Why Sex?
Darevskia answers

Susana M. R. N. de Freitas

Programa Doutoral em Biodiversidade, Genética e Evolução

Departamento de Biologia

2017

Orientador

Miguel A. Carretero, Professor Auxiliar e Investigador, Departamento de Biologia, Faculdade de Ciências da Universidade do Porto

Coorientador

D. James Harris, Professor Auxiliar e Investigador, Departamento de Biologia, Faculdade de Ciências da Universidade do Porto

Coorientador

Roger K. Butlin, Professor of Evolutionary Biology, Animal and Plant Sciences Department of the University of Sheffield

FCUP
Why Sex? *Darevskia* answers.

ai será que a vida quer ser vivida,
será que não,
será que o sexo é só pra procriar
ou é só para armar confusão
(J.P.S.)

ABSTRACT

Considering all costs associated to sex, sexual reproduction could be expected to be rare. However, we find sexual reproduction to be pervasive in nature, and only a reduced number of taxa are completely asexual. In fact, the variety of sex reproductive systems, mating techniques and traits, reveals that the organisms' reproduction has been an expressive target of selection and adaptation throughout evolution.

Sex is selected because it facilitates adaptation in complex and changing environments (Weismann, 1889), and this has been already shown empirically in several models. However, studies and theories developed concerning the paradox of sex will not be able to clarify the current distribution and abundance of asexual lineages in nature. The understanding of rates of origin and extinction of asexual lineages, together with the putative differential evolution rates in organisms with sexual or asexual reproduction, should be considered in the main "origin of sex" question.

The model used here, the genus *Darevskia,* is the first true parthenogenetic vertebrate described. It is a highly speciose genus, currently including 27 species, of which seven are parthenogenetic and of hybrid origin. Only a few of the sexual species contributed for the hybridization events which originated the parthenogens clonal lineages, and these were directional: the maternal species were always *Darevskia raddei or D. mixta*, and the paternal always *D. valentini* or *D. portshinskii.*

Using a set of microsatellite and mtDNA markers, we first start to study the role of hybridization in the origin of vertebrate parthenogens, the extent of gene flow with sexual relatives and the relation between hybridization, asexuality and polyploidy, in the context of the main theories of asexual evolution proposed to date. We focus on the *D. unisexualis*, *D. uzzelli* and *D. armeniaca* parthenogens and polyploid backcrosses found in sympatric locations between the parthenogens and its sexual parentals. We find that only specific parental pairs are responsible for the origin of vertebrate parthenogenesis, regardless of their phylogenetic distance. Despite the recurrent hybridization presently reported in *Darevskia*, asexuality originated multiple times but only in a single temporal event the past. Parthenogenetic females are capable of backcrossing with sexual males, but can only produce polyploid individuals with reduced fertility that do not contribute for gene flow between parthenogens and sexuals. The sexual reproduction machinery can be lost with time in the

parthenogenetic hybrids, which are most likely the result of a post-zygotic reproductive barrier to gene flow in the speciation continuum of *Darevskia* sexual species.

In order to reconstruct the phylogenetic inference of this group, a set of capture sequence probes were designed from a transcriptome de novo assembly. Phylogenetic inference reconstruction has been extensively used to ask several evolutionary biology questions. However, despite the exponential use of high-throughput sequencing technologies and its promising applications in phylogenetics, next generation sequencing (NGS) techniques are still far from being as widely used in this field as in other areas of evolutionary genetics. The workflow developed in this thesis has proved to be not only cost-effective, but also to produce a very high number of cross genome phylogenetically informative markers in this non-model species.

Finally, taking advantage of the hundreds of cross genome markers developed, a species tree was constructed and within genus relationships evaluated. Introgression tests showed a widespread pattern of gene flow across the genus, not only recent between recently diverged clades, but also both recent and ancient gene flow between early diverging *Darevskia* clades. Several evidences of hybridization with gene flow between diverging clades have been found in other systems. However, such widespread patterns where introgression has happened between all major clades, and in a very high number of species pairs, are not frequently found. It is also found that gene flow is completely absent between parthenogens sexual parental pairs. In groups with sexual-parthenogenetic reproducing species, pre- or post-zygotic reproductive barriers could be more difficult to attain and hybrid asexuality can, thus, effectively create reproductive barriers between diverging sexual taxa, that were likely faced with recurrent secondary contact.

The work presented in this thesis shows that only a few sexual species pairs were responsible for the origin of the parthenogens, that some species acted always as the maternal species and others always as the paternal, that the phylogenetic distances between each pair are included in a wide range interval, and that introgression is found widespread across the genus but absent between the parental pairs.

Regardless of the reproduction type, this study also shows the relevance and depth of hybridization during divergence. As has been shown intensively, gene flow is common during speciation and clades divergence. However, it would be interesting to analyse other groups together and test if deep branch gene flow is found only in some exclusive groups (such as sexual-asexual genera) or if it is something more widespread.

This work presented here moves *Darevskia* towards providing a model that can be used to understand the origin of asexuality in vertebrates. Here the basis for future studies about sex and asexuality is provided, so that a focus on the differential evolution rates genomes of both reproduction mode can be analysed.

Key words

Sex, parthenogenesis, *Darevskia*, hybridization, gene flow, introgression, incomplete lineage sorting, phylogenomics, transcriptome, capture sequence

RESUMO

Tendo em conta todas as desvantagens associadas à reprodução sexuada, podíamos esperar que esta fosse improvável. No entanto, não só é altamente abrangente na natureza, como a reprodução assexuada, apesar da sua capacidade mais elevada de produção de descendência, apenas se encontra num reduzido número de organismos. A variedade dos sistemas reprodutores sexuados, e das técnicas e comportamentos associados à copulação, revela que a reprodução sexual tem sido alvo de selecção e adaptação ao longo dos tempos.

Vários estudos empíricos demostraram que o sexo é seleccionado já que facilita a adaptação em ambientes complexos e instáveis. No entanto, os estudos e teorias sobre o paradoxo do sexo não explicam a distribuição e frequência de organismos que se reproduzem na ausência de sexo, mas estes não podem ser separados de estudos sobre a frequência de origem e extinção de linhagens assexuadas, bem como de estudos sobre potenciais diferenças nas taxas de evolução de organismos com os diferentes tipos reproductivos.

O modelo usado aqui, o género *Darevskia*, foi o primeiro vertebrado onde a partenogénese verdadeira foi descrita. É um género com um elevado número de espécies, cerca de 27, das quais sete são espécies híbridas com reprodução exclusivamente partenogenética. Estas foram originadas pelo cruzamento de *Darevskia* sexuadas, e apenas algumas contribuíram para estes cruzamentos: *Darevskia raddei* e *D. mixta* actuaram sempre como espécies maternais, e *D. valentini* e *D. portschinskii* como espécies paternais.

Recorrendo a um conjunto de microssatélites e ADN mitocondrial, analisou-se a importância da hibridação na origem destas espécies de vertebrados partenogenéticos, da dimensão do fluxo génico com as espécies sexuadas e a relação entre hibridação, assexualidade e poliploidia, no contexto das teorias conhecidas sobre a evolução da partenogénese usando as espécies partenogenéticas *D. unisexualis, D. uzzelli* e *D. armeniaca*, e híbridos poliplóides resultantes do cruzamento entre fêmeas partenogenéticas e machos de espécies sexuadas quando se encontram em simpatria. Com isto, confirmou-se a especificidade das combinações de espécies sexuadas parentais, que a distância filogenética entre elas não parecia ser relevante e que apesar da origem da partenogénese nas *Darevskia* ter resultado de múltiplos

cruzamentos entre as espécies sexuadas, estes apenas ocorreram num curto intervalo de tempo no passado, e não estão a ocorrer no presente. As fêmeas partenogenéticas são capazes de se cruzar sexualmente com machos de outras espécies, mas apenas dão origem a híbridos poliplóides com fertilidade quase negligente, já que não contribuem para fluxo génico entre as duas espécies. A maquinaria da reprodução sexuada pode ser comprometida ao longo do tempo nas linhagens partenogenéticas, que são muito provavelmente o resultado de barreiras reprodutivas pós-zigóticas que surgiram durante a especiação das *Darevskia* sexuadas.

De forma a estudar as relações filogenéticas deste grupo, foi desenhado um conjunto de sondas capazes de capturar determinadas regiões do genoma, a partir da reconstrução do transcriptoma. Apesar do aparecimento e rápida propagação de técnicas de sequenciação em massa em várias áreas da genética evolutiva, estas não têm sido frequentemente usadas em estudos filogenéticos. Isto deve-se ao facto destas técnicas estarem mais adaptadas à detecção de polimorfismos sem muita homologia nos diferentes indivíduos, e não à sequenciação de segmentos do genoma que podem ser então usados em análises filogenéticas mais comuns. Aqui apresenta-se um método economicamente viável, rápido e eficaz para a detecção de sondas e posterior sequenciação de várias centenas de loci variáveis e espalhados pelo genoma, num grupo de organismos pouco estudado.

Finalmente, recorrendo a estes novos marcadores, foi construída uma árvore filogenética para o género e as relações entre espécies sexuadas foram avaliadas. Testes de introgressão detectaram níveis muito elevados de fluxo génico distribuído por todo o grupo, com a excepção dos pares das espécies parentais dos híbridos partenogenéticos. O fluxo génico foi detectado não só em espécies que se separaram há pouco tempo, mas também entre clados que divergiram há mais tempo. Vários estudos demonstraram já indícios de fluxo génico entre diferentes organismos do mesmo grupo, apesar da frequência deste fluxo não ser normalmente tão abrangente. Para além disso, a ausência de fluxo génico entre os pares de espécies parentais das partenogenéticas demonstra que em grupos com espécies sexuadas e assexuadas, as barreiras reprodutivas pre- e pós-zigóticas podem não ser atingidas facilmente e a assexualidade híbrida surgir como uma barreira reprodutiva altamente eficaz entre espécies sexuais em diferentes estádios de divergência, e que possivelmente sofreram vários episódios de contacto secundário no passado.

Os resultados apresentados nesta tese demonstram que apenas algumas espécies sexuadas fizeram parte dos cruzamentos que originaram os híbridos partenogenéticos,

que algumas dessas espécies actuaram sempre como a espécie maternal, e outras sempre como a espécie paternal, que as suas distâncias filogenéticas não parecem seguir nenhum padrão e que houve fluxo génico entre várias espécies deste género, tanto recente como mais antigo. Independentemente do tipo de reprodução destes organismos, aqui também é demonstrada a importância e extensão da hibridação e fluxo génico durante a separação dos diferentes taxa. Seria interessante fazer este tipo de estudo noutros grupos de organismos, e não só centrando o estudo em algumas espécies, para testar se este padrão se encontra em grupos que apresentam diferentes tipos de reprodução, ou se, pelo contrário, é mais prevalente.

O trabalho desenvolvido nesta tese apresenta o género *Darevskia* como um modelo que poderá ser usado para estudar a origem da assexualidade nos vertebrados. Aqui foram criadas as bases para mais estudos sobre o paradoxo do sexo, para que se possa centrar esses estudos nas consequências evolutivas da sexualidade/assexualidade e na genómica evolutiva do sexo.

Palavras chave

Reprodução sexuada, partenogénese, *Darevskia,* hibridação, introgressão, fluxo génico, polimorfismo ancestral, filogenómica, transcriptoma

TABLE OF CONTENTS

LIST OF FIGURES AND TABLES

ABBREVIATIONS

bp – base paired-end

BI – Bayesian Inference

CF – Congruence Factor

DNA – Deoxyribonucleic Acid

F_{ST} – Fixation index

GPS – Global Positioning System

HWE – Hardy & Weinberg Equilibrium

ILS – Incomplete Lineage Sorting

indel – insert/deletion

kyrs – thousands of years

LGM – Last Glacial Maximum

LIG – Last Inter Glacial

Mya - million years ago

mtDNA – mitochondrial DNA

N_e - effective population size

NGS – Next Generation Sequencing

SNP – Single Nucleotide Polymorphism

ML – Maximum Likelihood

MNP – Multiple Nucleotide Polymorphism

RNA – Ribonucleic Acid

VCF – variant call format

yrs – years

FCUP
Why Sex? *Darevskia* answers.

CHAPTER 1

GENERAL INTRODUCTION

REPRODUCTION IN NATURE

Sex

Sex has been fascinating humans since the beginning of times. The variety of sexual reproductive systems, mating techniques and traits, reveal this fascination is not restricted to the human species, and we can largely assume sex has been an expressive target of selection and adaptation since it first was originated.

The word sex comes from the latin *sectus*, which means cut/divide, and derives from the verb *sectare* (cut, split, divide). Going further back, the proto indo european *sek*, which means cut, is thought to have been a primitive form of the latin form. Sex alludes to the separation of the human species into two groups, female and male.

Sex as is scientifically defined is the formation of a new individual by the union of two cells. However, and as the concept of *species,* the scientific concept of *sex* has still not arrived to a generally accepted agreement and there is still interchangeable use of the term sex with concepts such as automixis or meiosis (Archetti, 2010; Beukeboom and Perrin, 2014; Mogie, 1986). Here, *sex* (or sexual reproduction) is defined as the genetic exchange and the formation of a new individual by the union of two cells originally formed by meiosis. Following this definition, bacterial conjugation is thus not considered sex since it does not show any evolved process to randomize chromosomal alleles between different individuals (Redfield, 2001). Processes like conjugation, transduction and transformation, despite being responsible for the movement of alleles between individuals, did not evolve for sex but instead as mechanisms of infectious transfer (like plasmids and phages)(Redfield, 2001). The absence of a clear evidence of selection for the ability to cause homologous recombination shows 'sex', or 'non-canonical sex' (Schwander, 2016) is only a secondary trait resultant from adaptation of evolutionary tools not related with exchange of chromosomes and production of fused cell (Redfield, 2001).

If sex involves the fusion of reproductive specialized cells (gametes), they can either be morphologically similar (isogamy) or differentiated, one carrying most of the reserves for the future zygote (anisogamy). In the latter, female and male gametes, ovule and sperm, respectively, can be produced by the same individual (hermaphrodite) or by individuals with reproductive systems specialised in producing either ovules (females) or sperm (males), a gonochorist system. So, as is defined here, *sex* has to include two processes, 1) the fusion of two germinal cells (syngamy) which will double the genome, and 2) meiosis, which will reduce the genome and ensure recombination. It is the most common

mode of reproduction in eukaryotes.

BOX 1 - Asexual reproduction in animals

Apomixis: a normal meiosis is absent
- Mitotic apomixis: a normal mitosis
- Meiotic apomixis: a meiosis in which one division is supressed
  * Suppression of the first division
  * Suppression of the second division

Automixis: a normal meiosis followed by fusion of the nuclei:
  * Cleavage nuclei (generated by the same pronuclei)
  * Sister nuclei (generated by the same division)
  * Nonsister nuclei (generated by the other division)

Endomitosis: a normal meiosis preceded by a replication
  * Sister chromosom pairing (generated by the same replication)
  * Nonsister chromosome pairing (not generated by the same replication

Figure 1.1 - Summary of different types of asexual reproduction concerning meiosis found in animals.

On the other hand, all reproduction taking place in the absence of sex is defined as asexual reproduction. Several taxa reproduce asexually, either obligatory, facultative or in specific conditions (spontaneously). Asexual reproduction is present in all major clades in the tree of life but is found in only 0.1% of vertebrates (Vrijenhoek, 1989). This mode of reproduction can happen through different ways (summary of asexual reproduction in animals in Fig 1.1), which can be reduced into two types: clonal propagation (vegetative reproduction in plants and budding in animals and fungi) and asexual reproduction with egg formation. The first is a process by which new organisms arise by the detachment of a part of the main organism, producing clonal offspring and always in the absence of meiosis. The second uses the reproductive machinery to produce eggs only from the ovule, and meiosis may be involved or not, and in vertebrates is largely identified as *parthenogenesis*.

The origin and the evolution of sex are some of the biggest puzzles in biology and have been subject of scientific research through time (Bell, 1982; Maynard Smith, 1978; Otto, 2009). In a sexual population (with two sexes), if a mutant arises and causes females to produce daughters asexually it will double in frequency each generation, giving this mutation high advantage over the sexual individuals. The fact that this costly reproductive

mode is the most common mode of reproduction in Nature and that the existence of these mutations does not extinguish sexual individuals/populations are the basis of what Maynard Smith coined as the paradox of the “cost of sex” (Maynard Smith, 1971).

At what cost does sex come?

As we know it today, sexual reproduction can be very costly for the females, and in some cases also for the males. The most immediate cost of sex consists on the twofold disadvantage of producing males (Maynard Smith, 1971).

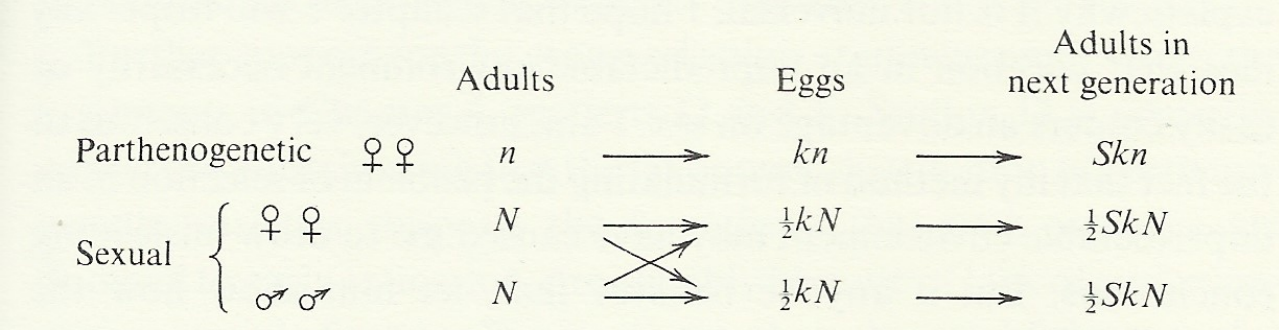


Figure 1.2: Adapted from "The Evolution of Sex", by J. Maynard Smith,1978, Cambridge University Press, Alden Press, Oxford

In a sexual population with equal number of females and males, *n*, if a mutation arises and causes females to produce asexually, the number of reproductive individuals in the following generation (considering non-overlapping generations) will be *n*. However, if that mutation does not occur the number of reproductive individuals in the next generation will be *n*/2. The number of offspring each female will bring, *k*, will not depend on the fact she reproduces sexually or asexually, but on the amount of resources she will be able to get. Also, the probability of an offspring to survive into adulthood, *S*, will also not depend on their reproductive mode. Thus, in the population where the mutation occurs, the population can double its number each generation. This gives asexual populations an immediate short-term advantage, and once it is established it can outcompete and replace the sexual ancestors. Several empirical evidences of asexual species outcompeting their sexual relatives have been found repeatedly in Nature (Barron et al., 2016; Lavanchy et al., 2016; Tarkhnishvili et al., 2010).

Other costs may affect the physiological level, since meiosis is highly more costly to the

cell than mitosis, it takes longer to achieve and produces fewer cells in the same amount of time and resources used (Lehtonen and Kokko, 2012). Mating can bring additional effects on female fitness, such as sexual harassment by males (Fig. 1.3), attraction of predators to the female (Han and Jablonski, 2010; Rowe, 1994), introduction of the risk of sexually transmitted diseases and parasitic genetic elements (Otto and Lenormand, 2002), and reduction of fitness and life expectancy (Huyghe et al., 2012). Finding and securing a mate can also represent a huge effort of resources allocation, such as for instance floral display and nectar rewards to ensure pollination (Otto and Lenormand, 2002).

BOX 2 - Sexual harassment by males

Bufo bufo - polyandry was suggested to have evolved in B. bufo and other herp species as to avoid male harassment (Lacerta vivipara: Fitze et al., 2005; Bufo bufo: Sztatecsny et al., 2006; Uller and Olsson, 2008). Multiply amplexing males can reduce fertilization success and multiple paternity does not appear to improve offspring fitness or be advantageous to females (Byrne and Robert, 2000). In anurans, the frequency in which polyandry occurs is dependant on the operational sex ratio (OSR), since it directly influences which sex is competing for access to mating partners: When in low densities, males attract females through advertisement calls and proportion of multiple amplexi is low. At male-biased OSR and during the peak of the breeding season, male pressure increases and attempts of multiple males to mate with a single female occur frequently (Höglund and Robertson, 1988). Females are frequently subjected to coercion and male-male competition, which can lead to several males amplexing one female and eventually leading to her drowning through the harassment of several competing males (Sztatecsny et al., 2006).

Thamnophis sirtalis parietalis - Red-sided garter snakes can be found in Manitoba (Canada) and due to very low winter temperatures they return to communal dens in autumn and emerge 8 months later in spring. Adult males normally remain around the den for 2 weeks to court females in "mating balls" that may contain more than 100 males (Shine et al., 2004). It has been shown that courtship slows the speed of female snakes, when leaving the den in spring, and adapt to avoid areas that are "male-scented", thus avoiding the presence of males and courtship (Shine et al., 2004). This study also proposes that juvenile females mate early as a way of dispersing from the den without harassment, since copulatory fluids serve as sex pheromones in some populations, discouraging male courtship. The sexual conflict can go as deep as suffocation of females by the male "mating balls" (Shine et al., 2001). Snakes do not commonly present male-female aggressive courtship elements (such as biting). The discovery that courting male garter snakes impair female dispersal and induce physiological stress, and that females attempt to evade males, suggests that sexual conflict may be widespread and important even in superficially nonconflictual mating systems.

Figure 1.3: Examples of sexual harassment on females caused by males during copulation

Finally, sex can also break apart favourable combinations of genes built by past selection

(Otto, 2009). Regarding all costs sex might bring to the individual, this is the most perverse since these breaking and re-shuffling of alleles is also put forward as an advantage sex brings to populations (Weismann, 1889). However, this is more likely to be advantageous in changing environments, since when in stable conditions organisms profit from maintaining favourable combinations of genes that worked well in the parent, and do not get re-shuffled in the offspring.

Taking into account all the costs sex might bring to the individual, we could expect sexual reproduction might be rare. However, we find sexual reproduction to be pervasive in nature, and only a reduced number of taxa are completely parthenogenetic (0.1% in vertebrates (Vrijenhoek, 1989) and despite being more common in plants, only ~1% in angiosperms (Whitton et al., 2008)). We also find asexual taxa mainly at the tips of the tree of life (Janko, 2014).

Therefore, if sex has so many costs and asexual species are known to outcompete asexual relatives, why do parthenogenetic varieties not replace sexual ones? *What selective forces maintain sexual reproduction and genetic recombination in nature?* (Maynard Smith, 1978).

What does sex bring?

To understand how sex is maintained in Nature and how it originated, we need to recognise what are the consequences of being sexual, and which are the possible advantages it might bring. We also need to distinguish between the origin of sex and the maintenance of sex, since forces that shape both events are likely to be different (Beukeboom and Perrin, 2014).

Early theoretical models on the evolution of sex studying the dynamics of modifier genes analysed the fate of any mutant modifier allele that would affect recombination in a population (Kimura, 1956; Nei, 1967). These modifier models allowed testing the evolution of two separate groups of individuals allowing for the variation of recombination and/or sex levels between them. They found that in populations at equilibrium under selection with heterozygote advantage, genetic mixing never evolved and only modifiers that reduced the amount of recombination spread (Kimura, 1956; Nei, 1967). These models only accounted for individual level selection, ignoring drift, mutation variation, inbreeding, sexual selection, and considered infinite populations. They also showed that, at equilibrium, there is no benefit to variation. If an individual (or population) have a genetic combination fit for a

specific set of environmental conditions, genetic admixture will tend to produce less fit offspring. Consequently, these mixed individuals will be negatively selected and take along the genetic modifiers that promote that genetic mixing (and recombination). This was later identified as the “reduction principle”, the tendency for sex and recombination to decline at equilibria under selection (Otto and Feldman, 1997).

More recent models tried to include more variables to better simulate real world conditions. So, more complex models including variation of selection over time and space, different rates of sex among individuals and selection in finite populations which necessarily include drift. When considering drift in the presence of selection, after a period of time the linkage between beneficial alleles with deleterious alleles tends to accumulate. Consequently, modifiers that increase the frequency of recombination and sex tend to spread together with the beneficial combination of alleles they were responsible for (Otto and Barton, 1997). These drift-based approach could then explain the evolution of sex even considering different variables, such as direction of epistasis (negative, positive or neutral), infinite (with many loci under selection and/or highly structured) or finite populations, and different forms of selection (directional selection, selection against deleterious mutations, Red Queen dynamics) (Otto, 2009). Sex, or higher recombination, spreads when drift and selection use up all variation in a population, leaving individuals with a mixture of high- and low-fitness genes. Sex and recombination allow the re-shuffling of these genes, bringing together beneficial combinations, and purging unfavourable ones. Only in models that simulate the complex and changing environments, sex and recombination are allowed to spread. In populations where sex is absent, selection is less effective due to Hill-Robertson effects and sex, even if only at some extent and costly, is selected for (Otto, 2009). Empirical evidence from studies with sexual/asexual systems has identified some of the situations sex can be selected for (Luijckx et al., 2017; McDonald et al., 2016).

The maintenance of sex

Muller's ratchet proposes that the accumulation of deleterious mutations over time in the genome of asexual organisms diminishes the fitness of asexual lineages, slowly leading them to extinction (Bachtrog and Gordo, 2004). Recombination can bring together mildly deleterious mutations, increasing the negative selective pressure and purging them out of the population, rescuing beneficial mutations from deleterious backgrounds and avoiding the Muller's ratchet.

In Red Queen dynamics the arms race between organisms and their parasites/predators/competitors increases selective pressure creating an unpredictably changing environment favouring the genetic shuffling and the genetic variation brought by sexual reproduction (Morran et al., 2011). Recombination can bring together beneficial mutations that arise on different genetic backgrounds, creating the new genotypes and accelerating co-evolution, relieving clonal interference and avoiding competition between equally favourable mutations.

Despite the different hypothesis put forward to explain the evolution of sexual reproduction, it all comes down to the potential of recombination in speeding adaptation by the generational re-shuffling of variation and the production of individuals with different combination of alleles, either favourable under a given environment (that are likely to be selected) or deleterious (that are likely to have decreased reproductive fitness and those alleles not be passed on to the following generations) (Cloney, 2016; McDonald et al., 2016).

The idea that sex evolves because it facilitates adaptation (Weismann, 1889) has been shown empirically in single cell organisms (McDonald et al., 2016) and macroinvertebrates (Luijckx et al., 2017). However, studies and theories developed concerning the paradox of sex will not be able to clarify the current distribution and abundance of asexual lineages in nature. The understanding of rates of origin and extinction of asexual lineages, together with the understanding of putative differential evolution rates in organisms with sexual or asexual reproduction, should be considered in the main “origin of sex” question.

THE MODEL SYSTEM: *Darevskia*

Ecology and distribution

The genus studied here is the first true parthenogenetic vertebrate described (Darevsky, 1967), *Darevskia* Arribas, 1997 (Lacertidae). This is a highly speciose genus, currently including 27 species (Ahmadzadeh et al., 2013; Tarkhnishvili et al., 2012; Uetz and Hošek), of which seven are parthenogenetic. On the basis of allozymes and Cyt-*b*, sexual *Darevskia* analysed so far are included in three genetically distinct clades (Murphy et al., 2000) and parthenogenetic species are hybrids resultant from inter-clade hybridization events (Fig. 1.4). Only some of the sexual species contributed for these events, which were directional: the maternal species were always from the Caucasica clade and the paternal species always from the Rudis clade (Freitas et al., 2016a; Murphy et al., 1996,

2000).

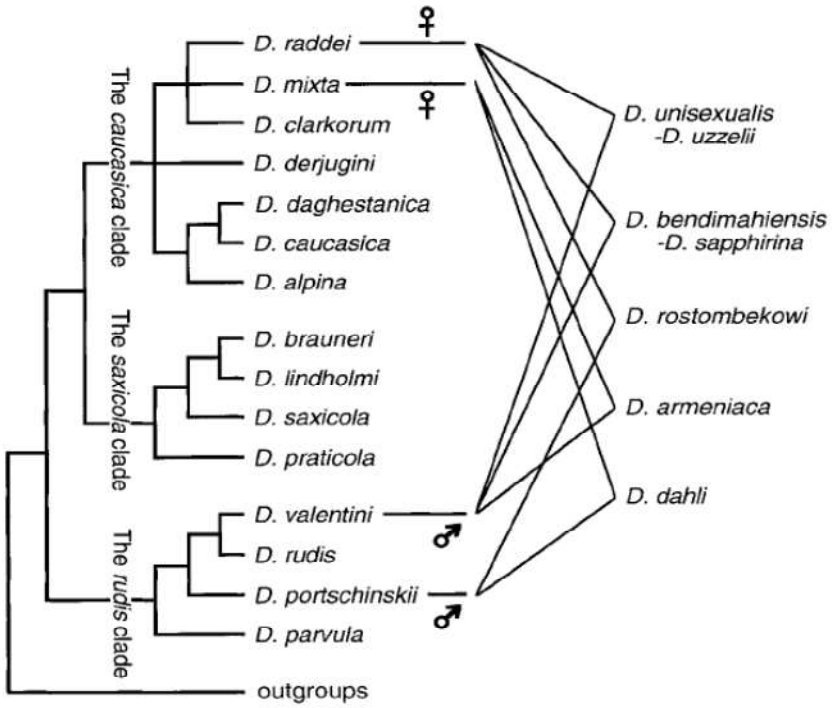


Fig. 1.4: Phylogenetic inference of the genus *Darevskia* using Cyt-*b* and allozyme markers. On the right are the parthenogenetic species and arrows indicate the putative parental sexual *Darevskia*. Adapted from "A fine line between sex and unisexuality: the phylogenetic constraints on parthenogenesis in lacertid lizards." by R. Murphy, 2000, Zool. J. Linn. Soc.

This genus mainly ranges in the Caucasus but also extends to the adjacent regions, including northern Turkey and the Balkans (Arnold, 2006; Darevsky, 1967). The clades Rudis and Caucasica are saxicolous lizards, with characteristically flat head and body (Arnold et al., 2007) distributed in the Caucasus (except its northwestern part) and parts of Anatolia. On the other hand, the clade Saxicola has only ground dwelling, lizards with characteristic higher head than the other species of the genus. This clade has a peculiar distribution, since most of its species are restricted to the western part of the Greater Caucasus and southern Crimea (Murphy et al., 2000; Tarkhnishvili et al., 2012), except *D. praticola* which has a allopatric distribution and can be found in the western part of the Greater Caucasus and in the Balkans. Questions have been raised about the species status of the Balkan meadow lizards (Tuniyev et al., 2011), since it is the westernmost *Darevskia* species and its distribution is not connected to the remaining *Darevskia.*

Asexuality in *Darevskia*

Most of the vertebrate parthenogenesis recorded is thought to have been originated via

interspecific hybridization (Avise, 2008). The Balance Hypothesis (Moritz et al., 1989) was put forward trying to explain how these events might trigger asexuality in hybrids, and it states that parthenogenetic vertebrates will only arise by the hybridization of two sexual species divergent enough to disrupt meiosis in the hybrids, yet not so divergent as to seriously compromise hybrid viability or fertility (Kearney et al., 2009). This hypothesis grounds on easily applicable premises and has been generally used as an explanation for the origin of vertebrate parthenogenesis.

However, another theory was put forward, the Phylogenetic Constraint hypothesis (Darevsky, 1967), which states that parthenogenetic lineages are originated by the hybridization between sexual species which present exclusive genetic peculiarities that allow them to interbreed and produce asexual viable reproducing hybrids, and that these hybridization events are directional, with species from different phylogenetic clades contributing either to the maternal or paternal ancestry. *Darevskia*, which underpinned the development of this hypothesis, fits its premises. However, the Phylogenetic Constraint Hypothesis has been found to fit not only *Darevskia* but also other parthenogenetic vertebrate systems, such as *Leiolepis* sp. (butterfly lizards), The parental species for these parthenogens are always the same, *L. reevesii* as the paternal and *L. guttata* as the maternal, and also belong to two different phylogenetic clades within the genus (Grismer et al., 2014). On the other hand, directional hybridizations are not always responsible for parthenogenetic hybrids, with other evidence showing otherwise: some parthenogenetic vertebrates, such as fish genus *Cobitis*, can even present both parthenogenetic hybrid lineages resultant from the cross between species from the same clade (hybrids between *C. elongatoides* and *C. tanaica*) or different clades (hybrids between *C. elongatoides* and *C. taenia*) (Janko et al., 2003). This study, however, reinforces the idea that only a few sexual species contributed to the hybridization events that originated the parthenogenetic vertebrates (either from *Leiolepis* or *Cobitis* genera mentioned, or *Darevskia*), and that the distance between them or the directionality of the hybridization was not as relevant as the specificity of the sexual ancestors, as the Phylogenetic Constraint Hypothesis states.

As already pointed out, only two species are thought to have contributed maternally to the hybridization events which originated the parthenogenetic hybrids, *D. raddei* and *D. mixta*. *D. raddei* is thought to be the maternal species for *D. unisexualis*, *D. uzzelli*, *D. bendimahiensis*, *D. sapphirina* and *D. rostombekowi* (Conflitti et al., 2014; Freitas et al., 2016b; Fu et al., 2000a; Murphy et al., 2000) while *D. mixta* is thought to be the maternal

species for the remaining two, *D. dahli* and *D. armeniaca* (Fu et al., 2000b; Murphy et al., 2000). On the other hand, the known paternal species are *D. portschinskii* (*D. rostombekowi* and *D. dahli*) and *D. valentini* (*D. uzzelli:* Freitas et al., 2016- paper II, paper III *; D. armeniaca:* paper III; *D. unisexualis:* Freitas et al., 2016- paper II, paper III), even though the contribution of *D. rudis* for the parental ancestry is still not completely clear. Parthenogenetic species are expected to be young in age (Freitas et al., 2016-paper II; Moritz et al., 1992), despite at least some of them presenting a considerable wide distribution range.

Darevskia parthenogenetic species can be found in sympatry with their sexual parentals, even though not commonly. Not only they have been shown to share the same habitat requirements of their parentals, specifically of their maternal species (*D. dahli:* Tarkhnishvili et al., 2010; and *D. unisexualis:* Freitas et al., 2016 – paper II), but in some situations they have even been found they are able to outcompete them (*D. dahli:* Tarkhnishvili et al. (2010)). When sympatry is between parthenogens and their putative paternal sexual species, polyploid backcrosses have been described (Arakelyan, 2013; Danielyan et al., 2008). These individuals are backcross hybrids between female parthenogens and males of the sexual species, and a reduced number of tetraploids have also been found, which together with the accounts of some polyploids with normally developed reproductive systems, reinforces the presumption of their fertility, albeit reduced.

PHYLOGENETICS IN THE AGE OF GENOMICS

Phylogenetic inference reconstruction has been extensively used to ask several evolutionary biology questions, and we have recently witnessed a steep change of analyses methods in phylogenetics (McCormack et al., 2012). Sanger DNA sequencing techniques had been used for decades given its easiness in the generation of sequences in a highly targeted way and with the possibility of application in several individuals, frequently cross species. However, this technique only allows for the sequencing of one locus per individual at a time, which can result in a limited and often biased estimation of the evolutionary history of the taxa studied. When studying groups with high levels of hybridization, and strong potential for introgression events, this bias can be particularly strong, with different genomic regions providing different ecolutionary histories. The rapid development of high-throughput sequencing technologies seen in the last years has facilitated the study of groups of species for which we lacked previous genetic data. These

techniques allowed the study of several markers across the genome for several individuals with a low cost-benefit ratio. The great amount of markers together with phylogenomic analyses that account for the discordance between loci allow us to study the divergence within a group considering also the amount of gene flow and introgression present during its evolutionary history.

However, phylogenetic inference analyses are constructed on the use of long homologous sequences across different individuals, but with many next generation sequencing (NGS) methods there is less control over which regions of the genome are sequenced, and over whether those regions are either homologous across the whole range of individuals analysed or phylogenetically informative (Carstens et al., 2012). NGS allows the sequencing of several individuals and markers simultaneously and very cost-effectively, exponentially increasing the amount of data gathered per run in terms of the number of markers and number of individuals assessed (Carstens et al., 2012; Metzker, 2010). NGS allows for the understanding of cases such as introgression, incomplete lineage sorting (ILS), fast radiations, hybridization and gene flow, that can only be resolved with sequences from many loci. In order to update phylogenetics to the use of NGS techniques we need to find a strategy that is not only cost-effective in terms of the number of orthologous sequences available across the individuals analysed, but which could also provide a wide range of long and phylogenetically informative sequences.

Most common NGS approaches use restriction digest and manual size selection to reduce the genome analysed, and thus “select” cross genome markers (Baird et al., 2008; Elshire et al., 2011). These methods are very cost effective (Stapley et al., 2010), and allow the sequencing of a large number of short markers distributed more or less randomly across the genome, many individuals at a time, contrary to Sanger techniques. However, these approaches have been widely used for population genomic studies at intra-specific level (McCormack et al., 2013), and not so much for phylogenetic inferences. That is because they produce short reads best suited for generating SNPs and often have low repeatability. Given they amplify sequences next to restriction sites they also often fail to sequence orthologous regions among the individuals analysed.

The best NGS approach to be applied in phylogenomic studies has to be capable of generating orthologous sequences in multiple individuals, such as targeting specific genomic regions (like target enrichment or sequence capture, (Brandley et al., 2015)). Sequence capture methods use probes targeted to specific regions of the genome and

can sequence hundreds of markers simultaneously in several cross-species individuals. Probe design can be performed from several sources (Bragg et al., 2015; Faircloth, 2016; Lemmon et al., 2012), but in the case of non-model taxa that do not have a genome assembly available, transcriptome can present as a good and accessible way of finding variable regions of the genome that are long and phylogenetically informative.

OBJECTIVES AND THESIS STRUCTURE

The main aim of this thesis is to understand how vertebrate asexuality originates and evolves in nature, identifying the rates of origin and extinction of parthenogenetic vertebrates together with the phylogenetic signal of sexual and asexual species, and the influence of asexual species on hybridization, gene flow and polyploidy. More specifically, the goals of the work are:

1) Recover the rates of origin and significance of asexuality in vertebrates using *Darevskia* lizards as a model. Here we will test the application of two general hypothesis put forward in relation to the hybrid origin of parthenogenesis in vertebrates.

2) Understand the role of backcrossing and polyploidization in *Darevskia*

3) Infer the extent of hybridization with gene flow between sexual clades and sexual-asexual individuals.

The research presented here is in the format of five original scientific articles, which have been already published (papers I and II) or are in the process of being submitted (papers III – IV).

The remaining chapters are structured as follows:

Chapter 2 – Phylogeographic patterns of selected *Darevskia* sp.

Chapter 2.1) The species status of *D. praticola* – paper I

- *Freitas, S., Vavakou, A., Arakelyan, M., Drovetski, S.V., Crnobrnja-isailović, J., Kidov, A.A., Cogălniceanu, D., Corti, C., Lymberakis, P., Harris, D.J., et al. (2016). Cryptic diversity and unexpected evolutionary patterns in the meadow lizard, Darevskia praticola (Eversmann, 1834). Systematics and Biodiversity 14, 184–197.*

This paper uses a three marker set for a phylogeographic analysis on a selected sexual *Darevskia.* This species, *D. praticola,* is not related to the parthenogenetic hybrids, however, given its disjunct distribution range, its evolutionary history provides an important insight on *Darevskia*'s biogeography.

Chapter 2.2) Age and hybrid origin of parthenogenetic *Darevskia* – paper II

- *Freitas, S., Rocha, S., Campos, J., Ahmadzadeh, F., Corti, C., Sillero, N., Ilgaz, Ç.,*

Kumlutaş, Y., Arakelyan, M., Harris, D.J., et al. (2016). Parthenogenesis through the ice ages: A biogeographic analysis of Caucasian rock lizards (genus Darevskia). Molecular Phylogenetics and Evolution 102, 117–127.

This paper uses mtDNA and nuclear markers to analyse the phylogeography of *D. raddei,* which is the sexual maternal species for most of the parthenogenetic *Darevskia* species. Three parthenogens are included in the analysis, and both maternal (mtDNA and nuclear) and paternal (nuclear) inferences of parentage are tested. Finally, niche modelling is used to compare the potential distribution range for *D. raddei* and one of the parthenogens with the widest distribution range, *D. unisexualis*.

Chapter 3 - Origin of Vertebrate Asexuality

- Freitas, S., Harris, D. J., Silero, N., Arakelyan, M., Butlin, R., Carretero, MA. (in prep) The origin of parthenogenesis and the roles of hybridization and polyploidy in a lizard model

This paper uses a set of microsatellite markers to try to interpret the origin of vertebrate parthenogenesis in the light of two hypotheses, the balance hypothesis and the phylogenetic constraint hypothesis. For that, the number of hybridization events that lead to the three parthenogenetic species included in this study, as well as the extent of recent hybridization between parthenogens and sexual species, are analysed.

Chapter 4 - Phylogenomics in non-model organisms

Freitas, S., Westram, A., Soria-Carrasco, V., Harris, D.J., Carretero, M.A., Butlin, R. (in prep). Development of cross-species, genome-wide capture sequence probes from the transcriptome of non-model species for phylogenetic analyses

This paper describes one approach used to design capture sequence probes from the transcriptome of two individuals, one sexual and one parthenogen. Three types of tissues are used and the transcriptome is assembled and annotated. Capture sequence probes are then tested for downstream phylogenetic inference.

Chapter 5 - Deep Branch Gene Flow

Freitas, S., Westram, A., Carretero, M.A., Harris, D.J., Butlin, R. (in prep). Deep branch gene flow in a hybridization rich diversification process

This paper uses capture sequence probes to sequence hundreds of cross genome markers in several individuals in order to assess the phylogenetic relationships within the

genus *Darevskia*. Past events of hybridization, introgression and incomplete lineage sorting is also tested among the sexual species of the genus.

Chapter 6 – General Discussion

This chapter includes a general discussion of the main results obtained during the course of this thesis and its significance for recent research.

The supplementary materials for all papers in this thesis are aggregated in the Appendix.

REFERENCES

Ahmadzadeh, F., Flecks, M., Carretero, M.A., Mozaffari, O., Böhme, W., Harris, D.J., Freitas, S., and Rödder, D. (2013). Cryptic Speciation Patterns in Iranian Rock Lizards Uncovered by Integrative Taxonomy. PLOS ONE *8*, e80563.

Arakelyan, M. (2013). The Study of Age, Growth, and Longevity in the Triploid Hybrids of Rock Lizards of the genus Darevskia in Armenia. Russ. J. Herpetol. *9*, 63–68.

Archetti, M. (2010). Complementation, Genetic Conflict, and the Evolution of Sex and Recombination. J. Hered. *101*, S21–S33.

Arnold, M.L. (2006). Evolution through genetic exchange (Oxford ; New York: Oxford University Press).

Arnold, E.N., Arribas, Ó., and Carranza, S. (2007). Systematics of the Palaearctic and Oriental lizard tribe Lacertini (Squamata:Lacertidae:Lacertinae), with descriptions of eight new genera (Auckland, N.Z.: Magnolia Press).

Avise, J.C. (2008). Clonality: the genetics, ecology, and evolution of sexual abstinence in vertebrate animals (Oxford ; New York: Oxford University Press).

Bachtrog, D., and Gordo, I. (2004). Adaptive Evolution of Asexual Populations Under Muller's Ratchet. Evolution *58*, 1403–1413.

