected squamates groups. This aim was achieved at the different squamate groups, from tribes, to genera and species and the findings in this thesis are relevant for the fields of phylogenetics and evolution, beyond its applications regarding squamates. In summary, the major conclusions of this work are:

i. The addition of fast evolving nuclear genes to the phylogenetic inference in all the different squamate groups had great impact, improving the resolution of relationships between species and genera and in the increase of the overall support of the nodes.

ii. The application of the species tree coalescent approach in the phylogenetic inference allowed the discrimination between contrasting phylogenetic hypotheses proposed by previous studies in the tribe Lacertini and the genus *Zamenis*.

iii. The results from the species tree and concatenation methods provided similar results in the tribe Lacertini and the genera *Zamenis*, *Psammodromus* and *Omanosaura*, almost certainly because of the inclusion of several independent nuclear genes, in addition to mitochondrial genes, as well as a comprehensive taxon sampling in all these groups to the phylogenetic analyses.

iv. Within the Lacertini tribe, the phylogenetic inference including nuclear genes and the species tree approach has not resolved the internal branching between the genera, corroborating the fast radiation hypothesis and demonstrating how the supermatrix approach can provide high support for nodes that conflict with alternative approaches.

v. Results from the Lacertini study recovered a new phylogenetic clade formed by four monotypic genera (*Archaeolacerta*, *Zootoca*, *Teira* and *Scelarcis*) and supported the monophyly of the genus *Algyroides*, which had been previously questioned.

vi. Phylogenetic inference including for the first time several nuclear genes, species tree approach and the six species of *Psammodromus*, together with the reconstruction of ancestral distribution areas have framed the phylogeographical history of this genus. Results show that over-sea dispersal played a major role in the intercontinental exchange and divergence within *Psammodromus*, with two dispersal towards Africa at different times, originating the African species and the *P. algirus* lineage of this genus.

vii. Phylogenetic results of *Psammodromus* show that the species *P. microdactylus*, for the first time included in a phylogenetic study, is sister to *P. blanci* and therefore these two African endemic species have diversified in Africa after an over-sea dispersal. Both these species display high levels of intraspecific diversity.

viii. The first phylogenetic assessment of the Eremiadini genus *Omanosaura* has showed a genetically homogeneous *O. jayakari*, while the smaller *O. cyanura* presents two well-differentiated lineages related with the geographical distribution of the samples, which shows similar levels of diversity across mitochondrial and nuclear markers.

ix. The phylogenetic inference in the ratsnake genera *Zamenis* and *Rhinechis*, performed with several fast evolving nuclear genes and the species tree has allowed the discrimination between contrasting hypotheses regarding the monophyly of *Zamenis* and intrageneric relationships. While the monophyly of *Zamenis* lacks statistical support, the clade formed by *Zamenis* and *Rhinechis* receives maximum support and *R. scalaris* was moved into *Zamenis*, designated *Zamenis scalaris* comb. nov.

x. Within *Zamenis*, the well resolved branching patterns and the ancestral area reconstruction supported an east to west dispersal of these snakes, since the oldest species *Z. hohenackeri* and *Z. persicus* are distributed in the Middle East, while most recent species have a more western Palaearctic distribution.

xi. An assessment of the effects of missing data in the *Zamenis* study revealed that the addition of missing data had no effect on the phylogenetic inference. However, the percentage of missing data in the final dataset (around 20%) is much lower than in previous studies with supermatrices, where the missing data can be as high as 80%, so direct comparisons cannot be drawn.

FUTURE PERSPECTIVES

The field of phylogenetics has been deeply transformed by the development of new statistical tools of phylogenetic inference and of data generation. The work developed for this thesis has showed the positive effects of using the coalescent species tree approach combined with the addition of highly informative nuclear genes for phylogenetic inference. This approach has contributed to an increase in the knowledge on the phylogenetic relationships of selected squamate groups of the families Lacertidae and Colubridae, but has also uncovered some questions which highlight directions for future research.

The application of fast evolving nuclear genes has been demonstrated to benefit the phylogenetic inference, particularly in groups with previous contrasting phylogenetic hypotheses. The application of NGS techniques with hundreds or thousands of nuclear loci will be clearly the next step for phylogenetic research in squamates and other groups in particular of those that represent old and fast radiations, extinction of old lineages and that have currently conflicting phylogenetic hypotheses mainly based on a few mitochondrial and slowly evolving nuclear genes.

Within the selected squamates of this thesis, although the use of fast evolving nuclear genes improved the phylogenetic resolution of relationships among genera and species, some questions still remain unanswered and new question raised. The Lacertini tribe represents a clear example of the benefits that the further addition of nuclear loci, particularly of the NGS approach, could bring resolution to the basal polytomy within this tribe. This constitutes a clear future line of research within the Lacertidae, since a few nuclear loci have recovered new clades, it is possible to expect that the NGS approach could further resolve the old cladogenesis events of this tribe.

Although cost-effective, NGS approaches such as those employed in phylogenomic studies (e.g. anchored hybrid enrichment; Lemmon et al., 2012) still currently require several thousands of euros and may be not well-suited for phylogenetic questions of narrow taxonomic focus such as within a small taxon set of a dozen of species. In the specific case of *Zamenis* ratsnakes we found that even a few number of highly informative nuclear genes can recover similar relationships as a phylogenomic approach. Therefore, as demonstrated in Chapters 2 and 3, in the cases where it is not possible to produce NGS datasets, the application of a few selected fast-evolving nuclear loci could still provide useful insights for difficult phylogenetic questions.

The cryptic diversity described for the endemic genus *Omanosaura* requires further research. A new phylogenetic and morphological assessment of the species *O.*

cyanura is necessary to understand if the highly divergent genetic lineages recovered for this species do reflect the true genetic diversity in *O. cyanura*. This requires a sampling effort in the whole distribution area of this species, particularly in the unsampled areas that separate the geographical regions associated to the recovered lineages. Both mitochondrial and nuclear genes should be applied in the analyses, to understand if the diversity is sorted in both genomes, and a species delimitation approach is likely to be required for the analyses, as well as a morphological characterisation of the individuals of the potential new lineages or species.

Aside from the taxonomical questions uncovered by our studies in the squamate study cases (see the section “Taxonomical implications”), an issue that certainly deserves further attention is the use of the Messinian Salinity Crisis as a calibration point for the divergence between Mediterranean taxa. The results from the *Psammodromus* study show strong evidence of speciation events by over-sea dispersal that occurred well before or long after the opening of the Strait of Gibraltar and the Atlantic flooding at the end of the MSC 5.3 Ma. This evidence, which is corroborated by similar dispersal events in many other groups, suggests caution on the use of the relatively short event of the end of the MSC and strongly indicates the need for a re-evaluation of its use as a calibration point for the divergence of Mediterranean taxa, which is very popular in literature.

REFERENCES

Arnold E.N. (1973) Relationships of the Palaearctic lizards assigned to the genera *Lacerta*, *Algyroides* and *Psammodromus* (Reptilia: Lacertidae). *Bulletin of the British Museum*, **25**, 289–366.

Arnold E.N., Arribas O., & Carranza S. (2007) Systematics of the Palaearctic and Oriental lizard tribe Lacertini (Squamata: Lacertidae: Lacertinae), with descriptions of eight new genera. Magnolia Press. *Zootaxa*, **1430**, 1–86.

Avcı A., Ilgaz Ç., Rajabizadeh M., Yılmaz C., Üzüm N., Adriaens D., Kumlutaş Y., & Olgun K. (2015) Molecular phylogeny and micro CT-scanning revealed extreme cryptic biodiversity in Kukri Snake, *Muhtarophis* gen. nov., a new genus for *Rhynchocalamus barani* (Serpentes: Colubridae). *Russian Journal of Herpetology*, **22**, 159–174.

Brehm A., Jesus J., Spínola H., Alves C., Vicente L., & Harris D.J. (2003) Phylogeography of the Madeiran endemic lizard *Lacerta dugesii* inferred from mtDNA sequences. *Molecular Phylogenetics and Evolution*, **26**, 222–230.

Burbrink F.T. & Lawson R. (2007) How and when did Old World ratsnakes disperse into the New World? *Molecular Phylogenetics and Evolution*, **43**, 173–189.

Busack S.D., Salvador A., & Lawson R. (2006) Two new species in the genus *Psammodromus* (Reptilia: Lacertidae) from the Iberian Peninsula. *Annals of Carnegie Museum*, **75**, 1–10.

Chang Y., Wang S., Sekimoto S., Aerts A.L., Choi C., Clum A., LaButti K.M., Lindquist E.A., Ngan C.Y., Ohm R.A., Salamov A.A., Grigoriev I. V., Spatafora J.W., & Berbee M.L. (2015) Phylogenomic analyses indicate that early fungi evolved digesting cell walls of algal ancestors of land plants. *Genome Biology and Evolution*, **7**, 1590–1601.

Chen X., Lemmon A.R., Lemmon E.M., Pyron R.A., & Burbrink F.T. (2017) Using phylogenomics to understand the link between biogeographic origins and regional diversification in ratsnakes. *Molecular Phylogenetics and Evolution*, **111**, 206–218.

Chiari Y., Cahais V., Galtier N., & Delsuc F. (2012) Phylogenomic analyses support the position of turtles as the sister group of birds and crocodiles (Archosauria). *BMC Biology*, **10**, 65.

Crawford N.G., Faircloth B.C., McCormack J.E., Brumfield R.T., Winker K., & Glenn T.C. (2012) More than 1000 ultraconserved elements provide evidence that turtles are the sister group of archosaurs. *Biology Letters*, **8**, 783–786.

Dunn C.W., Hejnol A., Matus D.Q., Pang K., Browne W.E., Smith S.A., Seaver E., Rouse G.W., Obst M., Edgecombe G.D., Sørensen M. V., Haddock S.H.D., Schmidt-Rhaesa A., Okusu A., Kristensen R.M., Wheeler W.C., Martindale M.Q., & Giribet G. (2008) Broad phylogenomic sampling improves resolution of the animal tree of life. *Nature*, **452**, 745–749.

Figueroa A., McKelvy A.D., Grismer L.L., Bell C.D., & Lailvaux S.P. (2016) A species-level phylogeny of extant snakes with description of a new colubrid subfamily and genus. *PLOS ONE*, **11**, e0161070.

Fontanella F.M., Olave M., Avila L.J., Sites Jr J.W., & Morando M. (2012) Molecular dating and diversification of the South American lizard genus *Liolaemus* (subgenus *Eulaemus*) based on nuclear and mitochondrial DNA sequences. *Zoological Journal of the Linnean Society*, **164**, 825–835.

Fu J. (1998) Toward the phylogeny of the family Lacertidae: implications from mitochondrial DNA 12S and 16S gene sequences (Reptilia: Squamata). *Molecular Phylogenetics and Evolution*, **9**, 118–130.

Fu J. (2000) Toward the phylogeny of the family Lacertidae - Why 4708 base pairs of mtDNA sequences cannot draw the picture. *Biological Journal of the Linnean Society*, **71**, 203–217.

Godinho R., Crespo E.G., & Ferrand N. (2008) The limits of mtDNA phylogeography: Complex patterns of population history in a highly structured Iberian lizard are only revealed by the use of nuclear markers. *Molecular Ecology*, **17**, 4670–4683.

Harris D.J., Arnold E.N., & Thomas R.H. (1998) Relationships of lacertid lizards (Reptilia: Lacertidae) estimated from mitochondrial DNA sequences and morphology. *Proceedings of the Roysal Society B Biological sciences*, **265**, 1939–1948.

Hipsley C.A., Himmelmann L., Metzler D., & Müller J. (2009) Integration of Bayesian molecular clock methods and fossil-based soft bounds reveals early Cenozoic origin of African lacertid lizards. *BMC Evolutionary Biology*, **9**, 151.

Jarvis E.D., Mirarab S., Aberer A.J., Li B., Houde P., Li C., Ho S.Y.W., Faircloth B.C., Nabholz B., Howard J.T., Suh A., Weber C.C., da Fonseca R.R., Li J., Zhang F., Li

H., Zhou L., Narula N., Liu L., Ganapathy G., Boussau B., Bayzid M.S., Zavidovych V., Subramanian S., Gabaldon T., Capella-Gutierrez S., Huerta-Cepas J., Rekepalli B., Munch K., Schierup M., Lindow B., Warren W.C., Ray D., Green R.E., Bruford M.W., Zhan X., Dixon A., Li S., Li N., Huang Y., Derryberry E.P., Bertelsen M.F., Sheldon F.H., Brumfield R.T., Mello C. V., Lovell P. V., Wirthlin M., Schneider M.P.C., Prosdocimi F., Samaniego J.A., Velazquez A.M. V., Alfaro-Nunez A., Campos P.F., Petersen B., Sicheritz-Ponten T., Pas A., Bailey T., Scofield P., Bunce M., Lambert D.M., Zhou Q., Perelman P., Driskell A.C., Shapiro B., Xiong Z., Zeng Y., Liu S., Li Z., Liu B., Wu K., Xiao J., Yinqi X., Zheng Q., Zhang Y., Yang H., Wang J., Smeds L., Rheindt F.E., Braun M., Fjeldsa J., Orlando L., Barker F.K., Jonsson K.A., Johnson W., Koepfli K.-P., O'Brien S., Haussler D., Ryder O.A., Rahbek C., Willerslev E., Graves G.R., Glenn T.C., McCormack J., Burt D., Ellegren H., Alstrom P., Edwards S. V., Stamatakis A., Mindell D.P., Cracraft J., Braun E.L., Warnow T., Jun W., Gilbert M.T.P., & Zhang G. (2014) Whole-genome analyses resolve early branches in the tree of life of modern birds. *Science*, **346**, 1320–1331.

Jiang W., Chen S.-Y., Wang H., Li D.-Z., & Wiens J.J. (2014) Should genes with missing data be excluded from phylogenetic analyses? *Molecular Phylogenetics and Evolution*, **80**, 308–318.

Leaché A.D. & Linkem C.W. (2015) Phylogenomics of Horned Lizards (Genus: *Phrynosoma*) using targeted sequence capture data. *Copeia*, **103**, 586–594.

Leaché A.D. & Oaks J.R. (2017) The utility of single nucleotide polymorphism (SNP) data in phylogenetics. *Annual Review of Ecology, Evolution, and Systematics*, **48**, .

Leaché A.D., Wagner P., Linkem C.W., Böhme W., Papenfuss T.J., Chong R. A., Lavin B.R., Bauer A.M., Nielsen S. V, Greenbaum E., Rödel M.-O., Schmitz A., LeBreton M., Ineich I., Chirio L., Ofori-Boateng C., Eniang E. A., Baha El Din S., Lemmon A.R., & Burbrink F.T. (2014) A hybrid phylogenetic-phylogenomic approach for species tree estimation in African *Agama* lizards with applications to biogeography, character evolution, and diversification. *Molecular Phylogenetics and Evolution*, **79**, 215–230.

Lemmon A.R., Emme S.A., & Lemmon E.M. (2012) Anchored hybrid enrichment for massively high-throughput phylogenomics. *Systematic Biology*, **61**, 727–744.

Lenk P., Joger U., & Wink M. (2001) Phylogenetic relationships among European rat snakes of the genus *Elaphe* Fitzinger based on mitochondrial DNA sequence comparisons. *Amphibia-Reptilia*, **22**, 329–339.

Mayer W. & Pavlicev M. (2007) The phylogeny of the family Lacertidae (Reptilia) based on nuclear DNA sequences: convergent adaptations to arid habitats within the subfamily Eremiainae. *Molecular Phylogenetics and Evolution*, **44**, 1155–1163.

McCormack J.E., Hird S.M., Zellmer A.J., Carstens B.C., & Brumfield R.T. (2013) Applications of next-generation sequencing to phylogeography and phylogenetics. *Molecular Phylogenetics and Evolution*, **66**, 526–538.

Meusemann K., von Reumont B.M., Simon S., Roeding F., Strauss S., Kück P., Ebersberger I., Walzl M., Pass G., Breuers S., Achter V., von Haeseler A., Burmester T., Hadrys H., Wägele J.W., & Misof B. (2010) A phylogenomic approach to resolve the Arthropod tree of Life. *Molecular Biology and Evolution*, **27**, 2451–2464.

Pavlicev M. & Mayer W. (2009) Fast radiation of the subfamily Lacertinae (Reptilia: Lacertidae): history or methodical artefact? *Molecular Phylogenetics and Evolution*, **52**, 727–34.

Perera A., Vasconcelos R., Harris D.J., Brown R.P., Carretero M.A., & Pérez-Mellado V. (2007) Complex patterns of morphological and mtDNA variation in *Lacerta perspicillata* (Reptilia; Lacertidae). *Biological Journal of the Linnean Society*, **90**, 479–490.

Pyron R.A., Burbrink F.T., Colli G.R., de Oca A.N.M., Vitt L.J., Kuczynski C.A., & Wiens J.J. (2011) The phylogeny of advanced snakes (Colubroidea), with discovery of a new subfamily and comparison of support methods for likelihood trees. *Molecular Phylogenetics and Evolution*, **58**, 329–342.

Pyron R., Burbrink F.T., & Wiens J.J. (2013) A phylogeny and revised classification of Squamata, including 4161 species of lizards and snakes. *BMC Evolutionary Biology*, **13**, 93.

Townsend T.M., Mulcahy D.G., Noonan B.P., Sites J.W., Kuczynski C. A, Wiens J.J., & Reeder T.W. (2011) Phylogeny of iguanian lizards inferred from 29 nuclear loci, and a comparison of concatenated and species-tree approaches for an ancient, rapid radiation. *Molecular Phylogenetics and Evolution*, **61**, 363–380.

Utiger U., Helfenberger N., Schätti B., Schmidt C., Ruf M., & Ziswiler V. (2002) Molecular systematics and phylogeny of Old World and New World ratsnakes, *Elaphe* Auct., and related genera (Reptilia, Squamata, Colubridae). *Russian Journal of Herpetology*, **9**, 105–124.

Verdú-Ricoy J., Carranza S., Salvador A., Busack S., & Díaz J. (2010) Phylogeography of *Psammodromus algirus* (Lacertidae) revisited: systematic implications. *Amphibia-Reptilia*, **31**, 576–582.

Weigert A., Helm C., Meyer M., Nickel B., Arendt D., Hausdorf B., Santos S.R., Halanych K.M., Purschke G., Bleidorn C., & Struck T.H. (2014) Illuminating the base of the Annelid tree using transcriptomics. *Molecular Biology and Evolution*, **31**, 1391–1401.

Zheng Y. & Wiens J.J. (2016) Combining phylogenomic and supermatrix approaches, and a time-calibrated phylogeny for squamate reptiles (lizards and snakes) based on 52 genes and 4162 species. *Molecular Phylogenetics and Evolution*, **94**, 537–547.

Appendices

Supporting Information

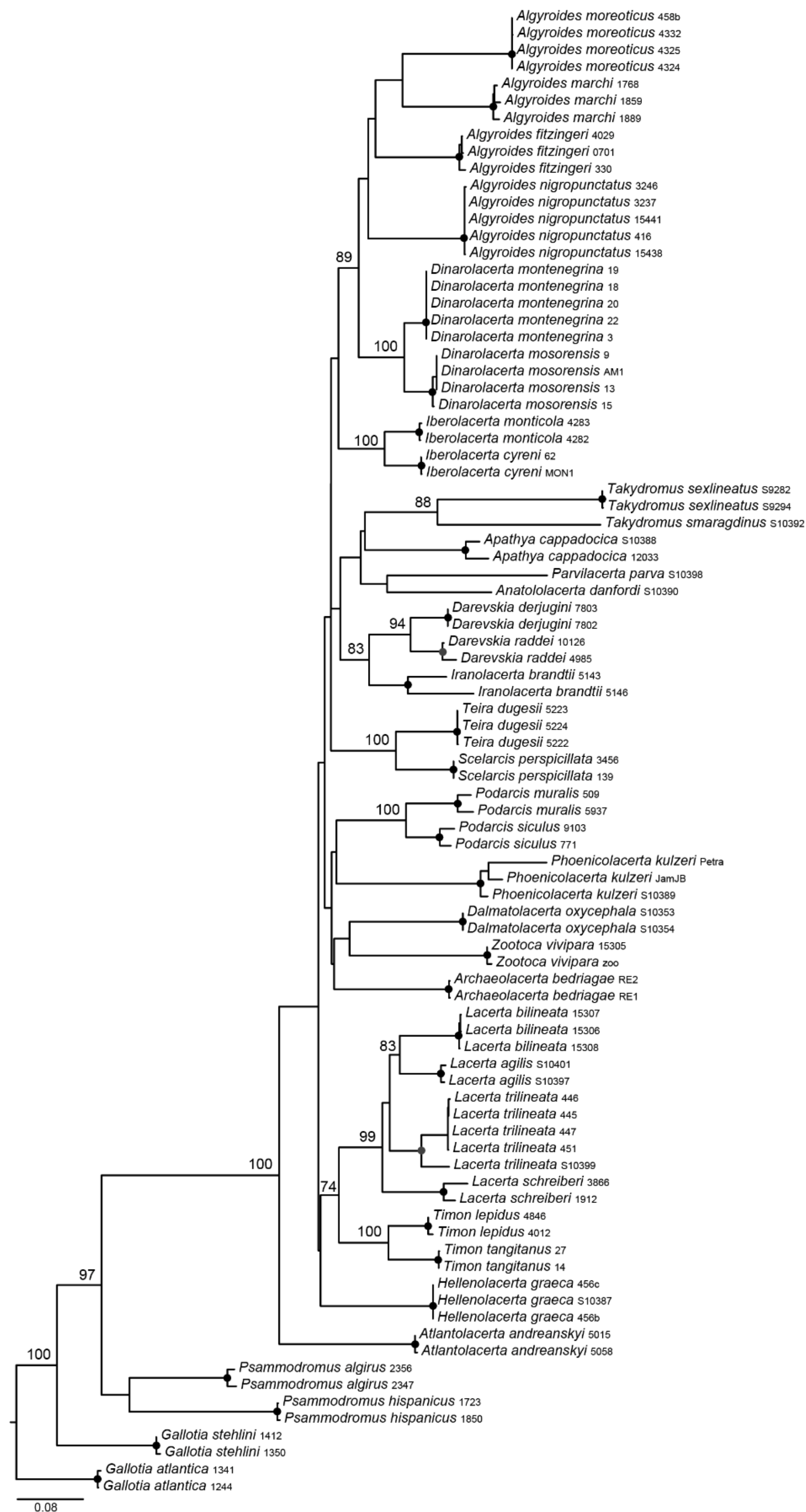


Figure S2.1. Phylogenetic relationships of Lacertini based on Maximum Likelihood analyses of concatenated tree based on mitochondrial DNA sequences (*12S* and *nd4*). Bootstrap values (BS) ≥70 are shown above branches. Within species, black dots represent BS = 100; grey dots represent BS of 70 ≥ 99.

Figure S2.2. Phylogenetic relationships of Lacertini based on Maximum Likelihood analyses of concatenated nuclear DNA sequences (*acm4*, *βfib*, *mc1r*, *pdc* and *reln*). Bootstrap values ≥70 are shown above branches. Within species, black dots represent BS = 100; grey dots represent BS of 70 ≥ 99.

Figure S2.3. Phylogenetic relationships of Lacertini based on Maximum Likelihood analyses of concatenated mitochondrial (*12S* and *nd4*) and nuclear (*acm4*, *βfib*, *mc1r*, *pdc* and *reln*). Bootstrap values ≥70 are shown above branches. Within species, black dots represent BS = 100; grey dots represent BS of 70 ≥ 99.