Baird, N.A., Etter, P.D., Atwood, T.S., Currey, M.C., Shiver, A.L., Lewis, Z.A., Selker, E.U., Cresko, W.A., and Johnson, E.A. (2008). Rapid SNP Discovery and Genetic Mapping Using Sequenced RAD Markers. PLOS ONE *3*, e3376.

Barron, J.N., Lawson, T.J., and Jensen, P.A. (2016). Analysis of potential factors allowing coexistence in a sexual/asexual minnow complex. Oecologia *180*, 707–715.

Bell, G. (1982). The masterpiece of nature: The evolution and genetics of sexuality (Berkeley).

Beukeboom, L., and Perrin, N. (2014). The Evolution of Sex Determination (Oxford, New

York: Oxford University Press).

Bragg, J.G., Potter, S., Bi, K., and Moritz, C. (2015). Exon capture phylogenomics: efficacy across scales of divergence. Mol. Ecol. Resour. n/a – n/a.

Brandley, M.C., Bragg, J.G., Singhal, S., Chapple, D.G., Jennings, C.K., Lemmon, A.R., Lemmon, E.M., Thompson, M.B., and Moritz, C. (2015). Evaluating the performance of anchored hybrid enrichment at the tips of the tree of life: a phylogenetic analysis of Australian Eugongylus group scincid lizards. BMC Evol. Biol. *15*, 62.

Carstens, B., Lemmon, A.R., and Lemmon, E.M. (2012). The Promises and Pitfalls of Next-Generation Sequencing Data in Phylogeography. Syst. Biol. *61*, 713–715.

Cloney, R. (2016). Molecular evolution: Friends with benefits — sex speeds up adaptation. Nat. Rev. Genet. *advance online publication*.

Conflitti, I.M., Shields, G.F., Murphy, R.W., and Currie, D.C. (2014). The speciation continuum: Population structure, gene flow, and maternal ancestry in the Simulium arcticum complex (Diptera: Simuliidae). Mol. Phylogenet. Evol. *78*, 43–55.

Danielyan, F., Arakelyan, M., and Stepanyan, I. (2008). Hybrids of Darevskia valentini, D. armeniaca and D. unisexualis from a sympatric population in Armenia. Amphib.-Reptil. *29*, 487–504.

Darevsky, I. (1967). Rock lizards of the Caucasus: systematics, ecology and phylogenesis of the polymorphic groups of Caucasian rock lizards of the subgenus Archaeolacerta. Leningrad. Nauka.

Elshire, R.J., Glaubitz, J.C., Sun, Q., Poland, J.A., Kawamoto, K., Buckler, E.S., and Mitchell, S.E. (2011). A Robust, Simple Genotyping-by-Sequencing (GBS) Approach for High Diversity Species. PLOS ONE *6*, e19379.

Faircloth, B.C. (2016). Identifying Conserved Genomic Elements and Designing Universal Probe Sets To Enrich Them.

Freitas, S., Rocha, S., Campos, J., Ahmadzadeh, F., Corti, C., Sillero, N., Ilgaz, Ç., Kumlutaş, Y., Arakelyan, M., Harris, D.J., et al. (2016a). Parthenogenesis through the ice

ages: A biogeographic analysis of Caucasian rock lizards (genus Darevskia). Mol. Phylogenet. Evol. *102*, 117–127.

Freitas, S., Vavakou, A., Arakelyan, M., Drovetski, S.V., Crnobrnja-isailović, J., Kidov, A.A., Cogălniceanu, D., Corti, C., Lymberakis, P., Harris, D.J., et al. (2016b). Cryptic diversity and unexpected evolutionary patterns in the meadow lizard, *Darevskia praticola* (Eversmann, 1834). Syst. Biodivers. *14*, 184–197.

Fu, J., Murphy, R.W., Darevsky, I.S., and McEachran, J.D. (2000a). Divergence of the cytochrome b gene in the Lacerta raddei complex and its parthenogenetic daughter species: evidence for recent multiple origins. Copeia *2000*, 432–440.

Fu, J., MacCulloch, R.D., Murphy, R.W., Darevsky, I.S., and Tuniyev, B.S. (2000b). Allozyme variation patterns and multiple hybridization origins: clonal variation among four sibling parthenogenetic Caucasian rock lizards. Genetica *108*, 107–112.

Grismer, J.L., Bauer, A.M., Grismer, L.L., Thirakhupt, K., Aowphol, A., Oaks, J.R., Wood, P.L., Onn, C.K., Thy, N., Cota, M., et al. (2014). Multiple origins of parthenogenesis, and a revised species phylogeny for the Southeast Asian butterfly lizards, Leiolepis. Biol. J. Linn. Soc. n/a – n/a.

Han, C.S., and Jablonski, P.G. (2010). Male water striders attract predators to intimidate females into copulation. Nat. Commun. *1*, 1–6.

Huyghe, K., Vanhooydonck, B., Herrel, A., Tadić, Z., and Van Damme, R. (2012). Female lizards ignore the sweet scent of success: Male characteristics implicated in female mate preference. Zoology *115*, 217–222.

Janko, K. (2014). Let Us Not Be Unfair to Asexuals: Their Ephemerality May Be Explained by Neutral Models Without Invoking Any Evolutionary Constraints of Asexuality. Evolution *68*, 569–576.

Janko, K., Kotlík, P., and Ráb, P. (2003). Evolutionary history of asexual hybrid loaches (Cobitis: Teleostei) inferred from phylogenetic analysis of mitochondrial DNA variation. J. Evol. Biol. *16*, 1280–1287.

Kearney, M., Fujita, M.K., and Ridenour, J. (2009). Lost Sex in the Reptiles: Constraints

and Correlations. In Lost Sex, I. Schön, K. Martens, and P. Dijk, eds. (Springer Netherlands), pp. 447–474.

Kimura, M. (1956). A Model of a Genetic System which Leads to Closer Linkage by Natural Selection. Evolution *10*, 278.

Lavanchy, G., Strehler, M., Llanos Roman, M.N., Lessard-Therrien, M., Humbert, J.-Y., Dumas, Z., Jalvingh, K., Ghali, K., Fontcuberta García-Cuenca, A., Zijlstra, B., et al. (2016). Habitat heterogeneity favors asexual reproduction in natural populations of grassthrips. Evolution *70*, 1780–1790.

Lehtonen, J., and Kokko, H. (2012). Positive feedback and alternative stable states in inbreeding, cooperation, sex roles and other evolutionary processes. Philos. Trans. R. Soc. B Biol. Sci. *367*, 211–221.

Lemmon, A.R., Emme, S.A., and Lemmon, E.M. (2012). Anchored Hybrid Enrichment for Massively High-Throughput Phylogenomics. Syst. Biol. *61*, 727–744.

Luijckx, P., Ho, E.K.H., Gasim, M., Chen, S., Stanic, A., Yanchus, C., Kim, Y.S., and Agrawal, A.F. (2017). Higher rates of sex evolve during adaptation to more complex environments. Proc. Natl. Acad. Sci. 201604072.

Maynard Smith, J. (1971). The origin and the maintenance of sex. In Group Selection, (Chicago: Aldine-Atherton), pp. 163–175.

Maynard Smith, J. (1978). The Evolution of Sex (New York: Cambridge University Press).

McCormack, J.E., Faircloth, B.C., Crawford, N.G., Gowaty, P.A., Brumfield, R.T., and Glenn, T.C. (2012). Ultraconserved elements are novel phylogenomic markers that resolve placental mammal phylogeny when combined with species-tree analysis. Genome Res. *22*, 746–754.

McCormack, J.E., Hird, S.M., Zellmer, A.J., Carstens, B.C., and Brumfield, R.T. (2013). Applications of next-generation sequencing to phylogeography and phylogenetics. Mol. Phylogenet. Evol. *66*, 526–538.

McDonald, M.J., Rice, D.P., and Desai, M.M. (2016). Sex speeds adaptation by altering

the dynamics of molecular evolution. Nature *531*, 233–236.

Metzker, M.L. (2010). Sequencing technologies — the next generation. Nat. Rev. Genet. *11*, 31–46.

Mogie, M. (1986). Automixis: its distribution and status. Biol. J. Linn. Soc. *28*, 321–329.

Moritz, C., Brown, W.M., Densmore, L.D., Vyas, S., Donnellan, S., Adams, M., and Baverstock, P. (1989). Genetic diversity and the dynamics of hybrid parthenogenesis in Cnemidophorus (Teiidae) and Heteronotia (Gekkonidae). In Evolution and Ecology of Unisexual Vertebrates, (Albany),.

Morran, L.T., Schmidt, O.G., Gelarden, I.A., Parrish, R.C., and Lively, C.M. (2011). Running with the Red Queen: Host-Parasite Coevolution Selects for Biparental Sex. Science *333*, 216–218.

Murphy, R.W., Darevsky, I.S., MacCulloch, R.D., Fu, J., and Kupriyanova, L.A. (1996). Evolution of the bisexual species of Caucasian rock lizards: a phylogenetic evaluation of allozyme data. Russ. J. Herpetol. *3*, 18–3.

Murphy, R.W., Fu, J., Macculloch, R.D., Darevsky, I.S., and Kupriyanova, L.A. (2000). A fine line between sex and unisexuality: the phylogenetic constraints on parthenogenesis in lacertid lizards. Zool. J. Linn. Soc. *130*, 527–549.

Nei, M. (1967). Modification of linkage intensity by natural selection. Genetics *57*, 625.

Otto, S.P. (2009). The Evolutionary Enigma of Sex. Am. Nat. *174*, S1–S14.

Otto, S.P., and Barton, N.H. (1997). The Evolution of Recombination: Removing the Limits to Natural Selection. Genetics *147*, 879–906.

Otto, S.P., and Feldman, M. (1997). Deleterious Mutations, Variable Epistatic Interactions, and the Evolution of Recombination. Theor. Popul. Biol. *51*, 134–147.

Otto, S.P., and Lenormand, T. (2002). Resolving the paradox of sex and recombination. Nat. Rev. Genet. *3*, 252–261.

Redfield, R.J. (2001). Do bacteria have sex? Nat. Rev. Genet. *2*, 634–639.

Rowe, L. (1994). The costs of mating and mate choice in water striders. Anim. Behav. *48*, 1049–1056.

Schwander, T. (2016). Evolution: The End of an Ancient Asexual Scandal. Curr. Biol. *26*, R233–R235.

Stapley, J., Reger, J., Feulner, P.G.D., Smadja, C., Galindo, J., Ekblom, R., Bennison, C., Ball, A.D., Beckerman, A.P., and Slate, J. (2010). Adaptation genomics: the next generation. Trends Ecol. Evol. *25*, 705–712.

Tarkhnishvili, D., Gavashelishvili, A., Avaliani, A., Murtskhvaladze, M., and Mumladze, L. (2010). Unisexual rock lizard might be outcompeting its bisexual progenitors in the Caucasus. Biol. J. Linn. Soc. *101*, 447–460.

Tarkhnishvili, D., Gavashelishvili, A., and Mumladze, L. (2012). Palaeoclimatic models help to understand current distribution of Caucasian forest species. Biol. J. Linn. Soc. *105*, 231–248.

Tuniyev, S.B., Doronin, I.A., Kidov, A.A., and Tuniyev, B.S. (2011). Systematic and geographical variability of meadow lizard, Darevskia praticola (Reptilia: Sauria) in the Caucasus. Russ J Herpetol *18*, 295–316.

Uetz, P., and Hošek, J. The Reptile Database.

Vrijenhoek, R.C. (1989). Genetic and ecological constraints on the origins and establishment of unisexual vertebrates. Evol. Ecol. Unisexual Vertebr. *466*, 24–31.

Weismann, A. (1889). The significance of sexual reproduction in the theory of natural selection. In Essays upon Heredity and Kin- Dred Biological Problems., (Oxford: Clarendon), pp. 251–332.

Whitton, J., Sears, C.J., Baack, E.J., and Otto, S.P. (2008). The Dynamic Nature of Apomixis in the Angiosperms. Int. J. Plant Sci. *169*, 169–182.

CHAPTER 2

PHYLOGEOGRAPHIC PATTERNS OF SELECTED *Darevskia* sp.

CHAPTER 2.1

The species status of *D. praticola* – paper I

Figure 2.1.0: *Darevskia praticola* from Turkey in its natural environment.

Cryptic diversity and unexpected evolutionary patterns in the meadow lizard, *Darevskia praticola*

Authors

Susana Freitas[1,2]

Anna Vavakou[3]

Marine Arakelyan[4]

Sergei V. Drovetski[5]

Jelka Crnobrnja-Isailović[6,7]

Artem A. Kidov[8]

Dan Coğalniceanu[9]

Claudia Corti[10]

Petros Lymberakis[11]

D. James Harris[1]

Miguel A. Carretero[1]

[1]CIBIO Research Centre in Biodiversity and Genetic Resources, InBIO, Universidade do Porto, Campus Agrário de Vairão, Rua Padre Armando Quintas, Nº 7. 4485-661 Vairão, Vila do Conde, Portugal

[2]Department of Animal and Plant Sciences, The University of Sheffield, Sheffield S10 2TN, UK

[3]Faculty of Biology, Bielefeld University, Bielefeld, Germany

[4]Yerevan State University, Alek Manoogian, 1, Yerevan 0025, Armenia

[5] Division of Birds, National Museum of Natural History, Smithsonian Institution 1000 Constitution Ave, NW, Washington, DC 20004

[6]Department of Biology and Ecology, Faculty of Sciences and Mathematics, University of Niš, Višegradska 33, 18000 Niš, Serbia

[7]Department of Evolutionary Biology, Institute for Biological Research "Siniša Stanković", University of Belgrade, Despota Stefana 142, 11000 Belgrade, Serbia

8 Department of Zoology, Faculty of Animal Science, Russian State Agrarian University — K. A. Timiryazev Moscow Agricultural Academy, 49 Timiryazevskaya str., Moscow 127550, Russia

[9]University Ovidius Constanta, Faculty of Natural Sciences, Bvd. Mamaia 124, Constanta, Romania

10Museo di Storia Naturale dell'Università di Firenze, Sezione di Zoologia "La Specola", Via Romana, 17, 50125 Firenze, Italy

[11] Natural History Museum of Crete, University of Crete, Knossos Ave., 71409 Herakleio, Crete, Greece

ABSTRACT

Darevskia praticola is the only species in the genus with a wide but disjoint distribution, found within the Balkan Peninsula and the Caucasus. Whereas most *Darevskia* species occupy saxicolous habits, *D. praticola* is found in meadows and forest environments. Our study explores the phylogeographic and phylogenetic relationships of *Darevskia praticola* sensu lato, assessing the identity of the taxa described within its range that have thus far been solely based on morphology. Analysing the evolutionary patterns of this group we also test the conspecificity of different populations and whether the shift to forest habitats happened only once (if this species is monophyletic), or multiple times. We sequenced samples collected across the entire range for two mtDNA markers (Cyt-*b* and ND4) and two nuclear markers (MC1R and ReLN). Our results support three main clades. The most divergent included all samples from the Balkans. The other two, more closely related, group samples from western Greater Caucasus and Transcaucasia respectively. *D. praticola* sensu lato appears to be monophyletic, but our findings are discordant with the taxonomic arrangement developed so far. Both mitochondrial and nuclear data show a deep divergence of the Balkans and two sister clades from the Caucasus region that was dated to the Late Pliocene approximately 2.5 Ma. Within the Caucasus, the incomplete sub-clade divergence is likely due to subsequent differentiation during the Pleistocene glaciations. Given its disjointed distribution, the geographic separation is likely due to a vicariance event associated with multiple climatic and vegetation shifts.

Keywords

Balkans, Caucasus, *Darevskia praticola,* phylogeny, mtDNA, nDNA

INTRODUCTION

Several geological and climate changes throughout times have given the Caucasus and the Balkans an irregular landscape and a wide scope of different habitats. During the

Pleistocene, both the Caucasus and the Balkans acted as a glacial refugia (Popov et al., 2006), and the post glacial contact between different biogeographic regions helped increasing their biodiversity. This lead to a high level of species richness and endemisms, such as both are currently considered hotspots of biodiversity (Myers et al., 2000).

Phylogeographic studies of the herpetofauna of the Black Sea region are scarce, compared to those available for other glacial refugia. Neverthless, the studies undercome so far already revealed considerable complexity (e.g. Fritz et al., 2009; García-Porta et al., 2009; Gvoždík et al., 2010; Lymberakis et al., 2007; Marosi et al., 2012; Mashkaryan et al., 2013; Recuero et al., 2011; Stöck et al., 2006, 2012; Tarkhnishvilli et al., 2013; Ursenbacher et al., 2008; Wielstra et al., 2010). Different phylogeographic patterns have been observed, suggesting a dual role for the Black Sea. While for some species the Black Sea was shown to act as a geographic barrier, dividing the distribution range into different clades, for others it seemed to have provided a corridor, linking the Balkans and the Caucasus (Fritz et al., 2009; Garcia-Porta et al., 2012; Gvoždík et al., 2010).

Darevskia is a genus of small lizards, present in a wide variety of environments, from rocks to meadows (Tarkhnishvilli 2012). The most widespread species is *Darevskia praticola* (Eversmann, 1834), which is found across the Caucasus from the Black Sea to northwestern Iran but also in the Balkans where it is the sole representative of the genus. Contrary to most *Darevskia* species, *D. praticola* is mainly a ground dweller, found across its wide range restricted to clearings with lush vegetation, meadows and glades within open broad-leaf woodland (Agasyan et al., 2009).

So far, the relationships beween the different forms of *D. praticola* were constructed with a single mitochondrial DNA marker (Cyt-*b*). Based on this, Murphy et al. (1996), and later Tarkhnishvilli (2012) and Ahmadzadeh et al. (2013), have placed *D. praticola* within the *saxicola* species group, as a sister taxa to *D. saxicola* and *D. brauneri*. Using morphological data, Roitberg (1999) inferred the phenetic relationships between *Darevskia caucasica*, *D. daghestanica*, and *D. praticola* (Caucasus), and proposed that *D. praticola* should have undergone a recent niche shift from rocky and stony habitats (where this genus is most frequently found), to forest habitat.

The taxonomy of *D. praticola sensu lato* is extremely complex and there were several taxonomic revisions addressing the Caucasian populations conducted during the last century (Tuniyev et al., 2011, 2013 and references therein). These studies were based solely on scalation and identified four forms with non-overlapping distributions (Ilgaz and

Kumlutaş 2005; Ljubisavljevic et al., 2006; Tuniyev et al., 2011). The nominotypical subspecies (*D. p. praticola*) is found in the northeastern Caucasus and Transcaucasia. *D. p. pontica* occupies the northwestern Caucasus and the Balkans (southeastern Romania, Serbia, Greece, Bulgaria and Turkish Thrace) (Ilgaz and Kumlutaş 2005; Ljubisavljevic et al., 2005; 2008; Tuniyev et al., 2011). *D. p. hyrcanica* is restricted to the Talysh mountains, Western Elburz mountains and Enzeli Bay of the Caspian Sea (Tuniyev et al., 2011). Finally, the fourth form was recently described in Armenia, *D. praticola loriensis* (Tuniyev et al., 2013).

Some authors argue that *D. praticola pontica* should be recognized as a different species. However, this taxonomic change is still poorly supported, since no genetic studies of the whole group have been conducted so far (Ljubisavljevic et al., 2006; Tuniyev et al., 2011).

Apart from the four recognised subspecies, *D. praticola hungarica* (Sobolevsky 1930) was also described. This form originates from the Transylvanian Alps in Romania and is currently synonymised with *D. praticola pontica.* As with most of the taxonomical assignments in the *Darevskia praticola* sensu lato, this was also based solely on morphological characters, lacking a genetic validation.

Tuniyev et al. (2011) helped reorganise the published data on *Darevskia praticola* sensu lato, with a major revision of all published data. In their work they assigned a lectotype and paralectotypes from the northwestern Caucasus for *D. pontica* and described the subspecies *D. praticola hyrcanica*. Also, a key for the identification of the species and subspecies was presented, exclusively based on a limited set of morphological characters, such as the number of pairs of chin shields and dorsoventral colouration.

Although the hypothesis of two different taxa in the Balkans and the Caucasus had been previously proposed (Stugren 1961; Bischoff 1976), later it was replaced by an arrangement with up to four lineages in the Caucasus with the westernmost of them also being present in the Balkans (Ljubisavljevic et al., 2006; Tuniyev et al., 2011, 2012). Although there are some morphological characters common to both Balkanic and northwestern Caucasian populations (Tuniyev et al., 2011, 2012), the lack of geographic correspondence (distant and unconnected populations are more similar than those in close proximity), the high risk of homoplasy in lacertid scalation and the absence of genetic data from the entire species range question any evolutionary hypothesis raised to date for this species complex. Therefore, following a multilocus approach and with a wide range sampling, this work aims to shed some light upon the phylogeographic structure within *D.*

praticola and its phylogenetic relationships with other *Darevskia* sp. Both levels of relationships are dated to allow the inference of the putative paleogeographic scenarios.

MATERIAL AND METHODS

A total of 31 samples of D. praticola were used in this study, collected between 1996 and 2012 across the distribution range of this species (Table 2.1.1, Figure 2.1.1). Tail tips of the lizards were taken and preserved in ethanol, while all animals were immediately released at their collection sites.

Code	Species	Subspecies	Country	Lat	Lon	MC1R	ReLN	Map
Balkans								
5201	*D. praticola*	*pontica*	Greece	41.33	24.88	5,5	-,-	4
5175	*D. praticola*	*pontica*	Romania	45.2180	23.3938	5,5	7,7	3
5176	*D. praticola*	*pontica*	Romania	45.2180	23.3938	5,5	7,7	3
5183	*D. praticola*	*pontica*	Romania	45.2180	23.3938	5,5	7,7	3
5204	*D. praticola*	*pontica*	Romania	45.2180	23.3938	-,-	7,7	3
13283	*D. praticola*	*pontica*	Serbia	44.6829	20.5528	3,4	5,5	1
13285	*D. praticola*	*pontica*	Serbia	44.6829	20.5528	3,3	-,-	1
13293	*D. praticola*	*pontica*	Serbia	44.6829	20.5528	4,4	5,6	1
13297	*D. praticola*	*pontica*	Serbia	44.6824	20.5528	3,3	5,6	1
13294	*D. praticola*	*pontica*	Serbia	44.6815	20.5518	3,3	6,6	1
13297	*D. praticola*	*pontica*	Serbia	44.6824	20.5528	3,3	5,6	1
5200	*D. praticola*	*pontica*	Serbia	44.6822	20.5521	3,3	5,6	1
5177	*D. praticola*	*pontica*	Serbia	43.2895	21.3070	-,-	5,7	2
16331	*D. praticola*	*pontica*	Turkey	41.9718	27.4232	5,5	7,7	5
Caucasus								
10360	*D. praticola*	*loriensis*	Armenia	40.7623	44.8885	1,1	1,1	9
6279	*D. praticola*	*loriensis*	Armenia	40.7500	44.8667	1,1	8,8	9
13916	*D. praticola*	*hyrcanica*	Azerbaijan	38.4667	48.5833	1,1	-,-	10
13917	*D. praticola*	*hyrcanica*	Azerbaijan	38.4667	48.5833	1,1	2,2	10
13918	*D. praticola*	*hyrcanica*	Azerbaijan	38.4667	48.5833	1,1	2,2	10
12191	*D. praticola*	*pontica*	Russia	44.0574	40.7342	1,2	3,3	7
12192	*D. praticola*	*pontica*	Russia	44.0574	40.7342	2,2	-,-	7
12193	*D. praticola*	*pontica*	Russia	44.0574	40.7342	1,1	-,-	7
12194	*D. praticola*	*pontica*	Russia	44.0574	40.7342	1,2	4,4	7
12195	*D. praticola*	*pontica*	Russia	44.0574	40.7342	1,1	3,3	7
12196	*D. praticola*	*pontica*	Russia	44.0574	40.7342	2,2	3,3	7
12197	*D. praticola*	*pontica*	Russia	44.0574	40.7342	1,1	3,3	7
12198	*D. praticola*	*pontica*	Russia	44.0574	40.7342	1,2	-,-	7
14664	*D. praticola*	*pontica*	Russia	43.9500	40.6833	1,1	3,3	8
14665	*D. praticola*	*pontica*	Russia	43.9500	40.6833	1,2	3,3	8
14666	*D. praticola*	*pontica*	Russia	43.9500	40.6833	1,2	3,3	8
6300	*D. praticola*	*pontica*	Russia	44.7667	38.2500	1,1	9,9	6

Table 2.1.1: Codes of the individuals sequenced, and corresponding nuclear haplotypes for MC1R and ReLN nuclear markers.[1] Ascription according to Tuniyev et al. (2011, 2013) and Ljubisavljevic et al. (2006).

Total genomic DNA was extracted from approximately 30 mg of each tail-tip following

standard high-salt protocols (Sambrook et al., 2000). DNA was eluted with Elution buffer (Qiagen). For older samples, DNeasy Blood and Tissue kit (Qiagen) was used, following the manufacturer's protocol.

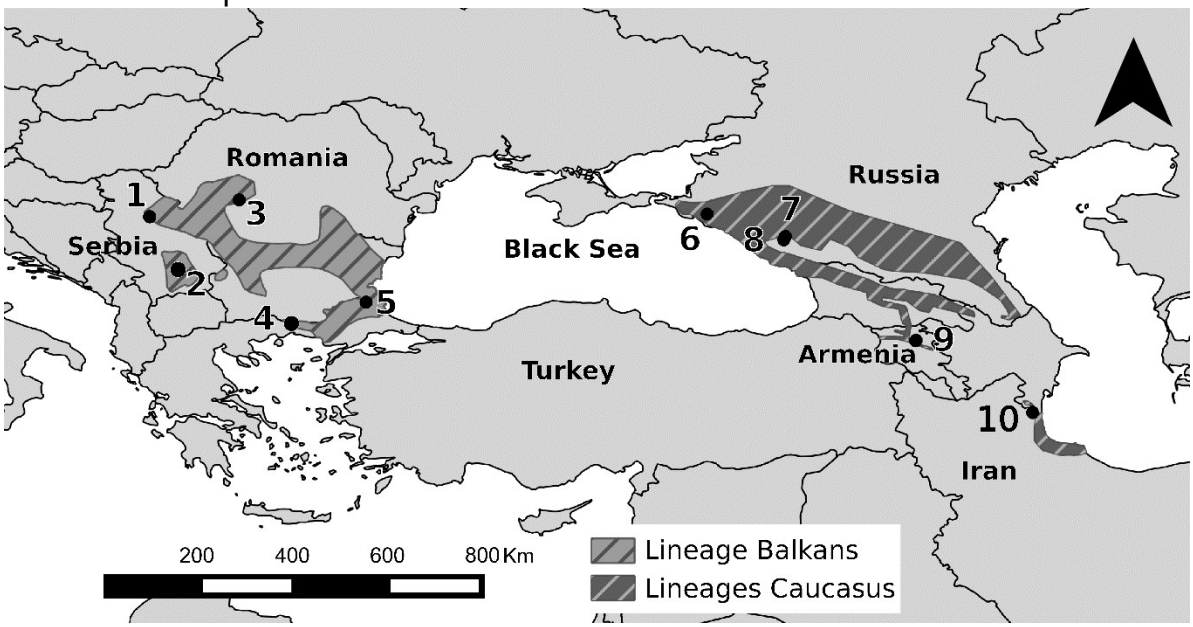


Figure 2.1.1: Distribution of *Darevskia praticola* sensu lato. The Balkan lineage is shown in light grey and the two Caucasus lineages in dark grey. Population numbers correspond to table I. Distribution is based on Valakos et al. (2008) and Agasyan et al. (2009).

For phylogenetic analyses two mtDNA genes Cytochrome-*b* (Cyt-*b*) and the 3′ end of the NADH dehydrogenase subunit 4 gene and adjacent tRNAs (ND4) and two nDNA genes (MC1R, ReLN) were selected. Primers for these genes were those described in Kocher et al. (1989), Arevalo et al. (1994) and Pinho et al. (2009). Amplifications were performed in 25 µL of 5x reaction buffer, 3.2 mM MgCl2, 1.6 mM each dNTP, 4.0 M each primer, 1U of Promega GoTaq DNA polymerase. PCRs consisted of an initial denaturation at 94 ºC for 2 min, followed by 40 cycles that included a denaturation step at 92 ºC for 30'', annealing at 57-65 ºC for 30'' and extension at 72 ºC for 1'. A final extension was conducted at 72 ºC for 5 min. Minor adjustments to conditions were required in some reactions. Sequencing was performed by an external facility (Macrogen® Europe).

Sequences were edited with Chromas Pro v1.7.6 (Technelysium Pty, Ltd) and aligned with the algorithms implemented in Mafft version 7 (Katoh et al., 2005), using default options for all the alignments except for RelN, for which was used the iterative refinement method E-INS-i. This algorithm is better suited for sequences with multiple conserved domains and long gaps, as we are expecting for the ReLN dataset. Heterozygous MC1R sequences were separated using PHASE 2.1.1 (Stephens et al., 2001; Stephens and Scheet 2005)

with the web tool SeqPhase (Flot 2010). Probability for every haplotype chosen was between 0.978 and 1.00. All new sequences were submitted to GenBank (accession numbers pending), and the alignments are available on Dryad (reference number pending).

Nucleotide diversity parameters were calculated with DNAsp (Librado and Rozas 2009). Genetic distances were calculated using MEGA5 (Tamura et al., 2011).

To reconstruct the phylogenetic relationships among groups, we performed a Bayesian analysis with each of the mitochondrial markers (Cyt-*b* and ND4) separately. Sequences from other *Darevskia* sp. were added and analysed together with those of *D. praticola* generated here. Such sequences were taken from our database, Ahmadzadeh et al. (2013) and available from GenBank (Table II), and chosen according to phylogenetic proximity to ingroup and availability. The best-fit evolutionary model for each gene was selected using the corrected Akaike information criterion in JMODELTEST v0.1.1 (Posada 2008) and PartitionFinder, for the partitioned loci Cyt-*b* and ND4 (Cyt-*b*/position 1: TVMef+I+G; ND4/position 3, Cyt-*b*/position 2: K81uf+I+G; ND4/position 1, Cyt-*b*/position 3: TIM+G; ND4/RNA subset, ND4/position 2: GTR+G; MC1R: TPM3uf+G).

Bayesian posterior probabilities were calculated with MrBayes v3.2 (Ronquist et al. 2011), using four heated Markov chains (default heating values) sampled every 1000 generations and run for 20 million generations. Each run was performed twice, to avoid being trapped in a local optimum, and these two sets were later combined. The first 6000 trees were discarded as burnin after analysing the convergence parameters for both runs in Tracer v1.4 (Rambaut and Drummond 2007). Maximum Likelihood analyses were performed using MEGA6 (Tamura et al, 2013) and node support was estimated using the bootstrap technique with 500 replicates (Felsenstein, 1985).

MC1R haplotype	DB Code	Species	Subspecies[1]	Country
1	10360	*D. praticola*	*loriensis*	Armenia
1	13916	*D. praticola*	*hyrcanica*	Azerbaijan
1	13917	*D. praticola*	*hyrcanica*	Azerbaijan
1	13918	*D. praticola*	*hyrcanica*	Azerbaijan
1	14664	*D. praticola*	*pontica*	Russia
1	14665	*D. praticola*	*pontica*	Russia
1	14666	*D. praticola*	*pontica*	Russia
1	6279	*D. praticola*	*loriensis*	Armenia
1	6300	*D. praticola*	*pontica*	Russia
1	12191	*D. praticola*	*pontica*	Russia
1	12193	*D. praticola*	*pontica*	Russia
1	12194	*D. praticola*	*pontica*	Russia
1	12195	*D. praticola*	*pontica*	Russia
1	12197	*D. praticola*	*pontica*	Russia
1	12198	*D. praticola*	*pontica*	Russia
2	13283	*D. praticola*	unknown	Serbia
2	13285	*D. praticola*	unknown	Serbia
2	13294	*D. praticola*	unknown	Serbia
2	13297	*D. praticola*	unknown	Serbia
2	5200	*D. praticola*	unknown	Serbia
3	13283	*D. praticola*	unknown	Serbia
3	13293	*D. praticola*	unknown	Serbia
4	14665	*D. praticola*	*pontica*	Russia
4	14666	*D. praticola*	*pontica*	Russia
4	12191	*D. praticola*	unknown	Russia
4	12192	*D. praticola*	unknown	Russia
4	12194	*D. praticola*	unknown	Russia
4	12196	*D. praticola*	unknown	Russia
4	12198	*D. praticola*	unknown	Russia
5	16331	*D. praticola*	unknown	Turkey
5	5175	*D. praticola*	unknown	Romania
5	5176	*D. praticola*	unknown	Romania
5	5183	*D. praticola*	unknown	Romania
5	5201	*D. praticola*	unknown	Greece
6	5392	*D. rudis*		Turkey
6	5402	*D. rudis*		Turkey
6	5403	*D. rudis*		Turkey
6	5250	*D. rudis*		Turkey
7	5402	*D. rudis*		Turkey
7	5403	*D. rudis*		Turkey
8	7802	*D. derjugini*		Georgia
9	7803	*D. derjugini*		Georgia
9	7846	*D. derjugini*		Georgia
10	7803	*D. derjugini*		Georgia
11	13797	*D. kopetdaghica*		Iran
11	13798	*D. kopetdaghica*		Iran
12	13807	*D. schaekeli*		Iran
13	13818	*D. kamii*		Iran

Table 2 (cont.)

MC1R haplotype	DB Code	Species	Subspecies[1]	Country
13	13819	*D. kamii*		Iran
13	6150	*D. caspica*		Iran
13	6151	*D. caspica*		Iran
14	13818	*D. kamii*		Iran
15	6146	*D. kamii*		Iran
16	6111	*D. steineri*		Iran
16	6112	*D. steineri*		Iran
16	6115	*D. steineri*		Iran
17	6117	*D. chlorogaster*		Iran
17	6118	*D. chlorogaster*		Iran
18	6200	*D. defilippii*		Iran
18	6201	*D. defilippii*		Iran
18	13799	*D. defilippii*		Iran
19	6316	*D. defilippii*		Iran
20	10197	*D. raddei*		Armenia
21	10197	*D. raddei*		Armenia
22	10019	*D. raddei*		Armenia
22	*LSH1*	*D. raddei*	*nairensis*	Armenia
23	*LSH1*	*D. raddei*	*nairensis*	Armenia
24	LSH3	*D. valentini*		Armenia
24	LSH4	*D. valentini*		Armenia
25	LSH3	*D. valentini*		Armenia
25	LSH4	*D. valentini*		Armenia
26	ST2	*D. portschinskii*	*portschinskii*	Armenia
26	10071	*D. portschinskii*	*portschinskii*	Armenia

Table 2.1.2 - Samples analysed for MC1R and Cyt-*b* analyses for other *Darevskia* species. Taxonomy according to Tuniyev et al. (2011, 2013) and Ljubisavljevic et al. (2006).

A multispecies coalescent model (Heled & Drummond 2010) as implemented in *BEAST was used to infer the species tree from multiple gene trees for all the lineages in *D. praticola*, In the multispecies coalescent model all model parameters are unlinked across loci, including the topology parameter, which allows the gene trees to differ in topology, even though being constrained by the species tree. Therefore, this approach allows for gene tree heterogeneity that can be due to incomplete lineage sorting, different evolutionary rates or even introgression. The dataset consisted of all the *D.praticola* individuals used in previous analyses.

Since the Cyt-*b* could possibly be a mix of mtDNA and nuclear pseudogene sequences (see results) it was discarded from the analyses and only the loci ND4 (mtDNA), MC1R (nuclear) and ReLN (nuclear) were used in our *BEAST analysis. The time of the most

recent common ancestor (tMRCA) was estimated with *BEAST (Drummond et al., 2012). This was set to run for 10^8 generations, sampling every 10^5 generation, and repeated 8 independent times. An uncorrelated lognormal relaxed clock was used for the mtDNA and nuclear markers. Individuals were constrained (traits option in *BEAST) as their taxonomy and geographical origin, meaning the *D. praticola* individuals were allocated to their lineages within the species and according to the previous ND4 Bayesian phylogenetic analysis in this study.

The prior for the ucld.mean parameter for the ND4 dataset was set as a normal distribution with a mean of 0.0226 and a standard deviation of 0.0031, so that mutation rate would be able to vary between 0.0278 and 0.0174 mutation/site/million years. This mutation rate interval was estimated for the same segment of the ND4 gene in genus *Podarcis* (Pinho et al., 2007). LogCombiner (Drummond et al., 2012) was used to compile the log and tree files of the eight cloned runs. 10% of the trees were discarded as burnin, following an analysis of convergence of individual run parameters in Tracer v1.4 (Rambaut and Drummond 2007). Evolutionary models applied were the most approximate to the Modeltest selected models available in BEAST (ND4: HKY+G; MC1R: GTR+GI; ReLN : HKY+I). The Yule process of speciation was selected as a tree prior with a random starting tree. High effective sample sizes were observed for all parameters in *BEAST analysis (posterior ESS values > 200 for the joined analyses). Convergence for all model parameters was determined in Tracer when sample size (ESS) > 200. Maximum clade credibility trees with divergence time means and 95% highest probability densities (HPDs) were produced using Tree Annotator (Drummond et al., 2012).

After phasing all haplotypes, a statistical parsimony network was built in TCS v1.21 (Clement et al., 2000) for the MC1R marker, with *D. praticola* and other *Darevskia* species for which data was available on GenBank. For the ReLN gene, a median joining approach was used to visualise the relationship among haplotypes using the software Network, considering gaps as a 5^{th} state (Bandelt et al., 1999).

RESULTS

Phylogenetic analysis

Bayesian algorithms produced different topologies for the mtDNA markers (Figure 2.1.2 and 2.1.3). In the Cyt-*b* tree (Figure 2.1.2), *D. praticola* is divided into two highly divergent

clades that are not sisters. Therefore, *D. praticola* appears paraphyletic. The lineage from the Caucasus occupies a basal position in the tree, with very low intralineal variation. The divergence between the *D. praticola* clades is approximately 12%.

It is noteworthy that in the Cyt-*b* tree (Figure 2.1.2) a published sequence (U88612) from Russia, vicinity of Sochi (43º35'N, 39º46'E) was placed in the Balkanic clade and not with the other Caucasian individuals as expected. However, the distance between this individual and the Balkanic *D. praticola* is 5%.

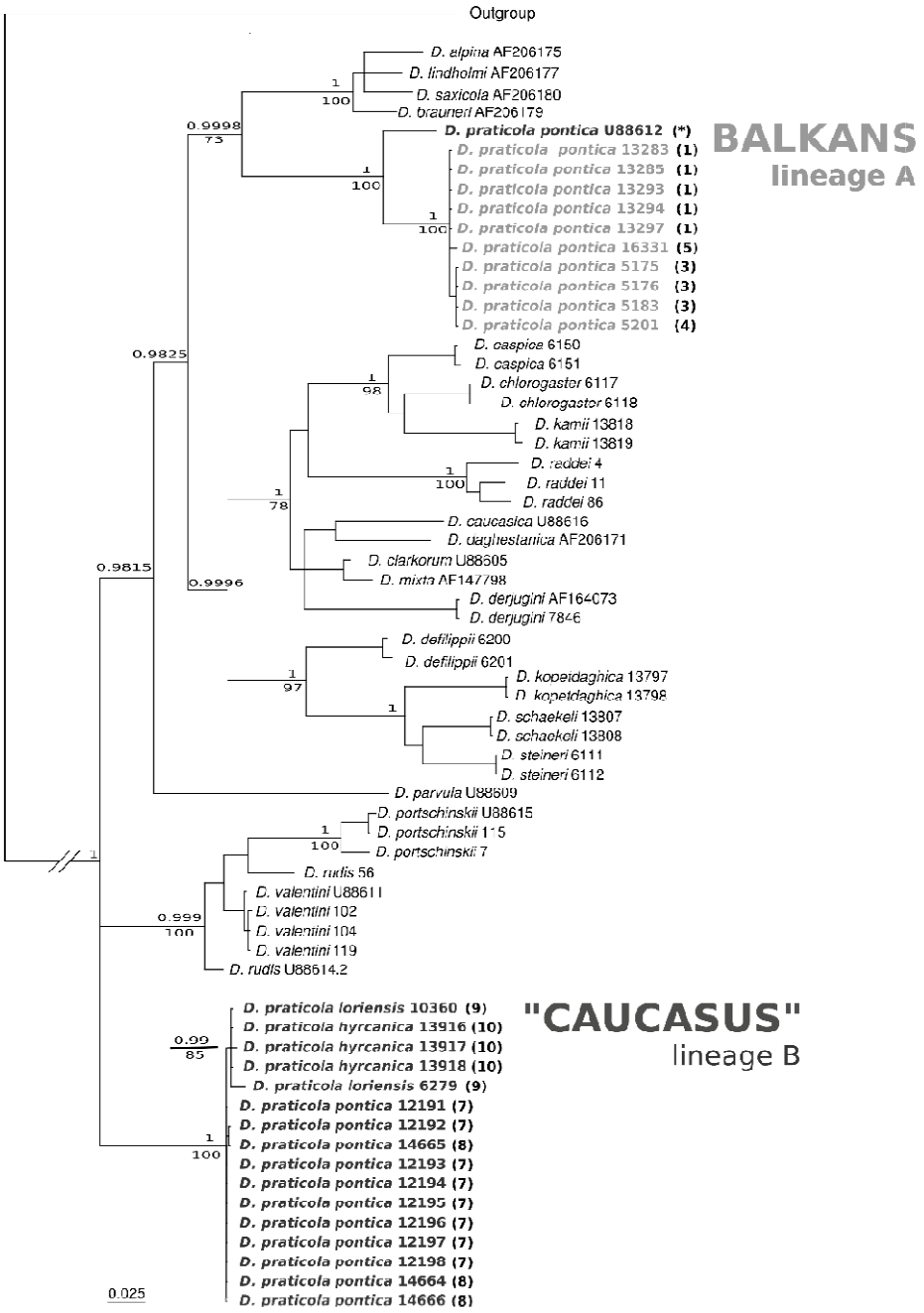


Figure 2.1.2 – Bayesian and ML phylogeny based on the mtDNA gene Cyt-*b* for an enlarged species dataset of the genus *Darevskia*. The Balkan lineage includes all the samples from the Balkans and is sister to the sample U88612 from the Caucasus. The other samples from the Caucasus belong to a single lineage of the basal trichotomy.

The topology of the ND4 tree (Figure 2.1.3) differs from that of the Cyt-*b* tree. It shows three clades within D. praticola that are monophyletic and less divergent than the Cyt-*b* clades. The ND4 tree recovered divergence not only between the Balkans and the Caucasus, but also between samples from the Greater Caucasus and from Transcaucasia (Armenia and Azerbaijan). The divergence between the Caucasian and the Balkanic lineage corresponds to 5.1-5.7%.