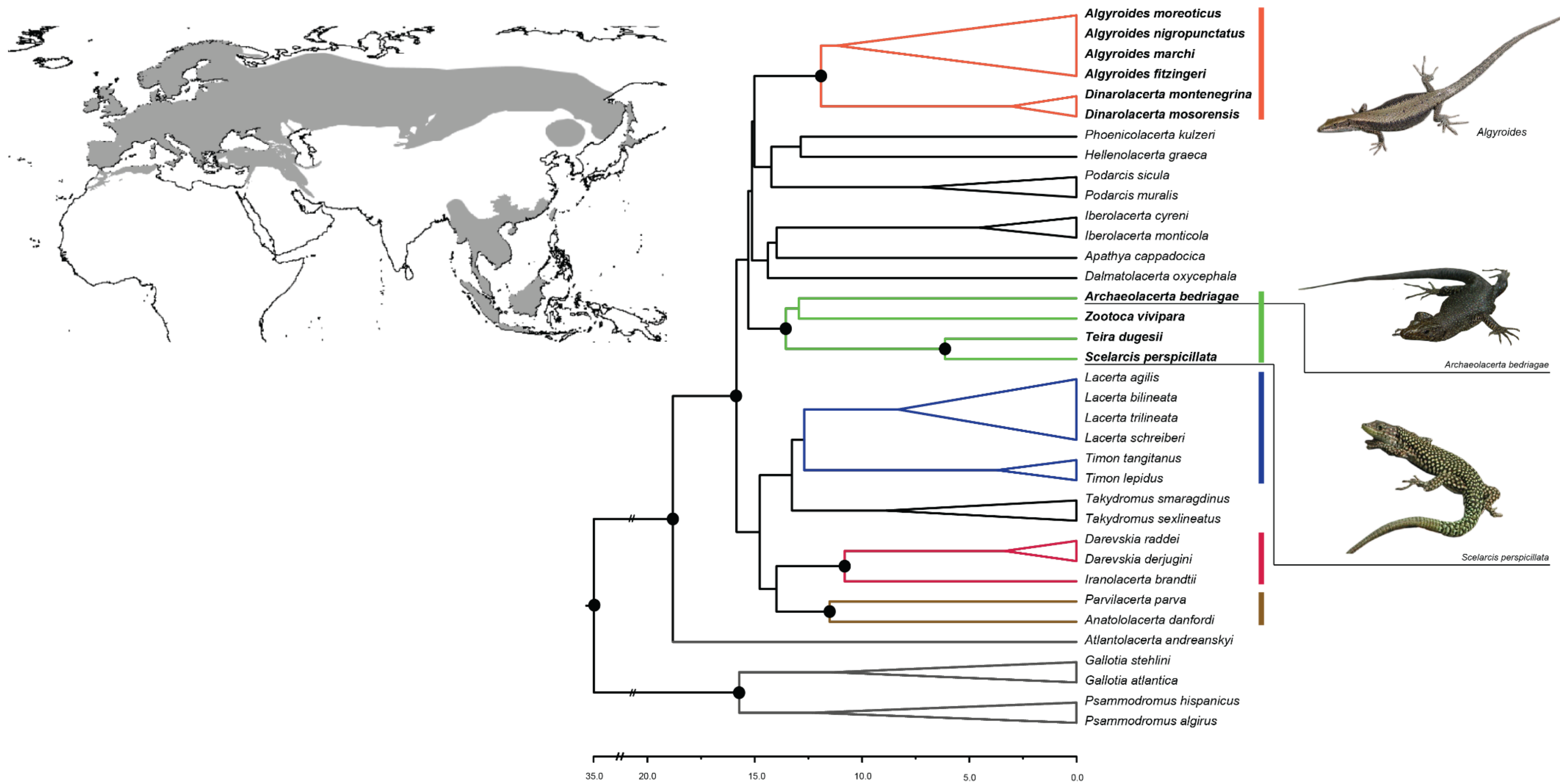


Figure S2.4. Distribution and phylogenetic relationships of the Lacertini genera. Graphical Abstract of the online version of the paper in Molecular Phylogenetics and Evolution.

Table S3.1. Divergence time estimates in million years based on rates from calibration I (Cal I) and calibration I (Cal II) with the *Zamenis* + ratsnakes taxon-set; 95% HPD intervals are provided in brackets. *Z. lin*: *Zamenis lineatus*, *Z. lon*: *Zamenis longissimus*, *Z. sit*: *Zamenis situla*, *Z. hoh*: *Zamenis hohenackeri*, *Z. per*: *Zamenis persicus*, *R. sca*: *Rhinechis scalaris*, *C. aus*: *Coronella austriaca*, *C. gir*: *Coronella girondica*, *He. alg*: *Hemorrhois algirus*, *Hi. vir*: *Hierophis viridiflavus*.

Node	***Zamenis* + ratsnakes**	
	Mean and 95% HPD (Ma)	
	Cal I	**Cal II**
(*Z. lin*, *Z. lon*)	2.94 (1.51 – 4.45)	5.28 (3.09 – 7.56)
(*Z. sit*, *Z. lin*, *Z. lon*)	3.69 (2.18 – 5.36)	6.62 (4.48 – 8.91)
(*Z. per*, *Z. sit*, *Z. lin*, *Z. lon*)	4.91 (2.81 – 7.32)	8.82 (5.79 – 12.13)
Zamenis spp.	6.34 (3.97 – 9.05)	11.42 (8.25 – 14.86)
(*R. sca*, *Zamenis* spp.)	7.45 (4.76 – 10.44)	13.39 (9.99 – 17.02)
(*Coronella* spp., (*R. sca*, *Zamenis* spp.))	9.04 (6.01 – 12.56)	18.20 (13.87 – 22.82)
Coronella spp.	5.89 (3.41 – 8.59)	10.59 (7.05 – 14.35)
Ratsnakes ingroup	10.08 (6.54 – 13.90)	18.20 (13.87 – 22.82)
(*He. alg*, *Hi. vir*)	11.57 (5.95 – 17.79)	20.93 (12.57 – 30.69)
Root	16.38 (10.12 – 23.72)	29.64 (20.80 – 39.56)

Table S3.2. Geographic location of the ancestors and lineages of *Rhinechis* and *Zamenis* species for each time slice (mean and 95% HPD interval of the latitude and longitude coordinates), based on the species tree applying the time calibration I. *Z. lin*: *Zamenis lineatus*, *Z. lon*: *Zamenis longissimus*, *Z. sit*: *Zamenis situla*, *Z. per*: *Zamenis persicus*.

Time slice	Ancestor/Lineage	Mean (lat/lon)	95%HPD (lat/lon)
a (8.0 Ma)	*Zamenis* and *Rhinechis* ancestor	39.3020 / 22.2972	37.2251 – 41.5373 / 1.4546 – 43.5169
b (7.0 Ma)	*Zamenis* ancestor	39.1294 / 25.8445	37.4215 – 40.9703 / 6.4762 – 44.8263
	R. scalaris	39.3504 / 18.9671	37.3755 – 41.5688 / -2.0889 – 44.1903
c (6.0 Ma)	*Z. per*, *Z. sit*, *Z. lin* and *Z. lon* ancestor	39.0797 / 28.1461	37.3156 – 40.6879 / 9.8227 – 47.0330
	Z. hohenackeri	39.0857 / 28.5675	37.3414 – 40.8380 / 9.4649 – 47.4450
	R. scalaris	39.5087 / 15.2740	37.2529 – 41.4938 / -7.7410 – 37.5427
d (5.0 Ma)	*Z. per*, *Z. sit*, *Z. lin* and *Z. lon* ancestor	38.9338 / 29.9140	37.4974 – 40.5097 / 11.71 – 45.8364
	Z. hohenackeri	39.0581 / 30.1360	37.3130 – 41.0071 / 11.0037 – 49.9636
	R. scalaris	39.5768 / 13.0522	37.3509 – 41.5709 / -9.5254 – 35.0871
e (4.0 Ma)	*Z. sit*, *Z. lin* and *Z. lon* ancestor	39.1954 / 29.1374	37.7825 – 40.5599 / 12.2192 – 45.5203
	Z. persicus	38.6477 / 33.3818	37.2968 – 40.2079 / 14.0528 – 49.6992
	Z. hohenackeri	39.0926 / 31.9442	37.1574 – 40.7704 / 11.9170 – 55.5023
	R. scalaris	39.7221 / 9.0447	37.5736 – 41.7560 / -15.6879 – 30.0375
f (3.0 Ma)	*Z. sit*, *Z. lin* and *Z. lon* ancestor	39.6093 / 26.3231	38.3965 – 40.7330 / 13.1778 – 38.3033
	Z. persicus	38.1387 / 37.8457	36.4911 – 39.6373 / 19.4412 – 53.5818
	Z. hohenackeri	38.9957 / 33.6155	37.4071 – 40.7381 / 13.9503 – 52.7411
	R. scalaris	39.7809 /	37.9486 – 41.5762 /

		6.2888	-14.9149 – 26.5754
g (2.0 Ma)	*Z. longissimus*	40.1946 / 27.6636	39.0710 – 41.4793 / 14.9311 – 41.3903
	Z. lineatus	39.5779 / 23.1189	38.4723 – 40.8839 / 9.6867 – 36.5691
	Z. situla	39.6099 / 25.2771	38.4207 – 40.8680 / 10.8953 – 38.7140
	Z. persicus	37.6885 / 41.6369	36.3295 – 39.0786 / 25.5155 – 55.9000
	Z. hohenackeri	38.9985 / 35.1886	37.4791 – 40.5535 / 15.9169 – 50.3554
	R. scalaris	39.9817 / 2.8080	38.3428 – 41.5248 / -13.8504 – 20.8862
h (1.0 Ma)	*Z. longissimus*	40.8636 / 30.1066	39.9237 – 41.8936 / 18.2479 – 41.6741
	Z. lineatus	39.3146 / 18.8258	38.3182 – 40.3590 / 7.7705 – 30.9904
	Z. situla	39.5690 / 24.5196	38.4720 – 40.7070 / 11.9389 – 35.8817
	Z. persicus	37.1173 / 46.9410	35.9619 – 38.1572 / 34.6619 – 59.6397
	Z. hohenackeri	38.9873 / 36.8432	37.8029 – 40.1997 / 22.6884 – 48.0475
	R. scalaris	40.0864 / -0.4674	38.8653 – 41.2867 / -13.7176 – 11.9374

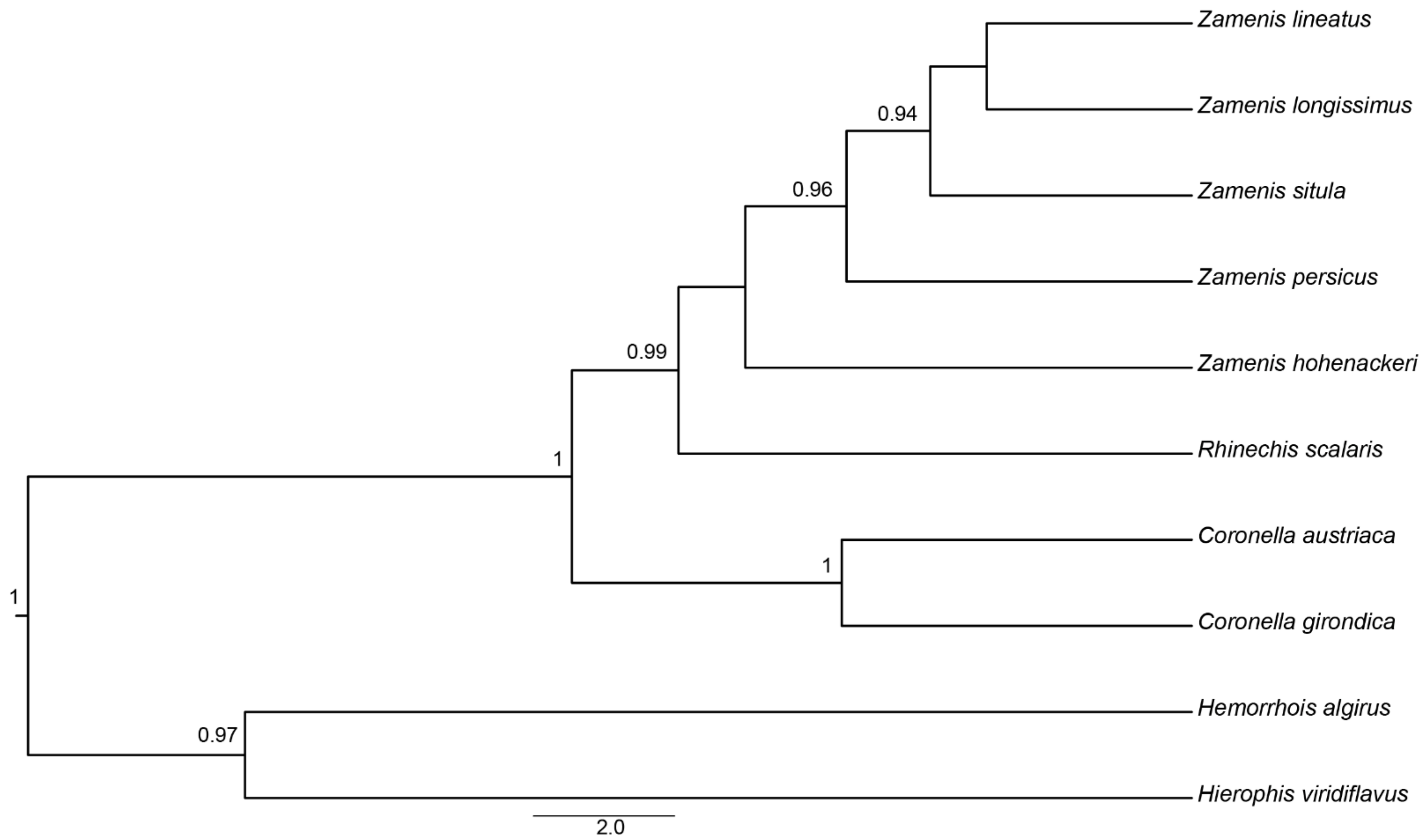


Figure S3.1. Bayesian tree topology of the BEAST analysis with one individual per species and the Yule prior of speciation. Bayesian posterior probabilities (BPP > 90) are represented above nodes.

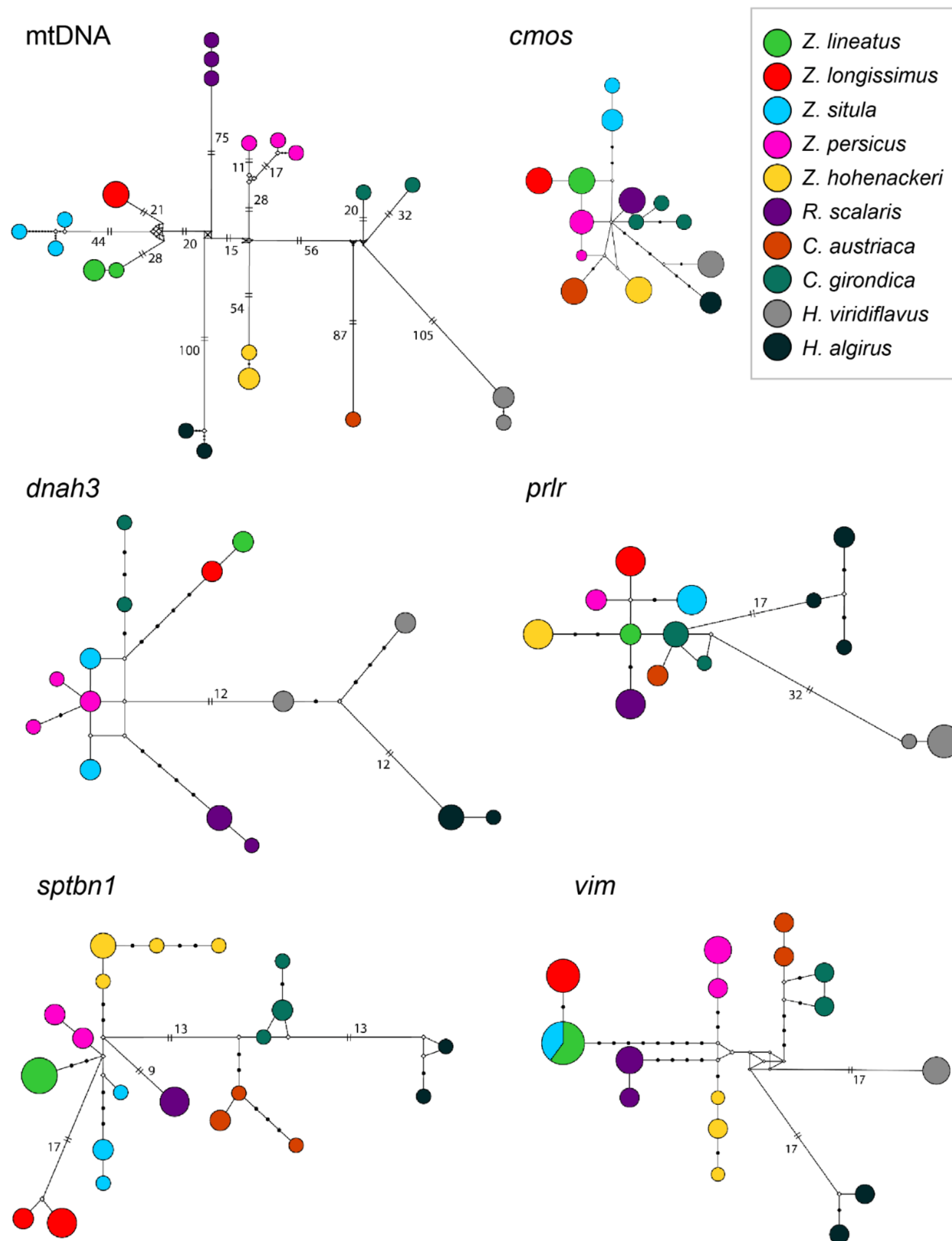


Figure S3.2. Median-joining haplotype networks for *Zamenis*, *Rhinechis*, *Coronella* and outgroups inferred with phased nuclear DNA sequences (*cmos, dnah3, prlr, sptbn1* and *vim*) and mitochondrial DNA sequences (concatenated *cytb* and *nd4*). Circles represent different haplotypes and the size is proportional to haplotype's frequency. Small white diamonds represent median vectors, mutations are represented by black dots or black lines with numbers on each connection.

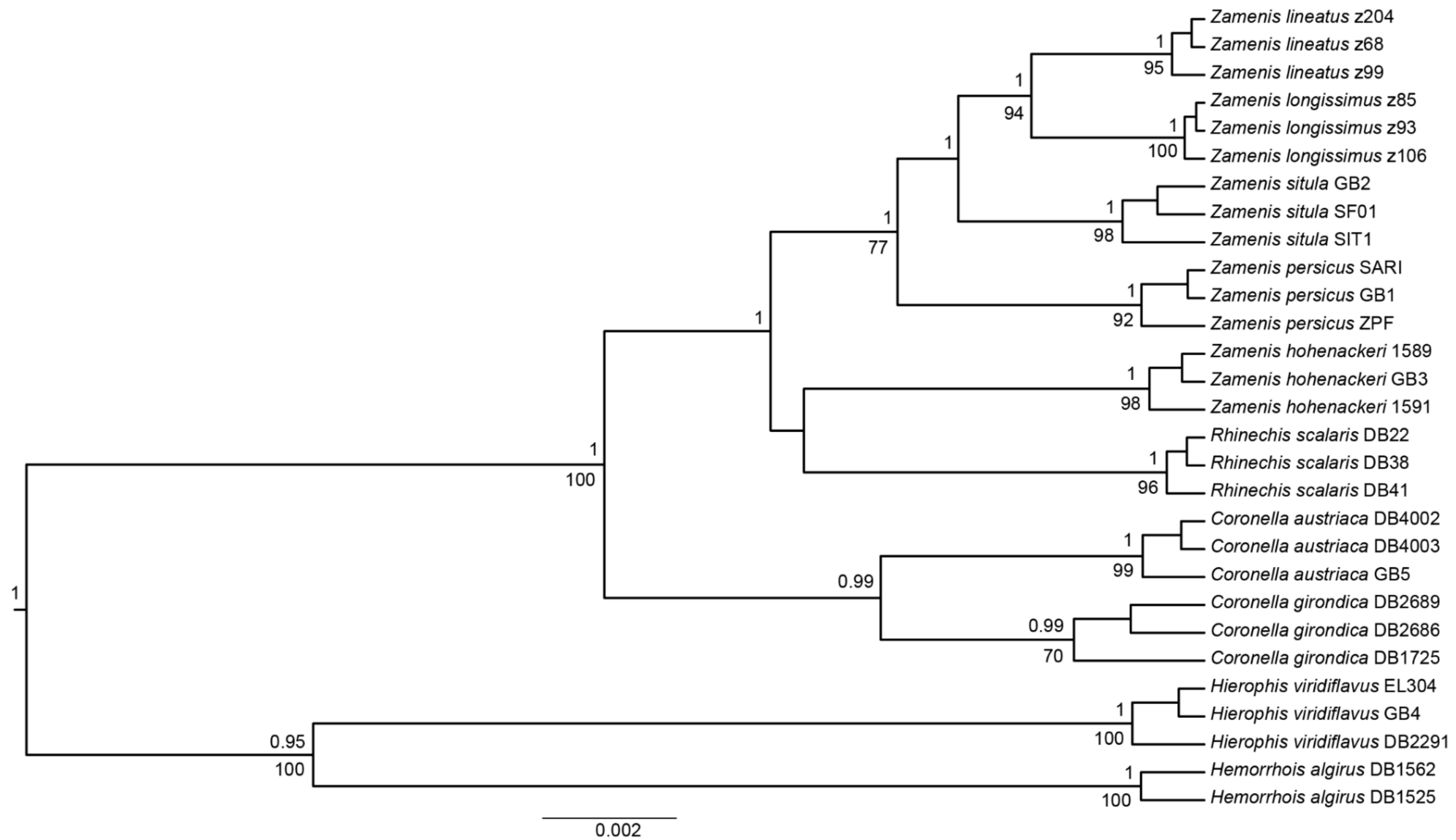


Figure S3.3. Bayesian tree depicting the phylogenetic relationships between *Zamenis*, *Rhinechis* and *Coronella* based on concatenated nuclear DNA sequences (*coms, dnah3, prlr, sptbn1* and *vim*). Bayesian posterior probabilities (BPP > 90) and bootstrap support (BS > 70) are represented above and below nodes, respectively.

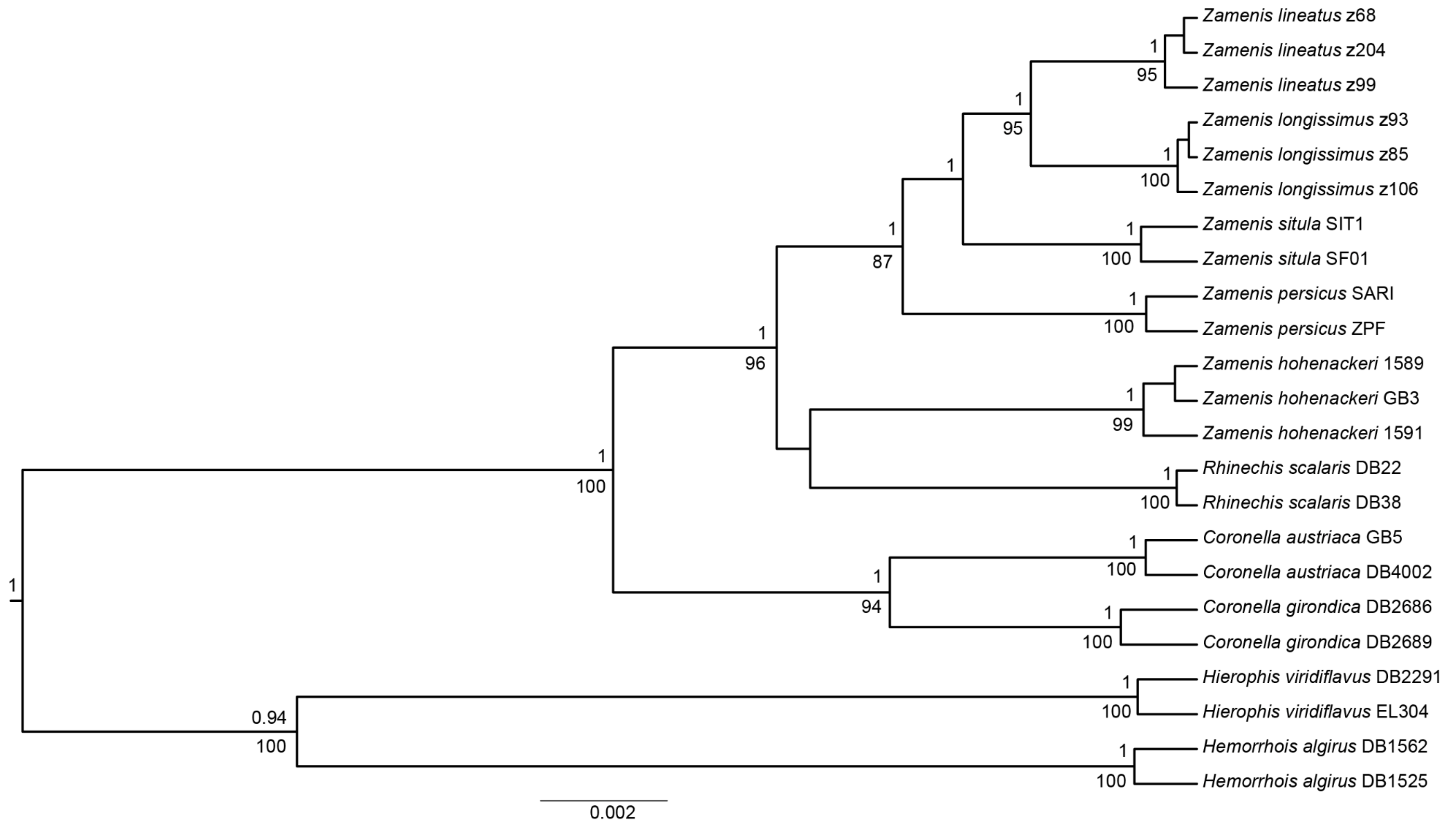


Figure S3.4. Bayesian tree depicting the phylogenetic relationships between *Zamenis*, *Rhinechis* and *Coronella* based on concatenated nuclear DNA sequences (*coms, dnah3, prlr, sptbn1* and *vim*) using the reduced dataset. Bayesian posterior probabilities (BPP > 90) and bootstrap support (BS > 70) are represented above and below nodes, respectively.