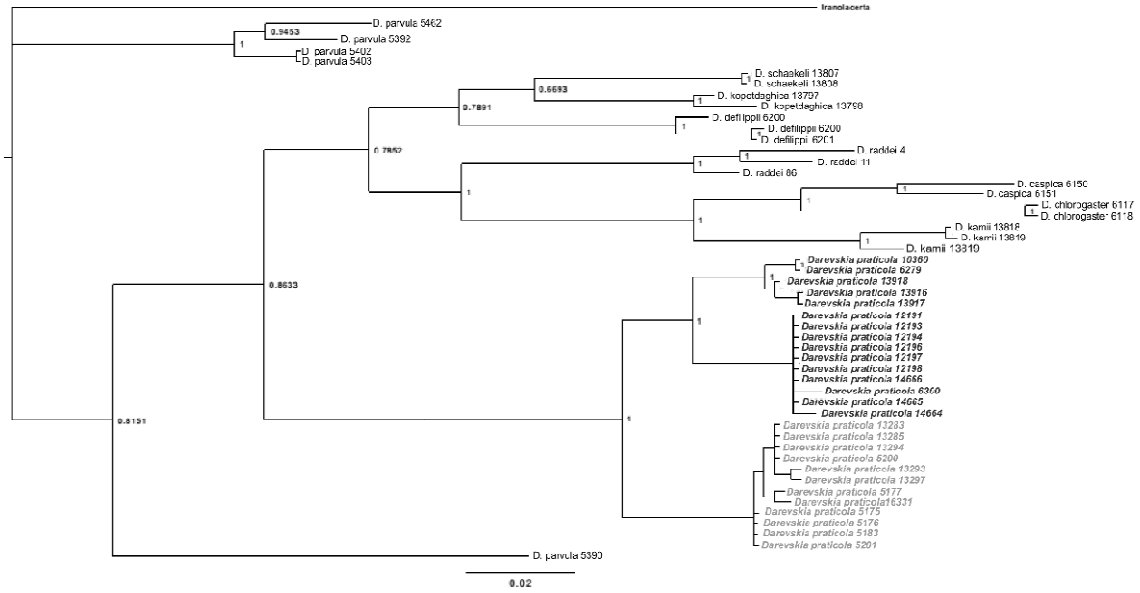


Figure 2.1.3 - Bayesian and ML phylogeny based on the mtDNA gene ND4 for *D.praticola* and outgroup species.

Given the topology of the Cyt-*b* tree, with a paraphyletic structure of the genus, with one of the lineages being a basal clade in the main tree, with very low differentiation, this suggests we may have amplified a pseudo-gene.

The divergence between clades in the ND4 tree is much lower than the same parameter for the Cyt-*b* tree (~5% vs 12%), even though the ND4 is a faster evolving gene. Still, the ND4 divergence between the Caucasian and Balkanic value (~5%) is concordant with the distance between the Genbank sequence (U88612) and the Balkanic sequences in the Cyt-*b* tree (5%).

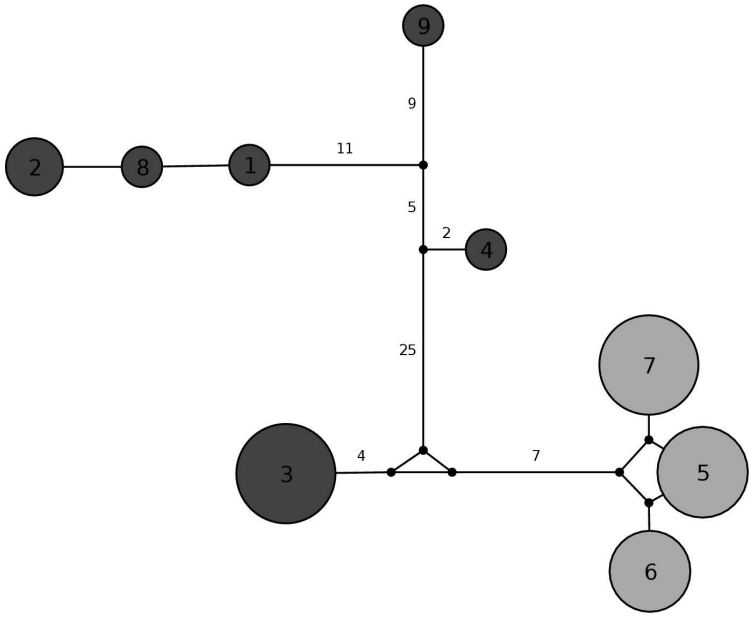


Figure 2.1.4 - Median-joining network for the nuclear gene RELN, based on the sequencing of 23 individuals of *D. praticola*. Number of substitutions is shown along the network branches. If branches are not noted, the corresponding haplotypes are separated by a single step. Small black circles correspond to the median vectors. Light grey represents the Balkan lineage and dark grey the Caucasian lineage. Circles correspond to haplotypes, numbered as in table I, with size reflecting the number of individuals per haplotype.

Finally, the MC1R phylogenetic tree showed a monophyletic *D.praticola* group, concordant with the ND4 tree topology (data not shown).

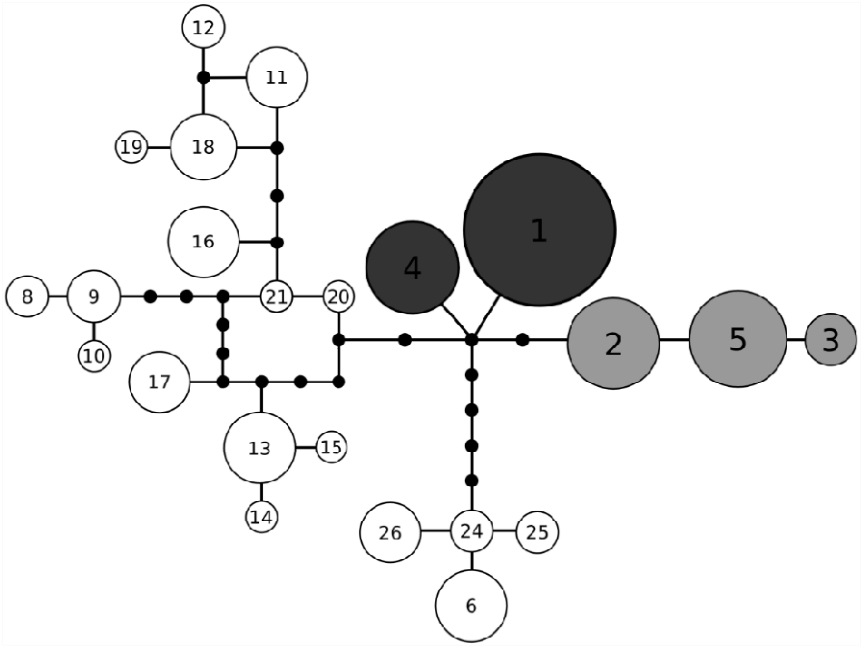
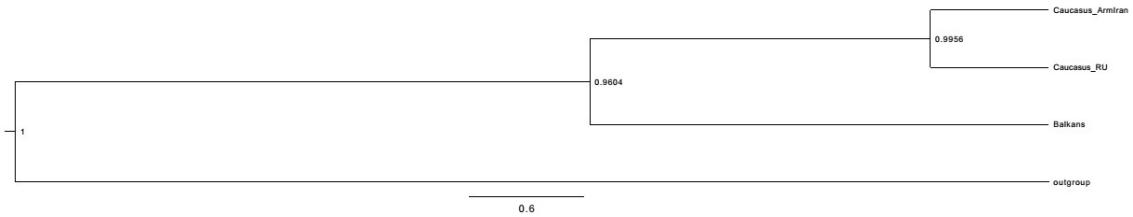


Figure 2.1.5 - MC1R statistical parsimony network of this nuclear gene for *D. praticola* sensu lato including outgroup species. Light grey represents the Balkan lineage and dark grey the Caucasian lineage. The circles in white correspond to Darevskia species that do not belong to the D. *praticola*-group. Numbered circles correspond to unique haplotypes, as in table II, their size reflecting the number of individuals per haplotype. Small black dots represent the median vectors.

Although we did not detect unusual amino-acid changes (results not shown), the strong conflict with ND4 and MC1R and the unexpected levels of variation led us to be conservative and not include this marker in the further analyses with the nuclear loci.



2.1.6 – Species tree recovered from *BEAST analysis of three genes (ND4, RELN and MC1R). The traits used (“Caucasus”, “Balkans”, “Armenia and Iran” and “Russian Caucasus”) were selected according to the ND4 tree lineages.

The species tree (Figure 2.1.6) recovers two sister clades within the Caucasus. One of these includes 12 individuals from the western Greater Caucasus (Krasnodarskiy Krai, Russian Federation). Individuals from that area are classified as *D. praticola pontica* (numbers 6, 7 and 8 in Figure 2.1.1). The other, Transcaucasian clade includes 3 individuals from the Talysh Mountains (Astara District, Azerbaijan) classified as *D. praticola*

hyrcanica (number 10 in Figure I) and 3 individuals from the Lesser Caucasus (Tavush Marz, Armenia) classified as *D. praticola loriensis* (number 9 in Figure I).

These two clades are much more distantly related to the Balkans clade that grouped all samples from the Balkans: including 7 individuals from two localities in Serbia (approximately 200-250 Km apart), 3 from one locality in Romania, 1 from Greece and 1 from Turkish Thrace (in Figure 2.1.1, numbers 1 and 2, 3, 4 and 5, respectively). Populations from these areas are currently assigned to *D. praticola pontica*. The sister relationship of the two Caucasian clades and their monophyly with the Balkanic clade were strongly supported (both posterior probability values > 0.96). In the species tree reconstruction, the group assignment of individuals followed the ND4 tree topology.

Individual gene trees from the *BEAST analysis overall show the same pattern as the species tree. The ReLN gene tree from *BEAST, however, groups some individuals of the Caucasus and the Balkans in the same lineage: haplotype 3 from the Caucasus is closer to the Balkanic haplotypes than to the remaining Caucasian haplotypes, as in figure 2.1.4. This could be due to incomplete lineage sorting, which is common with nuclear markers even across species, or an incorrect value given to the gaps in the analysis.

MC1R statistical parsimony network (Figure V) identified five haplotypes in *D. praticola* base on 6 polymorphic and 6 parsimony informative sites in a total of 629 bp. Haplotypes 1 was shared by individuals from the Russian Caucasus, Armenia, and Iran. Haplotype 4 was found only in the Russian Caucasus and differed from Haplotype 1 by two substitutions. Haplotypes 2, 3 and 5 were found only in the Balkans and differed from each other by one or two substitutions. The distance between Caucasian and Balkanic haplotypes varied from three to five substitutions. Therefore, overall geographic structure of MC1R haplotypes was concordant with that of the ND4 haplotypes and the species tree.

The median joining analysis on the ReLN gene recovered a network of nine haplotypes based on 37 polymorphic and 36 parsimony informative sites of 774 bp (Figure 2.1.4). Three haplotypes (5-7) were found only in the Balkans and differed by 2-4 substitutions from each other. Haplotypes 1-2 and 8, were found in Armenia and Iran and formed another closely related group of haplotyes that differed by one or two mutations from each other. Haplotypes 3-4 and 9 were found in the Russian Caucasus, even though they were a bit apart from each other.

The divergence time between the Balkanic and the common ancestor of the two Caucasian clades was estimated between 0.75 and 5.12 Ma with the mean of 2.5 Ma

(Table 2.1.3). The divergence between the Caucasus and Transcaucasia clades was estimated approximately 0.6 Ma (95% HPD interval 0.073 - 1.56 Ma. This split gave rise to a North-Caucasian lineage in the north (identified as “Caucasus_RU” in the tree) and another to the South of Caucasus, with the individuals from Armenia and Azerbaijan (identified as “Caucasus_ArmIran” in the tree).

DISCUSSION

(Dis)agreement of the phylogeographic signal across loci

Our results constitute the first phylogeographic analysis of *D. praticola* across its fragmented range. Both mitochondrial (ND4) and nuclear loci (MC1R and ReLN) identified the presence of a strong phylogeographic structuring concordant across them. On the contrary, they disagree with the current intraspecific taxonomy based on morphological characters of scalation and colouration (Ilgaz and Kumlutaş 2005; Ljubisavljevic et al., 2006; Tuniyev et al., 2011).

In our analysis, not all markers were concordant. The Cyt-*b* tree yielded unexpected results, discordant from the trees of other loci analysed. *D. praticola* Cyt-*b* haplotypes divided themselves into two very divergent clades. One of the clades had all samples from the Balkans and a previously published sequence from the northwestern Caucasus (U88612). The other clade included our samples from the northwestern Caucasus and represented a short branched basal clade in the tree.

ND4 and MC1R gene trees showed a different topology regarding the Cyt-*b* tree, but consistent with between them, with a monophyletic *D.praticola.* All these suggestions made us supect our Cyt-b sequences from the Caucasus could be of nuclear origin.

Finally, the divergence between the Balkanic and Caucasian lineages in the ND4 tree is much lower than in the Cyt-*b* tree (~5% vs 12%), even though ND4 is a faster evolving gene. However, an equivalent divergence (5%) was depicted between a Genbank sequence (U88612), originary from the Caucasus, and the Balkanic lineage. This could be an indication we were not amplifying the correct sequence, as there was a similar pattern between the previous published sequence and our Balkanic sequence data in comparison with the other loci (ND4 and MC1R).

Given all of this information, although no stop codons or double peaks were found in these

sequences, due to the similarity of this pattern with other pseudo-gene cases in lacertids (e.g. Podnar et al., 2007 we suspected that our Cyt-*b* sequences from Caucasus could be of nuclear origin, or a pseudo-gene. Thus, we removed Cyt-*b* from the multilocus analysis.
Our results provide a novel perspective on relationships within *Darevskia.* Most of the phylogenetic analyses of the genus published to date are based on Cyt-*b* data (Murphy et al., 1996; Tarkhnishvilli 2012). Since our data indicate commonly used, universal primers for Cyt-*b* may amplify nuclear copies in some taxa and lead to erroneous tree reconstruction, it highlights the limitations of approaches based on single molecular markers. The evolutionary history of *Darevskia* is complex. It is characterized by frequent hybridisation events between different species (Darevsky, 1967, Danielyan et al., 2008). Some of these hybridisations are responsible for the origin of parthenogenetic taxa (Murphy et al., 1996; Tarkhnishvilli 2012). Therefore, a more rigorous, multilocus analysis of the relationships within *Darevskia* is needed to better understand the complexity of its evolutionary history.

Phylogeographic patterns within *D. praticola sensu lato*

The molecular markers analysed support a division of the currently recognized *D. praticola* into two main clades concordant with the disjoint parts of the range - the Balkans and the Caucasus. The Balkanic clade includes the individuals from Serbia, Romania, Greece, and Turkish Thrace. The Caucasian clade includes individuals from Russian Caucasus, Armenia, Azerbaijan, and Iran. Even though the sample size is relatively small, the individuals included in this study reasonably cover the whole distribution range of *D. praticola* sensu lato. While regions such as Bulgaria, Central Isthmus or Georgia would deserve further analysis, this sampling is likely to provide a robust prediction of the whole group. The divergence between these clades is approximately 5% in the ND4 dataset. Nuclear loci had no shared haplotypes between the Balkans and Caucasus. Furthermore, in both loci, Balkanic haplotypes were more closely related to each other than to Caucasian haplotypes. Therefore, our data suggests the lack of gene flow and, as follows, an evolutionary independence of *D. praticola* populations inhabiting Balkanic and Caucasian parts of the species range. In other words, Balkanic and Caucasian parts of the species range represent independent “evolutionary significant units” (Moritz 1994), or even different species.
This study recovered divergence not only between Balkans and Caucasus, but also

divergence between samples within the Caucasus. The localtions from the Greater Caucasus and from Transcaucasia (Armenia and Azerbaijan) represent two closely related clades, a pattern first depicted in the ND4 tree (Figure 2.1.3), and further analysed in the species tree (Figure 2.1.6). One of these clades included all individuals from the western part of the Greater Caucasus range (Krasnodarskyi Kray, Russsian Federation). The other included individuals sampled in Transcaucasia (Armenia, Azerbaijan, Iran), here the two subspecies found in the area also possible to distinguish. This is partially consistent with phenotypic assignment of the Caucasian individuals to *D. praticola pontica* and the group formed by *D. praticola loriensis* (Armenia) and *D. praticola hyrcanica* (Talysh Mountains) (Figure 2.1.3). Unfortunately, our sampling is both geographically and sample size wise insufficient to reach firm conclusion about evolutionary independence of these clades. However, our data clearly indicates the presence of some geographic structuring of genetic variation within the Caucasus region that needs to be studied in a greater detail.

Evolutionary history

The divergence between the Balkans and the Caucasus has been estimated at approximately 2.50 Ma (95% HPD interval 0.75-5.12 Ma). Although the 95% HPD interval is large and covers all of Pliocene and half of Pleistocene, the divergence time estimate is concordant with other species that have similar distribution and is likely to be associated with the Pliocene-Pleistocene transition. This transition is characterized by a switch from a dry and warm period of moderate stability to a period with multiple abrupt glacial advances. Pleistocene glaciations facilitated genetic divergence in a variety of organisms (Hewitt 2000). South European peninsulas and mountains, such as the Iberian Peninsula and the Balkans (Salvi et al., 2013) as well as the Caucasus (Ursenbacher et al., 2006) played an important role in the maintenance and increase of regional biodiversity. These southern refugia acted as cradels of divergence and speciation and provided the pioners for interglacial expansion and colonisation of new habitats (Knowles 2001).

At the end of the Pliocene, the Balkans and Asia Minor were connected and the Black Sea was isolated from the Mediterranean (Popov et al., 2004). In the beginning of the Pleistocene, the sea level dropped even more due to the increase of the ice caps. However, the aridification of the climate apparently facilitated the vicariance between the Balkanic and Caucasian forests, despite the land connections along northern and southern

Black Sea coasts. Only during the late Pleistocene did forests expand, and the Black Sea level gradually rose due to the melting of glaciers. At that time, the Bosphorus strait was a narrow canyon not allowing a large flow of water from the Black Sea to the Marmara Sea, which led to the water accumulation and Black Sea level rise (Esin et al., 2014). These vicariant events likely shaped the phylogeography of *D. praticola*. Similar paleoclimatic scenarios have been proposed to give rise to similar lineage diversification in other reptile species in this region (Ahmadzadeh et al., 2013a, 2013b, 2013c).
Some authors had already suggested the existence of two different species within *D. praticola.* However, they grouped the Balkans with the northwestern Caucasus in one and the remaining range in the Caucasus into another (Tuniyev et al., 2011). This hypothesis requires a recent dispersal between the Balkans and northwestern Caucasus. Our data rejects this hypothesis suggesting instead that the divergence between the Balkans and the Caucasus is the deepest in the *D. praticola* tree.

Using morphological characters, Tuniyev *et al. (*2011) suggested the species status for *D. praticola sensu stricto* and *D. pontica* assigning the lectotype for the latter from the northwestern Caucasus. Balkanic and northwestern Caucasus individuals, treated as *D. pontica*, were distinguished from *D. praticola* individuals by the presence of a 3/3 arrangement of chin shields, longer head, distinct upper positioned masseteric plate in the temporal region (approaching toward tympanic plate and supratemporals), and tympanic plate separated from the supratemporals (Ljubisavljevic et al., 2006; Tuniyev et al., 2011). However, these authors conceded a degree of uncertainty regarding the characters in both taxa, especially due to a putative overlap in these characters in the Stravropol Plateau, where individuals with 3/3 ("*D. pontica*") and with 2/2 ("*D. praticola*") chin shields are found sympatrically. Although we did not have access to samples from this region, the fact that our samples west and east from this area all fell within the Caucasian clade suggests that, if there is any contact there, it should be between its subclades, and not between the Caucasian and the Balkanic clades. Ryabinina et al. (2011) had already proposed this hypothesis, although the limited sampling and the type of markers used (RAPD and inter-MIR-PCR) do now allow a full comparison with our results.

Taxonomical implications

Our results clearly disagree with the current intraspecific taxonomy of *D. praticola*.

Morphological data indicate that *D. p. pontica* inhabits both the Balkans and northwestern Caucasus, and differs from *D. p. praticola, D. p. hyrcanica* and *D. p. loriensis* in Transcaucasia (Tuniyev *et al.,* 2011). However, our multilocus molecular data indicate the presence of a deep divergence between the Balkanic and the Caucasus clades, possible dating back to the Late Pliocene, and that the northwestern Caucasus clade is much more closely related to the Transcaucasia clade. The northwestern Caucasus clade includes eastern populations of *D. p. praticola* and populations classified as *D. p. pontica,* suggesting that these subspecies are synonymous. Our data also suggest that *D. p. hyrcanica* and *D. p. loriensis* represent two less divergent mtDNA lineages, even though they share the nuclear loci haplotypes. This is concordant with the subspecies differentiation. The paraphyly of *D. p. pontica* and the deep divergence between its Balkanic and Caucasian populations renders applications of this subspecies to both parts of the range inappropriate. Populations from northwestern Caucasus should retain this name because a lectotype for *D. praticola* has been designated from Gagry, northwestern Caucasus (Tuniyev et al., 2011). As suggested by Ljubisavljevic et al. (2006), the name *D. p. hungarica* first described by Sobolevsky in 1930 (see Stugren 1961) with the type locality in the Transylvanian Alps would be available for the Balkanic clade.

In conclusion, and given the deep split found between the Caucasian and the Balkanic clades, it would be more objective to elevate them to a species level. Given this, *D. hungarica* would then include all individuals from the Balkans, and *D. praticola* the Caucasian indiduals. Here, nested structure would still discriminate the subspecies *D.p.pontica*, for the western Caucasus individuals, *D. p. loriensis* (for the Armenian individuals) and *D. p. hyrcanica* (for the Iranian individuals present in the Talysh mountains).

Author's contributions

SF performed the analyses and wrote the manuscript draft. SF and AV generated the molecular data. SF, SD, JH and MC were involved in most of the results discussion and manuscript writing. SF, MA, SD, JC-I, AK, DC, CC, PL and MC performed the sample collection. MC was responsible for the funding for the molecular work. All authors read, corrected and approved the final manuscript.

Acknowledgements

We thank S. Ferreira and S. Rocha for their valuable comments on the manuscript. Also N. Sillero, F. Jorge, A. Vardanyan, S. Ferreira, K. A. Matushkina, A. L. Timoshina and E. Galoyan for fieldwork assistance. This research was partially funded by FCOMP-01-0124-FEDER-007062 FCT through the project PTDC/BIA-BEC/101256/2008, by the project “Biodiversity, Ecology and Global Change” co-financed by North Portugal Regional Operational Program 2007/2013 (ON.2 – O Novo Norte), under the National Strategic Reference Framework (NSRF), through the European Regional Development Fund (ERDF) and by the project “Preserving Armenian biodiversity: Joint Portuguese – Armenian program for training in modern conservation biology.” of Gulbenkian Foundation (Portugal). JCI was funded by Grant No.173025 “Evolution in heterogenous environments: adaptation mechanisms, biomonitoring and biodiversity conservation” of Ministry of Education, Science and Technological Development of Republic of Serbia.

REFERENCES

Agasyan, A., Avci, A., Tuniyev, B., Crnobrnja-Isailović, J., Lymberakis, P., Andrén, C., ...Krecsák, L. (2009). *Darevskia praticola*. IUCN 2014: IUCN Red List of Threatened Species, Version 2014.1.

Ahmadzadeh, F., Flecks, M., Carretero, M. A., Böhme, W., Ilgaz, C., Engler, J. O., …Rödder, D. (2013a). Rapid Lizard radiation lacking niche conservatism: ecological diversification within a complex landscape. *Journal of Biogeography, 40,* 1807-1818.

Ahmadzadeh, F., Flecks, M., Carretero, M. A., Mozaffari, O., Böhme, W., Harris, D. J., ...Rödder, D. (2013b). Cryptic Speciation Patterns in Iranian Rock Lizards Uncovered by Integrative Taxonomy. *PLoS ONE, 8.* doi: 10.1371/journal.pone.0080563

Ahmadzadeh, F., Flecks, M., Rödder, D., Böhme, W., Ilgaz, C., Engler, J. O., ...Carretero MA (2013c). Multiple dispersal out of Anatolia: biogeography and evolution of oriental green lizards. *Biological Journal of the Linnean Society, 110,* 398-408.

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

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

Bandelt, H-J., Forster, P., & Röhl, A. (1999). Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution, 16,* 37-48.

Bishoff, W. (1976). Echsen des Kaukasus 11. Die Wieseneidechse, Lacerta praticola Eversmann, 1834. *Aquarien Terrarien, 12,* 415-417.

Cavazza, W., & Wezel, F. C. (2003). The Mediterranean region-a geological primer. *Episodes, 26*, 160-168.

Clement, M., Posada, D., & Crandall, K. A. (2000). TCS: a computer program to estimate

gene genealogies. *Molecular Ecology, 9,* 1657-1660.

Collar, D. C., Schulte, J. A. I., & Losos, J. B. (2011). Evolution of extreme body size disparity in monitor lizards (*Varanus*). *Evolution, 65,* 2664-2680.

Crochet, P-A., Chaline, O., Surget-Groba, Y., Debain, C., & Cheylan, M. (2004). Speciation in mountains: phylogeography and phylogeny of the rock lizards genus *Iberolacerta* (Reptilia:Lacertidae). *Molecular Phylogenetics and Evolution, 30,* 860-866.

Danielyan, F. D., Arakelyan, M., & Stepanyan, I. (2008). Hybrids of *Darevskia valentini*, *D. armeniaca* and *D. unisexualis* from a sympatric population in Armenia. *Amphibia-Reptilia, 29,* 487-504.

Darevsky, I. S. (1967). *Rock lizards of the Caucasus (Systematics, Ecology and Phylogenesis of the polymorphic groups of Rock lizards of the Subgenus Archaeolacerta).* Nauka Press, Leningrad, 1-276. [English translation: New Delhi: India National Scientific Documentation Centre]

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

Drummond, A. J., Suchard, M., Xie, D., Rambaut, A. (2012). Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution, 29,* 1969-1973.

Esin, N. V., & Esin, N. I. (2014). Mathematical modelling of Black Sea level change for the last 20000 years. *Quaternary International, 345,* 32-47.

Felsenstein, J. (1985). Confidence limits on phylogenies: an approach using the bootstrap. *Evolution, 39*, 783–791.

Fiz-Palacios, O., & Valcárcel, V. (2013). From Messinian crisis to Mediterranean climate: A temporal gap of diversification recovered from multiple plant phylogenies. *Perspectives in Plant Ecology, Evolution and Systematics, 15,* 130–137.

Fritz, U., Ayaz, D., Hundsdörfer, A. K., Kotenko, T., Guicking, D., Wink, M., ...Buschbom J (2009). Mitochondrial diversity of European pond turtles (*Emys orbicularis*) in Anatolia and the Ponto-Caspian Region: Multiple old refuges, hotspot of extant diversification and critically endangered endemics. *Organisms, Diversity & Evolution, 9,* 100-114.

Flot (2010). SeqPHASE: a web tool for interconverting PHASE input/output files and FASTA sequence alignments. *Molecular Ecology Resources, 10,* 162-166.

García-Castellanos, D., Estrada, F., Jiménez-Munt, I., Gorini, C., Fernàndez, M., Vergés J, & De Vicente, R. (2009). Catastrophic flood of the Mediterranean after the Messinian salinity crisis. *Nature, 462,* 778-782.

Garcia-Porta, J., Litvinchuk, S. N., Crochet, P. A., Romano, A., Geniez, P., Lo Valvo, M., ...Carranza S (2012). Molecular phylogenetics and historical biogeography of the west-palearctic common toads (*Bufo bufo* species complex). *Molecular Phylogenetics and Evolution, 63,* 113-130.

Guicking, D., Joger, U., & Wink, M. (2009). Cryptic diversity in a Eurasian water snake (*Natrix tesselata*, Serpentes: Colubridae): Evidence from mitochondrial sequence data and nuclear ISSR-PCR fingerprinting. *Organisms, Diversity and Evolution, 9,* 201-214.

Guindon, S., Dufayard, J. F., Lefort, V., Anisimova, M., Hordijk, W., & Gascuel, O. (2010). New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. *Systematic Biology, 59,* 307-21.

Gvoždík, V., Jandzik, D., Lymberakis, P., Jablonski, D., & Moravec, J. (2010). Slow worm, *Anguis fragilis* (Reptilia: Anguidae) as a species complex: Genetic structure reveals deep divergences. *Molecular Phylogenetics and Evolution, 55,* 460-472.

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. (2000). The genetic legacy of the Quaternary ice ages. *Nature, 405, 907-913.* doi:10.1038/35016000

Hillis, D. M. (1987). Molecular Versus Morphological Approaches to Systematics. *Annual Review of Ecology and Systematics, 18,* 23-42.

Ilgaz, Ç., & Kumlutaş, Y. (2005). The amphibian and Reptile Species of İğneada (Kırklareli) and its vicinity. *Pakistan Journal of Biological Sciences, 8,* 558-560.

Katoh K, Kuma K, Toh H, Miyata T (2005). MAFFT version 5: improvement in accuracy of multiple sequence alignment. Nucleic Acids Research 33: 511–518.

Kaya, S., Boztepe, Z., & Çiplak, B. (2013). Phylogeography of *Troglophilus* (Orthoptera: Troglophilinae) based on Anatolian members of the genus: radiation of an old lineage following the Messinian. *Biological Journal of the Linnean Society, 108,* 335-348.

Knowles, L. L. (2001). Did the Pleistocene glaciations promote divergence? Tests of explicit refugial models in montane grasshopprers. *Molecular Ecology, 10,* 691–701

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 USA, 86,* 6196-6200.

Lanfear, R., Calcott, B., Ho, S. Y. W., & Guindon, S. (2012). PartitionFinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Molecular Biology and Evolution, 29,* 1695-1701.

Librado, P., & Rozas, J. (2009). DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics, 25,* 1451-1452.

Ljubisaljevic, K., Dzukic, G., & Kalezic, M. L. (2008). Female reproductive life history traits of the meadow lizard, *Darevskia praticola* (Eversmann, 1834) from westernmost boundary of the species range. *Polish Journal of Ecology, 56,* 289 – 297.

Ljubisavljevic, K., Orlova, V. F., Dzukic, G. & Kalezic, M. L. (2006). Geographic patterns in morphological variation of the meadow lizard, *Darevskia praticola* (Lacertidae): taxonomical and biogeographical implications. *Periodicum biologorum, 108,* 47 – 55.

Llorente, G., Vidal-García, M., Garriga, N., Carranza, S., Pleguezuelos, J., & Santos, X. (2012). Lessons from a complex biogeographical scenario: morphological characters match mitochondrial lineages within Iberian *Coronella austriaca* (Reptilia: Colubridae). *Biological Journal of Linnean Society, 106,* 210-223.

Lymberakis, P., Poulakakis, N., Manthalou, G., Tsigenopoulos, C. S., Magoulas A., & Mylonas, M. (2007). Mitochondrial phylogeography of *Rana (Pelophylax) populations* in the Eastern Mediterranean region. *Molecular Phylogenetics and Evolution, 44,* 115-125.

Marosi, B., Zinenko, O. I., Ghira, I. V., Crnobrnja-Isailovic, J., Lymberakis, P., Sos, T., & Popescu, O. (2012). Molecular data confirm recent fluctuations of northern border of dice

snake (*Natrix tessellata*) range in Eastern Europe. *Northwestern Journal of Zoology, 8,* 374-377.

Mashkaryan, V., Vamberger, M., Arakelyan, M., Hezaveh, N., Carretero, M. A., Corti, C., ...Uwe Fritz (2013). Gene flow among deeply divergent mtDNA lineages of *Testudo graeca* (Linnaeus, 1758) in Transcaucasia. *Amphibia-Reptilia, 34,* 337-351.

Moritz, C. (1994). Defining 'evolutionary significant units' for conservation. *Trends in Ecololy & Evolution, 9*, 373–375.

Murphy, R. W., Darevsky, I. S., MacCulloch, R. D., Fu, J., & Kupriyanova, L. A. (1996). Evolution of the bisexual species of Caucasian Rock Lizards: a phylogenetic evaluation of allozyme data. *Russian Journal of Herpetology, 3,* 18 – 31.

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

Pepper, M., Doughty, P., Hutchinson, & M., Keogh, J. (2011). Ancient drainages divide cryptic species in Australia's arid zone: Morphological and multi-gene evidence for four new species of Beaked Geckos (*Rhynchoedura)*. *Molecular Phylogenetics and Evolution, 61,* 810-822.

Perea, S., Böhme, M., Zupancic, P., Freyhof, J., Sanda, R., Ozuluğ, M., ...Doadrio, I. (2010). Phylogenetic relationships and biogeographical patterns in Circum-Mediterranean subfamily Leuciscinae (Teleostei, Cyprinidae) inferred from both mitochondrial and nuclear data. *BMC Evolutionary Biology, 10,* 265.

Pinho, C., Harris, D. J., & Ferrand, N. (2007). Contrasting patterns of population subdivision and historical demography in three western Mediterranean lizard species inferred from mitochondrial DNA variation. *Molecular Ecology,* 16, 1191-1205.

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.

Podnar, M., Haring, E., Pinsker, W. & Mayer, W. (2007). Unusual origin of a nuclear pseudogene in the Italian wall lizard: intergenomic and interspecific transfer of a large

section of the mitochondrial genome in the genus *Podarcis* (Lacertidae). *Journal of Molecular Evolution, 64,* 308-320.

Popov, S., Shcherba, I., Ilyina, L., Nevesskaya, L., Paramonova, N., Khondkarian, S., & Magyar, I. (2006). Late Miocene to Pliocene palaeogeography of the Paratethys and its relation to the Mediterranean. Palaeogeography, Palaeoclimatology, Palaeoecology, 238: 91-106.

Posada, D. (2008). jModelTest: phylogenetic model averaging. *Molecular Biology and Evolution, 25,* 1253–1256.

Recuero, E., Canestrelli, D., Vörös, J., Szabó, K., Poyarkov, N. A., Arntzen, J. W., ...Martínez-Solano, I. (2011). Multilocus species tree analyses resolve the radiation of the widespread *Bufo bufo* species group (Anura, Bufonidae). *Molecular Phylogenetics and Evolution, 62,* 71-86.

Roitberg, E. S. (1999). Phenetic relationships between *Lacerta caucasica*, *L. daghestanica*, and *L. praticola* Reptilia, Lacertidae): a multivariate trend in external morphology. *Natura Croatica, 8,* 189 – 200.

Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D., Darling, A., Höhna, S., ...Huelsenbeck, J. P. (2011). MrBayes 3.2: Efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology 61,* 1-4.

Ryabinina, N. L., Bannikova, A. A., Kosushkin, S. A., Ciobanu, D., Milto, K. D., Tuniyev, B. S., ...Darevsky, I. S. (2011). Estimation of the subspecific level of differentiation in Caucasian lizards of the genus *Darevskia* (SYN. "*Lacerta saxicola* complex," Lacertidae, Sauria) using genome DNA markers. *Russian Journal of Herpetology, 9,* 185-194.

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 Evolutionary Biology, **13,*** 147.

Sambrook, J., & Green, M. R. (Ed) (2000). *Molecular Cloning - a Laboratory Manual*. Cold Spring Harbor Laboratory Press (Third Edition).

Scotland, R. W., Olmstead, R. G., & Bennett, J. R. (2003). Phylogeny reconstruction: the role of morphology. *Systematic Biology, 52,* 539–548.

Sobolevsky, N. I. (1930). New form from genus *Lacerta* (Reptilia) from Hungary. *Trudy Ass. Issl. Inst. Fiz.-mat. Fak. 1 MGU, 3,* 3–9.

Sos, T., Kecskes, A., Hegyeli, Z., & Marosi, B. (2012). New data on the distribution of *Darevskia pontica* (Lantz and Cyrén, 1919) (Reptilia: Lacertidae) in Romania: filling a significant gap. *Acta Herpetologica, 7,* 175-180.

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.

Stephens, M., Smith, N., & Donnelly, P. (2001). A new statistical method for haplotype reconstruction from population data. *American Journal of Human Genetics, 68,* 978-989.

Stöck, M., Moritz, C., Hicherson, M., Frynta, D., Dujsebayeva, T., Eremchenko, V., ...Wake, D. B. (2006). Evolution of mitochondrial relationships and biogeography of Palearctic green toads (*Bufo viridis* subgroup) with insights in their genomic plasticity. *Molecular Phylogenetics and Evolution, 41,* 663-689.

Stöck, M., Dufresnes, C., Litvinchuk, S. N., Lymberakis, P., Biollay, S., Berroneau, M., ...Perrin, N. (2012). Cryptic diversity among Western Palearctic tree frogs: Postglacial range expansion, range limits, and secondary contacts of three European tree frog lineages (*Hyla arborea* group). Molecular Phylogenetics and Evolution 65: 1-9.

Stugren, B. (1961). Systematik der Wieseneidchse *Lacerta praticola* Eversmann. *Zoolgische Beitrage, 6,* 379 – 390.

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.

Tan, D. S. H., Ang, Y., Lim, G. S., Ismail, M. R. B., & Meier, R. (2009). From “cryptic species” to integrative taxonomy: an iterative process involving DNA sequences, morphology and behavior leads to the resurrection of *Sepsis pyrrhosoma* (Sepsidae: Diptera). *Zoologica Scripta, 39,* 51-61.

Tarkhnishvilli, D. N. (2012). *Evolutionary history, habitats, diversification, and speciation in Caucasian rock lizards.* In: Jenkins O. P. (ed), Advances in Zoology Research Vol. 2, Nova

Science Publishers, Hauppeauge, 79-190.

Tarkhnishvilli, D. N., Murtskhvaladze, M., & Gavashelishvili, A. (2013). Speciation in Caucasian lizards: climatic dissimilarity of the habitats is more important than isolation time. *Biological Journal of the Linnean Society, 109,* 876-892.

Tuniyev, S. B., Doronin, I. A., Kidov, A. A. & Tuniyev, B. S. (2011). Systematic and Geographical variability of meadow lizard, *Darevskia praticola* (REPTILIA: SAURIA) in the Caucasus. *Russian Journal of Herpetology, 18,* 295 – 316.

Tuniyev, S. B., Doronin, I. V., Tuniyev, B. S., Aghasyan, A. L., Kidov, A. A., & Aghasyan, L. A. (2013). New subspecies of meadow lizard, *Darevskia praticola loriensis* ssp. nov. (Reptilia: Sauria) from Armenia. *Russian Journal of Herpetology, 20*, 223-237.

Ursenbacher, S., Schweiger, S., Tomovic, L., Crnobrnja-Isailovic, J., Fumagalli, L., & Mayer, W. (2008). Molecular phylogeography of the nose-horned viper (*Vipera ammodytes*, Linnaeus (1758)): Evidence for high genetic diversity and multiple refugia in the BalkanicPeninsula. *Molecular Phylogenetics and Evolution, 46,* 1116-1128.

Valakos, E. D., Pafilis, P., Sotiropoulos, K., Lymberakis, P., Maragou, P., & Foufopoulos, J. (2008). Chimaira (ed), *The Amphibians and Reptiles of Greece* (pp 480). Frankfut am Main .

Villesen, P. (2007). FaBox: an online toolbox for fasta sequences. *Molecular Ecology Notes, 7,* 965–968.

Wielstra, B., Themudo, G., Güçlü, O., Olgun, K., Poyarkov, N., Arntzen, J. W. (2010). Cryptic crested newt diversity at the Eurasian transition: The mitochondrial DNA phylogeography of Near Eastern *Triturus* newts. Molecular *Phylogenetics and Evolution, 56,* 888-896.

CHAPTER 2.2

Age and hybrid origin of parthenogenetic *Darevskia* - Paper II

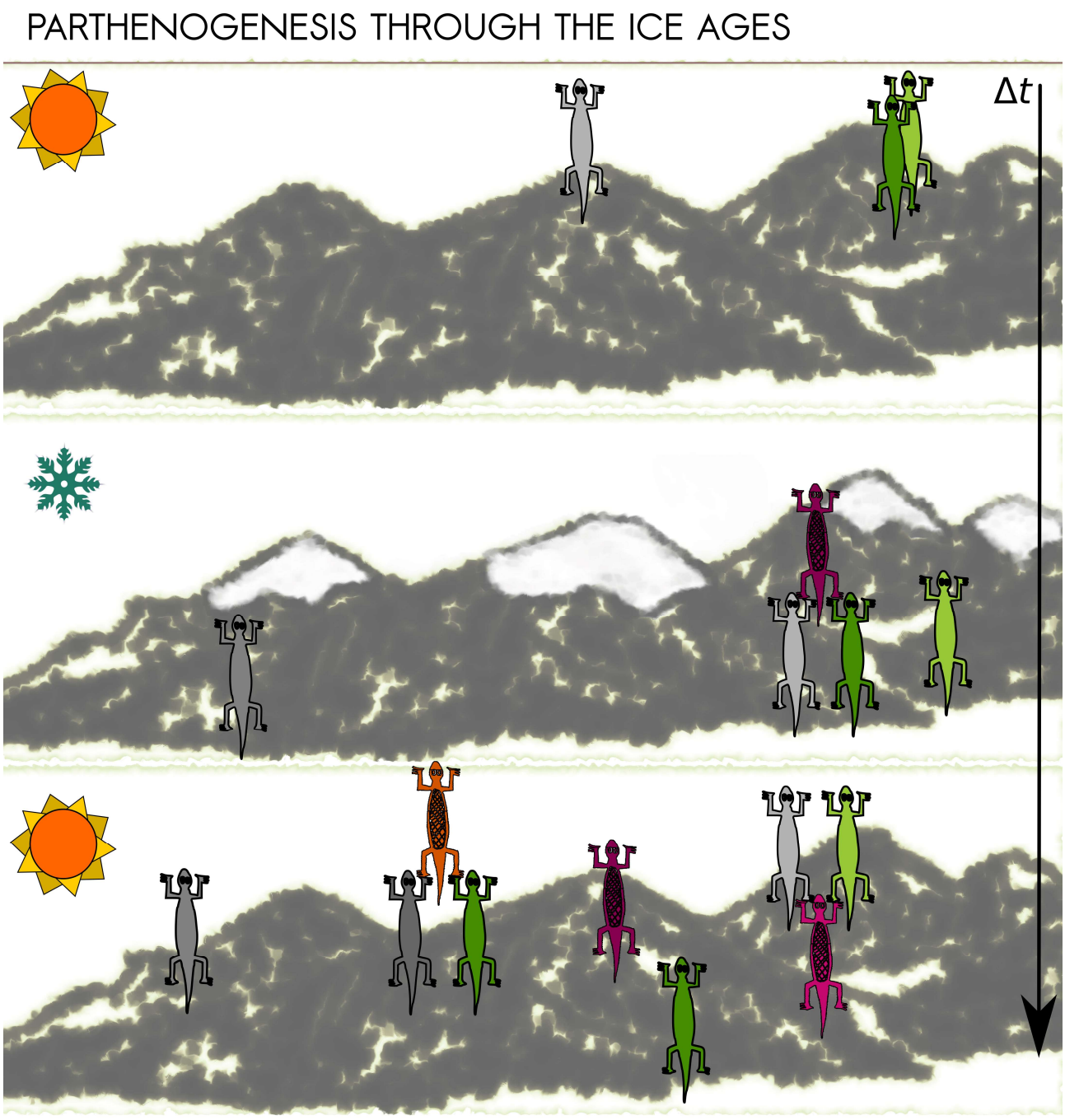


Figure 2.2.0: Graphical abstract

Parthenogenesis through the Ice Ages: a biogeographic analysis of Caucasian rock lizards (genus *Darevskia*)

Susana Freitas[1,2]

Sara Rocha[3]

João Campos[1]

Faraham Ahmadzadeh[4]

Claudia Corti[5]

Neftali Sillero[6]

Çetin Ilgaz[7]

Yusuf Kumlutaş[7]

Marine Arakelyan[8]

D. James Harris[1]

Miguel A. Carretero[1]

[1]CIBIO Research Centre in Biodiversity and Genetic Resources, InBIO, Universidade do Porto, Campus Agrário de Vairão, Rua Padre Armando Quintas, Nº 7. 4485-661 Vairão, Vila do Conde, Portugal

[2]Department of Animal and Plant Sciences, The University of Sheffield, Sheffield S10 2TN, UK

[3]Department of Biochemistry, Genetics and Inmunology, University of Vigo, 36310 Vigo, Spain

[4]Department of Biodiversity and Ecosystem Management, Environmental Sciences Research Institute, Shahid Beheshti University, G.C., Evin, Tehran, Iran

[5]Museo di Storia Naturale dell'Università di Firenze, Sezione di Zoologia "La Specola", Via Romana, 17, 50125 Firenze, Italy

[6]CICGE Centro de Investigação em Ciências Geo-Espaciais, Faculdade de Ciências da Universidade do Porto (FCUP), Observatório Astronómico Prof. Manuel de Barros, Alameda do Monte da Virgem, 4430-146, Vila Nova de Gaia, Portugal

[7]Dokuz Eylül University, Faculty of Science, Department of Biology, 35160 Buca, İzmir, Turkey

[8]Yerevan State University, Alek Manoogian, 1, Yerevan 0025, Armenia

ABSTRACT

Darevskia rock lizards include both sexual and parthenogenetic species, mostly distributed in the heterogeneous and ecologically diverse Caucasus. The parthenogenetic species originated via directional hybridogenesis, with only some of the sexual species known to serve as parentals. However, it remains unclear when and where these events happened and how many parental lineages were involved. A multilocus phylogeographic analysis was performed on the parthenogens *D. unisexualis*, *D. bendimahiensis* and *D. uzzeli*, and their putative maternal species *D. raddei*. Results show the parthenogenetic species all have relatively recent origins, approximately 200 - 70 kyr ago, and at least three hybridization events were involved in their formation. Ecological niche models identify the region where hybridization events leading to the formation of *D. unisexualis* took place, namely in the northeast of the current distribution. Models also suggest that the sexual *D. raddei* might have undergone a habitat shift between the Last Interglacial and the Last Glacial Maximum.