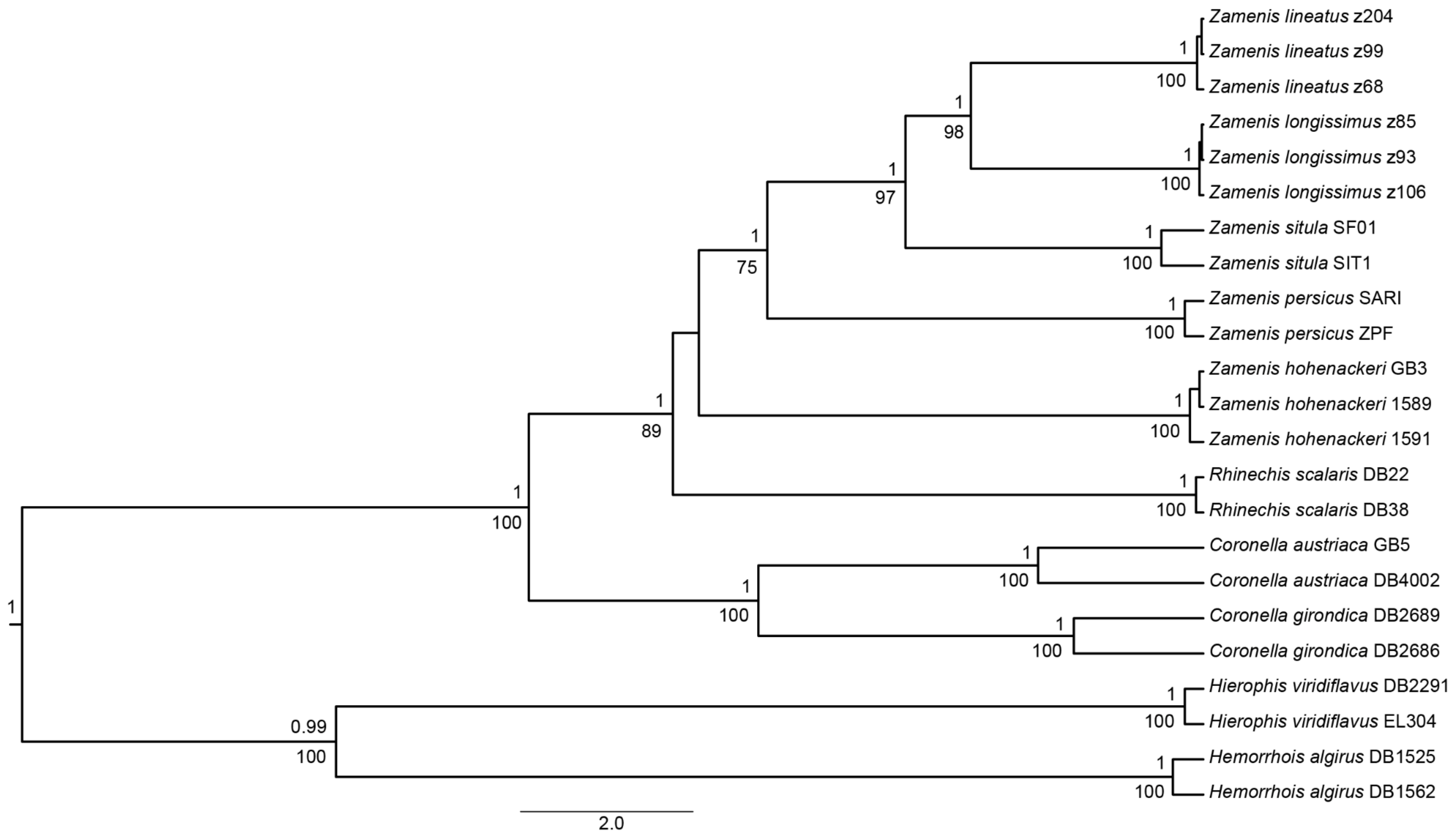


Figure S3.5. Bayesian tree depicting the phylogenetic relationships between *Zamenis*, *Rhinechis* and *Coronella* based on concatenated mitochondrial (*cytb* and *nd4*) and nuclear DNA sequences (*coms, dnah3, prlr, sptbn1* and *vim*) using the reduced dataset. Bayesian posterior probabilities (BPP > 90) and bootstrap support (BS > 70) are represented above and below nodes, respectively.

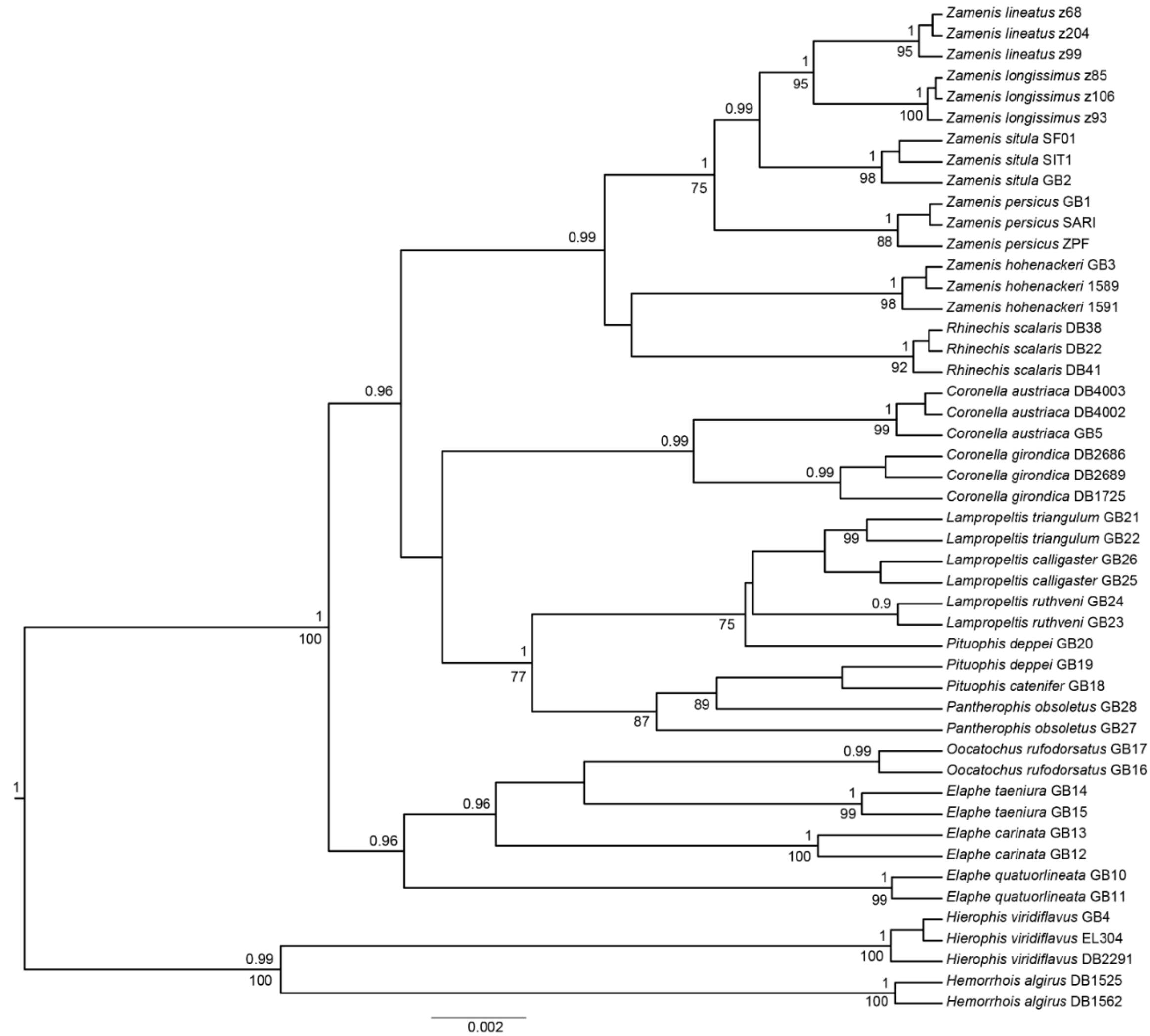


Figure S3.6. Bayesian tree depicting the phylogenetic relationships between *Zamenis*, *Rhinechis*, *Coronella* and representatives of five other ratsnakes' genera based on concatenated nuclear DNA sequences (*coms, dnah3, prlr, sptbn1* and *vim*). Bayesian posterior probabilities (BPP > 90) and bootstrap support (BS > 70) are represented above and below nodes, respectively.

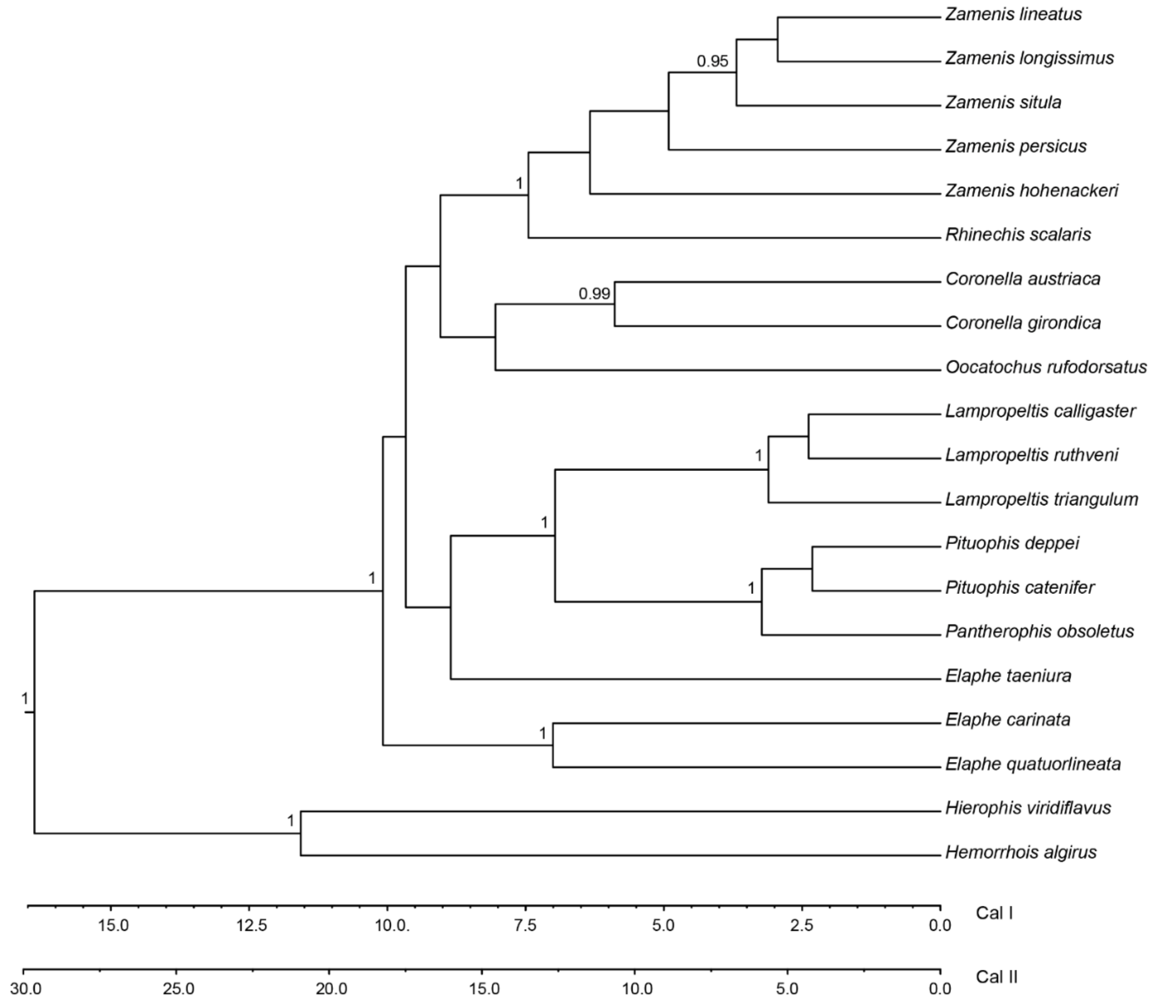


Figure S3.7. Species tree of *Zamenis*, *Rhinechis*, *Coronella* and representatives of five other ratsnakes' genera based on DNA sequences from mitochondrial (*cytb* and *nd4*) and nuclear (*cmos, dnah3, prlr, sptbn1* and *vim*) loci. Bayesian posterior probabilities (BPP>90) are represented above nodes. Divergence time estimates from Cal I and Cal II, in million years, are represented by each time axis.

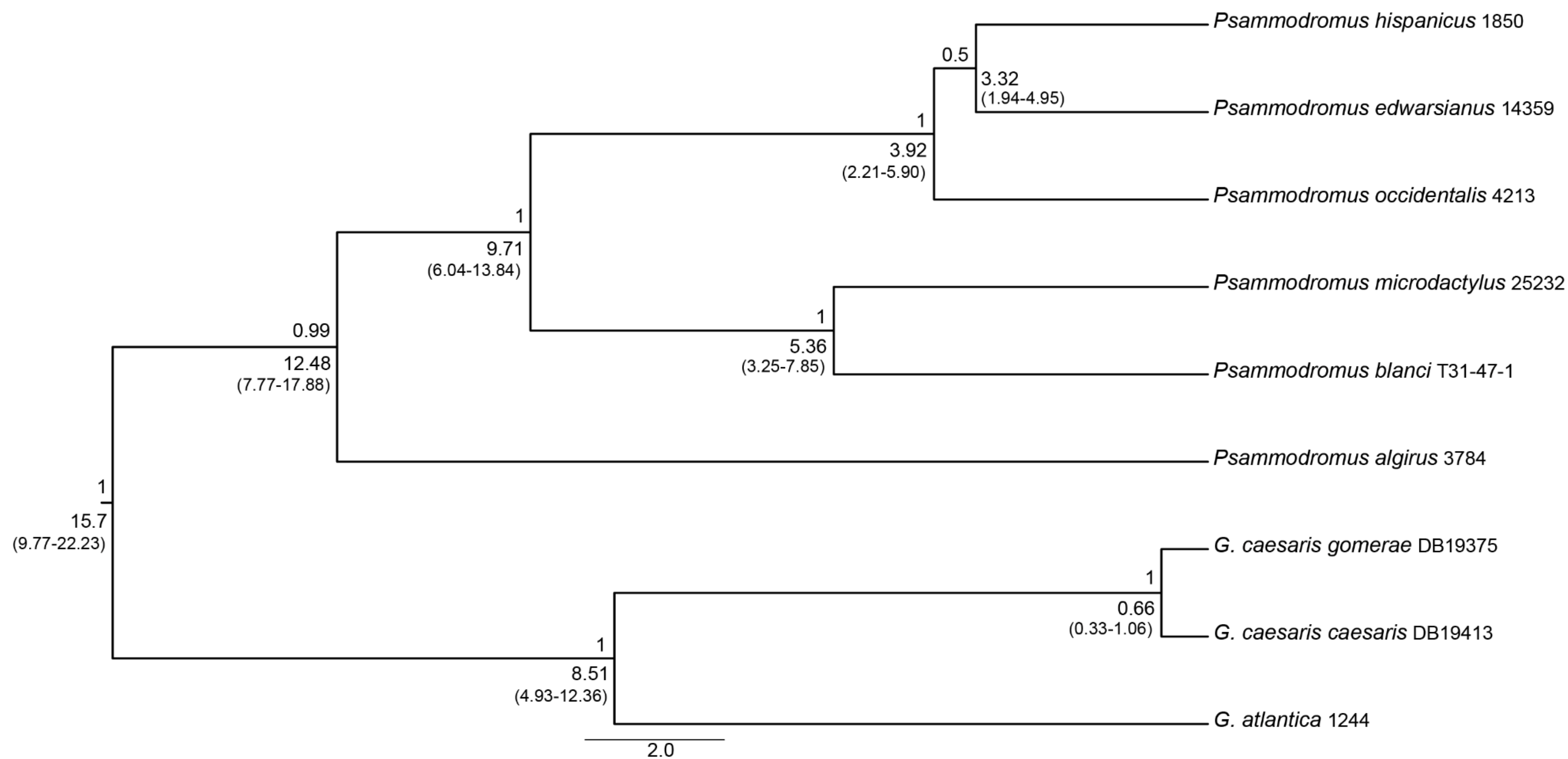


Figure S4.1. Bayesian tree depicting the phylogenetic relationships and divergence time estimates between *Psammodromus* species, inferred with only one individual per species, to test the sensitivity of the choice of the Yule tree prior, and based on concatenated mitochondrial (*12S*, *cytb*, *nd4*) and nuclear (*acm4*, *mc1r*, *pomc*) DNA sequences. Values above nodes represent Bayesian posterior probabilities (BPP). Values below nodes represent the estimated age of the node with associated 95% Highest Posterior Density interval (in parentheses) according to molecular dating analyses.

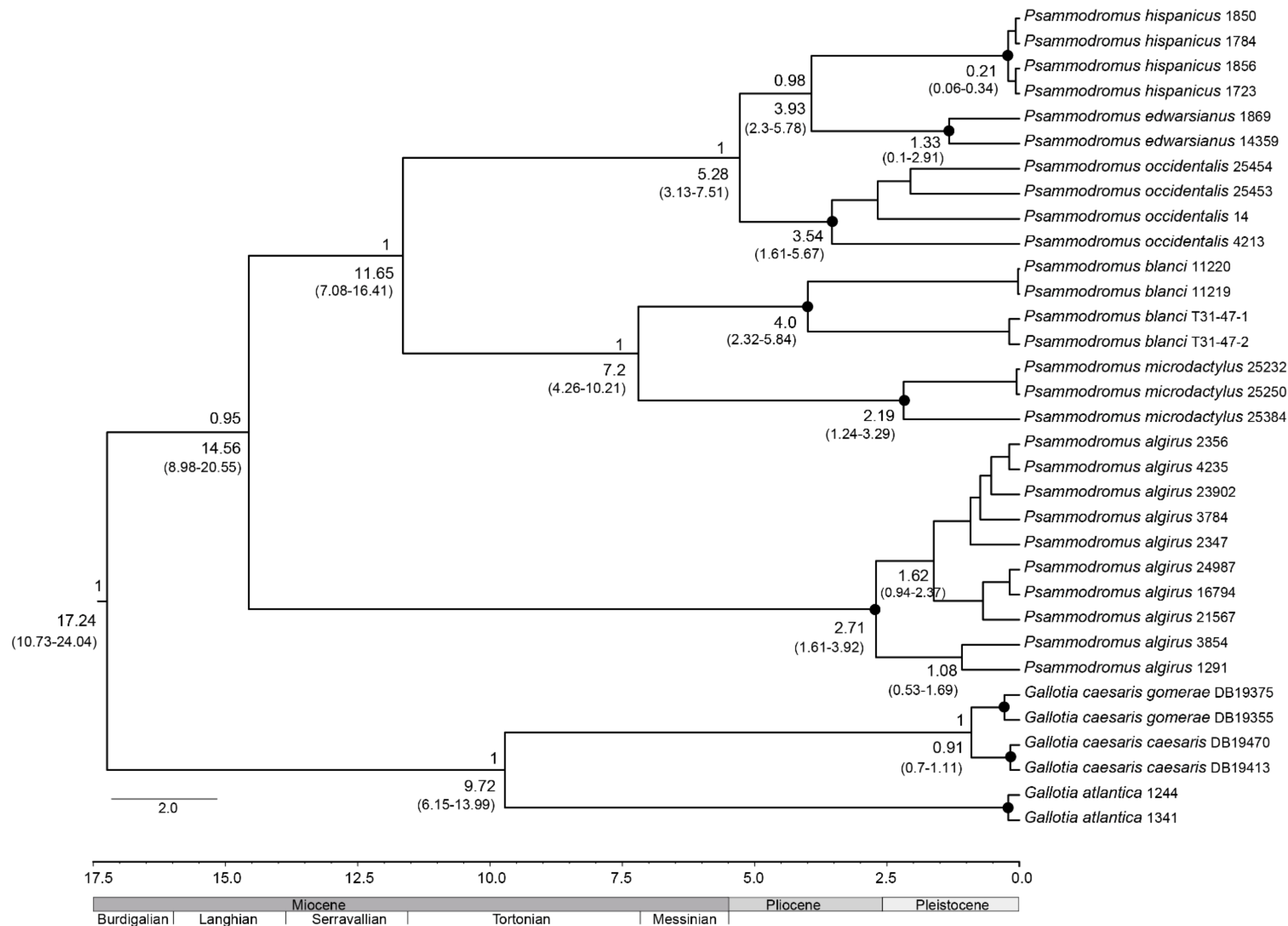


Figure S4.2. Bayesian tree depicting the phylogenetic relationships and divergence time estimates between *Psammodromus* species. Molecular dating was performed with *Gallotia* node priors only, as a cross-validation of the substitution rates applied in this study. Values above nodes represent Bayesian posterior probabilities. Node support for intra-specific clades is represented by black circles: BPP ≥ 0.98. Values below nodes represent the estimated age of the node with associated 95% Highest Posterior Density interval (in parentheses).