Keywords

Darevskia, parthenogenesis, mtDNA, phylogeny, ecological niche models, Glaciations.

INTRODUCTION

The study of parthenogenetic organisms, which reproduce in the absence of sex, provides an opportunity to understand the significance of sexual reproduction and the evolution of sex. In particular, taxa that present both sexual and parthenogenetic reproduction within the same clade, provide an opportunity to compare both reproductive forms and analyse their eventual ecological interactions (Gilabert et al., 2014; Otto and Nuismer, 2004). Reptiles are good model organisms for such studies due to the wide variety of reproductive modes and life history strategies, and lizards in particular are recurrent models used in studies of speciation, phylogeography and adaptation (Camargo et al., 2010). Several lizard families include parthenogenetic and sexual species, making them especially interesting for studying the evolution and function of sexual reproduction (Avise, 2008). Indeed, parthenogenesis was first described in vertebrates in the lizard genus *Darevskia* (Darevsky, 1967). Since then, at least 43 other cases of parthenogenetic

reproduction have been described in the Squamata (Kearney, 2003; Vrijenhoek, 1989). It is estimated that 0.6% of squamates (which comprise around 7000 species) can reproduce parthenogenetically, either obligatorily or facultatively (Kearney et al., 2009). Parthenogenesis is found across the squamate phylogeny and through a wide geographical range and ecological conditions. Most, but not all, parthenogenetic forms arose after hybridization between two related species, but the scenario for the origin of the parthenogenesis varies with the group and it is highly complex (Avise, 2008). Given this widespread distribution and the fact that parthenogenetic reproduction is frequently considered an "evolutionary dead-end" (Bell, 1982), it is still not fully understood whether new parthenogenetic lineages regularly appear and how they compete with sexual forms.

In this study, we focus on lizards of the genus *Darevskia* Arribas, 1997 of the family Lacertidae. This is a group of small lizards found across the Caucasus and adjacent regions, including Turkey, Iran and the Balkans (Arnold et al., 2007). Currently 32 species are recognised (Ahmadzadeh et al., 2013b; Uetz, 2015) which occupy a wide diversity of habitats, from forest and meadows to rocky habitat. Initial estimates of phylogenetic relationships based on partial Cytochrome-*b* (Cyt-*b*) mitochondrial DNA sequences and protein electrophoretic data suggest parthenogenetic lineages result from successful directional hybridization events between sexual *Darevskia* species. Only four parent species are thought to have been involved, *D. raddei* (Boettger, 1892) and *D. mixta* (Méhely, 1909) as the maternal donors and *D. valentini* (Boettger, 1892) and *D. portschinskii* (Kessler, 1878) as the paternal donors (Fu et al., 1997; Murphy et al., 2000). The sexual *Darevskia* species that most commonly contributes as a parental for the parthenogenetic lineages is *D. raddei*, being the proposed maternal species for at least five of them: *D. unisexualis* (Darevsky, 1966) (Armenia, northeastern Turkey and southern Georgia), *D. uzzelli* (Darevsky & Danielyan, 1977) (northeastern Turkey), *D. bendimahiensis* (Schmidtler, Eiselt & Darevsky, 1994) (northeast of Lake Van), *D. sapphirina* (Schmidtler, Eiselt & Darevsky, 1994) (north of Lake Van in the vicinity of Erciş) and *D. rostombekowi* (Darevsky, 1957) (northern Armenia and western Azerbaijan) (Fu et al., 1997; Baran et al., 2012). Nevertheless, *D. raddei* itself has been suggested to be a species-complex containing the forms *"raddei"*, *"nairensis"* and *"vanensis"* whose status and phylogenetic relationships are still a matter of debate (Grechko et al., 2007). As a consequence, it remains unclear if different *D. raddei* lineages may have been involved in

the hybridization events that led to the parthenogenetic lineages. The form "*raddei*" is distributed throughout the south and northeast of Armenia and Nagorno-Karabakh (Arakelyan et al., 2011), Azerbaijan and the northern part of the east Azerbaijan and Ardabil provinces of Iran (Anderson, 1999). The form "*vanensis*" is found in easternmost Anatolia, east of Lake Van and the west Azerbaijan Province of Iran (Baran et al., 2012). The differences between them are based on quantitative morphological traits that are not fully diagnostic (Anderson, 1999). The third form "*nairensis*" is restricted to the northeastern part of Armenia, along the western margin of the Sevan Lake. It is noteworthy that sympatry of *D. raddei* "*nairensis*" with one of the parthenogenetic forms (*D. unisexualis*) has been described for a single locality: Lchap (Gegharkunik province), on the west margin of the Sevan Lake in Armenia (Arakelyan et al., 2011; M. Arakelyan and F. Danielyan, unpubl. com.). Examining the diversity of Cyt-*b* sequences within the *D. raddei* complex (except "*vanensis*"), MacCulloch et al. (2000) concluded that the forms "*raddei*" and "*nairensis*" were conspecific due to the paraphyletic relationships found. A fourth form, *D. raddei "chaldoranensis*" has been recently described based on scalation and coloration characters, from a single locality of northern Zagros, western Azerbaijan Province of Iran (Rastegar-Pouyani et al., 2011, 2012), falling within the putative range of the form "*vanensis*".

The region where these forms occur, the Caucasus, includes a remarkable habitat and topographical heterogeneity likely to have promoted the formation of important biological barriers, and harbored multiple glacial refugia for sedentary species, including reptiles, during the last cold period (Ahmadzadeh et al., 2013a, 2013b; Tarkhnishvili et al., 2000, 2013). Nevertheless, evolutionary studies reveal heterogeneous biogeographic patterns for the biota in this region. While the Caucasus may have acted as a complex secondary contact zone for some species (Seddon et al., 2002), for others it appears to have acted as a barrier to expansion (Tarkhnishvili et al., 2000).

Here, we aim to infer the biogeographic patterns of parthenogenetic and bisexual rock lizards by addressing three questions: 1) Where and when did the parthenogenetic *Darevskia* species appear and could this be related to known biogeographic events? 2) How many parental lineages contributed for the parthenogenetic species under study? and 3) Have parthenogenetic species undergone identifiable periods of range expansion or contraction since their origin? We focus on the *Darevskia raddei* sensu lato sexual species

and the hybrid parthenogenetic daughter lineages, *D. unisexualis, D. uzzelli and D. bendimahiensis.*

To answer the first and second questions, a phylogenetic dating approach was employed. The molecular markers were used to determine the specific maternal lineage for each of the parthenogenetic forms analysed and, specifically, whether the parthenogenetic lineages come from single or multiple hybridization events. To try to infer the location of those events ecological niche modelling was performed based on the current environment and on projections to two different paleoscenarios, the Last Interglacial (LIG – 130 to 115 kyr ago) and the Last Glacial Maximum (LGM – 22 kyr ago), taking into account the age estimates for each species. If the origin of the parthenogenetic species occurred after the LIG, then comparisons of the potential distributions during the paleoscenarios analysed (LIG and LGM) with the present distribution model would allow inference regarding where these lineages could have been during the hybridization events. Regarding the last question, tests on population expansion/contraction were performed.

Furthermore, the current distribution ranges of the sexual species and of the parthenogen *D. unisexualis* were compared to the present habitat suitability model and to the projections for the estimated paleoscenarios as inferred by ecological niche modelling. With this we intend to infer how competition may influence the distribution of both parthenogenetic species and the sexual parentals. Due to their extremely restricted distribution, insufficient to infer ecological models, the other two parthenogenetic species, *D. uzzelli* and *D. bendimahinesis*, could not be included in this analysis.

MATERIAL AND METHODS

Study area and datasets

A total of 235 samples collected across the whole species ranges were used for the molecular analyses (Supplementary Table 1). *D. raddei* sensu lato individuals were selected from 90 localities covering the whole distribution range of the complex, *D. unisexualis* from 15 localities (N = 32), and *D. uzzelli* (N = 5) and *D. bendimahiensis* (N = 3) from one locality each, due to their locally restricted distribution (Figure 2.2.1, Supplementary Table 1). Presence records for 165 individuals (see Supplementary Table 1) were used to construct the ecological niche models (ENMs). In all cases, only records confirmed by molecular data were used. Geographic coordinates of sampling localities

were geo-referenced with a Global Positioning System (GPS) receptor on the WGS84 datum. The study area is a polygon which includes the global distribution of both species (*D. raddei* and *D. unisexualis*) as provided by IUCN, defined by the coordinates xMin,yMin 37.8275,34.8814:xMax,yMax 53.126,45.1208. This area was chosen in order to detect suitable habitats outside the distribution ranges of both species and to analyse the overlap between both ENMs, but taking into account their limited dispersal rate. Outgroup species used were sampled (*D. portchinskii, D. rudis* (Bedriaga, 1886) and *D. valentini*) or their sequences downloaded from Genbank (*Iranolacerta*). From all individuals sampled in the field, tail tips, photographs and basic measurements were also collected to associate to morphological descriptions of the species (Arakelyan et al., 2011).

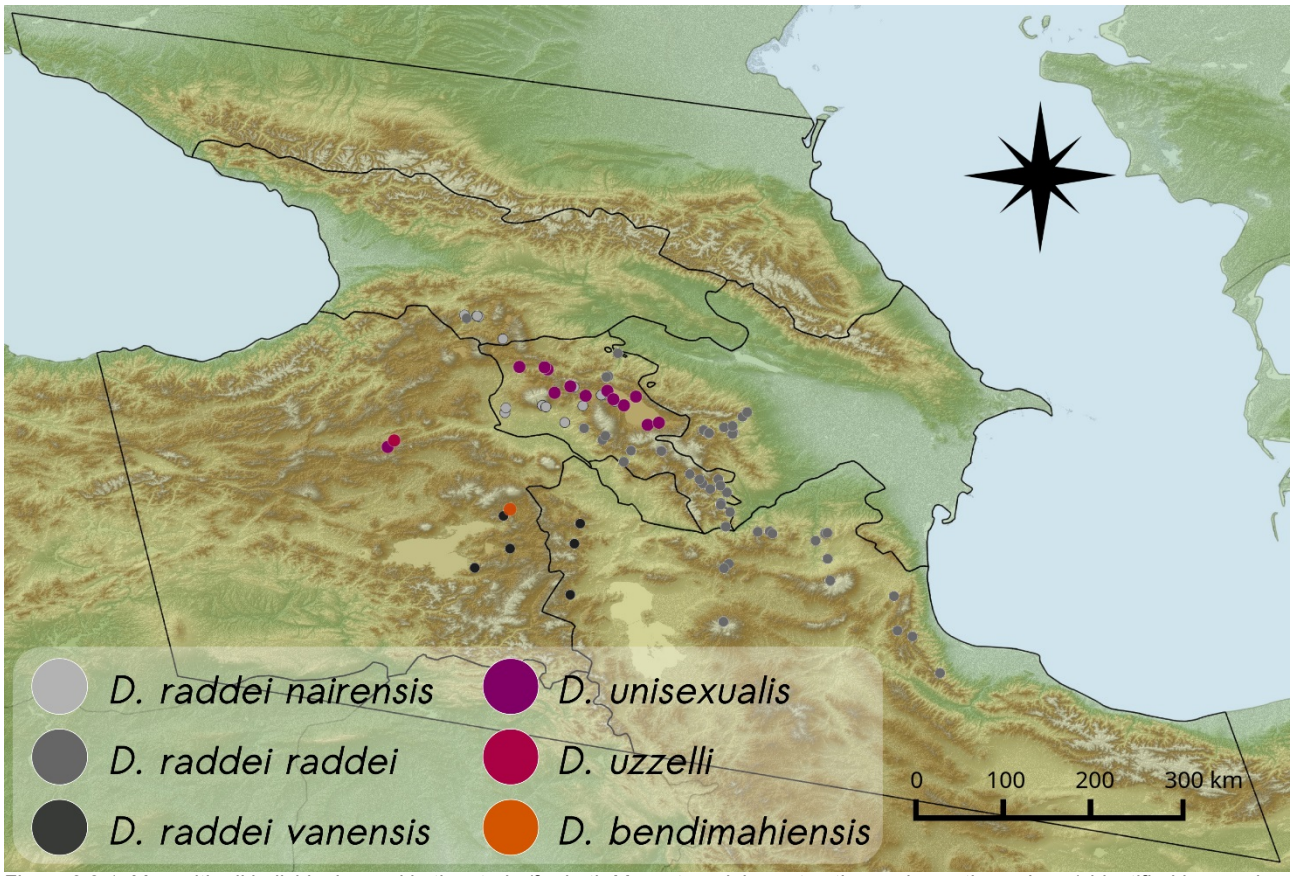


Figure 2.2.1: Map with all individuals used in the study (for both Maxent model construction and genetic analyses) identified by species-specific colour codes. Ecotypes of sexual species D. raddei are in different tones of grey (light grey, "nairensis"; medium grey, "raddei"; dark grey, "vanensis"). Parthenogenetic species are represented in purple (*D. unisexualis*), dark pink (*D. uzzelli*) and orange (*D. bendimahiensis*).

Molecular data

Total genomic DNA was extracted from approximately 30 mg of each tail-tip following

standard high-salt protocols (Sambrook and Russell, 2001). For phylogenetic analyses two partial mitochondrial genes; Cytochrome-*b* (Cyt-*b*) and NADH dehydrogenase-4 (ND4), and two partial nuclear genes; Melanocortin 1 receptor (MC1R) and oocyte maturation factor Mos (C-mos) were selected. Primers and PCR protocols are described elsewhere (Arévalo et al., 1994; Barata et al., 2012; Kocher et al., 1989; Pinho et al., 2007). Sequencing was conducted by a commercial facility (Macrogen Inc). Chromatograms were edited by eye in ChromasPro v1.7.4 (Technelysium), using ambiguity codes to represent heterozygous positions.

Phylogenetic analyses and divergence-time estimates

Sequence alignment was performed in MAFFT v6 (Katoh and Standley, 2013) using the automatic settings for the algorithm choice. For nuclear fragments, haplotypes phase was inferred with Phase version 2.1 (Stephens et al., 2001), and to reduce potential biases in downstream analyses only haplotype pairs with total posterior probabilities values above 0.6 were included in the analysis (Garrick et al., 2010) – this resulted in the exclusion of less than 1% of the sequences. Input files were prepared with SeqPHASE (Flot, 2010), which was also used to produce bi-allelic fasta files from PHASE outputs.

For the phylogenetic analyses, mtDNA fragments were concatenated but nuclear genes were analysed independently. Departing from an a-priori partitioning per coding position on each gene, PartitionFinder (Lanfear et al., 2012) was used to select the best-fit partitioning scheme and DNA substitution model(s). Phylogenetic analyses were performed using Bayesian (MrBayes v. 3.2, Huelsenbeck and Ronquist, 2001; Ronquist and Huelsenbeck, 2003) and Maximum Likelihood (PhyML 3.0, Guindon and Gascuel, 2003) inferences (BI and ML, respectively). In ML, nodal support was estimated through 1000 bootstrap replicates (Felsenstein, 1985). In BI, all analyses started with randomly generated trees and ran for $30x10^6$, with sampling at intervals of 1000 generations, producing 30,000 trees. Two independent runs were performed on each dataset. Burn-in was determined upon stabilisation of log likelihood using TRACER v1.5 (Drummond and Rambaut, 2007) and of the clades posterior probabilities with AWTY (Nylander et al., 2008). Individual mtDNA gene trees were also estimated with MrBayes, using the same strategy as with the concatenated mtDNA dataset. These gene trees were then compared to test for possible

incongruences (data not shown).

The age of the most recent common ancestor (tMRCA) was estimated for all lineages of the mitochondrial DNA dataset on a "species-tree" analysis using *BEAST 2.3.1 (Bouckaert et al., 2014) with the mtDNA dataset. This approach was preferred since "species-tree" analysis can provide accurate gene-tree estimates (Drummond et al, 2012) and more realistic assessments of posterior clade supports (Drummond and Bouckaert 2015). Even though only mtDNA markers were used the term "species-tree" analysis is used to identify the method in question. Both markers (ND4 and Cyt-*b*) were run with unlinked trees, sites and clock models so that each marker and respective priors used would not constrain the calculation of the parameters for the other marker, such as mutation rate, tree topology or branch length. DNA substitution models for both markers were searched again with PartitionFinder (Lanfear et al., 2012), but this time unpartitioned schemes per marker were selected since mutation rates used (Pinho et al., 2007) were developed for a non-partitioned marker (ND4). The models selected for each marker are GTR+I+G (Cyt-*b*) and HKY+G (ND4). Individuals were assigned to "species" based on their mtDNA lineages. Four independent searches were run for 10^7 generations. A lognormal relaxed molecular clock was assumed, using the mutation rate for ND4 estimated for the lacertid lizard genus *Podarcis* Wagler, 1830 (Pinho et al., 2007) and co-estimated for Cyt-*b*. Nuclear markers were not included in the tMRCA estimations given the hybrid origin of the parthenogens and respective uncertainty associated with phased haplotypes. The clock rate prior for the ucld.mean parameter for the ND4 dataset was set as a normal distribution with a mean of 0.0226 and a standard deviation of 0.0031, so that mutation rate varied between 0.0278 and 0.0174 mutation/site/million years. A uniform Yule prior was selected for the tree, with a random starting one. For the remaining parameters the default options were chosen. Convergence for all model parameters was determined in Tracer v.1.5. (Drummond and Rambaut, 2007) where high effective sample sizes (ESS) were observed for all parameters (> 200 for the combined analyses). LogCombiner 2.3.0 (Bouckaert et al., 2014) was used to combine the log and tree files of the four runs, with 20% of the trees of each one discarded as burn-in, following an analysis of convergence of individual run parameters in Tracer v1.4 (Drummond and Rambaut, 2007). A maximum clade credibility (MCC) tree with mean tree heights and 95% highest probability densities (HPDs) was produced using Tree Annotator (Bouckaert et al., 2014).

Population structure

Haplotype networks were constructed for both nuclear loci (MC1R and C-*mos*) using the statistical parsimony algorithm in TCS 1.21 (Clement et al., 2000). Analyses were performed with phased nuclear data of the species in study and additional sexual species expected to have acted as, or be closely related to, the paternal species. These were included to compare alleles of the parthenogens to the sexual species from which they potentially originated.

Diversity parameters for each gene were estimated only for the sexual species *D. raddei* and the parthenogen *D. unisexualis,* since sample sizes for the remaining (parthenogenetic) species were insufficient. Estimates of haplotype diversity (Hd), nucleotide diversity (π), neutrality tests Tajima's *D* and Fu's *Fs*, as well as Harpending's raggedness index (r) were calculated, as well as the significance of Tajima's *D* and Fu's *Fs* statistics, tested by generating 1000 random samples under the null hypothesis of selective neutrality and population equilibrium, using a coalescent simulation algorithm adapted from Hudson (1991) in DNAsp. Significance of r was tested using a parametric bootstrap approach (Schneider and Excoffier, 1999).

Environmental data

Climatic variables were retrieved from the WorldClim online data (Hijmans et al., 2005). The spatial resolution for current climate variables was 30 arc-seconds (approximately 1 km2) and for past climate variables 2.5 arc-minutes (approximately 5km^2). From the 19 Bioclim variables, those with a correlation lower than 0.7 and considered biologically relevant for both species were selected (supplementary table 3).

Three past climate scenarios were used: one scenario for the Last Interglacial (LIG: ~120-140 kyr years BP; Otto-Bliesner et al. 2008); and two scenarios (CCSM - the Community Climate System Model, and MIROC - the Model for Interdisciplinary Research on Climate) for the Last Glacial Maximum (LGM: ~22 kyr years BP)(Hijmans et al., 2005).

Ecological niche models

The realised niches (*sensu* Hutchinson, 1957) of *D. raddei* sensu lato and *D. unisexualis* were estimated using the Maximum Entropy method implemented in Maxent 3.3.2

(Phillips et al., 2004, 2006). All forms within *D. raddei* were considered as a single group given the low phylogenetic distance between them ("*vanensis*" vs "*raddei*") or the lack of differentiation ("*nairensis*" and "*raddei*"; see Results). The ecological niche models for the present were then projected to the three past climate scenarios selected.

Maxent runs were performed with autofeatures, selecting randomly 70% (number of points) of the presence records as training data and 30% (number of points) as test data for *D. raddei* locations, and all the presence records (number of points) as training data for *D. unisexualis*, due to the limited number of records for this species. Default parameters were used in order to compare the different models.

Models were evaluated with receiver operated characteristics (ROC) plots. The area under the curve (AUC) of the ROC plot was taken as a measure of the overall fit of the Maxent model (Liu et al., 2005) (random models have an AUC equal to 0.5) AUC was selected because it is independent of prevalence (the proportion of presence in relation to the total dataset size; see VanDerWal et al. (2009)). The importance of each climate variable for explaining the species distribution was determined by: (1) jackknife analysis of the average AUC with training and test data; and (2) average percentage contribution of each environmental factor to the models. The mean realised niche model and its projections to past scenarios were reclassified in presence-absence maps using the average value of the 10 percentile training presence logistic as the threshold. This would decrease the potential error associated to the dataset. So that we defined suitable habitat to include 90% of the data used to develop the model. Cells with values higher and lower than the threshold were considered either suitable or unsuitable for the presence of the species (in the latter case species were considered to be absent from these cells). Identification of areas of probable sympatry between species was determined by overlap analysis, multiplying the distribution model of each species in the "Raster Calculator" function of QGIS.

RESULTS

Phylogenetic analyses

Two mitochondrial DNA markers were analysed in this study comprising 270 concatenated sequences and 110 unique haplotypes within the concatenated dataset. In total, mtDNA markers correspond to 1753 bp (Cyt-*b*: 919bp, 143 parsimony informative sites; ND4: 834bp, 98 parsimony informative sites).

The mtDNA gene markers do not show any indels or stop codons when translated. Under the corrected Akaike information criteria (AICc), the best partition set and models chosen by PartitionFinder and applied to the dataset are as follows: Cyt-*b*/position 1 = TVMef+I+G; ND4/position 3, Cyt-*b*/position 2 = K81uf+I+G; ND4/position 1, Cyt-*b*/position 3 = TIM+G; ND4/RNA subset, ND4/position 2 = GTR+G. The individual gene trees recovered from both mtDNA gene regions are topologically concordant with no well supported conflict and both Bayesian and Maximum Likelihood analyses result in the same overall tree topology (data not shown). A previous study found evidence for a nuclear copy of the Cyt-*b* in another species of *Darevskia* (Freitas et al, 2016). However, given that both mtDNA markers produced concordant individual tree topologies for the major lineages, and the lack of stop codons and indels in these sequences, we have no reason to consider this issue further here. The mtDNA gene genealogy shows *D. raddei* sensu lato as monophyletic with maximum support (Figure 2.2.2), with the three parthenogens analysed being placed within the *D. raddei* lineage. According to our results, the form *"nairensis"* (Figure 2.2.2, light grey) does not correspond to a monophyletic lineage, and haplotypes from individuals morphologically assigned to this form are shared with individuals recognized as *"raddei"* (haplotype number 5 is shared by *D. raddei raddei* from Gosh and Pzorak and *D. raddei nairensis* from Hovk, all in Armenia). In contrast, the form *"vanensis"* (Figure 2.2.2, dark grey) does corresponded to a single lineage, which appears distinct from the rest of the *D. raddei* individuals analysed ("*raddei" and "nairensis"*) although still nested within *D. raddeii* sensu lato. Its haplotypes are shared only with the parthenogen *D. bendimahiensis.*

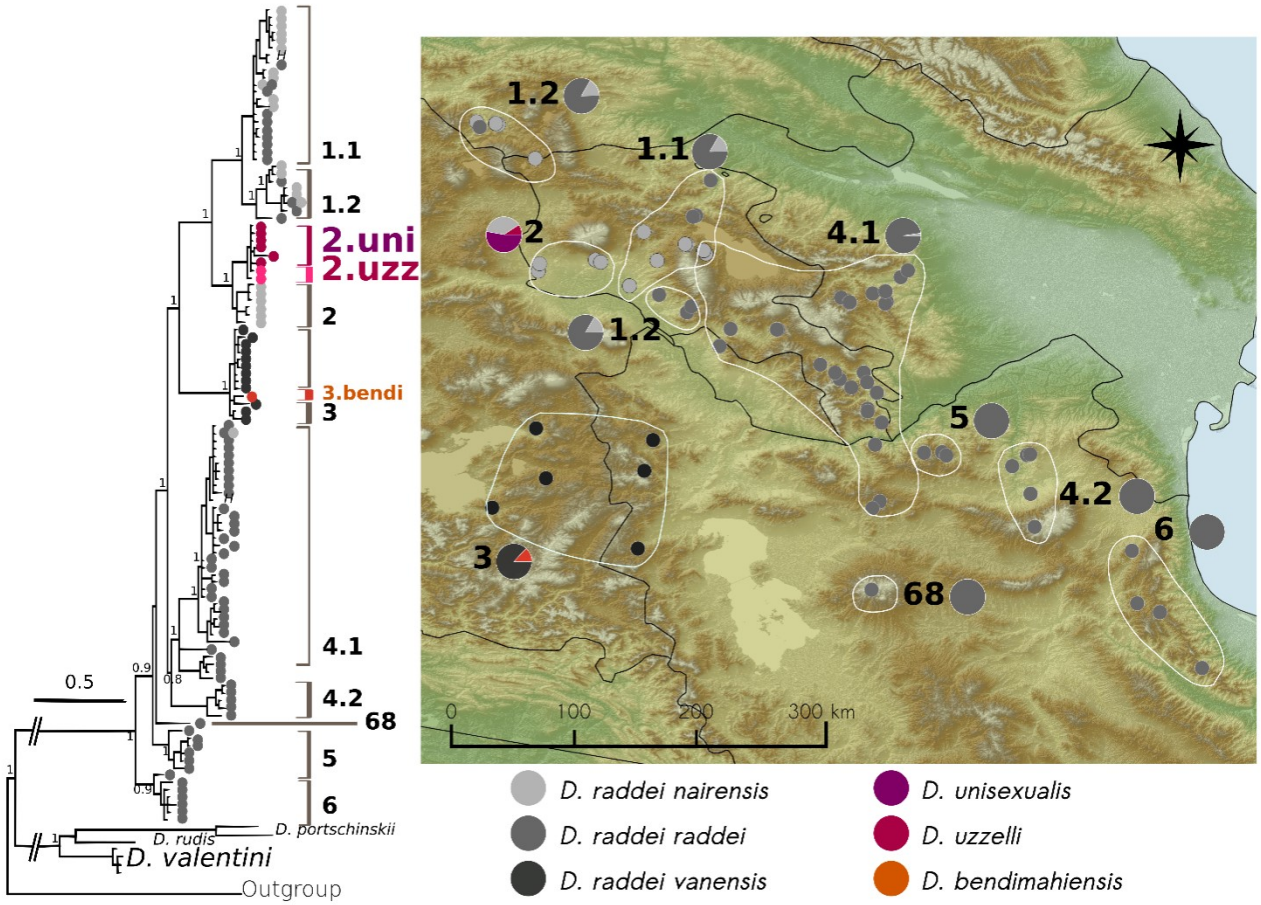


Figure 2.2.2 – 50% majority rule consensus of bayesian estimates of mtDNA (Cyt-*b* and ND4) trees for the *D. raddei* "complex", parthenogenetic species and outgroups. Lineages correspond to the lineages on the map (on the right) and tMRCA estimates (Table 3). Pie charts represent the different taxa clustered within each mtDNA lineage (dark grey: *D. raddei "vanensis"*, medium grey: *D. raddei "raddei"*, and light grey: *D. raddei "nairensis"*; orange: *D. bendimahiensis*, purple: *D. unisexualis*, dark pink: *D. uzzelli*). Only posterior probability values above 0.8 are presented.

Both Bayesian and ML mtDNA phylogenetic analyses show a Southeast-Northwest differentiation within the *D. raddei* group (Figure 2.2.2): the basal lineages 68 (a single haplotype), 5 and 6 contain the individuals from the southernmost part of the distribution, in the region of Ardabil and east Azerbaijan provinces of Iran (Figure 2.2.2). Lineages 4.1 and 4.2 include individuals from South Armenia and Ardabil region in Iran and are found South and North of the lineages 5 and 6. Lineage 3 contains the individuals located around Lake Van in Turkey and lineages 1.1 and 1.2, and 2 are found Northeast of the Geghama Mountains. Even though there is a clear Southeast-Northwest differentiation, most mtDNA lineages are in contact and geographical structure is only detected in some cases, such as lineage 3 whose samples are geographically isolated and genetically differentiated from the rest.

The “species-tree” inference (Figure 2.2.3) largely matches the Bayesian concatenated “gene-trees” except for the position of lineage 3. However, given the low posterior probability values, the position of this lineage relative to the other *D. raddei* lineages remains unclear. The estimated 95% HPD intervals of node heights show the differentiation in the *D. raddei* group dates back to around 0.7 Myr (1.1 – 0.5) and its split from *D. clarkorum* (the closest sexual species included in this study) dates from 3 – 1.2 Myr. The closest *D. raddei* lineage to the parthenogens *D. unisexualis* and *D. uzzelli* - lineage 2 - splits from these parthenogens 290 – 7 kyr ago while the differentiation within these two parthenogenetic species dated 61 – 2 kyrs ago. The origin of *D. bendimahiensis*, or the split between this parthenogenetic species and its closest *D. raddei* lineage, dates to 204 – 18 kyr.

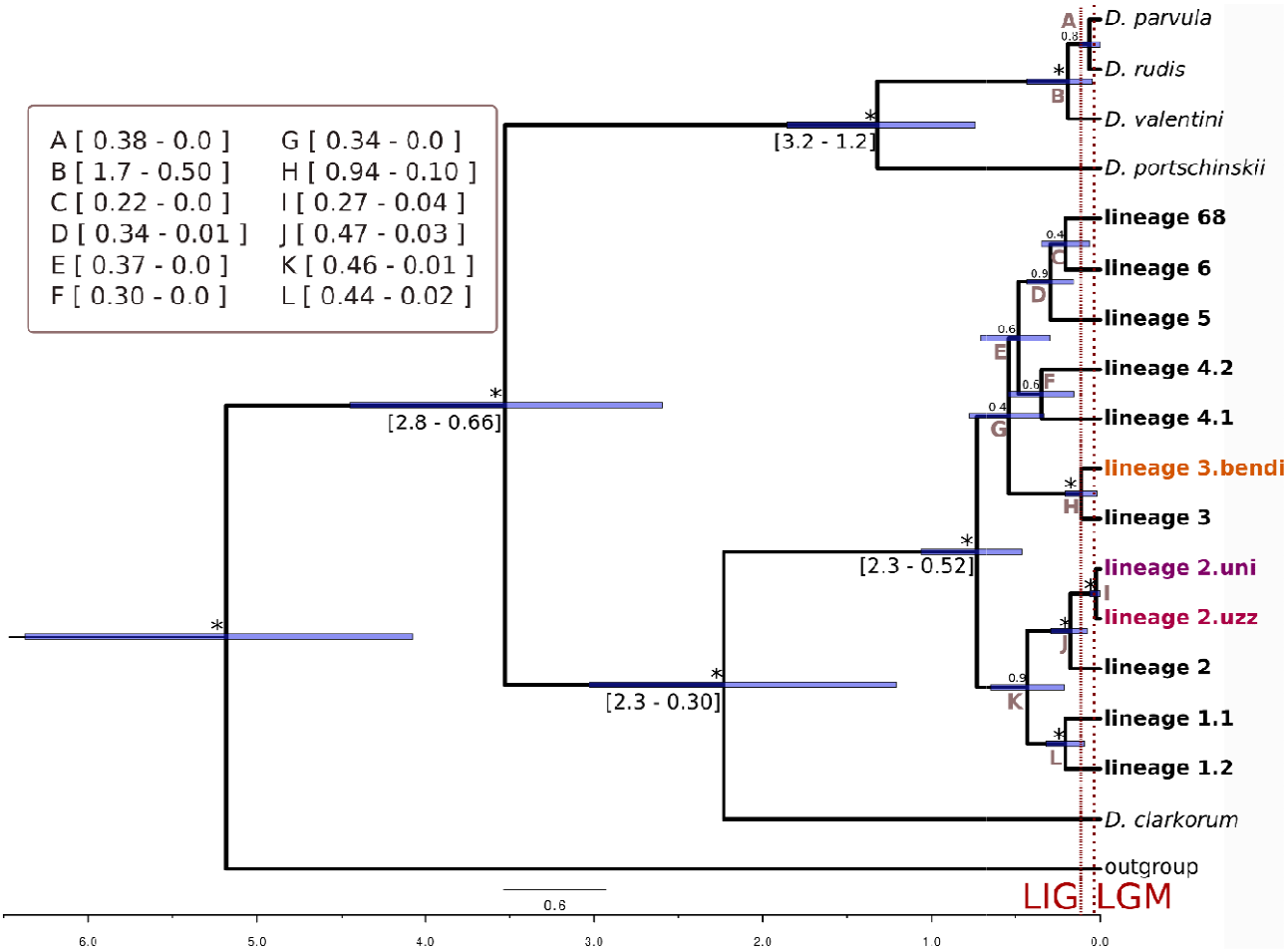


Figure 2.2.3. Species-tree estimate (MCC) of *D. raddei* sensu lato and the parthenogenetic species *D. unisexualis* (lineage 2.uni), D. uzzelli (lineage 2.uzz) and *D. bendimahiensis* (lineage 3.bendi). Divergence time intervals in Myrs. Posterior probabilities are presented for each split, stars represents posterior probability of 1. Parthenogenetic species are shown in different colours, similar as in other figures (orange: *D. bendimahiensis*, purple: *D. unisexualis*, dark pink: *D. uzzelli*).

Population structure

Analyses of nuclear DNA show all parthenogenetic individuals are heterozygous for both markers, with each allele shared with a different species. Some sexual individuals are also found to be heterozygous but their alleles are never shared between species as it is always the case with the parthenogens. To assess the maternal and paternal genomic contribution in the parthenogenetic individuals, samples of *D. portschinskii, D. valentini* and *D. rudis* individuals were incorporated to the analyses and haplotype networks were constructed (Figure 2.2.4). Each of the alleles found were expected to group with each parental group. Thus, the allele corresponding to the maternal contribution is considered to be the allele shared with the *D. raddei* individuals, the other allele corresponding to the paternal contribution (shared or closer to *D. portschinskii, D. rudis* and *D. valentini*). The expected paternal species for the three parthenogenetic species was *D. valentini* (Darevsky, 1967), although it was still pending confirmation by genetic data. In order to include other species closely related to *D. valentini* (Fu et al., 1997) and the diversity within the putative paternal group, we have also added samples of *D. portschinskii* and *D. rudis.*

Within *D. raddei* sensu lato*,* networks of nuclear haplotypes show a weak geographic structure, even though MC1R (644 bp, 43 variable positions), faster evolving than C-*mos* (550 bp, 14 variable positions) (Figure 2.2.4), shows a higher degree of diversity. However, some of the mitochondrial lineages have corresponding haplotype groups in the MC1R network: haplotypes 16, 17, 32 and 33 (of MC1R) is a group formed by samples found in the southernmost part of the species distribution, in the west Alborz region (Asalem, Hir, Khalkhal, Meskin Shahr) which corresponds to mtDNA lineage six. Haplotypes 9, 6, 20, 36, 39 correspond to individuals from South Armenia and NKR (mtDNA lineage 4.1 in Figure 2.2.2). Haplotypes 26, 27 and 29 are only found in individuals from the Lake Van area in Turkey, and western Azerbaijan (province of Iran) and Gollodja in the adjacent Iranian Azerbaijan (near the borderwith Turkey and the Lake Van region), which define mtDNA lineage three. The remaining haplotypes correspond to individuals ascribed either to *D. raddei "nairensis"* or to *D. raddei "raddei"*, distributed North and South of Mount Aragats in central Armenia, respectively.

Regarding the maternal contribution (*D. raddei* sensu lato), all of the parthenogenetic species present two haplotypes only, 1 and 27. Specimens identified as *D. unisexualis* and *D. uzzelli* share haplotype 1 with individuals from Mount Aragats, specifically Amberd Castle and Lchaschen. The only homozygous individuals for this haplotype are *D. raddei*

nairensis found in Amberd Castle, where the frequency of this haplotype is likely higher. Haplotype 27 is shared by *D. bendimahiensis* and individuals identified as *D. raddei "vanensis"*, located around Lake Van in Turkey and Iranian Azerbaijan. Haplotypes 5, 38, 25, 22, 23 and 24 corresponded to the putative paternal species. Alleles 4, 21 and 25 are found in the parthenogenetic species, and were therefore inherited from the paternal species that contributed to the original hybridization event. Regarding the paternal contribution, *D. unisexualis* presents two different haplotypes, and *D. bendimahiensis* and *D. uzzelli* only one each. While *D. unisexualis* and *D. uzzelli* share the same maternal haplotype, for the paternal contribution *D. unisexualis* shares its most common haplotype with *D. bendimahiensis*, while *D. uzzelli* shares its haplotype with individuals of *D. valentini*.

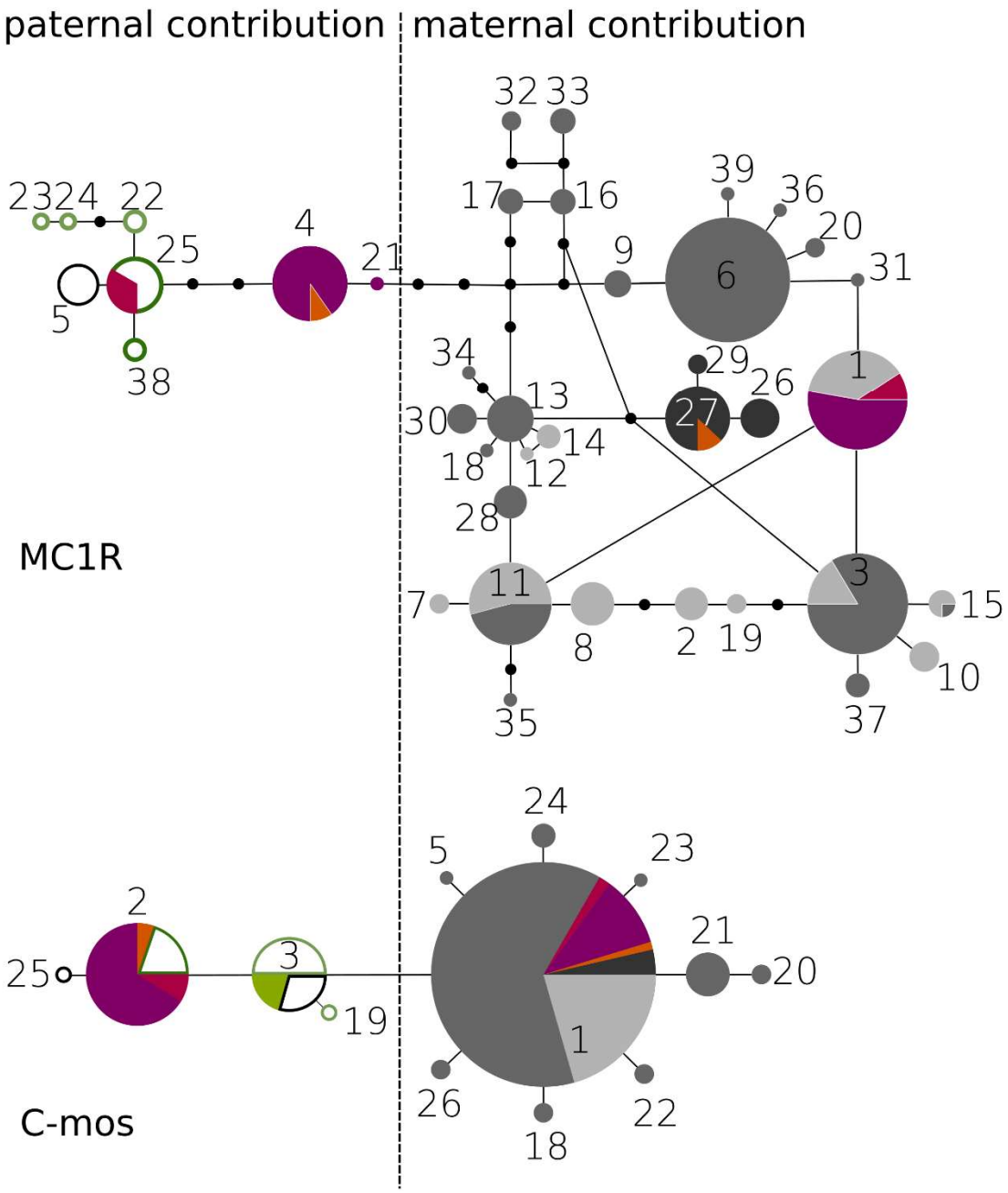


Figure 2.2.4 – Statistical parsimony networks for MC1R and C-mos in D. raddei group, parthenogenic descendant species and some individuals of the putative paternal species of those parthenogens. Small black circles represent missing or unsampled haplotypes. Grey colours correspond to the *D. raddei* sensu lato (dark grey: *D. raddei "vanensis"*, medium grey: *D. raddei "raddei"*, and light grey: *D. raddei "nairensis"*; orange: *D. bendimahiensis*, purple: *D. unisexualis*, dark pink: *D. uzzelli*; white with dark green outline: *D. valentini;* white with light green outline*: D. rudis; green: D. parvula; white with black outline: D. portschinskii*). Different parental contributions were identified with the position on the network of the parental species: *D. raddei* sensu lato as the maternal genomic contribution and *D. valentini*/*D. portschinskii* group as the paternal contribution. Circles correspond to haplotypes, numbered as in Supplementary Table 1, with size proportional to their frequency.