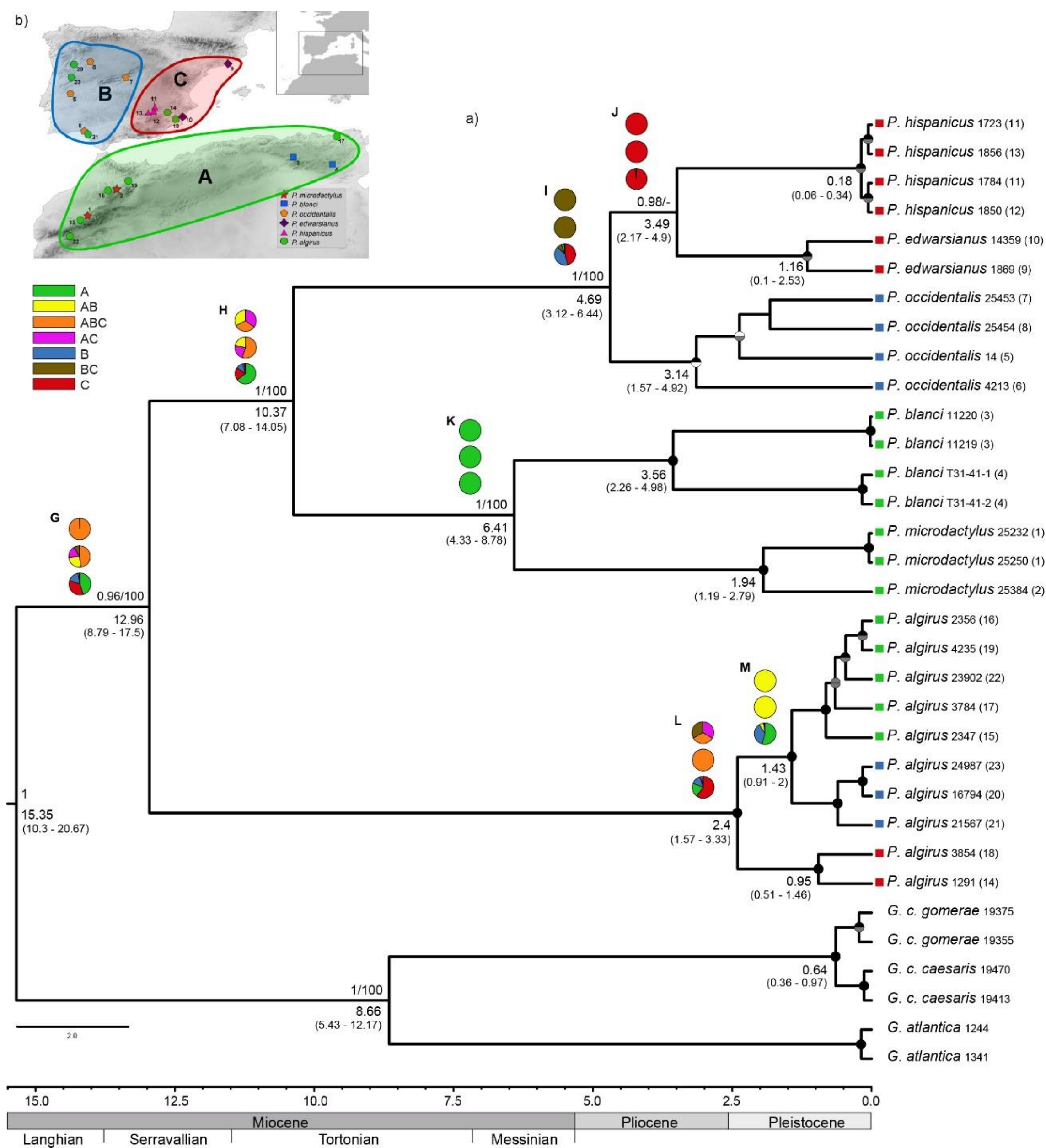


Figure S4.3. a) Bayesian tree depicting the phylogenetic relationships between *Psammodromus* lizards based on concatenated mitochondrial (*12S*, *cytb*, *nd4*) and nuclear (*acm4*, *mc1r*, *pomc*) DNA sequences. Values above nodes represent Bayesian posterior probabilities (BPP) ≥ 0.9 (left) and Maximum Likelihood bootstrap (BS) values ≥ 70 (right). Node support for intra-specific clades is represented by black circles: BPP ≥ 0.98 (upper half) and BS ≥ 95 (bottom half); grey circles: 0.9 ≤ BPP < 0.98 (up) and 70 ≤ BS < 95 (bottom) and empty circles for nodes with lower support. Values below nodes represent the estimated age of the node with associated 95% Highest Posterior Density interval (in parentheses) according to molecular dating analyses. Coloured squares represent the geographic origin of each tip sample in either Iberia or Africa according to the inset (b). Coloured pie charts in correspondence of nodes (which are named from G to M) represent results from ancestral range reconstructions based on S-DIVA (top), DEC (middle) and BBM (bottom) methods. Colour legend of ancestral areas represented in pie charts is at the top left. b) Map of the study area with sampling localities and the areas used for ancestral range reconstruction: A) Africa (green), B) Western Iberia (blue) and C) South-eastern Iberia (red).

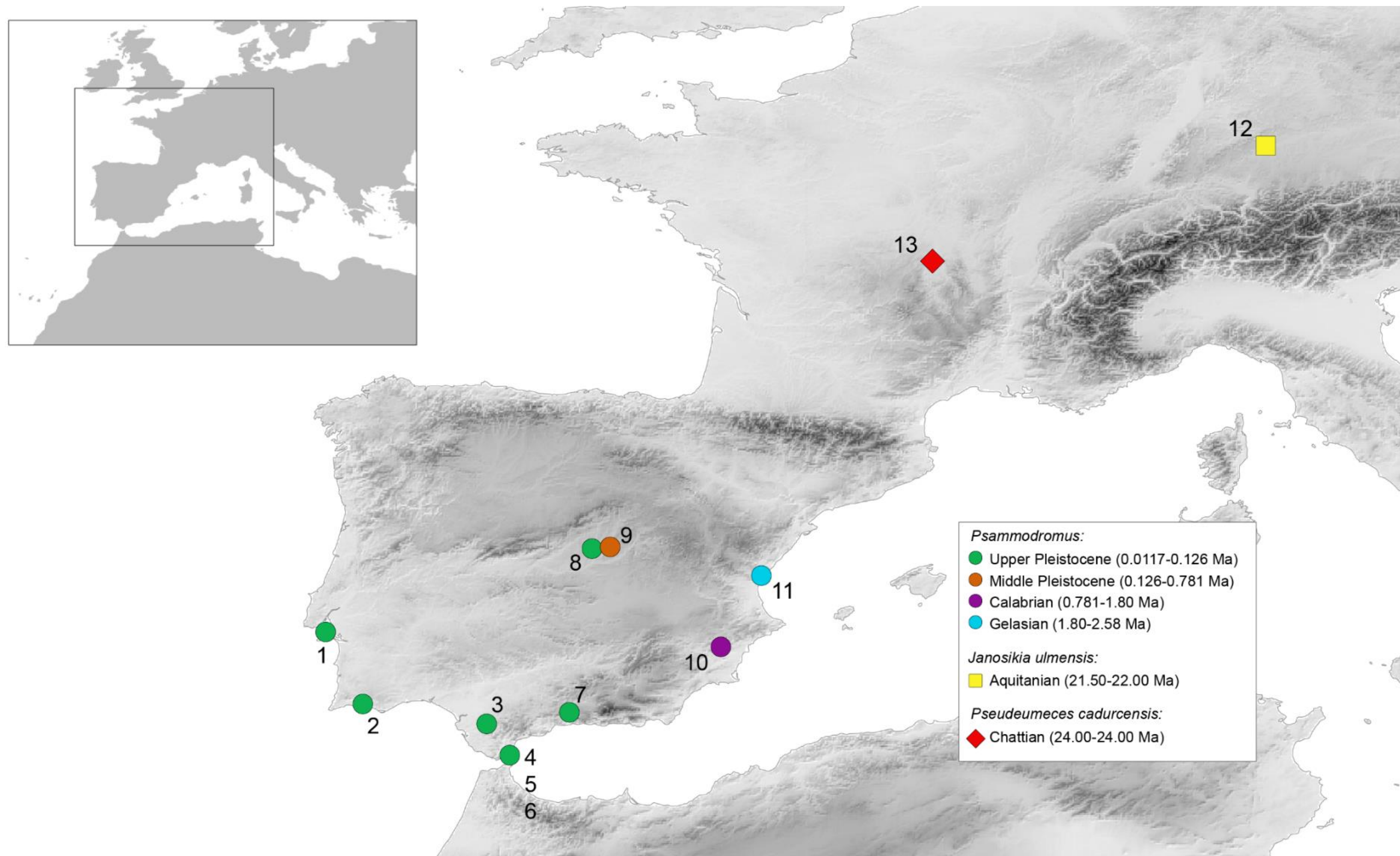


Figure S4.4. Distribution of the described fossils of *Psammodromus*, *Janosikia ulmensis* and one specimen of *Pseudeumeces cadurcensis.* Further information on the localities can be found in **Table S4.1.**

Table S4.1. Information of the described fossils of *Psammodromus* and of the species *Janosikia ulmensis* and one specimen of *Pseudeumeces cadursensis*, taxa from the total clade of *Gallotia*, including the map code, locality and geographic coordinates, minimum and maximum age in Million years (Ma), upper and lower geological stage and species. Data retrieved from the Database of vertebrates: fossil fishes, amphibians, reptiles, birds; available at www.wahre-staerke.com.

Map code	Locality	Latitude	Longitude	Ma min.	Ma max.	Upper Stage	Lower Stage	Species
1	Gruta da Figueira Brava, Arrábida, Portugal	38.568	-9.148	0.030	0.300	Upper Pleistocene	Upper Pleistocene	*Psammodromus algirus*
2	Guia, Albufeira, Portugal	37.136	-8.300	0.010	0.015	Upper Pleistocene	Upper Pleistocene	*Psammodromus* sp.
3	Higueral de Valleja Cave, Spain	36.750	-5.810	0.021	0.057	Upper Pleistocene	Upper Pleistocene	*Psammodromus* sp.
4	EnglandGorham´s Cave IIIb, Gibraltar Peninsula	36.120	-5.341	0.017	0.023	Upper Pleistocene	Upper Pleistocene	*Psammodromus* sp.
5	Vanguard Cave, Gibraltar Peninsula	36.120	-5.341	0.118	0.122	Upper Pleistocene	Upper Pleistocene	*Psammodromus* sp.
6	Vanguard Cave, Gibraltar Peninsula	36.120	-5.341	0.118	0.122	Upper Pleistocene	Upper Pleistocene	*Psammodromus algirus*
7	Cueva del Boquete de Zafarraya, Sierra de Alhama, Málaga, Spain	36.966	-4.133	0.010	0.127	Upper Pleistocene	Upper Pleistocene	*Psammodromus algirus*
8	Preresa, Perales del Rio, Getafe, 18 km SE Madrid, Spain	40.304	-3.573	0.079	0.089	Upper Pleistocene	Upper Pleistocene	*Psammodromus algirus*
9	Áridos 1, left bank Jarama river, Arganda, Madrid Procince, Spain	40.300	-3.433	0.374	0.424	Middle Pleistocene	Middle Pleistocene	*Psammodromus hispanicus*
10	Sierra de Quibas, Abanilla, Murcia, Spain	38.300	-1.050	1.300	1.400	Calabrian	Calabrian	*Psammodromus algirus*
11	Casablanca 1 (ACB-1), Almenara, Castellón, Spain	39.750	-0.216	1.800	1.900	Gelasian	Gelasian	*Psammodromus algirus*
12	Ulm-Westtangente, Germany	48.420	9.934	21.50	22.00	Aquitanian	Aquitanian	*Janosikia ulmensis*
13	Gannat, Allier, France	46.100	3.200	24.00	24.00	Chattian	Chattian	*Pseudeumeces cadurcensis*

(Table S5.1) Primers and PCR conditions used for the amplification of the genes in the *Omanosaura* study

DNA sequences were amplified through polymerase chain reaction (PCR) in 25µL volume, containing 1X PCR buffer (50mm Tris–HCl, 50mm NaCl, pH 8.5); 3mM $MgCl_2$; 0.6mM each dNTP, 2U of GoTaq DNA polymerase (Promega), 0.4µM each primer and approximately 50ng of genomic DNA. Amplification conditions consisted of a preliminary denaturation step at 94°C for 5 minutes, followed by 35 cycles of 30 seconds denaturation at 94°C, 40 seconds at varying annealing temperature according to the gene (48.5ºC for *cytb*; 50ºC for *12S*, *nd4* and *mc1r*; and 55ºC for *cmos*) and 60 seconds of extension at 72°C. A final extension was carried out at 72°C for 10 minutes. The primers used for amplification of each gene fragment are reported in Table S5.1.

Table S5.1. Name, sequence and reference of the primers used in this study.

Gene	Primer name	Primer sequence (5'-3')	Reference
12S	12SaGekko	CAA ACT AGG ATT AGA TAC CCT ACT ATG C	(Metallinou et al., 2015)
	12SbGekko	GAG GGT GAC GGG CGG TGT GTA C	
nd4	ND4	CAC CTA TGA CTA CCA AAA GCT CAT GTA GAA GC	(Arevalo, Davis, & Sites, 1994)
	LEU	CAT TAC TTT TAC TTG GAA TTT GCA CCA	
cytb	GluDG	TGA CTT GAA RAA CCA YCG TTG	(Palumbi et al., 1991)
	cytb2	CCC TCA GAA TGA TAT TTG TCC TCA	
cmos	S77	CAT GGA CTG GGA TCA GTT ATG	(Lawson, Slowinski, Crother, & Burbrink, 2005)
	S78	CCT TGG GTG TGA TTT TCT CAC CT	
mc1r	MC1RF	GGC NGC CAT YGT CAA GAA CCG GAA CC	(Pinho et al., 2009)
	MC1RR	CTC CGR AAG GCR TAG ATG ATG GGG TCC AC	

References:

Arevalo, E., Davis, S. K., & Sites, J. W. J. (1994). Mitochondrial DNA sequence divergence and phylogenetic relationships among eight chromosome races of the *Sceloporus grammicus* complex (Phrynosomatidae) in Central Mexico. *Systematic Bioloy*, 43, 387–418.

Lawson, R., Slowinski, J. B., Crother, B. I., & Burbrink, F. T. (2005). Phylogeny of the Colubroidea (Serpentes): New evidence from mitochondrial and nuclear genes. *Molecular Phylogenetics and Evolution*, 37, 581–601.

Metallinou, M., Červenka, J., Crochet, P.-A., Kratochvíl, L., Wilms, T., Geniez, P., Carranza, S. (2015). Species on the rocks: Systematics and biogeography of the rock-dwelling *Ptyodactylus* geckos (Squamata: Phyllodactylidae) in North Africa and Arabia. *Molecular Phylogenetics and Evolution*, 85, 208–220.

Palumbi, S. R., Martin, A., Romano, S., McMillan, S., Stice, W. O., & Grabowski, L. (1991). *The simple fool's guide to PCR* (1st ed.). Honolulu: Department of Zoology, University of Hawaii.

Pinho, C., Rocha, S., Carvalho, B. M., Lopes, S., Mourão, S., Vallinoto, M., Ferrand, N. (2009). New primers for the amplification and sequencing of nuclear loci in a taxonomically wide set of reptiles and amphibians. *Conservation Genetics Resources*, *2*, 181–185.

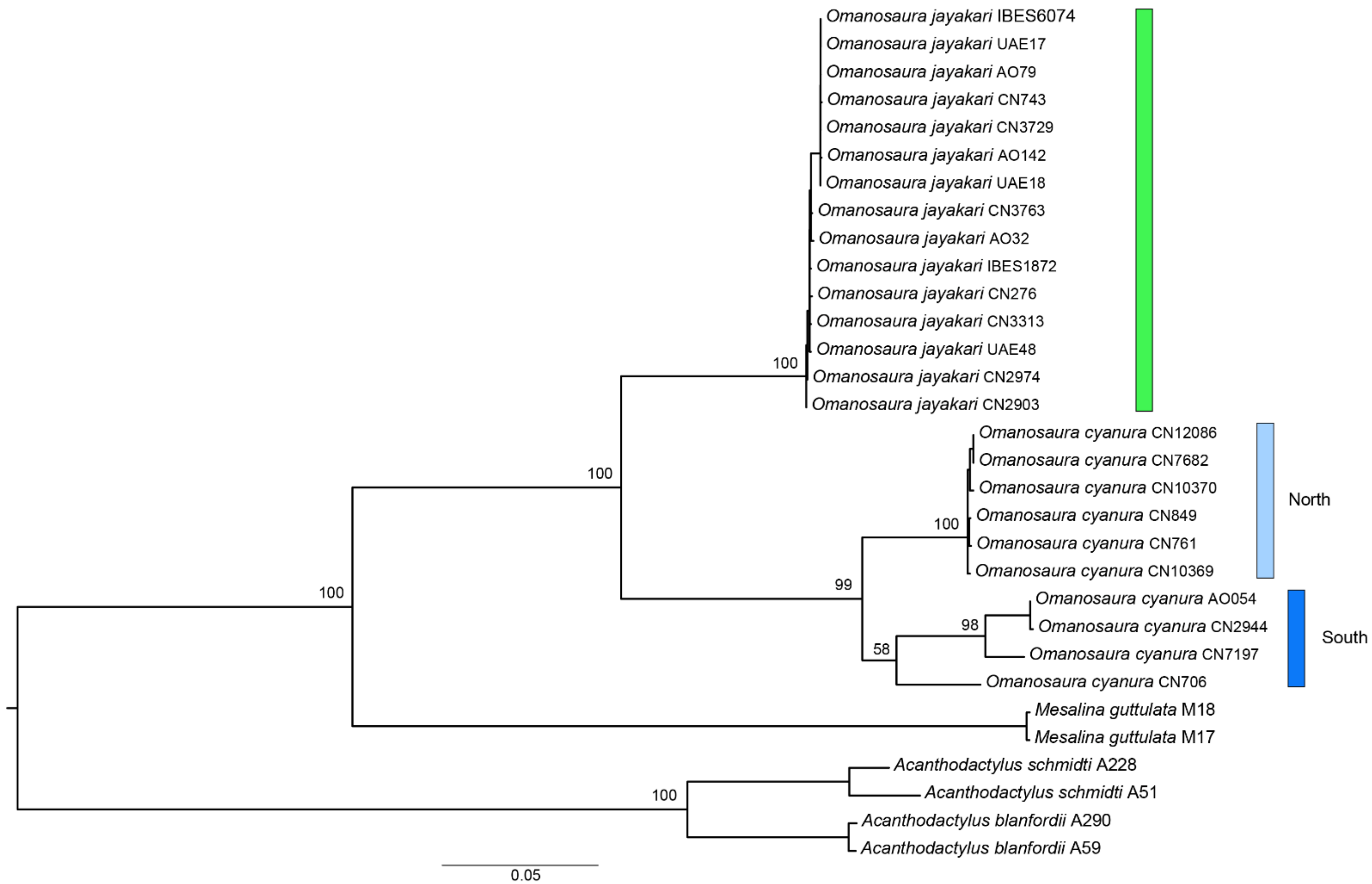


Figure S5.1. Phylogenetic relationships of *Omanosaura* inferred by maximum likelihood with the mitochondrial *12S*, *cytb* and *nd4*, and nuclear genes *cmos* and *mc1r*. Bootstrap support values are represented above nodes.

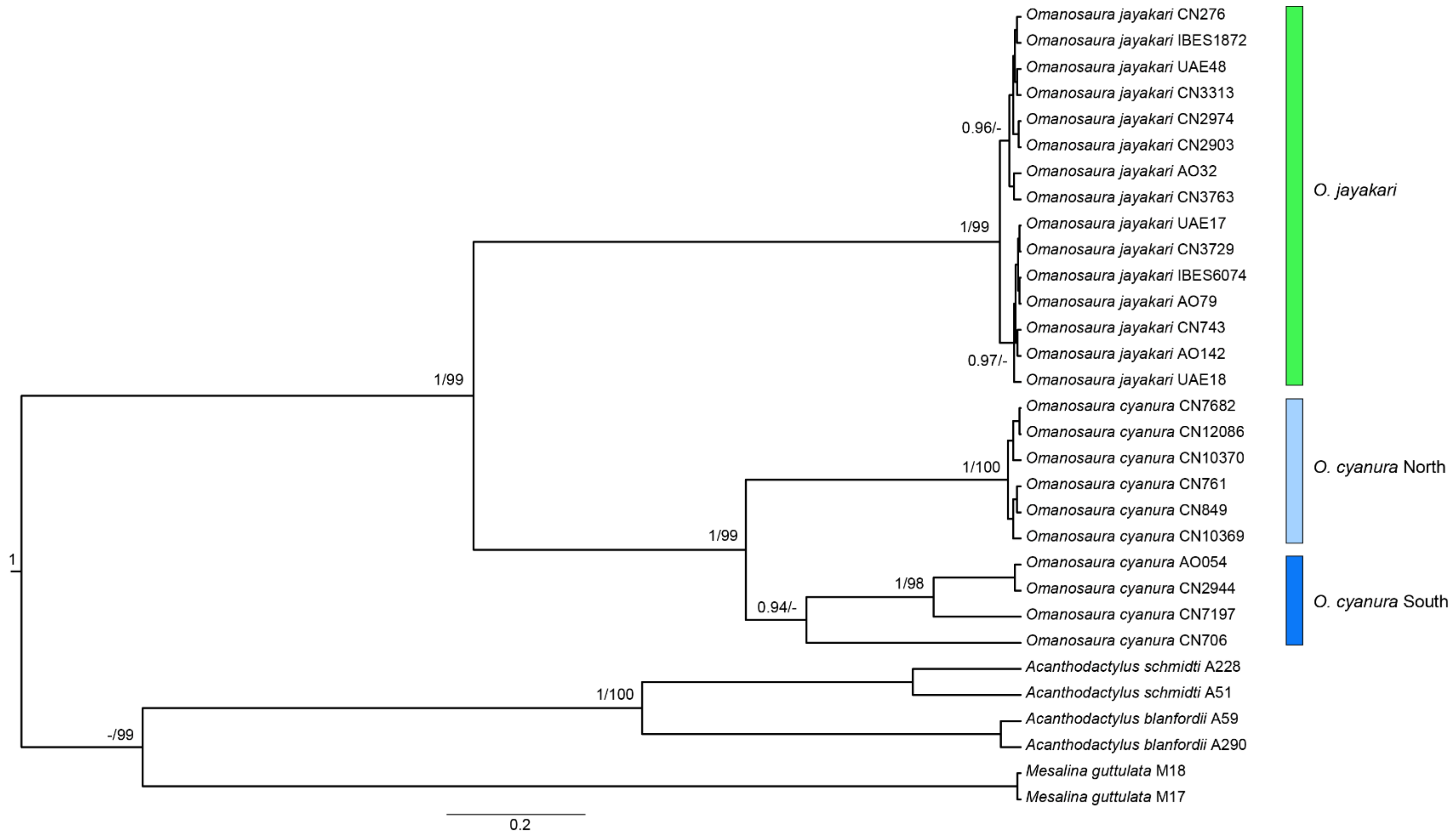


Figure S5.2. Bayesian tree depicting the relationships of *Omanosaura* inferred from concatenated mitochondrial sequences (*12S*, *cytb* and *nd4*). Bayesian Posterior Probabilities (left) and maximum likelihood Bootstrap Support (right) are represent above nodes.

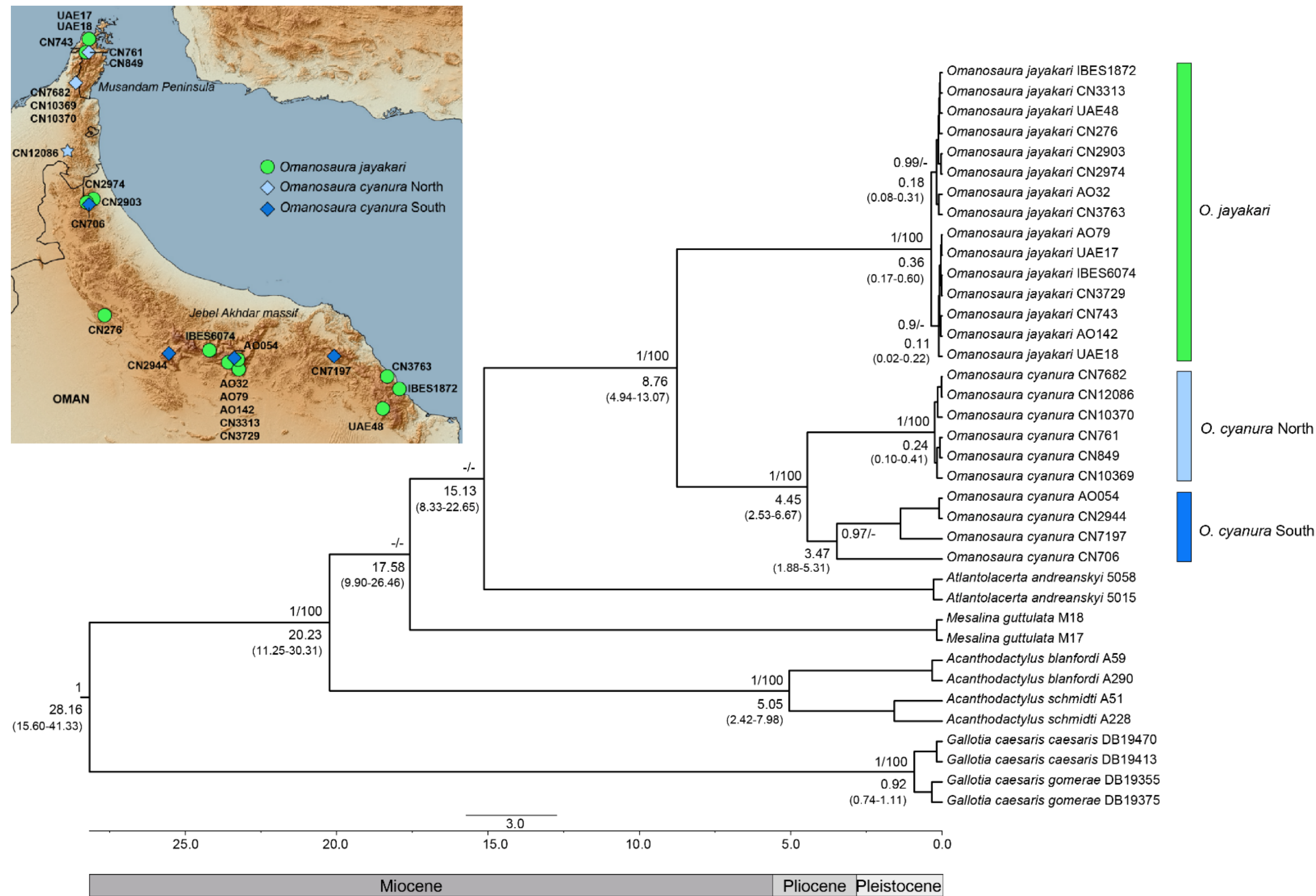


Figure S5.3. Bayesian tree depicting the phylogenetic relationships and divergence time estimates between *Omanosaura* species. Molecular dating was performed with *Gallotia caesaris* node priors only, as a cross-validation of the substitution rates applied in this study. Values above nodes represent Bayesian Posterior Probabilities (left) and Bootstrap Support (right). Values below nodes represent the estimated age of the node with associated 95% Highest Posterior Density interval (in parentheses).