The C-*mos* haplotype network (Figure 2.2.4, bottom) shows little variation and most *D. raddei* sensu lato individuals share haplotype 1 or one derived from it by one or two

mutation steps. There is no geographic structure reflected in this network. As with MC1R, C-*mos* sequences of the putative paternal species (*D. valentini, D. rudis* and *D. portschinskii*) were used to allocate the paternal contribution and to differentiate the maternal from the paternal alleles. Regarding the maternal contribution, all three parthenogenetic lineages analysed share the same allele. This is the most common allele, found also in all the individuals ascribed to *D. raddei "nairensis"* (and all homozygous), but also in those identified as *D. raddei "vanensis"* and most of the *D. raddei "raddei"*. Regarding the paternal contribution, the three parthenogens share the same haplotype, which is not found in any of the putative paternal species used. This network shows a slight star-like shape. Neutrality tests were calculated for each species (Table 1). Tajima's *D* shows significant negative values for *D. raddei* (ND4) and *D. unisexualis* (Cyt-*b* and ND4). Fu's *Fs* also shows negative values for all markers analysed in both species, even though none is significant. Both tests R_2 and raggedness r detected significant low positive values for all markers.

Species	Marker	n	sites	π	Tajima's D	F_S	H_d	θ W	Raggedness r	R2
D. raddei	cytb	169	921	29.27	-0.12119	-1.1866	0.96758	29.867	0.00595*	0.08399*
	ND4	169	839	20.47941	-0.10468*	-1.05767	0.95241	20.304	0.00984*	0.8582*
	MC1R	328	695	2.79	-0.049	-4.575	0.819	2.86997	0.054*	0.078*
	Cmos	248	552	0.00045	-0.04419	-0.21047	0.20499	0.248	0.49367*	0.09042*
D. unisexualis	cytb	30	921	0.14073	-0.013*	0.10483	0.11468	0.13151	0.41563*	0.15*
	ND4	30	841	0.71221	-0.02719*	0.057	0.427	0.74666	0.298*	0.139*
	MC1R	56	695	3.765	-0.06668	-0.13597	0.7857	3.77	0.081*	0.105*
	Cmos	54	552	0.983	0.024	0.01565	0.496	1.01	0.25*	0.11*

m, number of sequences, sites, number of sites analysed per sequence, π, nucleotide diversity, F_S, Fu's (1997) F_S, H_d, Haplotype diversity, R_2, Ramos-Onsins and Rozas' (2002) R_2. *significant at P <0.05

Table 2.2.1 - Summary statistics, tests of neutrality and growth for the sexual species *D. raddei* sensu lato and the parthenogen *D. unisexualis.*

Ecological niche models

Maxent models were generated only for species with a sufficient number of geographic records, *D. raddei* sensu lato and *D. unisexualis*. Given their restricted distribution range, *D. uzzelli* and *D. bendimahiensis* could not be included in this analysis.

Both Maxent ensemble models have mean AUC values higher than 0.9, for training data (*D. raddei*: 0.9131; *D. unisexualis*: 0.9792) and close to 0.9 for test data (*D. raddei*: 0.8714; *D. unisexualis*: 0.9734): thus, training AUC and test AUC are within the same

value range meaning the model is dependent on the record data but not on which subset of the record data is used. The variables that more strongly contribute to the model of *D. raddei* are BIO18, BIO17, BIO4 and BIO112 (Precipitation of the Warmest Quarter, Precipitation of the Driest Quarter, Temperature Seasonality (standard deviation * 100) and Mean Temperature of the Coldest Quarter, respectively) and mostly revolve around the availability of water in the warmest (and driest) months of the year. The model of *D. unisexualis* is more strongly affected by BIO9 and BIO1 (Annual Mean Temperature and Mean Temperature of the Driest Quarter) and similarly to *D. raddei*'s model, BIO17 and BIO18 equally affected by similar variables except for the BIO2 and BIO4 (Supplementary Figure 2.2.1). These patterns are concordant with the jack-knife analysis of AUC and gain values of training and test data, for models calculated with only one variable and models calculated without that variable (Supplementary Figure 2.2.1).

The present area suitable for *D. raddei* (Figure 2.2.6) mostly overlaps with its current distribution range, although some suitable unoccupied areas (Arakelyan *et al.,* 2011) are identified as suitable habitat, especially towards the west (Turkey) and northeast (Georgia-Azerbaijan). *D. raddei vanensis* individuals (located next to Lake Van in Turkey) fall outside the suitable habitat for *D. raddei*. When projected to the LGM most suitable habitat for *D. raddei* is shifted to the east of its current distribution, does not include mountain tops, and tends to be restricted to valleys and plains. Interestingly, no suitable habitat was found when projecting the distribution of *D. raddei* to the LIG.

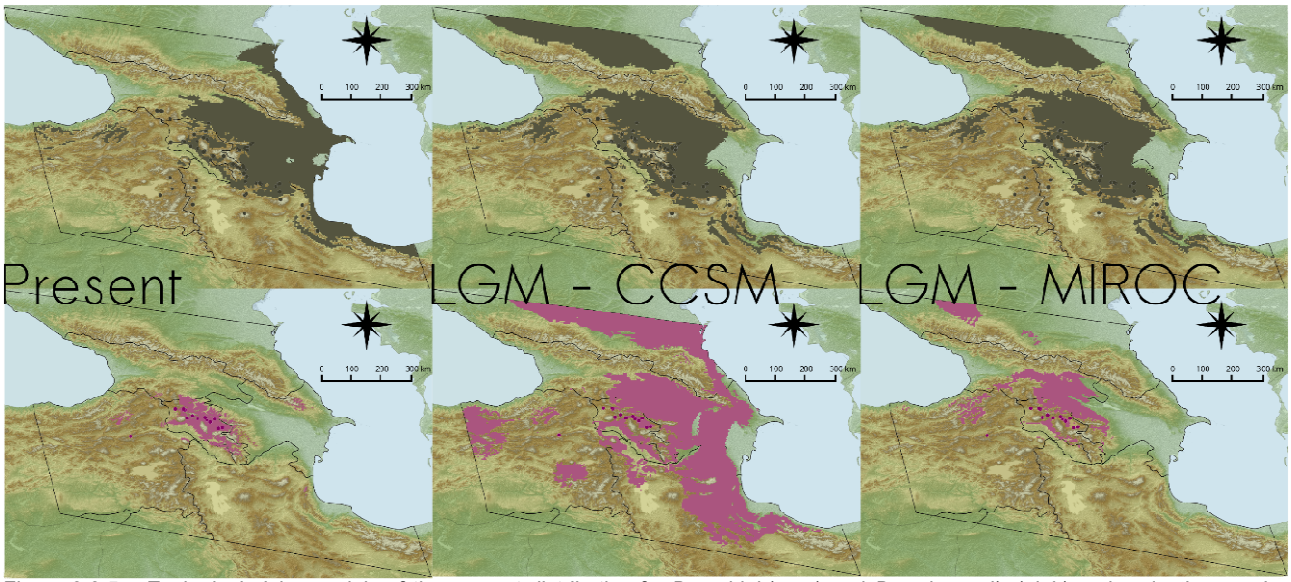


Figure 2.2.5 – Ecological niche models of the present distribution for D. raddei (grey) and *D. unisexualis* (pink) and projections to the past (Last Glacial Maximum, LGM). Projections to the past were performed using two scenarios, MIROC and CCSM. Details of the individuals used are in Supplementary Table 1. Points of the individuals used are in the maps in grey (*D. raddei*) and pink (*D. unisexualis*).

The *D. unisexualis* model (Figure 2.2.6) occupies a much more reduced area, concentrating within Armenia. *D. unisexualis* individuals found in Turkey (Horasan) do not fall inside the area predicted by the model. When projected to the LGM, both scenarios (CCSM and MIROC) produced slightly different results in terms of predicted suitable area, even though they both tend to find more suitable habitat to the east, as with the projection for *D. raddei*. No suitable habitat was found when projecting to the LIG. In all cases, when comparing the models of distribution for both species, the potential habitat of *D. unisexualis* falls within that of *D. raddei*.

DISCUSSION

Parthenogenesis is a rare reproductive mode that, despite being found in most animal groups, is observed in less than 0.5% of known species (Vrijenhoek, 1989). Given the low number of species where it is observed, the switch from sexual to parthenogenetic reproduction is expected to happen only rarely. Also, the twiggy distribution of parthenogenetic species in the tree of life suggests independent sexual to asexual events (Butlin, 2002), with most appearing to be recent species and with few deep branches - a measure of longevity - of the parthenogenetic forms.

Our results show multiple origins of parthenogenetic species resulting from recurrent

hybridization events in a very short time interval. Parthenogenetic species are expected to be short lived, with severe evolutionary constraints and the change to parthenogenesis is expected to happen rarely (Vrijenhoek, 1989). However, parthenogenetic species in *Darevskia* evolved multiple times in a reticulate pattern and different sexual lineages participated in the hybridization events that led to their origin. This had already been suggested from the limited evidence based on Cyt-*b* and proteins (Murphy et al., 2000) but is now placed in a robust spatiotemporal context by our multilocus analyses for those parthenogens that have *D. raddei* as the maternal species.

Parthenogenetic species origin

All parthenogenetic species analysed here are young in age (Figure 2.2.3). Despite this recent origin, *D. unisexualis* is distributed across a considerable range (Figure 2.2.1) and is rarely found in sympatry with its maternal species, *D. raddei* (Arakelyan et al., 2011). This is even more surprising considering that the suitable area of *D. unisexualis* predicted by the ecological niche modelling widely overlaps with that of *D. raddei* sensu lato. This niche overlap (Figure 2.2.6), together with the wide distribution of *D. unisexualis* despite its recent origin (Figures 2.2.1-2.2.2) and the obtained signals of population expansion (Table 2.2.1), suggest this parthenogen may even outcompete its maternal species within its range. Evidence that parthenogenetic *Darevskia* can outcompete their sexual parental species has already been shown for the parthenogenetic *D. dahli* (Tarkhnishvili et al., 2000).

In this study, the mtDNA was used to analyse the maternal ancestry of the parthenogenetic species and nuclear markers to assess their maternal and paternal contributions. For both nuclear markers, as expected, all parthenogens presented two different alleles, one representing the maternal ancestry (shared with *D. raddei*) and the other representing the paternal contribution (shared or closer to *D. portschinskii, D. rudis* and *D. valentini*). The parthenogens analysed here are allocated in two different lineages in the mtDNA tree. Thus, two different *D. raddei* lineages were apparently involved in the hybridization events that led to the origin of these parthenogens. *D. unisexualis* and *D. uzzelli* belong to the same mtDNA clade as individuals identified as *D. raddei "nairensis"* (Figure 2.2.2: mtDNA lineage 2), and it is likely that this is the maternal lineage for both *D. unisexualis* and *D. uzzelli*. On the other hand, *D. bendimahiensis* shares the same mtDNA haplotype with individuals identified as *D. raddei "vanensis"* found in Turkey east of Lake Van and

adjacent Iran (Figure 2.2.2: mtDNA lineage 3), and therefore this lineage is the most likely one to have contributed in situ as the maternal parental for this parthenogen.

Considering the maternal ancestry, two contributing maternal lineages for the three parthenogenetic species analysed are confirmed with the nuclear markers. *D. unisexualis* and *D. uzzelli* share the same MC1R haplotype with individuals from Mount Aragats. This haplotype was only found in homozygosity in individuals from Amberd Castle, and it is probable that this population reflects the original genetic maternal ancestry of both parthenogenetic species (*D. unisexualis* and *D. uzzelli*). The inferred maternal ancestry of *D. bendimahiensis* with nuclear markers is concordant with the mtDNA phylogenetic tree.

D. valentini and *D. portschinskii* were used to allocate the paternal ancestry in the analysis of the nuclear markers. In the MC1R network, *D. unisexualis* shares its most common haplotype with *D. bendimahiensis*, while *D. uzzelli* shares its haplotype with individuals of *D. valentini*. This contrasts with the maternal ancestry where *D. unisexualis* and *D. uzzelli* share the same haplotype, and *D. bendimahiensis* presents a different maternal allele. Therefore, even though the maternal lineage was the same, two different paternal alleles are identified and, hence, at least two different hybridization events were responsible for the (independent) origin of *D. unisexualis* and *D. uzzelli*. It is noteworthy that contrary to what had been previously reported by Fu et al. (2000a), we did not find evidence of reciprocal hybridization in *D. uzzelli*. In their work, Fu and collaborators suggested the initial hybridization leading to the origin of *D. uzzelli* was most likely reciprocal, since they found mtDNA of both parental species in these parthenogens. However, all individuals analysed in this study showed the same combination of haplotypes both for mtDNA and for nuclear markers, so it is unlikely that a reciprocal hybridization is at the origin of *D. uzzelli.*

Parthenogenetic reproduction can be performed via two ways: apomictic parthenogenesis or automictic parthenogenesis (Simon et al., 2003). While in the first the meiosis is suppressed and clonal offspring are produced under a mitosis-like cell division, the second retains meiosis (and recombination) and ploidy is restored by the duplication or fusion of the maternal gametes (Simon et al., 2003). Since all parthenogens analysed are heterozygous for the nuclear markers and considering the high number of individuals tested, apomictic parthenogenesis is here favoured. However, in some cases of automixis, the chromosomes are replicated prior to the normal meiosis, so diploidy and heterozygozity are restored in the egg (Simon et al., 2003). In such cases heterozygosity will only be lost in some parts of the genome and after some time. Thus, their consistent

heterozygosity can also be explained by their recent origin, and it would be interesting to perform a genome-wide analysis to clarify this question.

According to the placement of *D. unisexualis* and *D. uzzelli* within the phylogeny of the *D. raddei* complex, we estimate that these parthenogens split from the closest *D. raddei* lineage around 170 kyr (291 – 75 kyr), very close to or even during the LIG (130 – 115 kyr). Very likely, mild climate conditions may have facilitated population expansions of parental species increasing the probability of secondary contacts and opening the opportunity for the hybridization between the parental species. For *D. bendimahiensis*, the split with its closest *D. raddei* lineage (lineage 3) should have happened between 204 – 78 k yrs ago. This time interval practically overlaps with the split between *D. unisexualis* + *D. uzzelli* with *D. raddei* lineage 2. This could suggest the hybridization mediating the origin of *D. bendimahiensis* was concurrent with the hybridization event which led to the *D. unisexualis* + *D. uzzelli* lineage. The split between *D. uzzelli* and *D. unisexualis*, on the other hand, appeared to have happened later while the LGM was taking place. Given they share the same mitochondrial lineage and maternal alleles and differ only in the paternal allele it is not clear whether *D. uzzelli* and *D. unisexualis* originated from two different hybridization events between *D. raddei* and *D. valentini*, or if one was first originated and then backcrossed with a *D. valentini* male giving origin to the other. Since only MC1R could differentiate different paternal lineages, nuclear markers across the genome need to be analysed in order to understand the complex reticulate evolution history of these parthenogens and the relationship between them.

Phylogenetic relationships and historical range shifts

Even though the maternal contributions for the parthenogenetic species studied here were already proposed (Fu et al., 2000b, 2000c), phylogenetic relationships between the parental species were obscure. Here, a phylogeographic analysis of *D. raddei* with mtDNA and nuclear markers is performed, and the intraspecific diversity compared to the possible biogeographic barriers, either current or past, within the range of this species complex. Currently, *D. raddei* sensu lato is distributed along the mountain ranges in the Central Caucasus. Given the interconnectivity of these mountain ranges and the prevalence of these species in mountain habitats, mountains are not expected to represent current barriers to dispersal, but instead act as bridges facilitating expansion. However, arid lowlands and possibly deep river beds may act as geographic barriers to dispersal for

these species.

Under the current climatic conditions, no obvious strong barriers to dispersal are found within the occupied range with the possible exception of the Aras River. The Aras valley, the political border between Armenia, Azerbaijan, Iran and Turkey, has a temperate arid mountain climate and is likely to be a current barrier to dispersal between lineages 1, 2 and 4 (in Armenia) and 3 (Turkey). This barrier may have caused the lineage formed by the individuals morphologically identified as *D. raddei "vanensis"* to be geographically isolated and to have evolved in allopatry. This group is monophyletic in the mtDNA tree and also harbours a distinct group of MC1R haplotypes. This indicates at least a certain degree of isolation, and provides some support for the subspecies *D. raddei vanensis*. In contrast, *D. raddei "raddei"* and *D. raddei* "*nairensis*" are found to be paraphyletic. Not only do they form part of the same mtDNA lineages but they also share haplotypes (both nuclear and mitochondrial). Therefore these two taxa lack phylogenetic support. Given the divergence time estimates (Figure 2.2.3), *D. raddei* divergence started no earlier than 1.5 Myrs ago [1.53-0.0116]. Hence, the semi-isolated pattern found for the mtDNA lineages likely originated during the Pleistocenic ice-ages.

To estimate if *D. raddei* and *D. unisexualis* show deviations from neutrality and signals of population expansion, diversity parameters were calculated. Both species showed significant R_2 for all markers while Tajima's *D* was significantly negative only for ND4 (*D. raddei and D. unisexualis*) and Cyt-*b* *(D. unisexualis*). Negative values of Tajima's *D* (and Fu's *Fs*) and small positive values of R_2 are indicative of population growth (Aris-Brosou and Excoffier, 1996; Tajima, 1989). While Tajima's *D* uses information on mutation frequency, Fu's *Fs* test relies on haplotype distribution and has been shown under simulation to be the more powerful when analysing small populations (Ramos-Onsins and Rozas, 2002). Given the high number of samples for each "population", or in this case, species, this could explain why this test did not detect significant departures from neutrality while Tajima's *D* did.

Considering the recent origin of the parthenogenetic species analysed here (*D. unisexualis*), a recent expansion of this species is to be expected. Currently, *D. unisexualis* has a large distribution area resulting from expansion since its origin. The mtDNA lineages with short branches and several closely related mtDNA haplotypes of *D. raddei* sensu lato are indicative of potentially recent expansions which could match the deviations from neutrality found.

Trends from the past to the present

In the phylogenetic analysis of *D. raddei* sensu lato, we found very little geographical structure at both nuclear markers and sympatry of the mtDNA lineages. This suggests cyclical contact-isolation events, concordant with the complex biogeography of the Caucasus. During the Pleistocenic Glacial Periods, contrary to the current situation, the mountain ranges, (i.e. Geghama Mountains, a volcanic mountain range west to Lake Sevan spanning North to South and attaining 3567 m), may have acted as a barrier between the lineage from Northern Armenia and Georgia, and all the others. The mtDNA lineage 1 reaches this topographic barrier and is found in Northwest Geghama, while lineages 4, 5, 68 and 6 are found Southeast of this Mountain. A similar picture emerges for lineages 1 and 2. Mount Aragats, which currently represents a suitable habitat for these species, did not do so during the LGM. Therefore, lineage 2 may have been trapped on the western side of Aragats, and lineage 1 on the eastern side during this period.

After the LGM temperatures started to increase and the mountain environments became again suitable for these lizards, with lineages that had been previously separated able to come into contact, as currently is the case for lineage 1 and lineage 4. This pattern of isolation can be observed in the projections of the *D. raddei* models for the LGM. Here, both scenarios (MIROC and CCSM) show there was a decrease of suitable habitat around the mountain tops (Aragats and Geghama) and a general increase of potential distribution area and a geographical shift of the suitable habitat to the Azerbaijan lowlands, when compared with the present distribution model (Figure 2.2.6).

The cyclical ice ages and subsequent expansion-contraction of organisms in a habitat relatively small but with heterogeneous topography must have allowed for the secondary contact of sexual *Darevskia* lineages in incomplete stages of reproductive isolation (Vrijenhoek, 1989). This likely allowed repeated hybridization events in separate geographical areas that originated hybrids that could not cross-back with the parental groups (or species) but instead were able to reproduce parthenogenetically.

Since its origin, there was a decrease of the potential habitat of *D. unisexualis* from the LGM to the present day, according to both scenarios (Figure 2.2.6). However, considering that this parthenogen is likely expanding, and hence, not in equilibrium with the environment, the predicted model and its consequent projections will be probably underestimations (Wiens et al., 2009).

Given the present ecological model estimated in this study for *D. raddei* and *D.*

unisexualis, one may conclude that there was no appropriate habitat available for either of them across the whole region in the LIG projections. Since the origin of *D. unisexualis* was estimated to have happened after the LIG, this species simply would not have been present. On the other side, and given its present distribution, the lack of suitable habitat for *D. raddei* during the LIG may suggest this species may have suffered a recent niche shift or, alternatively, that the scale of the model was inappropriate to detect suitable habitat during this period.

The models for the present cover a much larger area than the known distribution for both *D. raddei* and *D. unisexualis.* Ecological niche models are estimated based on a dataset of presence points collected and have an error rate associated with them. Additionally, given the limited dispersal abilities of these species, their absence from the potential habitat could be due to a separation by unsuitable habitat or an incomplete expansion process, or even because of competitive exclusion by other *Darevskia* species with similar ecological niches.

CONCLUSION

Given the scarcity and distribution of parthenogenetic species in the tree of life, the switch from sexual to parthenogenetic reproduction is expected to arise rarely and independently (Butlin, 2002). Most parthenogenetic forms appear to have originated recently, as shown by the lack of parthenogenetic deep phylogenetic branches.

Most sexual-parthenogenetic complexes show a polyphyletic origin of parthenogenetic lineages (Crease et al., 1989; Grismer et al., 2014; Simon et al., 2003), where parthenogenesis has evolved more than once. Our results clearly support a polyphyletic origin of parthenogenesis in *Darevskia* lizards as well, dated back to the Pleistocene, with different parental lineages contributing to the hybridization events occurring several times in different geographical regions. The origin of the polyphyletic parthenogenetic *Darevskia* has to be interpreted as resulting from repeated secondary contacts between groups that did not developed complete reproductive isolation. The distribution of the different *Darevskia* groups (or species) likely underwent repeated contraction-expansion events in response to the Pleistocenic climate oscillations, colder periods interspersed with warmer interglacials, promoting secondary contacts. Thus, some lineages were divergent enough to produce hybrids with disrupted meiosis, yet not so divergent as to compromise hybrid

viability and fertility (Vrijenhoek, 1989).

We also show that even though sexual species and parthenogens overlap in their ecological niche, *D. unisexualis* is not found in sympatry with *D. raddei*. Since *D. unisexualis* derived from *D. raddei* and is therefore younger, this suggests that the first is outcompeting the second, as has been shown for other parthenogenetic *Darevskia* (Tarkhnishvili et al., 2010). Parthenogenetic species have some advantages over sexual species; they avoid the two-fold cost of males (Maynard Smith, 1978), having twice the reproductive output if other factors are excluded, they are not affected by the associated costs of sex as male-male competition, search and choice of mates (Galoyan, 2013) and in some cases the mechanics of meiosis (Lehtonen et al., 2012). In the short term, this may provide an advantage when in sympatry with sexual species (Burke et al., 2015; Tarkhnishvili et al., 2000). On the other hand, sexual reproduction is known as a driver of evolution and speciation. As such, parthenogenetic species, which lack the recombination benefits of sexual reproduction, are expected to be at disadvantage when in competition with sexual species in changing environments. The different stages of parthenogenetic species can help to understand the effect of asexuality (or the absence of sex) on the genome. Given the different ages and the polyphyletic and hybrid origin of their parthenogenetic species, *Darevskia* lizards provide a promising model for the study of the evolution of asexuality and why sexual reproduction is so widespread in the tree of life.

Acknowledgements

S. Drovetski, F. Jorge and A. Vardanyan assisted with fieldwork. We are especially grateful to two anonymous reviewers for their constructive comments on the manuscript. This research was partially funded by FCOMP-01-0124-FEDER-007062 FCT through the project PTDC/BIA-BEC/101256/2008, by the project "Biodiversity, Ecology and Global Change" co-financed by North Portugal Regional Operational Program 2007/2013 (ON.2 – O Novo Norte), under the National Strategic Reference Framework (NSRF), through the European Regional Development Fund (ERDF) and by the project "Preserving Armenian biodiversity: Joint Portuguese – Armenian program for training in modern conservation biology." of Gulbenkian Foundation (Portugal). Field work in Turkey was funded by the project supported by Dokuz Eylül University, Turkey (Project No. 2009.KB.FEN.003). NS was supported by a research contract (IF2013) from FCT.

REFERENCES

Ahmadzadeh, F., Flecks, M., Carretero, M.A., Böhme, W., Ilgaz, C., Engler, J.O., James Harris, D., Üzüm, N., and Rödder, D. (2013a). Rapid lizard radiation lacking niche conservatism: ecological diversification within a complex landscape. J. Biogeogr. *40*, 1807–1818.

Ahmadzadeh, F., Flecks, M., Rödder, D., Böhme, W., Ilgaz, Ç., Harris, D.J., Engler, J.O., Üzüm, N., and Carretero, M.A. (2013b). Multiple dispersal out of Anatolia: biogeography and evolution of oriental green lizards. Biol J Linn Soc Lond *110*, 398–408.

Anderson, S.C. (1999). The lizards of Iran (Ithaca, NY: Society for the Study of Amphibians and Reptiles).

Arakelyan, M., Danielyan, F., Corti, C., Sindaco, R., and Levinton, A.E. (2011). The Herpetofauna of Armenia and Nagorno-Karabakh (San Francisco).

Arévalo, E., Davis, S.K., and Sites, J.W. (1994). Mitochondrial DNA Sequence Divergence and Phylogenetic Relationships among Eight Chromosome Races of the Sceloporus Grammicus Complex (Phrynosomatidae) in Central Mexico. Syst Biol *43*, 387–418.

Aris-Brosou, S., and Excoffier, L. (1996). The impact of population expansion and mutation rate heterogeneity on DNA sequence polymorphism. Mol Biol Evol *13*, 494–504.

Arnold, E.N., Arribas, Ó., and Carranza, S. (2007). Systematics of the Palaearctic and Oriental lizard tribe Lacertini (Squamata:Lacertidae:Lacertinae), with descriptions of eight new genera. Zootaxa *1474*, 1–86.

Avise, J.C. (2008). Clonality: the genetics, ecology, and evolution of sexual abstinence in vertebrate animals (Oxford ; New York: Oxford University Press).

Barata, M., Carranza, S., and Harris, D.J. (2012). Extreme genetic diversity in the lizard Atlantolacerta andreanskyi (Werner, 1929): A montane cryptic species complex. BMC Evolutionary Biology *12*, 167.

Bell, G. (1982). The masterpiece of nature: The evolution and genetics of sexuality (Berkeley).

Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C.-H., Xie, D., Suchard, M.A., Rambaut, A., and Drummond, A.J. (2014). BEAST 2: A Software Platform for Bayesian Evolutionary Analysis. PLoS Comput Biol *10*, e1003537.

Burke, N.W., Crean, A.J., and Bonduriansky, R. (2015). The role of sexual conflict in the evolution of facultative parthenogenesis: a study on the spiny leaf stick insect. Animal Behaviour *101*, 117–127.

Butlin, R. (2002). The costs and benefits of sex: new insights from old asexual lineages. Nat Rev Genet *3*, 311–317.

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

Clement, M., Posada, D., and Crandall, K.A. (2000). TCS: a computer program to estimate gene genealogies. Molecular Ecology *9*, 1657–1659.

Crease, T.J., Stanton, D.J., and Hebert, P.D.N. (1989). Polyphyletic Origins of Asexuality in Daphnia pulex. II. Mitochondrial-DNA Variation. Evolution *43*, 1016–1026.

Darevsky, I.S. (1967). Rock lizards of the Caucasus: systematics, ecology and phylogenesis of the polymorphic groups of Caucasian rock lizards of the subgenus Archaeolacerta. Leningrad. Nauka.

Drummond A. J. & Bouckaert R. R. (2015) "Bayesian evolutionary analysis with BEAST", Cambridge University Press.

Drummond, A.J., and Rambaut, A. (2007). BEAST: Bayesian evolutionary analysis by sampling

trees. BMC Evolutionary Biology *7*, 214.

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(8), 1969-1973.

Felsenstein, J. (1985). Confidence Limits on Phylogenies: An Approach Using the Bootstrap. Evolution *39*, 783.

Flot, J.-F. (2010). SeqPhase: a web tool for interconverting PHASE input/output files and FASTA sequence alignments. Molecular Ecology Resources *10*, 162–166.

Freitas, S., Vavakou, A., Arakelyan, M., Drovetski, S.V., Crnobrnja-isailović, J., Kidov, A.A., Cogălniceanu, D., Corti, C., Lymberakis, P., Harris, D.J., et al. (2016). Cryptic diversity and unexpected evolutionary patterns in the meadow lizard, *Darevskia praticola* (Eversmann, 1834). Systematics and Biodiversity *14*, 184–197.

Fu, J., Murphy, R.W., and Darevsky, I.S. (1997). Toward the phylogeny of caucasian rock lizards: implications from mitochondrial DNA gene sequences (Reptilia: Lacertidae). Zoological Journal of the Linnean Society *120*, 463–477.

Fu, J., MacCulloch, R.D., Murphy, R.W., Darevsky, I.S., and Tuniyev, B.S. (2000a). Allozyme Variation Patterns and Multiple Hybridization Origins: Clonal Variation Among Four Sibling Parthenogenetic Caucasian Rock Lizards. Genetica *108*, 107–112.

Fu, J., Murphy, R.W., Darevsky, I.S., and McEachran, J.D. (2000b). Divergence of the cytochrome b gene in the Lacerta raddei complex and its parthenogenetic daughter species: evidence for recent multiple origins. Copeia *2000*, 432–440.

Garrick, R.C., Sunnucks, P., and Dyer, R.J. (2010). Nuclear gene phylogeography using PHASE: dealing with unresolved genotypes, lost alleles, and systematic bias in parameter estimation. BMC Evolutionary Biology *10*, 118.

Gilabert, A., Simon, J.-C., Dedryver, C.-A., and Plantegenest, M. (2014). Do ecological niches differ between sexual and asexual lineages of an aphid species? Evolutionary Ecology *28*, 1095–1104.

Grechko, V.V., Bannikova, A.A., Kosushkin, S.A., Ryabinina, N.L., Milto, K.D., Darevsky, I.S., and Kramerov, D.A. (2007). Molecular genetic diversification of the lizard complex Darevskia raddei (Sauria: Lacertidae): Early stages of speciation. Molecular Biology *41*, 764–775.

Grismer, J.L., Bauer, A.M., Grismer, L.L., Thirakhupt, K., Aowphol, A., Oaks, J.R., Wood, P.L., Onn,

C.K., Thy, N., Cota, M., et al. (2014). Multiple origins of parthenogenesis, and a revised species phylogeny for the Southeast Asian butterfly lizards, Leiolepis. Biol J Linn Soc Lond n/a – n/a.

Guindon, S., and Gascuel, O. (2003). A Simple, Fast, and Accurate Algorithm to Estimate Large Phylogenies by Maximum Likelihood. Syst Biol *52*, 696–704.

Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G., and Jarvis, A. (2005). Very high resolution interpolated climate surfaces for global land areas. International Journal of Climatology *25*, 1965–1978.

Hudson, R.R. (1991). Gene geneologies and the coalescent process. Oxford Surveys in Evolutionary Biology *7*, 1–44.

Huelsenbeck, J.P., and Ronquist, F. (2001). MRBAYES: Bayesian Inference of phylogenetic trees. Bioinformatics *17*, 754–755.

Hutchinson, G.E. (1957). Concluding Remarks. Cold Spring Harb Symp Quant Biol *22*, 415–427.

Katoh, K., and Standley, D.M. (2013). MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. Mol Biol Evol *30*, 772–780.

Kearney, M.R. (2003). Why is sex so unpopular in the Australian desert? Trends in Ecology & Evolution *18*, 605–607.

Kearney, M., Fujita, M.K., and Ridenour, J. (2009). Lost Sex in the Reptiles: Constraints and Correlations. In Lost Sex, I. Schön, K. Martens, and P. Dijk, eds. (Springer Netherlands), pp. 447–474.

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

Lanfear, R., Calcott, B., Ho, S.Y.W., and Guindon, S. (2012). PartitionFinder: Combined Selection of Partitioning Schemes and Substitution Models for Phylogenetic Analyses. Molecular Biology and Evolution *29*, 1695–1701.

Liu, C., Berry, P.M., Dawson, T.P., and Pearson, R.G. (2005). Selecting thresholds of occurrence in the prediction of species distributions. Ecography *28*, 385–393.

MacCulloch, R.D., Fu, J., Darevsky, I.S., and Murphy, R.W. (2000). Genetic evidence for species status of some Caucasian rock lizards in the Darevskia saxicola group. Amphibia-Reptilia *21*, 169–

176.

Maynard Smith, J. (1978). The Evolution of Sex (New York: Cambridge University Press).

Murphy, R.W., Fu, J., MacCulloch, R.D., Darevsky, I.S., and Kupriyanova, L.A. (2000). A fine line between sex and unisexuality: the phylogenetic constraints on parthenogenesis in lacertid lizards. Zoological Journal of the Linnean Society *130*, 527–549.

Nylander, J.A.A., Wilgenbusch, J.C., Warren, D.L., and Swofford, D.L. (2008). AWTY (are we there yet?): a system for graphical exploration of MCMC convergence in Bayesian phylogenetics. Bioinformatics *24*, 581–583.

Otto, S.P., and Nuismer, S.L. (2004). Species Interactions and the Evolution of Sex. Science *304*, 1018–1020.

Phillips, S.J., Dudík, M., and Schapire, R.E. (2004). A maximum entropy approach to species distribution modeling. In Proceedings of the Twenty-First International Conference on Machine Learning, (ACM), p. 83.

Phillips, S.J., Anderson, R.P., and Schapire, R.E. (2006). Maximum entropy modeling of species geographic distributions. Ecological Modelling *190*, 231–259.

Pinho, C., Harris, D.J., and Ferrand, N. (2007). Contrasting patterns of population subdivision and historical demography in three western Mediterranean lizard species inferred from mitochondrial DNA variation: historical demography in a latitudinal gradient. Molecular Ecology *16*, 1191–1205.

Ramos-Onsins, S.E., and Rozas, J. (2002). Statistical Properties of New Neutrality Tests Against Population Growth. Mol Biol Evol *19*, 2092–2100.

Rastegar-Pouyani, N., Faizi, H., Oraei, H., Khosravani, A., Fathinia, B., Heidari, N., Karamiani, R., and Rastegar-Pouyani, E. (2011). A brief history and current status of herpetology in Iran.

Rastegar-Pouyani, N., Karamiani, R., Oraei, H., Khosrawani, A., and Rastegar-Pouyani, E. (2012). A New Subspecies of Darevskia raddei (Boettger, 1892) (Sauria: Lacertidae) from the West Azerbaijan Province, Iran: A New Subspecies of Darevskia raddei (Boettger, 1892) (Sauria: Lacertidae) from the West Azerbaijan Province, Iran. Asian Herpetological Research *2*, 216–222.

Ronquist, F., and Huelsenbeck, J.P. (2003). MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics *19*, 1572–1574.

Rull, V. (2010). On microrefugia and cryptic refugia. Journal of Biogeography *37*, 1623–1625.

Sambrook, J., and Russell, D.W. (2001). Molecular Cloning: A Laboratory Manual (CSHL Press).

Schneider, S., and Excoffier, L. (1999). Estimation of Past Demographic Parameters From the Distribution of Pairwise Differences When the Mutation Rates Vary Among Sites: Application to Human Mitochondrial DNA. Genetics *152*, 1079–1089.

Seddon, J.M., Santucci, F., Reeve, N., and Hewitt, G.M. (2002). Caucasus Mountains divide postulated postglacial colonization routes in the white-breasted hedgehog, Erinaceus concolor. Journal of Evolutionary Biology *15*, 463–467.

Simon, J.-C., Delmotte, F., Rispe, C., and Crease, T. (2003). Phylogenetic relationships between parthenogens and their sexual relatives: the possible routes to parthenogenesis in animals. Biological Journal of the Linnean Society *79*, 151–163.

Stephens, M., Smith, N.J., and Donnelly, P. (2001). A New Statistical Method for Haplotype Reconstruction from Population Data. The American Journal of Human Genetics *68*, 978–989.

Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics *123*, 585–595.

Tarkhnishvili, D., Thorpe, R.S., and Arntzen, J.W. (2000). Pre-pleistocene refugia and differentiation between populations of the Caucasian salamander (Mertensiella caucasica). Mol. Phylogenet. Evol. *14*, 414–422.

Tarkhnishvili, D., Gavashelishvili, A., Avaliani, A., Murtskhvaladze, M., and Mumladze, L. (2010). Unisexual rock lizard might be outcompeting its bisexual progenitors in the Caucasus. Biological Journal of the Linnean Society *101*, 447–460.

Tarkhnishvilli DN, Murtskhvaladze M, Gavashelishvili A (2013) Speciation in Caucasian lizards: climatic dissimilarity of the habitats is more important than isolation time. Biological Journal of the Linnean Society 109:876-892

Uetz, P., and Hošek, J. The Reptile Database, http://www.reptile-database.org, accessed December 4, 2015.

VanDerWal, J., Shoo, L.P., Johnson, C.N., and Williams, S.E. (2009). Abundance and the Environmental Niche: Environmental Suitability Estimated from Niche Models Predicts the Upper Limit of Local Abundance. The American Naturalist *174*, 282–291.

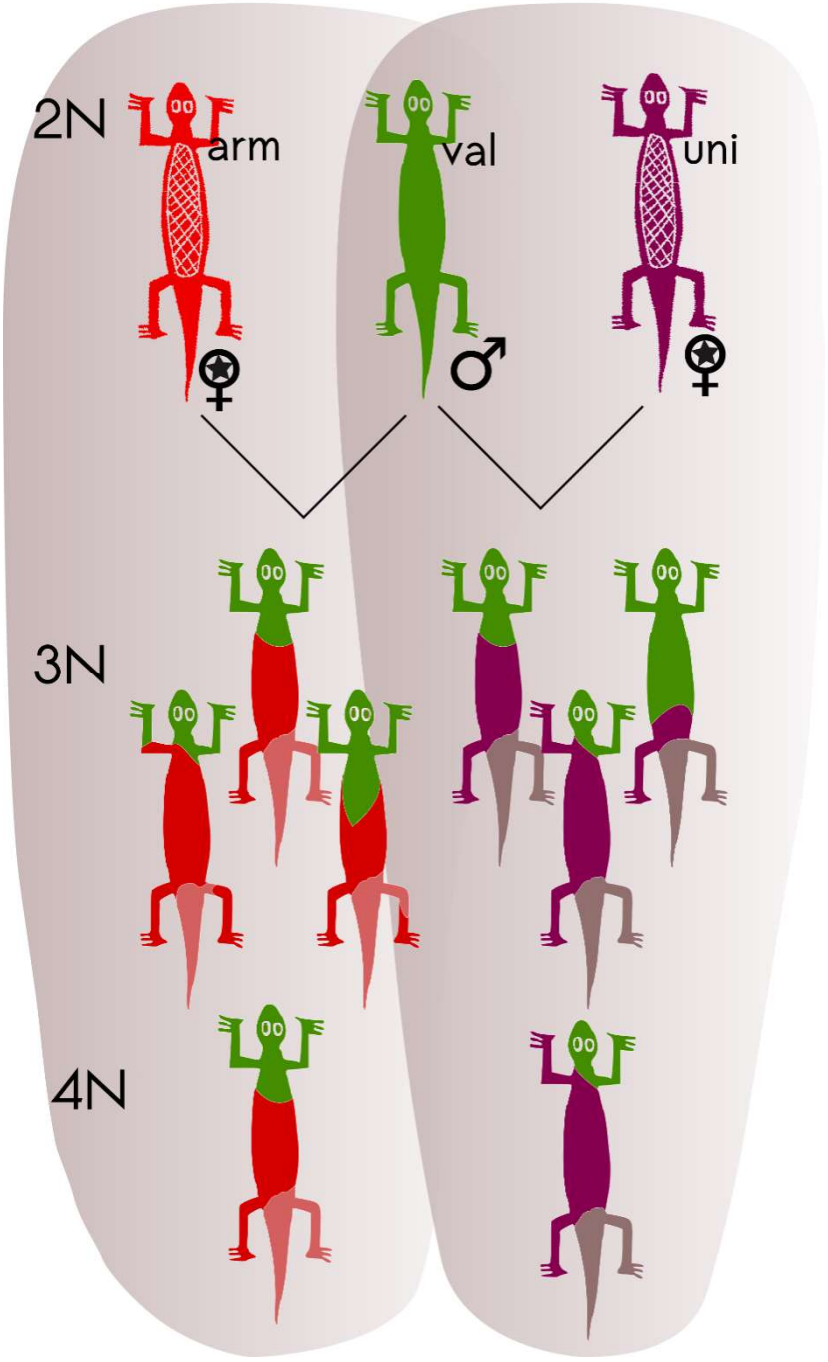
Vrijenhoek, R.C. (1989). Genetic and ecological constraints on the origins and establishment of

unisexual vertebrates. Evolution and Ecology of Unisexual Vertebrates *466*, 24–31.

Wiens, J.A., Stralberg, D., Jongsomjit, D., Howell, C.A., and Snyder, M.A. (2009). Niches, models, and climate change: Assessing the assumptions and uncertainties. PNAS *106*, 19729–19736.

CHAPTER 3

ORIGIN OF VERTEBRATE ASEXUALITY - paper III



The origin of parthenogenesis and the roles of hybridization and polyploidy in a lizard model

Susana Freitas[1,2]

D. James Harris[1]

Neftali Silero[1]

Marine Arakelyan[3]

Roger Butlin*[2]

Miguel A. Carretero*[1]

*these authors contributed equally

[1]CIBIO Research Centre in Biodiversity and Genetic Resources, InBIO, Universidade do Porto, Campus de Vairão, Rua Padre Armando Quintas, Nº 7. 4485-661 Vairão, Vila do Conde, Portugal

[2]Department of Animal and Plant Sciences, The University of Sheffield, Sheffield S10 2TN, UK

[3]Yerevan State University, Alek Manoogian, 1, Yerevan 0025, Armenia

ABSTRACT

Obligate parthenogenesis is found in only 0.1% of vertebrate species and thought to be relatively short lived and typically with a hybrid origin. However, neither the evolutionary persistence of asexuality in vertebrates, nor the conditions that allow the generation of new parthenogenetic lineages are currently well understood. It has been proposed that in vertebrates, parthenogenetic lineages arise from the hybridization between two divergent taxa, within a specific phylogenetic distance. Moreover, parthenogenetic species often maintain a certain level of hybridization with their closest sexual relatives, potentially generating new polyploid hybrid lineages. Here we use a set of microsatellite and mtDNA markers to examine the role of hybridization in the origin of vertebrate parthenogens, the extent of gene flow with sexual relatives and the relationship between hybridization, asexuality and polyploidy, in the context of the main theories regarding the origins of asexuality. The model used, *Darevskia*, is a group of rock lizards from the Caucasus that includes at least seven hybrid parthenogenetic species. We focused on the three most widespread parthenogens, and their polyploid backcross hybrids. Our results show that the recurrent backcrossing between sexual and parthenogenetic *Darevskia* has not led to gene flow, reinforcing the idea that parthenogenesis in vertebrates acts as a reproductive

barrier. We also find parthenogens result from the hybridization of lineage-specific species, regardless of their phylogenetic distance, and that these hybridization events should have happened multiple times in the past, but are not occurring in the present.

INTRODUCTION

Asexual reproduction is distributed across all major clades of the tree of life. Exclusively asexual species are expected to be short lived (Lehtonen and Kokko, 2014), and typically have a twiggy phylogenetic distribution (Bell, 1982), while originating from sexual ancestors (Avise, 2008). Many studies have focused on the evolution of sex, trying to understand how such a costly reproductive mode is so successful in nature (Maynard Smith, 1978; Otto, 2009; Weismann, 1889), and this paradox of sex has been tentatively explained by the superior potential for adaptation to changing environments of sexual species compared to asexual ones (Luijckx et al., 2017; McDonald et al., 2016). Given the putative young age of (most) asexual species, we can deduce that sexual species have the potential to constantly give rise to new asexual lineages, but the predicted balance between generation and extinction of asexual lineages is not frequently empirically studied.

The maintenance of asexual species (which constitute clusters of lineages of similar genotype) depends on the balance between the generation of asexual lineages from sexual progenitors and their loss, either through accumulation of deleterious mutations (Haigh, 1978), failure to adapt (Lively, 2010) or even neutral processes (Janko, 2014; Schwander and Crespi, 2009). The proportion of parthenogenetic taxa is higher in invertebrates than in vertebrates (Beukeboom and Vrijenhoek, 1998), generally resulting from spontaneous thelytokous parthenogenesis (Bullini, 1994). On the other hand, parthenogenetic vertebrates are less common (Beukeboom and Vrijenhoek, 1998), generally originated through hybridization (but see (Noonan et al., 2013)), are obligate, mostly recent (Avise, 2008; Beukeboom and Vrijenhoek, 1998) and often include polyploids (Bullini, 1994). The hybridization events that originate them are generally multiple, as has already been described in the whiptail lizards *Aspidoscelis* sp. (Reeder et al., 2002), the gecko *Lepidodactylus lugubris* (Trifonov et al., 2015) and the clonal hybrid fish *Cobitis elongatoides-taenia* (Choleva et al., 2012).

A hybrid origin of asexuality has been reported recurrently in several groups, including plants (Beck et al., 2011), fish (where hybridization may lead to gynogenesis and

polyploidy) (Choleva et al., 2012), and lizards (Lutes et al., 2011). Frequently, these sexual-asexual complexes are better described as a network of species that recurrently hybridize, possibly allowing for gene flow between asexual and sexual species (Menken et al., 1995), or even the origin of new parthenogenetic species (Taylor et al., 2015). Cases where gene flow occurs between proposed evolutionary units or where there is a recurrent origin of new asexual lineages from sexual parents question the current definition and applicability of some species concepts (Birky and Barraclough, 2009; Coyne et al., 1988).

Two general hypotheses have been put forward regarding the conditions of inter-species hybridization on the origin of parthenogenetic vertebrates: the Balance Hypothesis (Moritz et al., 1989) and the Phylogenetic Constraint Hypothesis (Darevsky, 1967). Although not mutually exclusive (Avise, 2008), they describe two different situations. The Balance Hypothesis suggests that parthenogenetic vertebrates arise by the hybridization of two sexual species divergent enough to disrupt meiosis in the hybrids, yet not so divergent as to seriously compromise hybrid viability or (parthenogenetic) fertility (Kearney et al., 2009). The Phylogenetic Constraint Hypothesis suggests that asexual lineages originate by the hybridization between sexual species that possess lineage-dependent genetic peculiarities that allow them to interbreed and produce viable, hybrids capable of reproducing parthenogenetically. Since these peculiarities may be specific to the female parent, the hybridization events are expected to be directional, with species from different phylogenetic clades contributing either the maternal or paternal ancestry (Avise, 2008).

The model system used in this study, Darevskia lizards, has a hybridization-rich evolutionary history, and thus is a very good model to study the correlation between hybridization and parthenogenesis. All of its asexual lineages are reported to be of hybrid origin (Freitas et al., 2016; Murphy, 2000), there is evidence for recurring mating between asexual females and sexual males when in sympatry generating polyploid backcrosses (Danielyan et al., 2008; Darevsky and Danielyan, 1968), and for frequent interspecific hybridization between sexual species (Darevsky, 1967). Given the restriction to a few sexual species as maternal and paternal ancestors of parthenogenetic *Darevskia*, this model fulfils the predictions of the Phylogenetic Constraint Hypothesis. However, current evidence for the origin of the different parthenogenetic species and the identification of the polyploids is based only on a reduced set of markers, and requires confirmation using additional genetic markers.

Initial estimates of the phylogenetic relationships of sexual species (Murphy, 2000; Murphy

et al., 1996) suggest *Darevskia* is divided into three main clades, caucasica, saxicola and rudis (Fu et al., 1997). Considering the asexual species and their putative parents (Figure 3.1), *D. armeniaca* is thought to have resulted from hybridization between *D. mixta* and *D. valentini (Fu et al., 2000b)*, while both *D. unisexualis* and *D. uzzelli* are thought to have resulted from hybridization between *D. raddei* and *D. valentini* (Fu et al., 2000a). The maternal species was always *D. raddei* or *D. mixta* (from caucasica clade) and the paternal was *D. portschinskii, D. rudis* or *D. valentini* (rudis clade). Hybridization between parthenogenetic *Darevskia* females and males of their sexual parent species has been reported previously based mainly on morphology and karyology of a limited set of individuals (Danielyan et al., 2008). Reported triploid hybrids (3n = 57) are either female or male with unknown fertility and different levels of reproductive organ development (Danielyan et al., 2008). Since eggs of different ploidy may develop in the same oviduct, it is not clear if hybrids between sexual males and asexual females are always polyploid or can also be diploid.

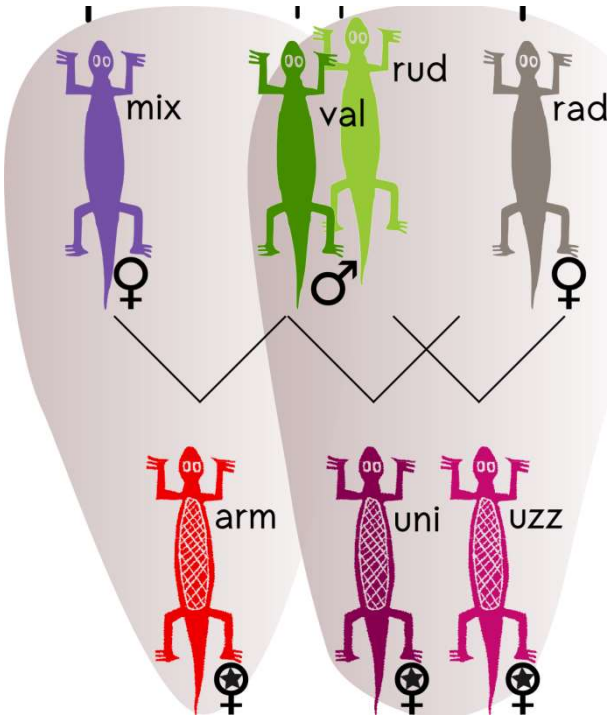


Figure 3.1 - Relationships between sexual species *D. mixta* (mix), *D. valentini* (val), *D. rudis* (rud) and *D. raddei* (rad), with their hybrid parthenogenetic descendant lineages, *D. armeniaca* (arm), *D. unisexualis* (uni) and *D. uzzelli* (uzz).

The goal of this study was to address three questions related to the origin of parthenogenesis in vertebrates using *Darevskia* as a model: (i) Are all sexual parentals restricted to specific phylogenetic clades and/or do they share the same genetic distance between parents; (ii) Do parthenogenetic lineages from the same parental taxa originate repeatedly through time or is there a single temporal event in each case; and (iii) Can parthenogens still backcross with their sexual parentals and generate fertile offspring?

To address these issues, we used microsatellite markers and genotyped individuals from a wide range of localities from Armenia, Turkey, Georgia and Iran. We reassess the inferences of parentage of three parthenogenetic *Darevskia: D. unisexualis*, *D. uzzelli* and *D. armeniaca*, comparing matrilineal (mtDNA) and nuclear (maternal + paternal) lineages of the putative parentals and parthenogens. Current sympatric localities were examined and the evolutionary significance of hybrids is discussed. With this work, we test whether the Phylogenetic Constraint Hypothesis holds for this group as proposed previously (Darevsky, 1967), either alone or in addition to the Balance Hypothesis.

MATERIAL AND METHODS

Sample collection

For this study, 378 *Darevskia* individuals were analysed (Table S3.1). These samples were collected during field expeditions between 2007 and 2011 in Armenia, Turkey, Georgia and Iran. Seven species were included in the analyses, four sexual, putative-parental species (*D. mixta*, *D. raddei*, *D. valentini* and *D. rudis*) and three parthenogenetic species (*D. armeniaca*, *D. unisexualis* and *D. uzzelli*). Individuals were provisionally identified in the field based on overall morphology, size (for putative backcross hybrids), colour pattern and scutellation (Arakelyan and Danielyan, 2011).

Four sympatric localities where sexual and parthenogenetic species coexist were sampled. In the first, Kuchak, previous studies had already reported backcross individuals between parthenogens and sexual species (Danielyan et al., 2008). There, three species are found: two parthenogenetic (*D. armeniaca* and *D. unisexualis*) and one sexual (*D. valentini*), the putative paternal species for both parthenogens found here. Backcrossed individuals, *D. armeniaca x D. valentini* and *D. unisexualis x D. valentini,* are generally identified based on morphology, but polyploidy of some individuals has already been confirmed by karyology, identifying both triploids and tetraploids (Danielyan et al., 2008). In the second locality,

Sotk, the sexual *D. valentini* and the parthenogen *D. armeniaca* are found in sympatry, and individuals morphologically similar to *D. armeniaca x D. valentini* backcrosses have also been reported based on large body size and intermediate colouration (Arakelyan, M. pers. comm.). In the remaining two localities, Lchap and Lchashen, the sexual *D. raddei* is found together with the parthenogen *D. unisexualis* (Lchap) or with the sexual *D. valentini* (Lchashen). *D. raddei* and *D. valentini* are the putative parental species of the hybrid parthenogen *D. unisexualis*.

Genotyping

Genomic DNA was extracted from 30 mg of tail-tip tissue following standard high-salt protocols (Sambrook and Russell, 2001). From a total of 74 tested microsatellite loci that had been developed previously for other lacertid lizards, 12 polymorphic markers were selected on the basis of reliable amplification and heterozygosity: cross-species markers D119, C118, C113 (Remón et al., 2008), Pb55 (Pinho et al., 2004), Lv-4-72 (Boudjemadi et al., 1999), P011, P054 (Wellenreuther et al., 2009), Ph39, Ph124, Ph128, Ph170 (NCBI accession numbers: KC869962, KC869964, KC869956, KC869961) and Du323, Du47, Du418 (Korchagin et al., 2007) (for more information see Table S3.2).

PCR amplifications were carried out using the Multiplex PCR Kit (QIAGEN) following the manufacturer's instructions in a final 10-µl volume, including a negative control. Amplicons were separated by size on an ABI3130xl Genetic Analyser. Allele sizes were scored against the GeneScan500 LIZ Size Standard using GENEMAPPER 4.0 (Applied Biosystems) and manually checked twice, independently. To control against allelic dropout, which is expected to be higher in polyploids due to the greater number of amplicons in a given PCR reaction, 35-45% of genotypes per marker were repeated, including all the putative polyploids, in independent PCR reactions (Table S3.2). For the repeated samples, loci were genotyped individually to confirm that the third (or fourth) allele scored was not an artefact of interaction between the different primer pairs in the multiplex.

To search for the presence of null alleles, genotyping errors and allelic dropout, we used Microchecker 2.2.3 (Van Oosterhout et al., 2004). These tests were performed only on *D. valentini* (excluding individuals from sympatric localities) because it was the species with most individuals, and the assumptions of this analysis are not appropriate for parthenogenetic species or for mixed species samples. Tests for Hardy–Weinberg equilibrium (HWE), linkage disequilibrium (LD) and standard genetic diversity measures,

observed (Ho) and expected (He) heterozygosities, allele frequencies and allelic richness were obtained using Cervus/Fstat v2.9.3.2 (Goudet, 1995). The critical probability for each test was adjusted with a sequential Bonferroni correction (Rice, 1989). Population genetic differentiation was analysed using Fisher's exact test and by estimating F_{ST} (Weir and Cockerham, 1984) in Fstat.

Sympatric localities: Ploidy determination and cluster analyses

Sympatric localities were analysed in pairs on the basis of the ancestry of the species found, grouping localities with similar arrangement of species. Individuals from Kuchak (with *D. valentini*, *D. armeniaca* and *D. unisexualis*) and Sotk (*D. valentini* and *D. armeniaca*) were analysed together, while those from Lchap (*D. raddei* and *D. unisexualis*) and Lchashen (*D.raddei* and *D. valentini*) were examined in a separate analysis.

Ploidy level was identified as the maximum number of alleles found among the markers used (*ploidy* function in Polysat; (Clark and Jasieniuk, 2011)). All individuals characterized as triploids had at least two markers with three different alleles. Given the multiple ploidy levels in the sympatric localities (see Results), traditional diversity measures and population genetic analyses could not be applied to these datasets. Instead, we calculated inter-individual genetic distances using the Bruvo method (Bruvo et al., 2004) implemented in Polysat (detailed information in SI).

Cluster analyses

In analyses restricted to diploid individuals we adopted a hierarchical approach, following the cluster analysis of the full diploid dataset by running a discriminant analysis of principal components (DAPC) using the ADEGENET package (Jombart et al., 2010) on subsets of the data, each addressing specific questions. To determine the number of clusters (*K*) in the total sexual species dataset, DAPC was performed and the optimum number of clusters was chosen with the *find.clusters* option, by comparing the different clustering solutions using a Bayesian Information Criterion (BIC). Individuals were then assigned to clusters and DAPC was used to ordinate individuals according to axes that maximise cluster distances relative to variation within clusters. For similar BIC values, the optimal *K* value was selected based on the concordance between the clusters and described taxa. In order to distinguish individuals from sexual and parthenogenetic species and to visualise their relationships, another DAPC was performed on the dataset of all diploid individuals,

including the diploids from the sympatric localities. The optimal *K* value was chosen as previously described. The pairwise F_{ST} values for the species (and populations) in this dataset were calculated with *pairwise.fst* in the package hierfstat, version 0.04-22 (Goudet, 1995) and used for neighbor-joining tree construction. To avoid effects on F_{ST} due to small sample size, the pairwise F_{ST} was estimated only when four or more individuals were available per population (or species). Finally, to determine the number of clonal lineages and hybrid origins, we conducted assignments using DAPC membership probability values for each parthenogenetic species alone. Isolation by distance analyses (IBD) were performed on the *D. raddei* and *D. valentini* sexual taxa, which were sampled over a wide area, using the ADEGENET package in R (Jombart et al., 2010)(Jombart, Devillard, and Balloux 2010) (Figures S3.3 and S3.4).

Population structure and the ancestry of the parthenogens were further investigated using the Bayesian multilocus clustering analysis implemented in STRUCTURE v2.3.4 (Hubisz et al., 2009; Pritchard et al., 2000) (more details on the runs can be found in the SI). The sexual species were used as “learning samples” (PopFlag = 1) (Murgia et al., 2006) to define the cluster membership when diploid parthenogenetic individuals were included in the dataset. Given their hybrid origin, parthenogens are expected to have half of their ancestry from the maternal species and the other half from the paternal species, with little variation given their clonal reproduction. Using the same number of clusters determined by the sexual dataset in DAPC, the ancestry of the asexual individuals from sympatric and allopatric localities (and two diploid individuals of uncertain status) was determined (PopFlag = 0) (Pritchard et al., 2000). Some level of misclassification was allowed with the MIGRPRIOR set at 0.01.

Phylogenetic inference

For phylogenetic analyses one partial mitochondrial gene, NADH dehydrogenase-4 (ND4), was selected. Sequenced individuals are identified in Table S1 and sequences from (Freitas et al., 2016) were added to the analyses. Primers and PCR protocols are described elsewhere (Freitas et al., 2016). Chromatograms were edited by eye in ChromasPro v1.7.4 (Technelysium) and sequence alignment was performed in MAFFT v6 (Katoh and Standley, 2013) using the automatic settings for the algorithm of choice. The selection of the DNA substitution model and MrBayes run parameters were described elsewhere (Freitas et al., 2016).

RESULTS

Analysis of the complete dataset

All analyses were performed with 12 markers in total, nine of which were discriminant for the polyploids. The remaining three markers never presented more than two alleles in any of the polyploids, despite their high diversity and heterozygosity in the parthenogenetic females. Low levels of uncertainty were found for the 12 selected markers (less than 1% disagreement between repeats) and uncertain genotypes were eliminated from further analyses. Polyploid individuals were found only in three of the sympatric localities (several individuals in Kuchak and Sotk, one in Lchaschen but none in Lchap) and classified as hybrids between parthenogens and sexuals (hereafter 'PS hybrids'). A few individuals presented three alleles for only one marker (pink coloured in Table S3.1: four *D. armeniaca*, one *D. mixta*, two *D. raddei*, two *D. unisexualis* and four *D. uzzelli*). These individuals were kept in all downstream analyses, but the specific markers were considered genotyping errors and assigned as missing data.

Sexual and parthenogenetic species analyses

Ploidy determination and an exploratory DAPC analysis on the diploid individuals identified major clusters corresponding to the recognized species, but also indicated some putative misidentification (three *D. valentini* females had been morphologically identified as hybrids, one PS hybrid *D. unisexualis x D. valentini* had been identified as *D. valentini,* and one *D. armeniaca* identified as *D. valentini* male, which was an obvious labelling error). These individuals were kept in the analysis and reclassified according to their genetic determination, along with the diploids (from sympatric localities) that had been identified as PS hybrids. After re-classification, revised DAPC analyses were performed on the sexual data set and on the full diploid dataset including both the sexual and the asexual species (Figure S3.1).

After BIC comparisons, Delta *K* and the rate of change of the log probability of the data between successive *K* values, *K* = 7 was selected for the sexual species dataset. This level of clustering resulted in consistent groups both in the DAPC and STRUCTURE analyses, and it separated the described species but also distinguished groups within species that may be relevant to determining the parentage of parthenogens. When adding the three parthenogenetic species, individuals were consistently divided into four clusters

in DAPC (two clusters for *D. armeniaca*), so a *K* value of 11 was used when analysing the total diploid dataset, sexual and parthenogenetic species together.

F_{ST} distances (Figure S3.2; Table 3.1) grouped diploid species into three main clusters, the *D. raddei* group (*D. raddei, D. uzzelli* and *D. unisexualis*), the *D. valentini* group (*D. valentini, D. mixta*-2) and the *D. rudis* group (*D. rudis, D. mixta* and *D. armeniaca*). Note that the two groups within *D. mixta* fell into different clusters (F_{ST} = 0.25) whereas other within-species groups were much less distinct.

	D. mixta2	*D. mixta*	*D. raddei –1*	*D. raddei –2*	*D. valentini –1*	*D. valentini –2*	*D. rudis*
D. mixta2	0.00	0.25	0.11	0.19	0.08	0.16	0.14
D. mixta	0.25	0.00	0.14	0.23	0.14	0.23	0.15
D. raddei –1	0.11	0.14	0.00	0.07	0.19	0.15	0.13
D. raddei –2	0.19	0.23	0.07	0.00	0.25	0.23	0.20
D. valentini –1	0.08	0.14	0.19	0.25	0.00	0.06	0.12
D. valentini –2	0.16	0.23	0.15	0.23	0.06	0.00	0.13
D. rudis	0.14	0.15	0.13	0.20	0.12	0.13	0.00

Table 3.1 - Pairwise F_{ST} distances calculated for all clusters of the diploid sexual species.

Asexual parentage inferences

Parthenogens (*D. unisexualis*, *D. uzzelli* and *D. armeniaca*) had higher observed heterozygosity than the expected heterozygosity, as predicted due to their hybrid origin (Figure 3.2). All parthenogenetic species had private alleles not present in the sexual taxa, and sometimes in high frequencies (e.g. allele 258 represents 53% of the diversity of C113 marker in *D. uzzelli;* Table S3.3).

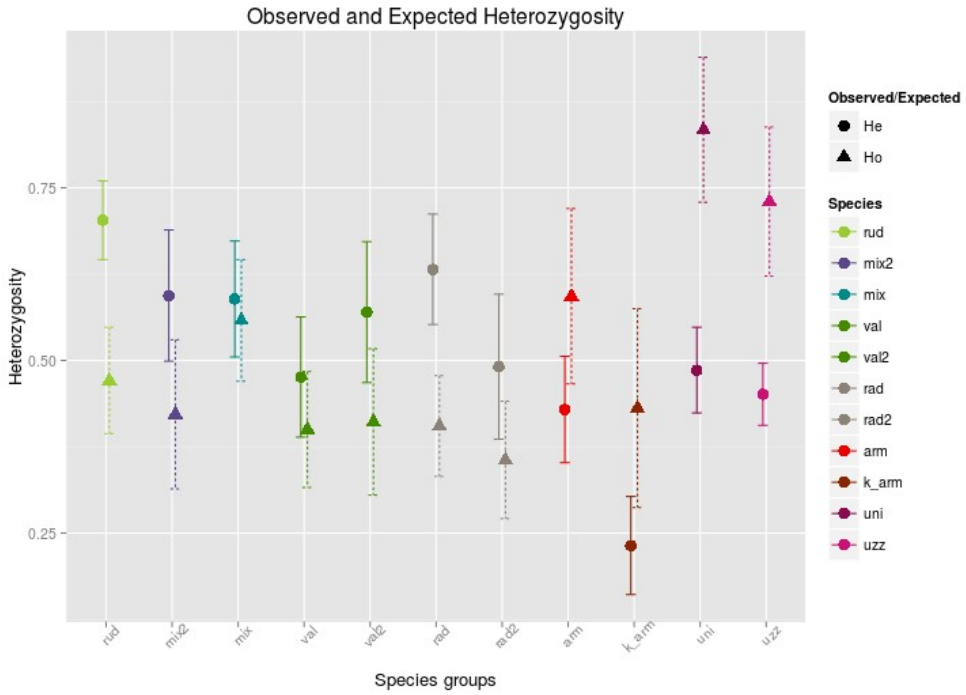


Figure 3.2 - Expected and observed heterozigosity for all diploid species. Codes are as follows: rud – *D. rudis*, mix2 – *D. mixta*2, mix – *D. mixta*, val – *D. valentini*1, val2 – *D. valentini*2, rad – *D. raddei*1, rad2 – *D. raddei*2, arm – *D. armeniaca*, uni – *D. unisexualis*, uzz – *D. uzzelli*. Parthenogens are arm, uni and uzz.

STRUCTURE analysis trained on the sexual species clusters (*K* = 7) and with the parthenogenetic species assigned as admixed, clearly shows a shared ancestry for each of the parthenogens (Figure 3.3). *D. armeniaca* shared half of its genomic composition with *D. mixta* (specifically *D. mixta-1*) but the other half was not exclusively from *D. valentini* as expected, but rather from both *D. valentini* and *D. rudis* (predominantly from one cluster in each case). Similarly, *D. uzzelli* shared half of its ancestry with *D. raddei* (both clusters) and the other half with a mix of backgrounds rather than *D. valentini* alone. *D. unisexualis* ancestry also shared half of its genes with *D. raddei* (both clusters but more strongly from cluster 2) and the other half from a mix of sources, similar but not identical to the composition of *D. uzzelli*. Interestingly, the parthenogens presented a clear 50% contribution from the putative maternal species, causing the parthenogenetic species to group with their maternal parents in the neighbor-joining tree (Figure S3.2).

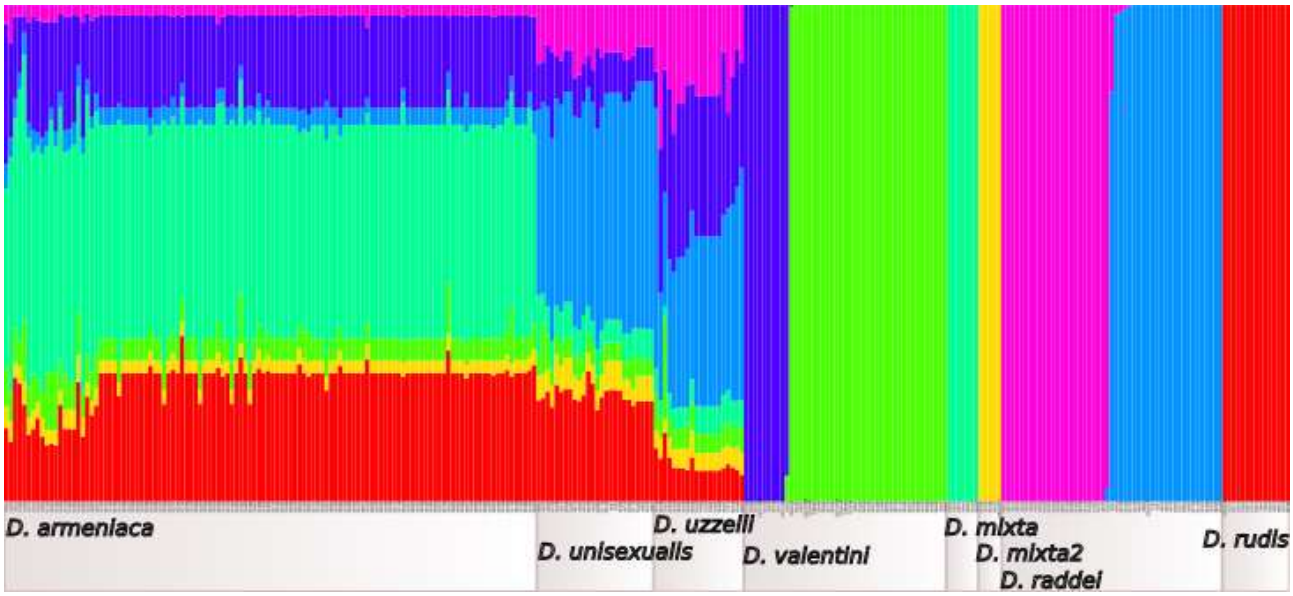


Figure 3.3 - Bayesian clustering results for the sexual (*D. raddei, D. mixta, D. valentini* and *D. rudis*) and asexual (*D. armeniaca, D. unisexualis* and *D. uzzelli*) species for *K* =7 and with set populations parameter. *D. raddei* and *D. valentini* were divided into two groups as in Table 1. *D. mixta* is also distinguished from *D. mixta* – 2.

Bayesian inference on the mtDNA marker shows matrilineal relationships of the parthenogenetic hybrids with their sexual ancestors (Figure 3.4). The parthenogens analysed here coalesced to one single lineage per maternal species, *D. unisexualis* and *D. uzzelli* descended from *D. raddei*, and *D. armeniaca* from *D. mixta-1*, even though previous studies have shown that several lineages within *D. raddei* contributed to asexual origins when other parthenogens were included (Freitas et al., 2016).

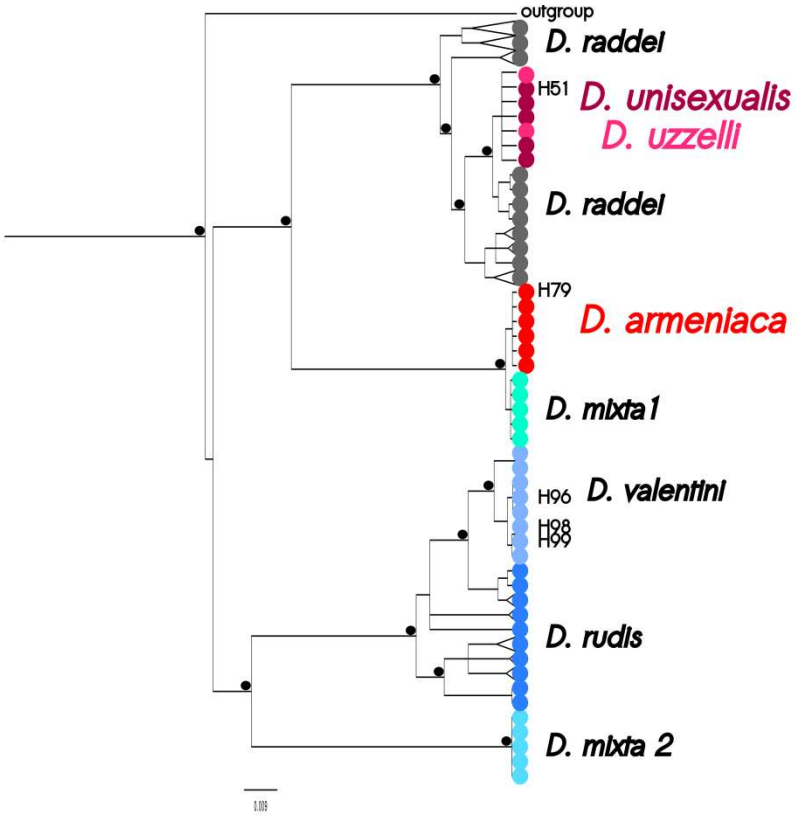


Figure 3.4 – Bayesian inference of the phylogenetic relationships among *Darevskia sp*. based on ND4 partial sequences. Sexual species names are in black, parthenogenetic species in colour. Black dot whenever posterior probability is one.

Further DAPC analyses were performed to assess the clonal composition of the parthenogenetic species (Figure 3.5). *D. unisexualis* and *D. uzzelli* were considered together since they share mtDNA. When plotting the densities of individuals on a single discriminant function, a parthenogenetic species with a single origin is expected to display a unimodal distribution whose variance increases with lineage age (and may become fragmented due to extinction of lineages). On the other hand, more than one origin would generate a multimodal distribution dependent on sampling of genotypes from the parental populations. *D. armeniaca* was the species with the largest interval of discriminant factor (DF) scores among the three parthenogens, consistent with the two groups in the previous analysis. Its distribution was bimodal, with one peak markedly wider than the other. *D. uzzelli* presented one peak only, clearly distinct from *D. unisexualis*, smaller peaks

corresponding to two samples containing a high proportion of missing data. Finally, *D. unisexualis* also presented two peaks, although less separated than those of *D. armeniaca*.

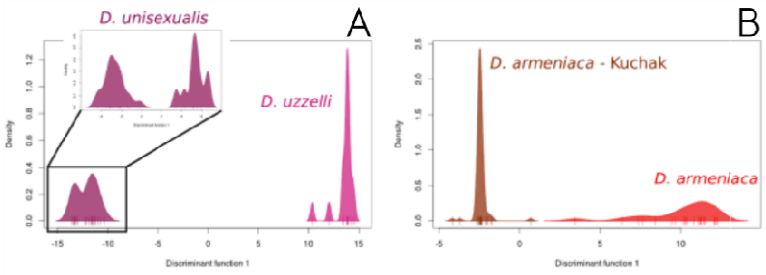


Figure 3.5 – Distribution of the Discriminant Factor (DF) values (from DAPC) calculated for each parthenogenetic species.

Sympatric localities

Individuals found in the sympatric localities were analysed for their ploidy and parentage, testing the outcomes of the hybridization between parthenogenetic females and sexual males. Polyploid individuals were found in the sympatric localities Kuchak and Sotk, and one in Lchap. In Kuchak, individuals were collected randomly, avoiding bias towards putative PS hybrids. Following ploidy genetic assignment, 17% (27/160) of the Kuchak individuals were polyploids, 3% (5/160) *D. armeniaca x D. valentini* and 14% (22/160) *D. unisexualis x D. valentini.* Diploid individuals from Kuchak corresponded to the parthenogenetic *D. unisexualis* and *D. armeniaca* and the sexual *D. valentini.* One tetraploid was found in Kuchak and another in Sotk, both PS hybrids *D. armeniaca x D. valentini* (IDs 12176 and 9910, respectively: Table S3.1). In Sotk, 11 individuals were *D. armeniaca x D. valentini* triploids and five *D. valentini* sexual diploids.

In the PCoA analysis for Kuchak and Sotk (Figure 3.6a), a total of 277 individuals were included: 160 from Kuchak and 25 from Sotk, the remainder from the general distribution range. Only *D. valentini* and *D. armeniaca x D. valentini* hybrids were available from Sotk, where no individual morphologically identified as *D. armeniaca* was sampled. *D. valentini* formed a different cluster from each of the two parthenogens, *D. armeniaca* and *D.*

unisexualis, and the three clusters were approximately equidistant. Diploid individuals from Kuchak (*D. valentini*, *D. armeniaca* and *D. unisexualis*) and Sotk (*D. valentini*) mostly overlapped with the general distribution of the same species. The spread of the total variation held by the sexual species was greater than for the parthenogens, as expected.

Triploid *D. armeniaca x D. valentini* PS hybrids from Kuchak and Sotk generally fell between their proposed maternal (the parthenogen *D. armeniaca*) and paternal species (the sexual *D. valentini*). As expected, these clusters were closer to the parthenogenetic parent, which contributed two out of three alleles. Triploid putative *D. unisexualis x D. valentini* PS hybrids from Kuchak behaved similarly, clustering between their proposed maternal (the parthenogen *D. unisexualis*) and paternal species (the sexual *D. valentini*). Lynch distances showed a similar arrangement of clusters and distances between species and hybrids (data not shown). Both tetraploid individuals fell among the remaining PS hybrids of the same cross.

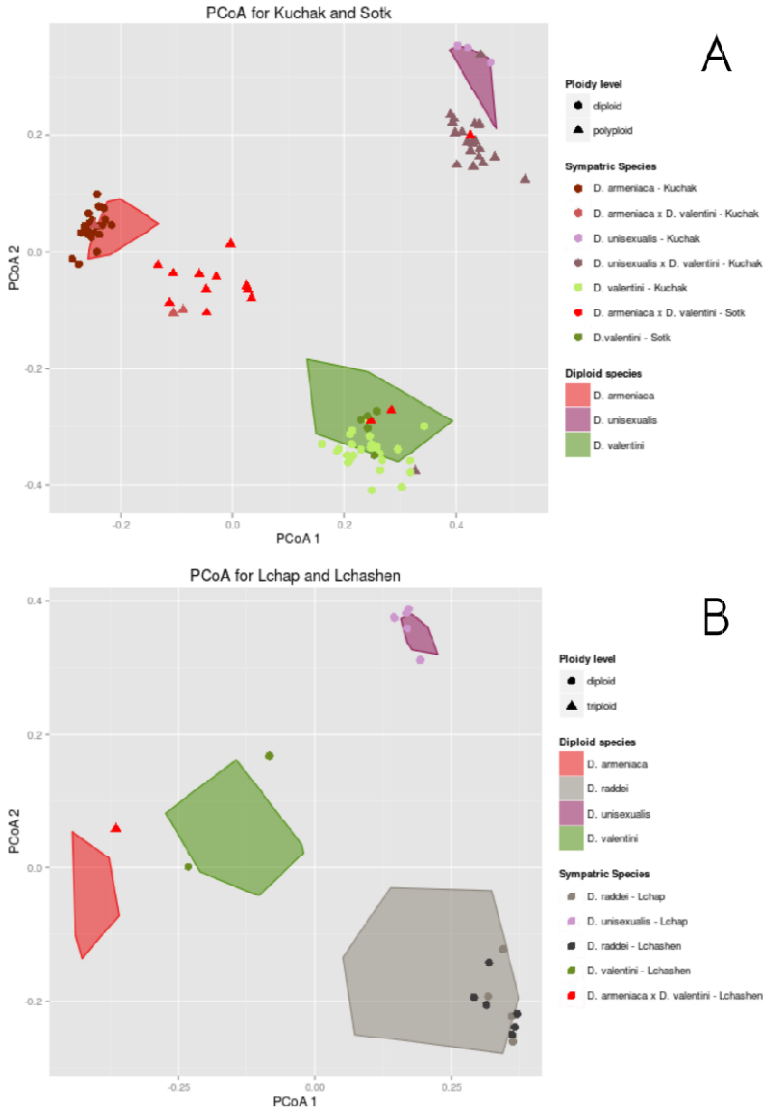


Figure 3.6 - (previous page) Principal Coordinate analysis of the Bruvo distances between the individuals from the sympatric localities including polyploid hybrids. Polyploid hybrids are identified as triangles and diploids as full circles. In this analysis, diploid individuals from non-sympatric localities are represented in the convex hulls, one for each species: *D. armeniaca, D. unisexualis* and *D. valentini*. Sympatric localities Sotk and Kuchak are in panel A, and Lchap and Lchaschen in panel B.

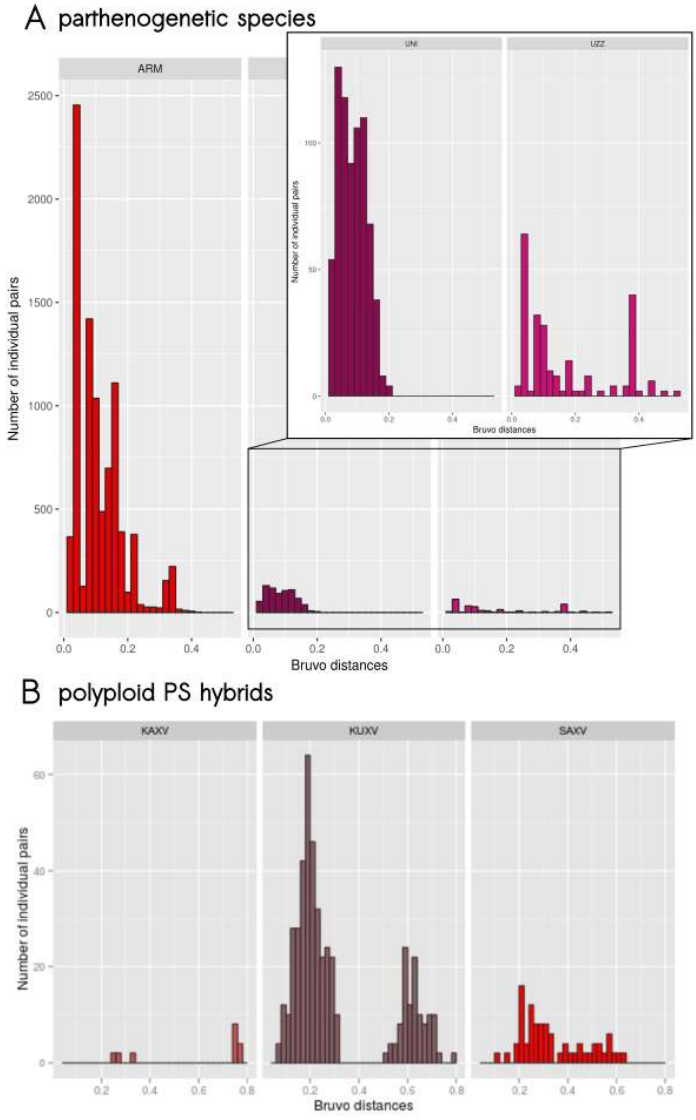


Figure 3.7 – Bruvo pairwise distances among each of the parthenogenetic species (all localities included) and among each of the polyploid backcrosses types.

Bruvo pairwise genetic distances among the *D. unisexualis* x *D. valentini* polyploid PS hybrids were distributed in a much larger interval than for the *D. armeniaca* x *D. valentini* polyploid PS hybrids (Figure 3.7). Also, the pairwise distance interval was much larger for the polyploids than for each of the diploid parthenogens (which ranged between 0.0 – 0.4, Figure 3.7), and a pairwise distance of naught was never found among the polyploids.

Triploid PS hybrids are expected to present only alleles that are shared with both putative parental species. Because their parthenogenetic maternal species are also of hybrid origin, *D. armeniaca* and *D. unisexualis*, these share alleles with their sexual parental species (*D. valentini*). Therefore, we found that in the PS hybrids, one allele was unique to *D. valentini* and the other (one or) two either unique to the parthenogen (*D. unisexualis* or *D. armeniaca)* or shared by the parthenogen and *D. valentini*. The allelic combination for each marker was normally such that the source of each allele could easily be determined.

If the PS hybrids are fertile and capable of backcrossing and producing viable offspring, we will find polyploids with a different distribution of alleles, specifically markers with alleles exclusive of only one of the species (and not shared by both), and eventually mtDNA haplotypes shared with *D. valentini* (if a male PS hybrid backcrosses with a female *D. valentini*). In our dataset, we found two triploid *D. unisexualis x D. valentini* PS hybrids which presented more than 50% of their markers with *D. valentini* (ID 12454: six out of nine) or *D. unisexualis* (ID 12572: 8 out of 12) unique alleles (clustering in the PCoA with the *D. valentini* and *D. unisexualis* clusters, respectively (Figure 3.6a)). Regarding the mtDNA, the first (ID 12454) shared its haplotype with *D. valentini* (H96 in Figure 3.4) and the second (ID 12572) with *D. unisexualis* (H51 in Figure 3.4). The same applies to the *D. armeniaca* x *D. valentini* PS hybrids: two (10048 and 10050) had five in 12 markers unique to *D. valentini* and another (9875) had five in 12 markers unique to *D. armeniaca* (clustering in the PCoA with the *D. valentini* and *D. armeniaca* clusters, respectively (Figure 3.6a)). Again, the first two (individuals 10048 and 10050) shared their mtDNA haplotype with *D. valentini* (H98 and H99 in Figure 3.4) and the later (9875) with *D. armeniaca* (H79 in Figure 3.4). Some triploid PS hybrids had unique alleles, that were absent not only from the diploids of the sympatric localities, but also from all other species (Table S3.3).

For the PCoA with the Lchap and Lchashen (Figure 3.6b), only 19 individuals in total were

available. All *D. raddei* individuals from these localities clustered with the *D. raddei* from elsewhere in its range, and separately from the *D. unisexualis* cluster found in Lchap. *D. valentini* individuals fell close to, but not within, the cluster formed by *D. valentini* individuals from the whole range, and unexpectedly not very close together. One PS hybrid *D. armeniaca x D. valentini* fell within the cluster formed by the PS hybrids *D. armeniaca x D. valentini* from Sotk when analysed together (data not shown).

Diploids from the sympatric localities were included in the STRUCTURE analysis and the genomic contribution of these individuals was concordant with the same species from other non-sympatric localities.

DISCUSSION

In nature, asexuality is maintained by the balance between the generation of new asexual lineages from sexual progenitors and their loss by extinction. Vertebrate parthenogenesis is found only in a reduced number of species (Beukeboom and Vrijenhoek, 1998) and they have mostly a hybrid origin (Avise, 2008). However, how these hybridization events take place, and the role and extent of hybridization in the sexual-asexual interactions is still not clear, as is the extent to which hybridization contributes to gene flow or to new parthenogenetic lineages throughout time.

The origin of hybrid vertebrate parthenogenesis has been tentatively explained by two hypotheses, the Balance Hypothesis and the Phylogenetic Constraint Hypothesis. If parthenogenetic *Darevskia* follow the Balance Hypothesis (Moritz et al., 1989), then their parental species pairwise distances will fall within a short range interval, different from the other sexual species pairwise distances (Kearney et al., 2009). On the other hand, if the second hypothesis better explains the origin of parthenogenetic lineages in *Darevskia*, then the pairwise distances of the parental species pairs will be more variable, and other species pairs within the same distance range and potential to hybridize will not produce any parthenogenetic lineage. They will also maintain the directionality in the parentage of the *Darevskia* parthenogens. If sexual species are capable of originating parthenogenetic lineages due to lineage dependent genetic peculiarities, than it is likely that those peculiarities are maintained and parthenogenetic females could still backcross and produce viable offspring with the sexual males of their parental species.

(i) Are all sexual parentals restricted to specific phylogenetic clades and/or do they share the same genetic distance between them?

Sexual diploid individuals analysed here were distinguished into different groups, concordant with the initial ascription based on phenotype, by all analyses performed. Nevertheless, we find some phylogenetic incongruence in the morphological identification of one species, specifically the *D. mixta* individuals. These were consistently divided into two different groups, one containing Georgia samples (*D. mixta-1*) and the other samples from Turkey *(D. mixta-2*). Pairwise F_{ST} between the clusters was 0.25, similar to the distance between other sexual species pairs. This is concordant with the division of *D. mixta* into two species already suggested by (Gabelaia et al., 2015) and based on mtDNA alone.

The reassessment of the parentage inference confirmed the hybrid genetic profile of all parthenogenetic species studied here. In the STRUCTURE analyses, the maternal contribution (confirmed by mtDNA BI analysis) for each parthenogen is clear (either *D. raddei* or *D. mixta*), and accounts for at least 50% of the nuclear genome of the parthenogenetic species. On the other hand, the paternal contribution is less clear, with different species contributions varying in the parthenogens. The STRUCTURE analysis shows recent admixture using the current genomic constitutions of the putative parental populations. However, the hybridization event that originated the parthenogens analysed here took place in the past, ~100 kyrs ago (Freitas et al., 2016), when the genetic constitutions of the parental species, and populations within species, were different and their divergence possibly incomplete. While *D. raddei* and *D. mixta* diverged earlier in the past, the split between *D. rudis* and *D. valentini* seems to be more recent (Freitas et al, in prep). Thus, the paternal individual that was part of that hybridization event was likely an ancestor state of the *D. rudis* / *D. valentini* clade, before the speciation process was completed, and shared alleles with extant populations that were not directly involved in the hybridization events. Another possibility for the mixed paternal ancestry could be that the true parental population was not sampled (and could even be extinct), however this is not as likely since the maternal ancestry was correctly identified, even when a reduced number of localities (individuals) was used (e.g. *D. mixta*).

The STRUCTURE analysis confirmed that *D. armeniaca* resulted from the hybridization between *D. mixta* (from Georgia) and *D. valentini,* while *D. uzzelli* resulted from the hybridization between *D. raddei* and *D. valentini. D. unisexualis* most likely resulted from

the multiple hybridization between *D. raddei* and *D. rudis* or *D. valentini* (Figure 3.3).

If the origin of parthenogenetic *Darevskia* is restricted to the hybridization between sexual species at a specific divergence as postulated by the Balance Hypothesis, then other sexual lineages at least as divergent as the parents of known parthenogens also have had the potential to generate asexual lineages. Moreover, if the distance between sexual parentals is more relevant to the origin of parthenogenetic lineages than which species pairs are involved, then these distances are expected to fall within a narrow interval. F_{ST} pairwise distances calculated with this set of markers can only give us a rough idea of the overall genomic divergence in the present, and may not accurately represent distances at the time of origin of the parthenogens. Nevertheless, they tell us that parthenogens originated at some point between zero divergence and the current, which considering the young age of the parthenogenetic lineages and the reduced time scale sexual parentals had to diverge, should not be very different than the distance at the time of the hybridization event. Other species pairs have distances within the same interval than the sexual parentals, and overlap geographically (e.g. *D. mixta* / *D. rudis* in Western Caucasus), but no parthenogenetic hybrids have been reported. Besides, the directionality of the hybridization events were consistent, with *D. mixta* and *D. raddei* always acting as maternal species, even though the distance between these species is similar to the parental pairs, and their distribution ranges adjoin in the Western Caucasus. In relation to other species, despite no inference can be made about their phylogenetic distance, the remaining 23 sexual species in the genus have produced no parthenogenetic hybrid species (Murphy, 2000). This considered, it is questionable whether phylogenetic distances between sexual parentals is an important factor at all.

The Balance Hypothesis remains the most widely accepted explanation for the origin of vertebrate parthenogenesis (Avise, 2008). However, evidence from *Darevskia* and other hybrid parthenogenetic vertebrates favours the Phylogenetic Constraint Hypothesis. For instance, the parental species for all parthenogens in the genus *Leiolepis* (butterfly lizard) are always the same and belong to two different phylogenetic clades within the genus (Grismer et al., 2014). In other cases, such as fish genus *Cobitis,* directional hybridizations are not always responsible for parthenogenetic hybrids, but parthenogenetic lineages may result from the cross between both distant and closely related species (Janko et al., 2003). Overall, our results and these studies demonstrate that only a few sexual species contributed to the hybridization events that originated parthenogenetic vertebrates, and

that the distance between parental species is not as relevant as the specificity of the sexual ancestors, and in some cases as the directionality of the hybridization events. These factors are in accordance with the propositions of the Phylogenetic Constraint Hypothesis.

(ii) Do parthenogenetic lineages from the same parental taxa originate repeatedly through time or is there a single temporal event in each case?

In localities where the sexual parentals are currently in sympatry (like in the the sympatric locality included here, Lchashen, where the parental species for *D. uzzelli, D. raddei* and *D. valentini,* are found), no evidence of new parthenogenetic lineages (or any parthenogenetic individual) was found. Instead, the only hybrid identified there was a PS triploid hybrid between the parthenogen *D. armeniaca* and the sexual *D. valentini*, which likely resulted from a recent expansion and range overlap of this parthenogen with the sexual *D. valentini*. So, even though some sexual parental pairs can be currently found together, no evidence that they are originating new parthenogenetic hybrids at the present was found. Instead, the origin of parthenogenetic *Darevskia* was likely associated with a temporal event in the past, specifically the last glaciations, as has been previously suggested based on phylogeographic patterns and ecological models (Freitas et al., 2016).

Nuclear and mtDNA analyses performed here on the parthenogenetic individuals showed that even if the hybridization events which originated the parthenogenetic species happened in the past, they likely have happened multiple times. Despite many parthenogenetic individuals sharing the same mtDNA ancestry, multimodal DF distributions of the nuclear markers support the existence of at least two hybridization events for *D. armeniaca*, and possibly for *D. unisexualis*.

(iii) Can parthenogens still backcross with their sexual parentals and originate fertile individuals?

When hybridizing with their sexual ancestors, parthenogenetic x sexual hybrids frequently form new asexual reticulate clades with polyploidization induced parthenogenesis (e.g. whiptail lizards *Aspidoscelis* sp.; (Cole et al., 2014; Taylor et al., 2015)). *Darevskia* PS hybrids pairwise distance distribution (calculated separately for *D. unisexualis x D. valentini* and *D. armeniaca x D. valentini* PS hybrids) (Figure 3.7) was much wider than the

distribution of any of the parthenogenetic diploids (0.0 – 0.4, Figure 3.7). Besides, we did not recover any PS hybrid pair with the same combination of genotypes for all markers, as we did in all populations of parthenogenetic diploid species analysed here. Therefore, in contrast with *Aspidoscelis*, the possibility of a triploid parthenogenetic lineage in *Darevskia* is not supported, and the backcross hybridization found at present is likely a temporary event that will end when the sexual species goes extinct: if sexual males and females backcross with parthenogenetic females and triploid males, respectively, then less sexual diploid offspring is produced each generation.

When organisms evolve to be asexual, sexual reproductive traits will likely be lost. In *Darevskia*, the parthenogenetic lineages have been proposed to be quite recent and given the generation of triploids, they still present all the sexual machinery necessary to mate and produce fertilized zygotes with a paternal contribution. However, it is also expected that older lineages will have less possibility of forming viable hybrids with the sexual males due to mutation accumulation and/or Dobzhansky-Muller incompatibilities. In Kuchak, a higher ratio of *D. unisexualis x D. valentini* PS hybrids relative to *D. armeniaca x D. valentini* was found, concordant with previous findings (4:1, (Danielyan et al., 2008)). However, in Sotk the proportion of PS hybrids was so high that no diploid parthenogens were collected. Given the reproductive pressure inferred from the intensity of copulation marks of *D. valentini* males on each parthenogenetic species present in Kuchak is the same (Carretero et al, in press), one possible explanation for the high proportion of PS hybrids found in Sotk, and low proportion found in Kuchak in relation to the other PS hybrid, is that the *D. armeniaca* lineage present in Kuchak is older and has lost part of the sexual reproduction machinery. For asexually reproducing organisms, sex associated machinery becomes useless and can even potentially encourage harassment by males, or other sex related liabilities (Shine et al., 2004; Sztatecsny et al., 2006). Consequently, in the long-term parthenogenetic females are expected to suffer high selection to lose these characters. If one *D. armeniaca* lineage is in fact older than the other, it had more time for selection to act upon it and lose sex related traits. This is concordant with the multimodal DF distribution for *D. armeniaca,* suggesting that it had more than one origin through hybridization events separated in time and/or space.

As triploids are likely to result from the hybridization between parthenogenetic diploid females and males of the sexual species, the tetraploid found is proposed to result from the cross between a “fertile” triploid and another individual. Various lines of evidence

support the sexual fertility of triploids, even if at a very reduced level and only infrequently: the presence of at least some triploid males with apparently fertile reproductive organs, the maintenance of alleles that are only present in the triploids and absent from the parentals in the sympatric localities, the varying degree of contribution from *D. valentini* and parthenogenetic alleles in some PS hybrids, and PS hybrids with mitochondrial DNA shared with the sexual species *D. valentini.* This considered, triploid hybrids are the result of the hybridization a parthenogenetic female and a sexual male, or between a triploid and a diploid individual from any of the diploid species (sexual *D. valentini,* or parthenogen *D. unisexualis* and *D. armeniaca*). Previous studies on the reproductive organs of PS hybrids from the populations studied here reported individuals with undeveloped reproductive organs, that were therefore deemed sterile (Danielyan et al., 2008). However, some of the triploids did show apparently normal reproductive organs or evidence of having already laid eggs (such as triploid female ID 10034 from Sotk).

Evolutionary consequences of hybridization and polyploidy in *Darevskia*

This study identified a considerable number of polyploid hybrids (17%) in sympatric locations of parthenogenetic species with their paternal sexual ancestors. Asexuality in vertebrates has been presented as a possible stage in the speciation continuum and an effective barrier to gene flow, when other forms of pre- and post-reproductive isolation mechanisms are absent (Janko et al., 2016). Given the high number of known hybrids (either diploid or polyploid) between *Darevskia* species, gene flow was expected. However, despite the high hybridization and backcrossing rates there was no clear evidence of recent gene flow, reinforcing the theory of asexuality as a reproductive isolation mechanism.

Sympatric populations such as Kuchak and Sotk are likely to be the scenario of a sporadic event, where it is possible to witness the reproductive isolation between the parthenogenetic species and their parental species. Observations from these localities could be used to determine the potential of the parthenogenetic species in occupying new habitats, in particular when in competition with the sexual individuals. It has been shown that parthenogenetic *Darevskia* species may outcompete their sexual ancestors (Tarkhnishvili et al., 2010), or even solely occupy habitats suitable for both themselves and their parentals (Freitas et al., 2016). When in contact, hybridizing with sexual males would mean a lower number of sexual viable offspring each generation. This way, sexual

populations could reach levels that would most likely lead to extinction, while parthenogens with the presence of only one individual could still reproduce and maintain the population. Together with the double reproductive output of parthenogens compared to sexuals, hybridization between both could contribute for the apparent out-performance of some parthenogenetic species in relation to their sexual ancestors in nature.

Finally, this model of vertebrate parthenogenesis questions whether the Balance Hypothesis suggested as a general theory on the origin of hybrid asexuality is applicable in this case. Parthenogenesis in vertebrates is rare and generally originates from the hybridization between specific species pairs with highly variable phylogenetic distances. This fits the proposal that the parthenogenetic parentals must hold some lineage-dependent specificity that allows them, when hybridizing, to originate a hybrid capable of reproducing asexually, as the Phylogenetic Constraint Hypothesis states. Identifying which are the constraining factors underlying the origin of parthenogenetic vertebrates gives us a better chance at understanding how these hybrids use the sexual reproduction machinery to reproduce asexually, escaping the limitations sex might bring. This study identifies major factors that are relevant to the origin of parthenogenesis in *Darevskia,* as in other vertebrate genus, but also reinforces the proposal that vertebrate asexuality acts as a strong reproductive barrier, as already described for other systems.

REFERENCES

Arakelyan, M., Danielyan, F., 2011. Syntopic populations of Darevskia biparental species in Armenia and Nagrono-Karabakh.

Avise, J.C., 2008. Clonality: the genetics, ecology, and evolution of sexual abstinence in vertebrate animals. Oxford University Press, Oxford ; New York.

Beck, J.B., Windham, M.D., Pryer, K.M., 2011. Do Asexual Polyploid Lineages Lead Short Evolutionary Lives? A Case Study from the Fern Genus Astrolepis. Evolution 65, 3217–3229. doi:10.1111/j.1558-5646.2011.01362.x

Bell, G., 1982. The masterpiece of nature: The evolution and genetics of sexuality, University of California Press. ed. Berkeley.

Beukeboom, L.W., Vrijenhoek, R.C., 1998. Evolutionary genetics and ecology of sperm-dependent parthenogenesis. J. Evol. Biol. 11, 755–782. doi:10.1046/j.1420-9101.1998.11060755.x

Birky, C.W., Barraclough, T.G., 2009. Asexual Speciation, in: Schön, I., Martens, K., Dijk, P. (Eds.), Lost Sex: The Evolutionary Biology of Parthenogenesis. Springer Netherlands, Dordrecht, pp. 201–216. doi:10.1007/978-90-481-2770-2_10

Boudjemadi, K., Martin, O., Simon, J.-C., Estoup, A., 1999. Development and cross-species comparison of microsatellite markers in two lizard species, Lacerta vivipara and Podarcis muralis. Mol. Ecol. Notes 8, 513–525.

Bruvo, R., Michiels, N.K., D'souza, T.G., Schulenburg, H., 2004. A simple method for the calculation of microsatellite genotype distances irrespective of ploidy level. Mol. Ecol. 13, 2101–2106. doi:10.1111/j.1365-294X.2004.02209.x

Bullini, L., 1994. Origin and evolution of animal hybrid species. Trends Ecol. Evol. 9, 422–426. doi:10.1016/0169-5347(94)90124-4

Choleva, L., Janko, K., De Gelas, K., Bohlen, J., Šlechtová, V., Rábová, M., Ráb, P., 2012. Synthesis of clonality and polyploidy in vertebrate animals by hybridization between two sexual species: synthesis of clonality and polyploidy in vertebrate animals. Evolution 66, 2191–2203. doi:10.1111/j.1558-5646.2012.01589.x

Clark, L.V., Jasieniuk, M., 2011. polysat: an R package for polyploid microsatellite analysis. Mol. Ecol. Resour. 11, 562–566. doi:10.1111/j.1755-0998.2011.02985.x

Cole, C.J., Taylor, H.L., Baumann, D.P., Baumann, P., 2014. Neaves' Whiptail Lizard: The First Known Tetraploid Parthenogenetic Tetrapod (Reptilia: Squamata: Teiidae). Breviora 1–20. doi:10.3099/MCZ17.1

Coyne, J.A., Orr, H.A., Futuyma, D.J., 1988. Do We Need a New Species Concept? Syst. Zool. 37, 190–200. doi:10.2307/2992276

Danielyan, F., Arakelyan, M., Stepanyan, I., 2008. Hybrids of Darevskia valentini, D. armeniaca and D. unisexualis from a sympatric population in Armenia. Amphib.-Reptil. 29, 487–504.

Darevsky, I., 1967. Rock lizards of the Caucasus: systematics, ecology and phylogenesis of the polymorphic groups of Caucasian rock lizards of the subgenus Archaeolacerta. Leningrad. Nauka, Leningrad.

Darevsky, I.S., Danielyan, F., 1968. Diploid and Triploid Progeny Arising from natural mating of parthenogenetic Lacerta armeniaca and L. unisexualis with Bisexual L. saxicola valentini. J. Herpetol. 2, 65. doi:10.2307/1563104

Freitas, S., Rocha, S., Campos, J., Ahmadzadeh, F., Corti, C., Sillero, N., Ilgaz, Ç., Kumlutaş, Y., Arakelyan, M., Harris, D.J., Carretero, M.A., 2016. Parthenogenesis through the ice ages: A biogeographic analysis of Caucasian rock lizards (genus Darevskia). Mol. Phylogenet. Evol. 102, 117–127. doi:10.1016/j.ympev.2016.05.035

Fu, J., MacCulloch, R.D., Murphy, R.W., Darevsky, I.S., Tuniyev, B.S., 2000a. Allozyme variation patterns and multiple hybridization origins: clonal variation among four sibling parthenogenetic Caucasian rock lizards. Genetica 108, 107–112.

Fu, J., Murphy, R.W., Darevsky, I.S., 1997. Toward the phylogeny of caucasian rock lizards: implications from mitochondrial DNA gene sequences (Reptilia: Lacertidae). Zool. J. Linn. Soc. 120, 463–477. doi:10.1111/j.1096-3642.1997.tb01283.x

Fu, J., Murphy, R.W., Darevsky, I.S., McEachran, J.D., 2000b. Divergence of the cytochrome b gene in the Lacerta raddei complex and its parthenogenetic daughter

species: evidence for recent multiple origins. Copeia 2000, 432–440.

Gabelaia, M., Murtskhvaladze, M., Tarkhnishvili, D., 2015. Phylogeography and morphological variation in a narrowly distributed Caucasian rock lizard, Darevskia mixta. Amphib.-Reptil. 36, 45–54. doi:10.1163/15685381-00002975

Goudet, J., 1995. FSTAT (Version 1.2): A Computer Program to Calculate F-Statistics. J. Hered. 86, 485–486.

Grismer, J.L., Bauer, A.M., Grismer, L.L., Thirakhupt, K., Aowphol, A., Oaks, J.R., Wood, P.L., Onn, C.K., Thy, N., Cota, M., Jackman, T., 2014. Multiple origins of parthenogenesis, and a revised species phylogeny for the Southeast Asian butterfly lizards, Leiolepis. Biol. J. Linn. Soc. n/a-n/a. doi:10.1111/bij.12367

Haigh, J., 1978. The accumulation of deleterious genes in a population—Muller's Ratchet. Theor. Popul. Biol. 14, 251–267. doi:10.1016/0040-5809(78)90027-8

Hubisz, M.J., Falush, D., Stephens, M., Pritchard, J.K., 2009. Inferring weak population structure with the assistance of sample group information. Mol. Ecol. Resour. 9, 1322–1332. doi:10.1111/j.1755-0998.2009.02591.x

Janko, K., 2014. Let Us Not Be Unfair to Asexuals: Their Ephemerality May Be Explained by Neutral Models Without Invoking Any Evolutionary Constraints of Asexuality. Evolution 68, 569–576. doi:10.1111/evo.12293

Janko, K., Kotlík, P., Ráb, P., 2003. Evolutionary history of asexual hybrid loaches (Cobitis: Teleostei) inferred from phylogenetic analysis of mitochondrial DNA variation. J. Evol. Biol. 16, 1280–1287. doi:10.1046/j.1420-9101.2003.00627.x

Janko, K., Pačes, J., Wilkinson-Herbots, H., Costa, R.J., Röslein, J., Drozd, P., Iakovenko, N., Rídl, J., Kočí, J., Reifová, R., Šlechtová, V., Choleva, L., 2016. Hybrid asexuality as a primary reproductive barrier: on the interconnection between asexuality and speciation (No. biorxiv;038299v2).

Jombart, T., Devillard, S., Balloux, F., 2010. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. BMC Genet. 11, 94.

Katoh, K., Standley, D.M., 2013. MAFFT Multiple Sequence Alignment Software Version 7:

Improvements in Performance and Usability. Mol. Biol. Evol. 30, 772–780. doi:10.1093/molbev/mst010

Kearney, M., Fujita, M.K., Ridenour, J., 2009. Lost Sex in the Reptiles: Constraints and Correlations, in: Schön, I., Martens, K., Dijk, P. (Eds.), Lost Sex. Springer Netherlands, pp. 447–474.

Korchagin, V.I., Badaeva, T.N., Tokarskaya, O.N., Martirosyan, I.A., Darevsky, I.S., Ryskov, A.P., 2007. Molecular characterization of allelic variants of (GATA)n microsatellite loci in parthenogenetic lizards Darevskia unisexualis (Lacertidae). Gene 392, 126–133. doi:10.1016/j.gene.2006.11.020

Lehtonen, J., Kokko, H., 2014. Sex. Curr. Biol. 24, R305–R306. doi:10.1016/j.cub.2014.01.060

Lively, C.M., 2010. A Review of Red Queen Models for the Persistence of Obligate Sexual Reproduction. J. Hered. 101, S13–S20. doi:10.1093/jhered/esq010

Luijckx, P., Ho, E.K.H., Gasim, M., Chen, S., Stanic, A., Yanchus, C., Kim, Y.S., Agrawal, A.F., 2017. Higher rates of sex evolve during adaptation to more complex environments. Proc. Natl. Acad. Sci. 201604072. doi:10.1073/pnas.1604072114

Lutes, A.A., Baumann, D.P., Neaves, W.B., Baumann, P., 2011. Laboratory synthesis of an independently reproducing vertebrate species. Proc. Natl. Acad. Sci. 108, 9910–9915. doi:10.1073/pnas.1102811108

Maynard Smith, J., 1978. The Evolution of Sex. Cambridge University Press, New York.

McDonald, M.J., Rice, D.P., Desai, M.M., 2016. Sex speeds adaptation by altering the dynamics of molecular evolution. Nature 531, 233–236. doi:10.1038/nature17143

Menken, S.B.J., Smit, E., Nijs, H. (J. . C.M.D., 1995. Genetical Population Structure in Plants: Gene Flow between Diploid Sexual and Triploid Asexual Dandelions (Taraxacum Section Ruderalia). Evolution 49, 1108. doi:10.2307/2410435

Moritz, C., Brown, W.M., Densmore, L.D., Vyas, S., Donnellan, S., Adams, M., Baverstock, P., 1989. Genetic diversity and the dynamics of hybrid parthenogenesis in Cnemidophorus (Teiidae) and Heteronotia (Gekkonidae), in: Evolution and Ecology of Unisexual

Vertebrates. Albany.

Murgia, C., Pritchard, J.K., Kim, S.Y., Fassati, A., Weiss, R.A., 2006. Clonal Origin and Evolution of a Transmissible Cancer. Cell 126, 477–487. doi:10.1016/j.cell.2006.05.051

Murphy, R., 2000. A fine line between sex and unisexuality: the phylogenetic constraints on parthenogenesis in lacertid lizards. Zool. J. Linn. Soc. 130, 527–549. doi:10.1006/zjls.1999.0241

Murphy, R.W., Darevsky, I.S., MacCulloch, R.D., Fu, J., Kupriyanova, L.A., 1996. Evolution of the bisexual species of Caucasian rock lizards: a phylogenetic evaluation of allozyme data. Russ. J. Herpetol. 3, 18–3.

Noonan, B.P., Pramuk, J.B., Bezy, R.L., Sinclair, E.A., Queiroz, K. de, Sites Jr., J.W., 2013. Phylogenetic relationships within the lizard clade Xantusiidae: Using trees and divergence times to address evolutionary questions at multiple levels. Mol. Phylogenet. Evol. 69, 109–122. doi:10.1016/j.ympev.2013.05.017

Otto, S.P., 2009. The Evolutionary Enigma of Sex. Am. Nat. 174, S1–S14. doi:10.1086/599084

Pinho, C., Sequeira, F., Godinho, R., Harris, D.J., Ferrand, N., 2004. Isolation and characterization of nine microsatellite loci in Podarcis bocagei (Squamata: Lacertidae). Mol. Ecol. Notes 4, 286–288. doi:10.1111/j.1471-8286.2004.00644.x

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

Reeder, T.W., Cole, C.J., Dessauer, H.C., 2002. Phylogenetic Relationships of Whiptail Lizards of the Genus Cnemidophorus (Squamata: Teiidae): A Test of Monophyly, Reevaluation of Karyotypic Evolution, and Review of Hybrid Origins. Am. Mus. Novit. 3365, 1–61. doi:10.1206/0003-0082(2002)365<0001:PROWLO>2.0.CO;2

Remón, N., Vila, M., GaláN, P., Naveira, H., 2008. Isolation and characterization of polymorphic microsatellite markers in Iberolacerta monticola, and cross-species amplification in Iberolacerta galani and Zootoca vivipara. Mol. Ecol. Resour. 8, 1351–1353. doi:10.1111/j.1755-0998.2008.02271.x

Rice, W.R., 1989. Analyzing Tables of Statistical Tests. Evolution 43, 223–225. doi:10.2307/2409177

Sambrook, J., Russell, D.W., 2001. Molecular Cloning: A Laboratory Manual. CSHL Press.

Schwander, T., Crespi, B.J., 2009. Multiple direct transitions from sexual reproduction to apomictic parthenogenesis in Timema stick insects. Evol. Int. J. Org. Evol. 63, 84–103. doi:10.1111/j.1558-5646.2008.00524.x

Shine, R., Phillips, B., Langkilde, T., Lutterschmidt, D.I., Waye, H., Mason, R.T., 2004. Mechanisms and consequences of sexual conflict in garter snakes (Thamnophis sirtalis, Colubridae). Behav. Ecol. 15, 654–660. doi:10.1093/beheco/arh058

Sztatecsny, M., Jehle, R., Burke, T., Hödl, W., 2006. Female polyandry under male harassment: the case of the common toad (Bufo bufo). J. Zool. 270, 517–522. doi:10.1111/j.1469-7998.2006.00120.x

Tarkhnishvili, D., Gavashelishvili, A., Avaliani, A., Murtskhvaladze, M., Mumladze, L., 2010. Unisexual rock lizard might be outcompeting its bisexual progenitors in the Caucasus. Biol. J. Linn. Soc. 101, 447–460. doi:10.1111/j.1095-8312.2010.01498.x

Taylor, H.L., Walker, J.M., Cole, C.J., Dessauer, H.C., 2015. Morphological Divergence and Genetic Variation in the Triploid Parthenogenetic Teiid Lizard, Aspidoscelis neotesselata. J. Herpetol. 49, 491–501. doi:10.1670/14-057

Trifonov, V.A., Paoletti, A., Barucchi, V.C., Kalinina, T., O'Brien, P.C.M., Ferguson-Smith, M.A., Giovannotti, M., 2015. Comparative Chromosome Painting and NOR Distribution Suggest a Complex Hybrid Origin of Triploid Lepidodactylus lugubris (Gekkonidae). PLOS ONE 10, e0132380. doi:10.1371/journal.pone.0132380

Van Oosterhout, C., Hutchinson, W.F., Wills, D.P.M., Shipley, P., 2004. micro-checker: software for identifying and correcting genotyping errors in microsatellite data. Mol. Ecol. Notes 4, 535–538. doi:10.1111/j.1471-8286.2004.00684.x

Weir, B.S., Cockerham, C.C., 1984. Estimating F-Statistics for the Analysis of Population Structure. Evolution 38, 1358–1370. doi:10.2307/2408641

Weismann, A., 1889. The significance of sexual reproduction in the theory of natural

selection., in: Essays upon Heredity and Kin- Dred Biological Problems. Clarendon, Oxford, pp. 251–332.

Wellenreuther, M., Runemark, A., Svensson, E.I., Hansson, B., 2009. Isolation and characterization of polymorphic microsatellite loci for the Skyros wall lizard Podarcis gaigeae (Squamata: Lacertidae). Mol. Ecol. Resour. 9, 1005–1008. doi:doi: 10.1111/j.1755-0998.2009.02550.x

1

CHAPTER 4

PHYLOGENOMICS IN NON-MODEL TAXA – paper IV

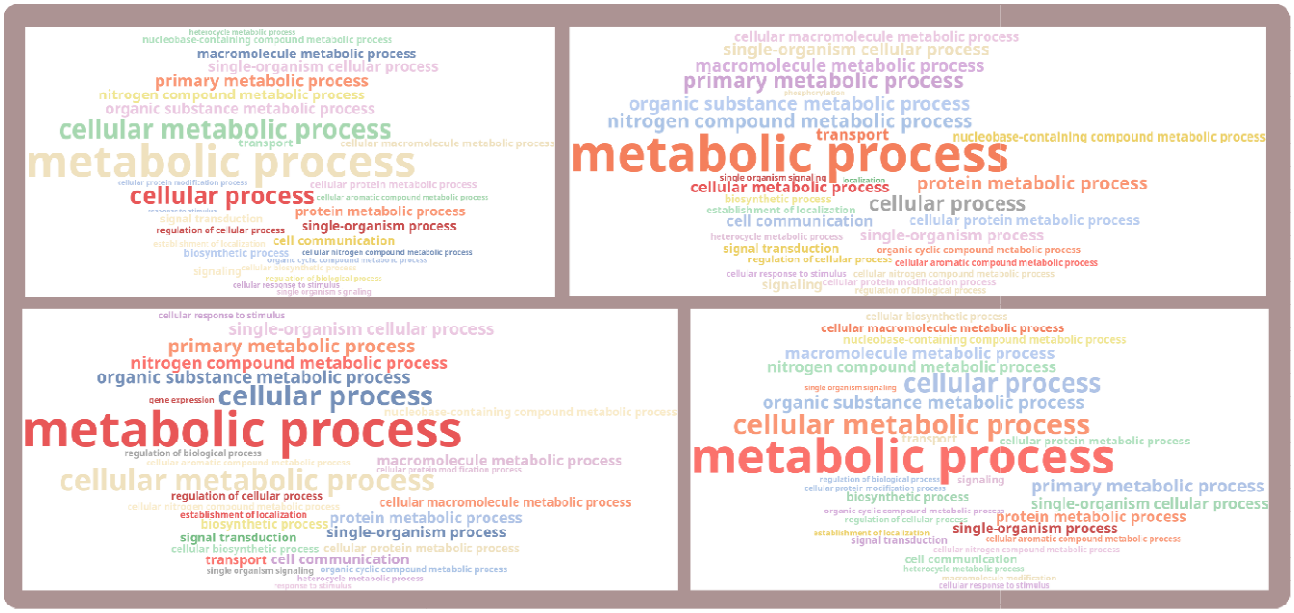


Figure 4.0 - Word cloud generated for the GO terms found in each annotated tissue (*D. unisexualis* - brain: top left; *D. unisexualis* - ovaries: top right; *D. unisexualis* - tail: bottom left; *D. valentini* - tail: bottom right)

Development of cross-species, genome-wide capture sequence probes from the transcriptome of non-model species for phylogenetic analyses

AUTHORS
Susana Freitas
Anja Westram
Victor Soria-Carrasco
D. James Harris
M. A. Carretero
R. Butlin

ABSTRACT
Phylogenetic inference reconstruction has been extensively used to ask several evolutionary biology questions. Recently we have witnessed a steep change of analyses methods in phylogenetics, and single marker sequencing is quickly being replaced for high-throughput sequencing technologies. These allow the study of several markers across the genome for several individuals with a low cost-benefit ratio, but despite the promising applications in phylogenetics, next generation sequencing (NGS) techniques are still far from being as widely used in this field as in other areas of evolutionary genetics.
Here we develop and test a workflow to design probes from transcritptome sequence for downstream capture sequence analyses. This approach has proved to be not only cost-effective, but also to produce a very high number of cross genome phylogenetically informative markers in a non-model species.

INTRODUCTION
Phylogenetic reconstruction is widely used in evolutionary biology as a tool to recover evolutionary patterns and to interpret processes in biogeography (e.g., (Ahmadzadeh et al., 2013; Wielstra et al., 2013), species diversification (e.g., (Mendes et al., 2016)), rates of speciation (e.g., (Pyron and Burbrink, 2013; Rabosky et al., 2013), and conservation (e.g., (Rosauer et al., 2016; Tonini et al., 2016), among others. Sanger DNA sequencing techniques have been used extensively in phylogenetic studies and, for decades,

researchers have been building new data on a few already published markers, creating a platform that allows comparative and phylogeographic studies, and the study of non-model species using cross-species markers. Entire fields of research, such as the use of "DNA barcoding" using CO1 sequences are built on this basis (Smith et al., 2008). Sanger methods generate sequences in a highly targeted and straightforward way via primer design and PCR amplification, one locus per individual at a time. But the phylogenetic analysis of a taxon can be very biased to the history and evolution of the one or few markers used. Cases such as introgression, incomplete lineage sorting (ILS), fast radiations, hybridization and gene flow can only be resolved with sequences from many loci.

The recent advent of next generation sequencing (NGS) techniques and the increasingly fast and cost-effective generation of very large amounts of data allows researchers to address ecological and evolutionary questions on a much wider scale and with deeper precision (Ekblom and Galindo, 2011; Stapley et al., 2010). However, despite the potential NGS techniques presents, phylogenetics as a field was not as quick to adopt them as other fields in evolutionary biology (Carstens et al., 2012; McCormack et al., 2013). NGS allows the sequencing of libraries of templates, exponentially increasing the amount of data gathered per run, not only in terms of the number of markers but also regarding the number of individuals assessed (Carstens et al., 2012; Metzker, 2010). However, with many NGS methods there is less control over which regions of the genome are sequenced, and over whether those regions are either homologous across the whole range of individuals analysed or phylogenetically informative (Carstens et al., 2012). This may pose a strong constraint on adopting NGS in phylogenetic reconstruction.

Phylogenetic analyses and coalescent-based methods are typically based on the construction of gene trees, which ideally require long and highly informative non-recombining sequences (e.g., (Bouckaert et al., 2014; Mirarab and Warnow, 2015). The challenge in shifting from phylogenetics to phylogenomics and embracing NGS techniques has been in the development of a strategy that is not only cost-effective in terms of the number of orthologous sequences available across the individuals analysed, but which could also provide a wide range of sequences, each as long and phylogenetically informative as possible.

Most commonly used NGS approaches accomplish genome reduction with restriction

digest and manual size selection (Baird et al., 2008; Elshire et al., 2011). These methods are very cost effective (Stapley et al., 2010), and allow sequencing a large number of short markers distributed more or less randomly across the genome, many individuals at a time. While they have been used widely for evolutionary studies, they are better applied in population genomic studies at intra-specific level (McCormack et al., 2013) since they produce short reads that are best suited for generating SNPs and often have low repeatability. Moreover, as they amplify sequences next to restriction sites, they often fail to sequence orthologous regions among the individuals analysed. On the other hand, methods that reduce the genome by targeting specific genomic regions (like target enrichment or sequence capture, (Brandley et al., 2015)) can be applied to multiple individuals, they generate orthologous sequences and are replicable, showing a higher potential for deep-level phylogenomic studies.

Sequence capture methods use probes targeted to specific regions of the genome and can sequence hundreds of markers simultaneously in individuals from multiple species. Because they provide orthologous sequences that can be reduced to SNP data, they have started to be used not only in phylogenomic studies (Brandley et al., 2015; Leaché and Linkem, 2015; Mandel et al., 2014), but also in other evolutionary questions (Tennessen et al., 2014; Westram et al., 2016), proving their flexibility and efficiency. Despite their wide applications and interesting results in phylogenomics, designing probes to suitable sequences is so far the biggest hurdle when working with under-studied taxa.

Transcriptome analysis has been demonstrated to be a good first approach for gathering sequence information in non-model species. In large genomes, such as in the case of humans (Treangen and Salzberg, 2012) or maize (Schnable et al., 2009), non-coding and repetitive elements account for large parts of the genome (~50 and 85%, respectively). Transcriptomes are smaller than genomes and have fewer repetitive elements making them easier to assemble and requiring less computational resource. Also, an assembled transcriptome can be annotated by comparison with protein sequences from more or less distant organisms, taking advantage of the already intensively studied genomes and proteomes of model organisms. Due to its feasibility, the number of publications with the words “transcriptome” + “non-model species” has been rapidly increasing since the first paper in 2007 to about 50 papers in 2016.

Despite these advantages, transcriptome analysis also carries some drawbacks when

compared to other phylogenetic resources. First, sequences recovered from transcriptomes are mainly protein coding sequences (even though other regions such as 5' and 3' UTRs can be recovered), which are more conserved than other regions such as introns or intergenic sequences (McCormack et al., 2013). Second, the coverage of any given sequence in a transcriptome is dependent on the number of transcripts sequenced and the expression rate, which is highly variable across genes, tissues and life/age stages. The unequal coverage can be a problem since transcripts with low representation may remain undetected, but also cannot be used in de novo assemblies of transcriptomes to resolve repeated motifs (such as gene duplications) as in the case of genomes. Third, alternative splicing can also complicate the assembly since splicing isoforms are made of different combinations of exons belonging to the same genes. Last but not least, due to the high number of studies and applications of transcriptome analysis and RNAseq, there was also a high turnout of software and pipelines that are available which perform very cohesive de novo assemblies of transcriptomes only with the read information despite the possible limitations the transcriptome may present (Peng et al., 2013; Robertson et al., 2010; Schulz et al., 2012).

Here we use transcriptome sequencing to select sequence information needed for probe design and hence the sequencing of many targeted orthologous regions. Genome sequencing has been widely used to design probes in understudied taxa (Faircloth, 2016; Lemmon et al., 2012). However, probe selection from transcriptomes may be a more cost effective procedure to sequence several regions of high phylogenetic information in non-model organisms. We performed a de novo assembly of RNAseq data to first find regions of interest and design the probes used downstream for a capture sequencing approach.

Given their under-representation in genomic studies, reptiles are ideal for testing our approach. In terms of evolutionary interest they represent a wide diversity of life histories, reproduction types, sex-determining systems and physiologies. However, apart from the anole lizards, non-avian reptiles are rarely considered model organisms and therefore are poorly represented at the genomic level: there are six sequenced reptilian genomes, compared to the 11 of birds and 73 of mammals (Ellegren, 2014).

One of the peculiarities of reptiles is the presence of different reproductive modes. While most of the species are sexual, some are found to be parthenogenetic. Parthenogenesis is typically obligatory, although there have been cases of spontaneous parthenogenesis in

captive females that have called into question the number of vertebrate species that can reproduce in the absence of sex (Avise, 2008). The genus studied here, *Darevskia*, is a genus of rock lizards found in the Caucasus. They were the first vertebrates to be described as true parthenogenetic (Darevsky, 1967). In this group all parthenogens are hybrids of two sexual species belonging to two different clades (Freitas et al., 2016; Murphy et al., 2000). Seven parthenogenetic species have been described and their hybrid origins are expected to have been between species of different clades and directional (maternal and paternal species are always from the same clades) (Fu et al., 2000; Murphy et al., 2000). Given both the hybrid origin and the unusual reproductive mode, transcriptome analysis of these species is of great interest.

In this study we describe the workflow used to design sequence capture probes from transcriptomes. We sequenced the transcriptome of two species, one sexual and other parthenogenetic of hybrid origin, using one individual of each species, enough to detect phylogenetically informative loci. De novo assembly of the transcriptomes was then used for development of capture sequence probes for downstream phylogenomic analyses across the whole genus (Chapter 5 – Freitas et al, in prep).

METHODS

RNA extraction

Total RNA was extracted from four samples; tail-tip from a male individual of the sexual *D. valentini*, brain and ovaries from one individual and tail-tip from another individual of the parthenogenetic *D. unisexualis*. Given the expected differential expression depending on the tissue, in the parthenogenetic individual tissue samples were analysed separately. The tail-tips were collected in the field and preserved in RNAlater. Ovaries and brain were used from a sacrificed animal, and were preserved in RNAlater immediately after the animal's dissection. All tubes were kept at -80º C prior to extraction, except during the fieldwork.

RNA was extracted using a Trizol (Invitrogen) extraction method, following (Westram et al., 2014). All extractions were performed in duplicate. RNA was preserved at -80ºC until ready to proceed with library preparation. Total RNA quality and concentration were determined using the Agilent 2100 BioAnalyzer and the sample used for library preparation was selected according to quality parameters.

Reads processing

RNAseq libraries for the four samples were prepared in Edinburgh Genomics, Edinburgh, UK where they were then sequenced in a single lane, using an Illumina HiSeq 2000 machine (100 bp paired-end reads; insert size around 80 bp). Read quality, before and after filtering, was assessed with FastQC (Andrews, 2010). Filtering was performed with Scythe (Buffalo, 2011) and Sickle (Joshi and Fass, 2011) discarding all reads shorter than 50 bp and/or with Ns, leaving the remaining parameters as the default option.

De novo assembly

A *de novo* assembly for the contigs selection was constructed with Velvet (1.2.10)(Zerbino and Birney, 2008) and Oases (version 0.2.08; http://www.ebi.ac.uk/~zerbino/oases/) (Schulz et al., 2012). Two assemblies were done, one with the reads from the tail-tip of *D. valentini* (VT) and the other with all reads from the ovaries and the brain of *D. unisexualis* (UBUO). The same parameters were used in both *de novo* assemblies, with initial Velvet assemblies using k-mer length values from 25 to 65 and a step increase of 8 [25, 33, 41, 49, 57, 65], and a final assembly of the initial runs was performed with Oases with a k-mer length of 31 and a coverage cut-off of 4. Redundancy of the assembly was decreased with CD-Hit (version:1.3.1)(Li and Godzik, 2006), setting a minimum similarity of 0.95 and a word length of 8.

For a more detailed annotation, reads from each sample were later used to construct four de novo assemblies, one per tissue/species. Given the extremely rapid turnout of softwares for transcriptome assembly, a different software was used for these de novo assemblies, more convenient to our dataset than the one used so far, which had not been available at the time of the reference contigs selection. IDBA-tran (Peng et al., 2010, 2013) is an iterative De Bruijn Graph De Novo short read assembler for transcriptome reads. It used local assembly to reconstruct missing k-mers in low-expressed transcripts, and separated the graph into components, which are one gene in most cases and do not contain many transcripts. Isoforms are searched by a heuristic algorithm based on paired-end reads. These assemblies were only used for annotation purposes.

SNP calling

To map the reads against the reference transcriptome, Stampy (version 1.0.23)(Lunter and

Goodson, 2011) was selected due to its high sensitivity for indels and divergent reads. Platypus (version 0.7.9.1) (Rimmer et al., 2014) was used to call the variants mapped.

We aimed to select contigs to give a total capture target of approximately 500kbp, with sufficient variable sites to give good phylogenetic signal within *Darevskia*. Given the high expected heterozygosity of the hybrid parthenogenetic *D. unisexualis* (Freitas et al., 2016), the transcriptome of the tail tissue of *D.valentini* was preferred as the reference for SNP calling. The longest isoform per contig was selected. To simplify downstream analyses, contigs with indels were eliminated, and only contigs with SNPs were selected. The *D. unisexualis* de novo assembly was also used to select variable loci and the same method was applied. The most variable contigs were selected from both assemblies and added to the target set until the required total target size was reached.

After contig selection from the VT and UBUO assemblies, the sets were blasted against each other using megablast option in blastn (Camacho et al., 2013) with a minimal value of $1e^{-3}$ in order to eliminate possible repeated sequences selected simultaneously from both assemblies.

Probe design for the reference loci and capture sequence was outsourced to RAPiD Genomics (FL, USA). Probes were designed to span across the two transcript contigs data sets (575,561 bp), to be 120 bp long, tiled across the whole length of the contig, with a 40 bp shift so giving 3x coverage. 10,000 probes are needed to cover the ~500kb capture target intended. A total of 75 individuals belonging to 14 sexual *Darevskia* species and three outgroup individuals (two *Iranolacerta brandtii* and one *Podarcis atrata*) were sequenced after genomic DNA fragmentation and ligation with barcoded Illumina-compatible adapters to the resulting fragments. Enrichment PCR and capture reactions with the custom-designed probes preceded the libraries sequencing.

Phylogenetic inference

Read filtering, read mapping against the reference contigs and SNP calling is described elsewhere (Chapter 5 – Freitas et al, in prep). Heterozygosity was calculated in the vcf outputs using vcftools -het option (Danecek et al., 2011).

Fasta files for each marker were made from the vcf file using a python script (https://github.com/SusanaFreitas/capture-data). Fasta files were then trimmed to remove missing data allowing only for maximum of 10% missing genotypes for each position in

each alignment. Species tree was constructed with SVDQuartets for the most variable 300 markers (Chifman and Kubatko, 2015).

Annotation of assembled contigs

The de novo assemblies performed per sample, four in total (tail of *D. valentini*, and brain, ovaries, and tail of *D. unisexualis*), were annotated. Annotation was performed with Interproscan v5 (Jones et al., 2014; Zdobnov and Apweiler, 2001). Functional annotation (GO and KEGG) was performed using BLAST2GO (Conesa et al., 2005) using the default parameters (BLAST e-value threshold of $1e^{-06}$, Gene Ontology annotation threshold of 55).

RESULTS

Reads and SNP detection

Between 36 and 42 million paired-end reads were generated per library and 94% were kept after trimming, adaptor removal and size and quality filtering. The reads of the *D. valentini* tail-tip sample and the reads of the *D. unisexualis* ovaries together with *D. unisexualis* brain samples were used to construct two *de novo* assemblies.

De novo assembly of the sexual *D. valentini* tail tissue (VT) generated 245,539 transcripts distributed over 26,866 contigs with an average length of 1588 bp and N50 length of 2315. Oases assembles transcripts and groups then according to similarity in contig groups. After CD-Hit redundancy analysis, 103,605 transcripts distributed over 26,899 contigs were left with a mean contig length of 1447 bp and N50 length of 2095 (Table 4.1). De novo assembly of the brain and ovaries tissue of the parthenogen *D. unisexualis* (UBUO) originated 690,263 transcripts distributed over 58,106 contigs with an average length of 1588 bp and N50 length of 2469. After CD-Hit redundancy analysis, 289,870 transcripts distributed by 54,563 contigs were left with and average contig length of 1556 bp and a maximum length of 24,289 bp and N50 length of 2544. N50 length was used to evaluate the quality of the contigs. This is a measure widely used in genomics and is defined as the length N for which 50% of all bases in the sequences are in a sequence of length L < N. The longest transcript per contig was selected for each assembly to be used as a reference for the mapping and downstream analyses.

Table 4.1: Statistics for the *D. valentini* tail de novo assembly (VT) and the *D. unisexualis* brain + ovaries de novo assembly (UBUO).

Assembly test	nr of transcripts	nr of loci	Kmer interval	steps	merge Kmer	coverage cutoff	insert length	N50	max contig length	average contig length	GC content
VT	245539	26899	25-65	8	31	4	80	2315	28740	1588.12	47.52
VT-cdhit0.95	103605	26899						2095	28740	1446.77	47.46
VT-cdhit0.80	60706	26894						2116	28740	1470.33	47.20
UBUO	690263	58106	25-65	8	31	4	80	2469	24289	1500.10	47.20
UBUO-cdhit0.95	289869	54465						2544	24289	1555.63	47.15

Contigs selection

To select the most variable contigs, all reads were mapped against both the *D. valentini* (VT) and the *D. unisexualis* (UBUO) *de novo* assemblies. Only reads mapped with PHRED quality over 100 were selected, so that there is 1% or less probability of a wrongly identified SNP.

Around 80% of the *D. valentini* reads mapped against the VT reference, with a bit less (between 69 – 72%) for the *D. unisexualis* reads. The yield of mapping regarding the UBUO assembly was significantly lower than for the VT assembly, with only around 50% of the reads successfully mapping against that reference (Table 4.1 - SI).

The SNP calling of all reads against the VT assembly resulted in a total of 721,238 variants. Only reference contigs that presented SNPs but not indels, in the reference allele were selected. In the end, 773 contigs were selected, presenting in total 2372 SNPs. The same approach was followed in the UBUO assembly. For this de novo assembly a total of 3765 contigs were selected, presenting a total of 7981 variants.

From these lists of contigs, the most variable contigs with a minimum length of 500 bp were selected from each (Table 4.2). Given that contigs were selected from two different assemblies that could contain orthologous sequences, the selected contig sets were blasted against each other and the shorter of any pair of duplicated contigs was removed. This resulted in 193 (215) contigs from the VT assembly (163,534 bp in total) and 400(410) contigs (265,335 bp) from the UBUO assembly. We also added random sequences that did not present any variation in the backward alignment, 23 from the VT assembly and 10 from the UBUO assembly. Finally we had 215 contigs from the VT assembly and 410 from the UBUO assembly.

Probe design was outsourced to RAPiD Genomics Inc. (FL, USA) and, out of a total of 625

sequences, probes were successfully designed for 624.

Table 4.2: Selected contigs from VT and UBUO de novo assemblies: statistics about length and number of variants.

de novo assembly		variant count	length	prop (var/length)	variant count	length	prop (var/length)
VT	193 seqs > 163534 bp				188 seqs > 160288.8 bp		
	Min.	1	500	0.000464	1	500	0.000464
	1st Qu.	2	561	0.002597	2	561.8	0.002575
	Median	4	697	0.005340	4	698.5	0.005297
	Mean	4.953	847.3	0.006382	4.973	852.6	0.006383
	3rd Qu.	7	863	0.009682	7	875.5	0.00969
	Max.	35	3408	0.021708	35	3408	0.021708
UBUO	400 seqs > 370240 bp						
	Min.	2	600	0.00216	in green > corrected after the BLAST hits.		
	1st Qu.	2	679	0.003003			
	Median	3	794.5	0.003954			
	Mean	4.295	925.6	0.004711			
	3rd Qu.	5	999	0.005597			
	Max.	21	4065	0.015458			

SNP calling and Phylogenetic inference

SNPcalling of the *Darevskia* species against the reference contigs resulted in a total of 64150 SNPs. Probes successfully captured all contigs, however one recovered no SNP variation. SNP count varied from 1 – 379 per contig (0.12 – 24%). A total of 10318 SNPs from 453 contigs were recovered in all individuals. If a maximum of 10% missing data was allowed, 37092 SNPs from 582 contigs were recovered.

The 300 more variable contigs were used to calculate the species tree with SVDQuartets. Three major clades are recovered in the species tree, and most phylogenetic relationships confirm previous studies (Murphy et al. 1996; Murphy et al. 2000; Ahmadzadeh et al. 2013; Freitas, et al. 2016; Freitas, et al. 2016) (Figure 4.1).

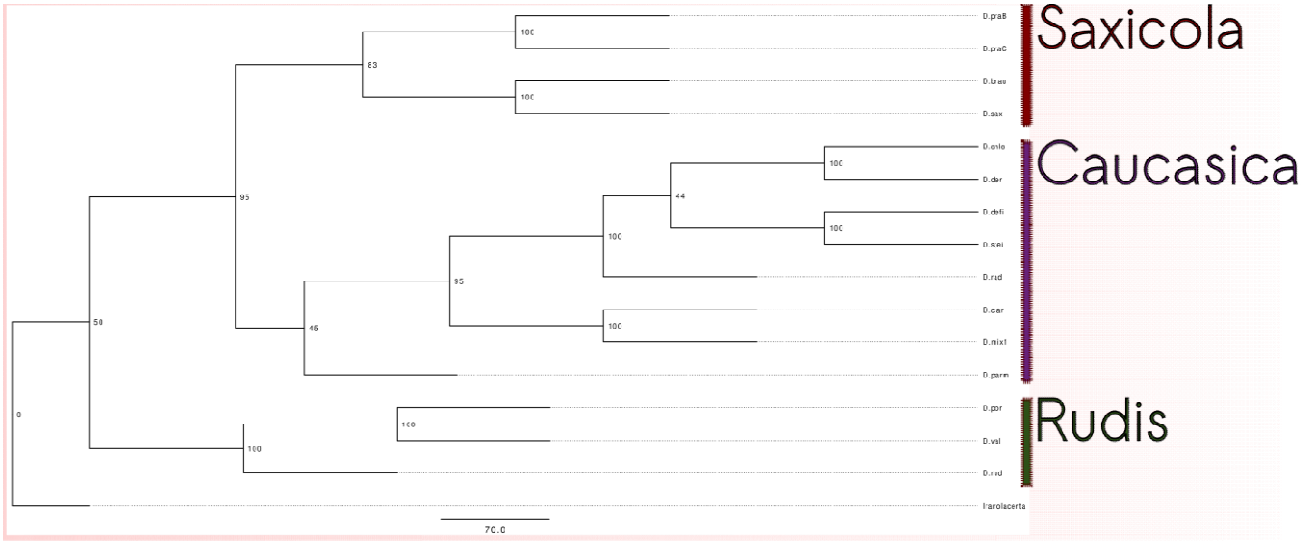


Figure 4.1: Phylogenetic relationships within genus Darevskia, with three major clades: Saxicola, Caucasica and Rudis.

Annotation

Annotation was performed on the de novo assemblies done per sample (four in total: tail of *D. valentini*, and brain, ovaries, and tail of *D. unisexualis*). Only a small percentage of transcripts were successfully annotated using Interproscan (Figure 4.2), and even fewer could be assigned to a gene ontology class: 17%, 43%, 5% and 35% for UB, UO, UT and VT de novo assemblies, respectively. GO terms for Biological Processes, Cell Components and Molecular Function are summarised on (Figure 4.3). We note, however, that the assemblies with higher proportion of annotated transcripts were the de novo assemblies for the *D. unisexualis* ovaries and *D. valentini* tail, which were also the tissues that presented the reads with the lowest heterozygosity values (Table 4.3). High heterozygosity values can be a factor disturbing assemblies, which could explain why the assembler performed better in these tissues (Ruttink et al., 2013).

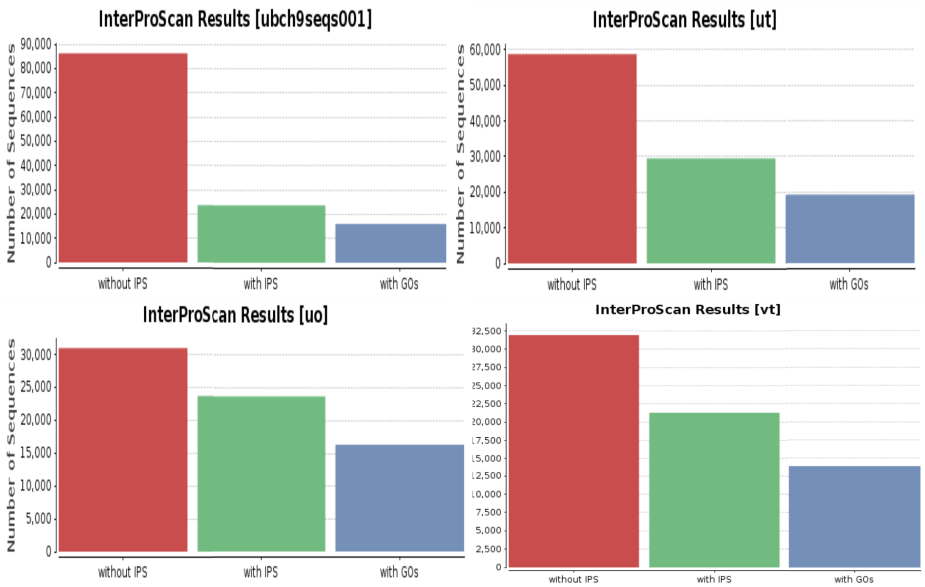


Figure 4.2: Proportion of annotated transcripts (IPS) and with corresponding Gene Ontology classification (GO), when available.

DISCUSSION

The molecular study of non-model species can be very restricted in terms of the scientific questions posed due to the lack of comparative literature. Recently, with the advent of NGS techniques, this disadvantage could be overcome and several studies are emerging on the application of these techniques to scarcely known species. Published studies using NGS techniques are no longer restricted to comparative and descriptive analyses of reptilian transcriptomes and genomes (Eckalbar et al., 2013; Georges et al., 2015; Hutchins et al., 2014; Tzika et al., 2011). In particular, NGS data are now being used for phylogenomic studies, either employing genome reduction methods with enzyme restriction (Cariou et al., 2013; Herrera and Shank, 2016) or with methods which target specific regions of the genome (Leaché and Linkem, 2015; Newman and Austin, 2016).

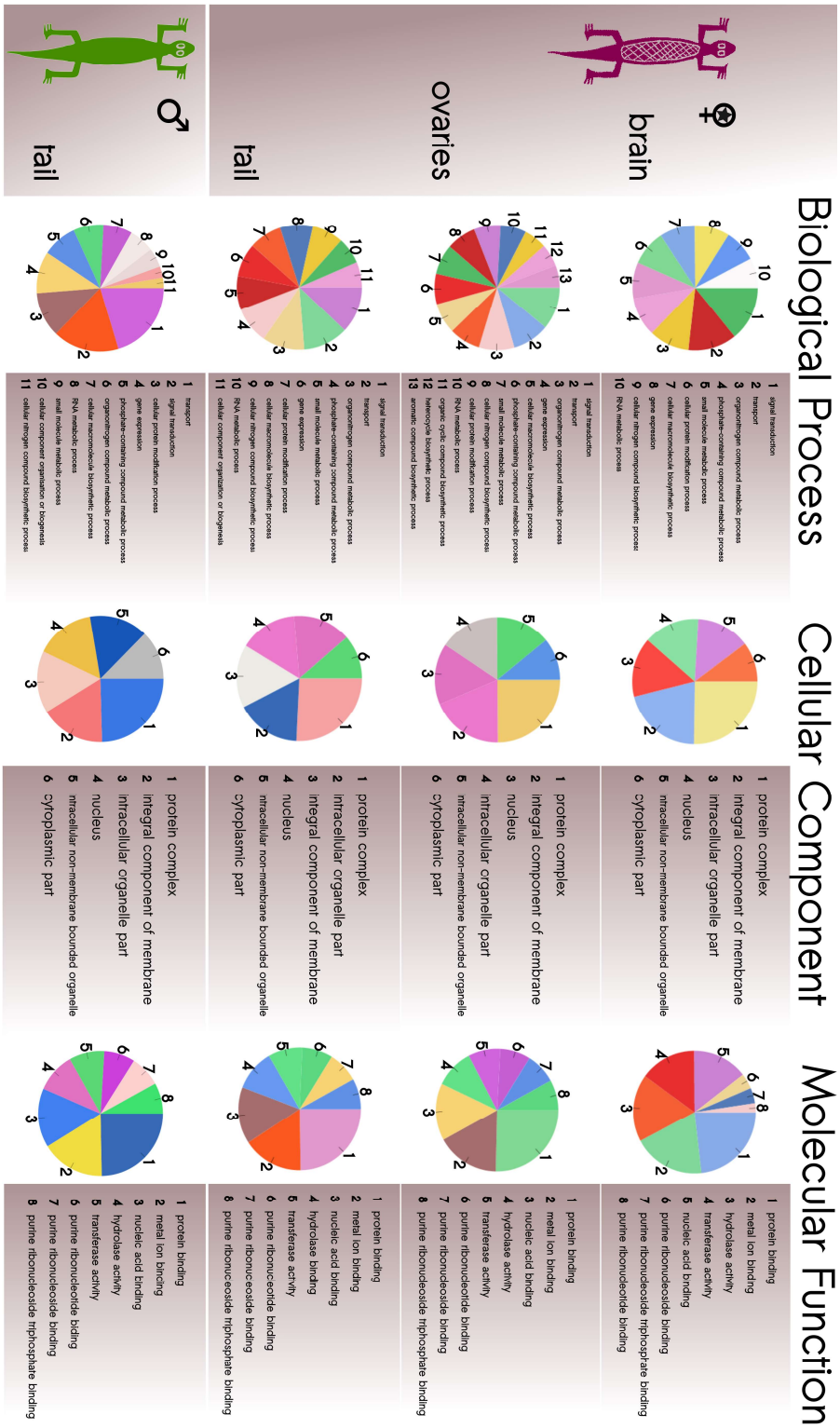


Figure 4.3: Gene ontology (GO) terms for the annotated transcripts to each tissue. Biological process, Cellular component and Molecular function classes are highlighted for each tissue.

The coalescent-based methods currently employed can produce robust phylogenetic inferences but require long sequences present in most of the individuals analysed (Hosner et al., 2016; Roure et al., 2012). The approach developed here was able to detect hundreds of regions with high variability and successfully sequence these homologous DNA regions from many species within the same genus. The amount of data generated could then be used downstream for phylogenetic inferences.

Table 4.3: Heterozygosity calculated for each of tissue mapped against the reference transciptome used for the probe design. Values for observed homozygous positions (Ho), observed heterozygous positions (He), Dxy (Dxy = -(3/4) * ln(1- 4/3 * D), from (Nei, 1978)) and observed heterozygosity (Observed He = He / (He + Ho)). Codes for the tissues analysed are as follow: UB (*D. unisexualis* – brain), UO (*D. unisexualis* – ovaries), UT (*D. unisexualis* – tail), VT (*D. valentini* – tail).

	UB	UO	UT	VT
He	7214	4319	5803	4316
Ho	421128	326091	388613	350613
Dxy	0.0132	0.0119	0.0127	0.0117
Observed He	0.0168	0.0131	0.0147	0.0122

Here we propose a workflow for the development of markers across the genome of non-model taxa from transcriptome de novo assembly. These markers can be used for capture sequence probe design and these probes can then be used in capture sequence analysis. Given the probe design was performed on transcriptome data, the markers selected could be identified to the protein level, once the transcriptome is annotated. Given the high degree of allelic diversity found in protein level coding sequence, such as obtained from the transcriptome (with sequences with high diversity and others more conserved), this approach can be applied not only at a species level, but also at a generic level.

Gene ontology terms from the annotation (GO) provide information on the function, locus of action and possible interactions of the transcripts assembled from our transcriptomes. From those we can identify gene product properties and describe them in terms of their

associated biological processes, cellular components and molecular functions. Here we could assign a total of 15978, 16337, 4050 and 13898 for the UB, UO, UT and VT de novo assemblies, respectively. This is the first annotation performed on *Darevskia* sp. or any other member of the lizard family Lacertidae.

We also found that the tail tissue presented lower heterozygosity in *D. valentini* than in *D. unisexualis*. This was expected given *D. unisexualis* is known to have a recent hybrid origin (Freitas et al., 2016; Murphy et al., 2000). Less expected was the high variation of heterozygosity across tissues within *D. unisexualis*, with the ovaries being the tissue with the lowest heterozygosity values (Table 4.3). This is likely the result of the biased expression of only one of the parental genomes, which could be due to genetic incompatibilities in the sex organs. However, a specific study of the RNA expression in the different tissues of parthenogens and sexual species has to be performed to determine either if the different genomic contributions in the hybrid parthenogens are generally differentially expressed, or if this is specific to the parthenogen individual we sequenced.

This work provides a way of using NGS techniques in non-model taxa that is efficient and cost-effective. The probes developed allow sequencing hundreds of markers across the genome and can be successfully used in different species of the same genus simultaneously and in several individuals. The data set provided can then be used downstream for phylogenomics, population genomics and other evolutionary ecology studies.

REFERENCES

Ahmadzadeh, F., Flecks, M., Carretero, M.A., Mozaffari, O., Böhme, W., Harris, D.J., Freitas, S., and Rödder, D. (2013). Cryptic Speciation Patterns in Iranian Rock Lizards Uncovered by Integrative Taxonomy. PLOS ONE *8*, e80563.

Andrews, S. (2010). FastQC: A quality control tool for high throughput sequence data.

Avise, J.C. (2008). Clonality: the genetics, ecology, and evolution of sexual abstinence in vertebrate animals (Oxford ; New York: Oxford University Press).

Baird, N.A., Etter, P.D., Atwood, T.S., Currey, M.C., Shiver, A.L., Lewis, Z.A., Selker, E.U., Cresko, W.A., and Johnson, E.A. (2008). Rapid SNP Discovery and Genetic Mapping Using Sequenced RAD Markers. PLOS ONE *3*, e3376.

Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C.-H., Xie, D., Suchard, M.A., Rambaut, A., and Drummond, A.J. (2014). BEAST 2: A Software Platform for Bayesian Evolutionary Analysis. PLOS Comput. Biol. *10*, e1003537.

Brandley, M.C., Bragg, J.G., Singhal, S., Chapple, D.G., Jennings, C.K., Lemmon, A.R., Lemmon, E.M., Thompson, M.B., and Moritz, C. (2015). Evaluating the performance of anchored hybrid enrichment at the tips of the tree of life: a phylogenetic analysis of Australian Eugongylus group scincid lizards. BMC Evol. Biol. *15*, 62.

Buffalo, V. (2011). Scythe - A very simple adapter trimmer.

Camacho, C., Madden, T., Ma, N., Tao, T., Agarwala, R., and Morgulis, A. (2013). BLAST Command Line Applications User Manual.

Cariou, M., Duret, L., and Charlat, S. (2013). Is RAD-seq suitable for phylogenetic inference? An in silico assessment and optimization. Ecol. Evol. *3*, 846–852.

Carstens, B., Lemmon, A.R., and Lemmon, E.M. (2012). The Promises and Pitfalls of Next-Generation Sequencing Data in Phylogeography. Syst. Biol. *61*, 713–715.

Chifman, J., and Kubatko, L. (2015). Identifiability of the unrooted species tree topology under the coalescent model with time-reversible substitution processes, site-specific rate variation, and invariable sites. J. Theor. Biol. *374*, 35–47.

Conesa, A., Götz, S., García-Gómez, J.M., Terol, J., Talón, M., and Robles, M. (2005). Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics *21*, 3674–3676.

Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T., et al. (2011). The variant call format and VCFtools. Bioinformatics *27*, 2156–2158.

Darevsky, I. (1967). Rock lizards of the Caucasus: systematics, ecology and phylogenesis of the polymorphic groups of Caucasian rock lizards of the subgenus Archaeolacerta. Leningrad. Nauka.

Eckalbar, W.L., Hutchins, E.D., Markov, G.J., Allen, A.N., Corneveaux, J.J., Lindblad-Toh, K., Di Palma, F., Alföldi, J., Huentelman, M.J., and Kusumi, K. (2013). Genome reannotation of the lizard Anolis carolinensis based on 14 adult and embryonic deep transcriptomes. BMC Genomics *14*, 49.

Ekblom, R., and Galindo, J. (2011). Applications of next generation sequencing in molecular ecology of non-model organisms. Heredity *107*, 1–15.

Ellegren, H. (2014). Genome sequencing and population genomics in non-model organisms. Trends Ecol. Evol. *29*, 51–63.

Elshire, R.J., Glaubitz, J.C., Sun, Q., Poland, J.A., Kawamoto, K., Buckler, E.S., and Mitchell, S.E. (2011). A Robust, Simple Genotyping-by-Sequencing (GBS) Approach for High Diversity Species. PLOS ONE *6*, e19379.

Faircloth, B.C. (2016). Identifying Conserved Genomic Elements and Designing Universal Probe Sets To Enrich Them.

Freitas, S., Rocha, S., Campos, J., Ahmadzadeh, F., Corti, C., Sillero, N., Ilgaz, Ç., Kumlutaş, Y., Arakelyan, M., Harris, D.J., et al. (2016). Parthenogenesis through the ice ages: A biogeographic analysis of Caucasian rock lizards (genus Darevskia). Mol. Phylogenet. Evol. *102*, 117–127.

Fu, J., Murphy, R.W., Darevsky, I.S., and McEachran, J.D. (2000). Divergence of the cytochrome b gene in the Lacerta raddei complex and its parthenogenetic daughter

species: evidence for recent multiple origins. Copeia *2000*, 432–440.

Georges, A., Li, Q., Lian, J., O'Meally, D., Deakin, J., Wang, Z., Zhang, P., Fujita, M., Patel, H.R., Holleley, C.E., et al. (2015). High-coverage sequencing and annotated assembly of the genome of the Australian dragon lizard Pogona vitticeps. GigaScience *4*, 45.

Herrera, S., and Shank, T.M. (2016). RAD sequencing enables unprecedented phylogenetic resolution and objective species delimitation in recalcitrant divergent taxa. Mol. Phylogenet. Evol. *100*, 70–79.

Hosner, P.A., Faircloth, B.C., Glenn, T.C., Braun, E.L., and Kimball, R.T. (2016). Avoiding Missing Data Biases in Phylogenomic Inference: An Empirical Study in the Landfowl (Aves: Galliformes). Mol. Biol. Evol. *33*, 1110–1125.

Hutchins, E.D., Markov, G.J., Eckalbar, W.L., George, R.M., King, J.M., Tokuyama, M.A., Geiger, L.A., Emmert, N., Ammar, M.J., Allen, A.N., et al. (2014). Transcriptomic Analysis of Tail Regeneration in the Lizard Anolis carolinensis Reveals Activation of Conserved Vertebrate Developmental and Repair Mechanisms. PLoS ONE *9*, e105004.

Jones, P., Binns, D., Chang, H.-Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., et al. (2014). InterProScan 5: genome-scale protein function classification. Bioinformatics *30*, 1236–1240.

Joshi, N., and Fass, J. (2011). Sickle: A sliding-window, adaptive, quality-based trimming tool for FastQ files.

Leaché, A.D., and Linkem, C.W. (2015). Phylogenomics of Horned Lizards (Genus: *Phrynosoma*) Using Targeted Sequence Capture Data. Copeia *103*, 586–594.

Lemmon, A.R., Emme, S.A., and Lemmon, E.M. (2012). Anchored Hybrid Enrichment for Massively High-Throughput Phylogenomics. Syst. Biol. *61*, 727–744.

Li, W., and Godzik, A. (2006). Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics *22*, 1658–1659.

Lunter, G., and Goodson, M. (2011). Stampy: A statistical algorithm for sensitive and fast mapping of Illumina sequence reads. Genome Res. *21*, 936–939.

Mandel, J.R., Dikow, R.B., Funk, V.A., Masalia, R.R., Staton, S.E., Kozik, A., Michelmore, R.W., Rieseberg, L.H., and Burke, J.M. (2014). A target enrichment method for gathering phylogenetic information from hundreds of loci: An example from the Compositae1. Appl. Plant Sci. *2*.

McCormack, J.E., Hird, S.M., Zellmer, A.J., Carstens, B.C., and Brumfield, R.T. (2013). Applications of next-generation sequencing to phylogeography and phylogenetics. Mol. Phylogenet. Evol. *66*, 526–538.

Mendes, J., Harris, D.J., Carranza, S., and 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. Mol. Phylogenet. Evol. *100*, 254–267.

Metzker, M.L. (2010). Sequencing technologies — the next generation. Nat. Rev. Genet. *11*, 31–46.

Mirarab, S., and Warnow, T. (2015). ASTRAL-II: coalescent-based species tree estimation with many hundreds of taxa and thousands of genes. Bioinformatics *31*, i44–i52.

Murphy, R.W., Fu, J., Macculloch, R.D., Darevsky, I.S., and Kupriyanova, L.A. (2000). A fine line between sex and unisexuality: the phylogenetic constraints on parthenogenesis in lacertid lizards. Zool. J. Linn. Soc. *130*, 527–549.

Nei, M. (1978). Estimation of Average Heterozygosity and Genetic Distance from a Small Number of Individuals. Genetics 89, 583–590.

Newman, C.E., and Austin, C.C. (2016). Sequence capture and next-generation sequencing of ultraconserved elements in a large-genome salamander. Mol. Ecol. *25*, 6162–6174.

Peng, Y., Leung, H.C., Yiu, S.-M., and Chin, F.Y. (2010). IDBA–a practical iterative de Bruijn graph de novo assembler. In Research in Computational Molecular Biology, (Springer), pp. 426–440.

Peng, Y., Leung, H.C.M., Yiu, S.-M., Lv, M.-J., Zhu, X.-G., and Chin, F.Y.L. (2013). IDBA-tran: a more robust de novo de Bruijn graph assembler for transcriptomes with uneven

expression levels. Bioinformatics *29*, i326–i334.

Pyron, R.A., and Burbrink, F.T. (2013). Phylogenetic estimates of speciation and extinction rates for testing ecological and evolutionary hypotheses. Trends Ecol. Evol. *28*, 729–736.

Rabosky, D.L., Santini, F., Eastman, J., Smith, S.A., Sidlauskas, B., Chang, J., and Alfaro, M.E. (2013). Rates of speciation and morphological evolution are correlated across the largest vertebrate radiation. Nat. Commun. *4*.

Rimmer, A., Phan, H., Mathieson, I., Iqbal, Z., Twigg, S.R.F., Wgs500 Consortium, Wilkie, A.O.M., McVean, G., and Lunter, G. (2014). Integrating mapping-, assembly- and haplotype-based approaches for calling variants in clinical sequencing applications. Nat. Genet. *46*, 912–918.

Robertson, G., Schein, J., Chiu, R., Corbett, R., Field, M., Jackman, S.D., Mungall, K., Lee, S., Okada, H.M., Qian, J.Q., et al. (2010). De novo assembly and analysis of RNA-seq data. Nat. Methods *7*, 909–912.

Rosauer, D.F., Blom, M.P.K., Bourke, G., Catalano, S., Donnellan, S., Gillespie, G., Mulder, E., Oliver, P.M., Potter, S., Pratt, R.C., et al. (2016). Phylogeography, hotspots and conservation priorities: an example from the Top End of Australia. Biol. Conserv. *204, Part A*, 83–93.

Roure, B., Baurain, D., and Philippe, H. (2012). Impact of Missing Data on Phylogenies Inferred from Empirical Phylogenomic Data Sets. Mol. Biol. Evol. mss208.

Ruttink, T., Sterck, L., Rohde, A., Bendixen, C., Rouzé, P., Asp, T., Van de Peer, Y., and Roldan-Ruiz, I. (2013). Orthology Guided Assembly in highly heterozygous crops: creating a reference transcriptome to uncover genetic diversity in Lolium perenne. Plant Biotechnol. J. *11*, 605–617.

Schnable, P.S., Ware, D., Fulton, R.S., Stein, J.C., Wei, F., Pasternak, S., Liang, C., Zhang, J., Fulton, L., Graves, T.A., et al. (2009). The B73 Maize Genome: Complexity, Diversity, and Dynamics. Science *326*, 1112–1115.

Schulz, M.H., Zerbino, D.R., Vingron, M., and Birney, E. (2012). Oases: robust de novo RNA-seq assembly across the dynamic range of expression levels. Bioinformatics *28*,

1086–1092.

Smith, M.A., Poyarkov, N.A., and Hebert, P.D.N. (2008). DNA BARCODING: CO1 DNA barcoding amphibians: take the chance, meet the challenge. Mol. Ecol. Resour. *8*, 235–246.

Stapley, J., Reger, J., Feulner, P.G.D., Smadja, C., Galindo, J., Ekblom, R., Bennison, C., Ball, A.D., Beckerman, A.P., and Slate, J. (2010). Adaptation genomics: the next generation. Trends Ecol. Evol. *25*, 705–712.

Tennessen, J.A., Govindarajulu, R., Ashman, T.-L., and Liston, A. (2014). Evolutionary Origins and Dynamics of Octoploid Strawberry Subgenomes Revealed by Dense Targeted Capture Linkage Maps. Genome Biol. Evol. *6*, 3295–3313.

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

Treangen, T.J., and Salzberg, S.L. (2012). Repetitive DNA and next-generation sequencing: computational challenges and solutions. Nat. Rev. Genet. *13*, 36–46.

Tzika, A.C., Helaers, R., Schramm, G., and Milinkovitch, M.C. (2011). Reptilian-transcriptome v1.0, a glimpse in the brain transcriptome of five divergent Sauropsida lineages and the phylogenetic position of turtles. EvoDevo *2*, 19.

Westram, A., Galindo, J., Alm Rosenblad, M., Grahame, J.W., Panova, M., and Butlin, R.K. (2014). Do the same genes underlie parallel phenotypic divergence in different Littorina saxatilis populations? Mol. Ecol. *23*, 4603–4616.

Westram, A.M., Panova, M., Galindo, J., and Butlin, R.K. (2016). Targeted resequencing reveals geographical patterns of differentiation for loci implicated in parallel evolution. Mol. Ecol. *25*, 3169–3186.

Wielstra, B., Crnobrnja-Isailović, J., Litvinchuk, S.N., Reijnen, B.T., Skidmore, A.K., Sotiropoulos, K., Toxopeus, A.G., Tzankov, N., Vukov, T., and Arntzen, J.W. (2013). Tracing glacial refugia of Triturus newts based on mitochondrial DNA phylogeography and species distribution modeling. Front. Zool. *10*, 13.

Zdobnov, E.M., and Apweiler, R. (2001). InterProScan – an integration platform for the signature-recognition methods in InterPro. Bioinformatics *17*, 847–848.

Zerbino, D.R., and Birney, E. (2008). Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. Genome Res. *18*, 821–829.

CHAPTER 5

DEEP BRANCH GENE FLOW – paper V

Figure 5.0 - *D. unisexualis*

Deep-branch gene flow in a hybridization-rich diversification process

AUTHORS
Susana Freitas
Anja Westram
Miguel A. Carretero
D. James Harris
Roger Butlin

ABSTRACT
Recent evolutionary evidence highlights the influence and extent of introgressive hybridization during species divergence. However, the emerging studies demonstrating hybridization with gene flow during divergence generally focus on the relationships between only a limited number of species, missing the wider perspective of the larger group to which these species belong. Moreover, the extent of hybridization with gene flow is unknown in groups of species with parthenogenetic lineages of hybrid origin, and how that is related with possible barriers or opportunities for gene flow. Here we show evidence of frequent past hybridization in a species-rich genus of rock lizards, with two different outcomes. We found widespread gene flow between all major clades in the group, including those that diverged recently and anciently, except between the parental species for the parthenogenetic hybrids. While this widespread introgressive hybridization is uncommon in other groups, the role of hybrid asexuality as a reproductive isolation mechanism in the absence of other pre- and post- zygotic barriers to gene flow in diverging taxa is demonstrated here.

INTRODUCTION
The role, frequency and relevance attributed to hybridization in evolution has changed markedly in recent decades. When the biological species concept was first developed (Dobzhansky, 1940; Mayr, 1942), hybridization between different populations was seen as mostly ineffective, hybrid zones were expected not only to be transient but also were considered evidence of reproductive barriers between the diverging groups, and could even be responsible for disrupting stable populations where natural selection had been

reinforcing mating preferences (Abbott et al., 2013; Soltis, 2013). Recently, and especially with the development of high-throughput sequencing techniques and the exponential increase of analytical power, extensive evidence has been provided of ongoing hybridization within and between diverging clades (Abbott et al., 2016). This allowed the investigation of the role of hybridization in the origin of new hybrid species (Alves et al., 2001; Freitas et al., 2016; Mavárez et al., 2006), in eroding barriers to gene exchange during speciation events (Abbott et al., 2013), and in offering evolutionary novelties as drivers of local adaptation, accelerating speciation by adaptive introgression (Cronk and Yang, 2016; Pereira et al., 2014; Stankowski and Streisfeld, 2015).

In vertebrates, interspecific hybridization is the most common process known to be responsible for the origin of new hybrid parthenogenetic species. However, it is not known how the hybrid origin of parthenogenesis fits the overall history of hybridization within a given group. Also, it is still not clear whether the origin of vertebrate parthenogenesis is induced by the hybridization of species at a specific degree of divergence (Balance Hypothesis: (Moritz et al., 1989)) or by the hybridization of specific pairs of phylogenetic clades (Phylogenetic Constraint Hypothesis: (Avise, 2008; Darevsky, 1967)). Nevertheless, the same sexual parental species are likely to hybridize not only to the extent of originating new parthenogenetic species, but possibly allowing for fertile sexual hybrids to vector gene flow between different clades. Depending on whether interspecific hybridization originates new parthenogenetic species due to their phylogenetic distance (Moritz et al., 1989) or phylogenetic constraints (Avise, 2008; Darevsky, 1967), we can expect gene flow either between sexual species younger than those that generated the parthenogens, or only between similarly aged and older lineages in some specific clades.

The group studied here is a genus of rock lizards present in the Caucasus, *Darevskia*, that has a wide distribution range and includes saxicolous and ground-dwelling species. The species *D. praticola*, *D. derjugini*, *D. steineri, D. clarkorum* and *D. chlorogaster* are ground-dwelling, mostly found in forest habitat, but some, like *D. praticola*, inhabit both forest and grasslands. On the other hand, *D. raddei*, *D. defilippii*, *D. portschinskii, D. saxicola* are exclusively saxicolous, and *D. valentini* and *D. rudis* (both belonging to the Rudis clade) can be found both in rocky habitats (saxicolous) as in grasslands (ground-dwelling) (Ahmadzadeh et al., 2013; Tarkhnishvili et al., 2013). Given this diversity of habitats occupied, it is not clear whether lineages in this group can shift their niche readily and occupy new habitats, or if niche specializations took place early in *Darevskia* evolution and

remained conserved, allowing for allopatric isolation.

Seven hybrid asexual species have been described in *Darevskia*, each resulting from the interspecific hybridization between a specific sexual species pair, and only some sexual species participated in these hybridization events. In particular, *D. raddei* and *D. mixta* alone contributed as the maternal species, and *D. portschinskii* and *D. valentini* as the paternal species. Parthenogenetic lineages are recent hybrids between pairs of sexual species (Freitas et al., 2016; Murphy, 2000), and polyploid backcross hybrids are known to occur where parthenogenetic populations are in sympatry with sexual species (Danielyan et al., 2008; Freitas et al., in prep).

In this study we assess the extent of introgressive hybridization events, and secondary contact, *Darevskia* sexual species underwent during divergence. If gene flow is detected, we predict different patterns of hybridization with gene flow in the genus depending on whether interspecific hybridization that originated the parthenogens was limited by the sexual species' phylogenetic distance or due to some specific phylogenetic constraints. Under the Balance Hypothesis, we predict gene flow between pairs of lineages younger than those that generated the parthenogens, and/or gene flow between the parental pairs of the parthenogens until these were originated. In contrast, under the phylogenetic hypothesis, we predict gene flow either concentrated in the lineages that produce parthenogens and at earlier stages, if hybridization is a specific property of the lineage pair, or widespread gene flow that is sometimes older than the lineages that generate parthenogens but absent between the parental pairs, if parthenogenesis rather than hybridization is the property of the lineage pair required for the origin of asexual lineages. Moreover, where gene flow is, in fact, detected between sexual species, we test whether it is restricted to a small proportion of the genome or whether there are "hybrid species" with even mixtures of alleles across the genome from different clades.

To answer these questions we used targeted re-sequencing to analyse over 600 loci in a phylogenetic framework. These loci were previously selected from transcriptome data of two *Darevskia* species (Freitas et al, in prep). A species tree was built to determine the divergence times between the different clades and tree topology. The markers were also analysed for discordance in tree topology, for evidence of introgression, and using migration tests to date the admixture events.

MATERIAL AND METHODS

Species selection

In total, 69 individuals belonging to 14 sexual species are included in this study (Table 1-SI). All samples were collected from the field, where tail tips were preserved and photographs and basic measurements were recorded for later assessment of species identification. An outgroup species, *Iranolacerta brandtii*, was chosen due to its phylogenetic proximity to our study group (Mendes et al., 2016) and DNA was extracted from fresh tissue samples collected from the field for all individuals. Total genomic DNA was extracted from approximately 30 mg of each tail tip following standard high-salt protocols (Sambrook and Russell, 2001). For older samples the Qiagen DNeasy Blood and Tissue kit was used, following manufacturer's instructions.

Capture sequence and genotyping

Probes for the capture sequence were previously designed using a de novo assembly of a *Darevskia* transcriptome (Freitas et al, in prep: Chapter 4). Library preparation and Illumina paired-end sequencing of the captured fragments was performed at Rapid Genomics Inc (Florida 32601 USA), resulting in reads 100 bp long with a 80 bp insert size.

Read filtering was performed with Cutadapt (Martin, 2011) eliminating all reads shorter than 40 bp and with PHRED quality lower than 20. Stampy (Lunter and Goodson, 2011) was used to map the reads against the reference *de novo* assembly used to design the probes. A substitution rate was set regarding the distance of the different species to the reference transcriptome, higher for phylogenetically distant species and lower for phylogenetically close species. Accordingly, individuals were divided into three groups (tests were repeated after correcting for misidentified individuals): Rudis group (with SR=0.01), non-Rudis group (SR = 0.03) and the outgroup (SR = 0.1). In cases of multiple mappings that are almost (but not exactly) the same, the SR parameter helps to choose the best mapping by influencing the mapping quality (MAPQ) score. Best SR values were chosen when comparing number of variants (after SNP calling) and flagstat values for the alignments. Samtools (Li et al., 2009) was used to convert to bam, sort, index and calculate basic statistics of the alignment (flagstat option). The MarkDuplicates option in Picard tools was used to remove PCR duplicates, which can be high given the capture sequence protocol. SNP calling was used with FreeBayes (Garrison and Marth, 2012) using minimum coverage of two per position, calling the four best alleles (for high

substitution rate loci, given we are using different species and expect more variation than within population) and without the population priors. Post-vcf filtering was performed with VCFtools (Danecek et al., 2011) and vcflib (https://github.com/vcflib/vcflib) selecting variants with genotype quality > 30, coverage between 5 and 400 for ingroup species. Ingroup individuals *D. parvula* [12738], *D. valentini* [12641] and *D. valentini* [16280] were filtered separately since they had higher coverage; the selected coverage interval was 5-800 read depth. Outgroup individuals (*Iranolacerta brandtii*) were filtered with the same parameters as the ingroup except for coverage (DP = 2-400). Only SNPs were used in the downstream analyses, so indels and other types of variants were removed. Multiple nucleotide positions (MNP) identified by FreeBayes were deconstructed with vcfallelicprimitives option from vcflib.

Phylogenetic analyses

Fasta files for each marker were made from the vcf file using a python script (https://github.com/SusanaFreitas/capture-data). Fasta files were then trimmed to remove missing data allowing only for a maximum of 10% missing genotypes for each position in each alignment. Phylogenetic analysis on each marker were performed using Bayesian inference (BI) (MrBayes v. 3.2, (Huelsenbeck and Ronquist, 2001; Ronquist and Huelsenbeck, 2003) and the resultant output trees were later used as input for Bucky (Larget et al., 2010) analysis (see bellow for detailed information about this analysis). Bucky looks for tree incongruence, thus if the number of taxa is too high the tree space is very large, reducing the probability of detecting the same tree more than once. To increase Bucky's sensitivity, taxon number was reduced my merging some of the most recently diverged species (after *BEAST2 analysis). Species merged were *D. brauneri* - *D. saxicola*, *D. mixta*2 - *D. parvula* and the two *D. praticola* clades, resulting in 13 taxa plus the ougroup. Only one randomly selected sequence per taxon was used in BI runs. In MrBayes, all analyses started with randomly generated trees and ran for $30x10^6$, with sampling at intervals of 1000 generations, producing 30,000 trees. A GTR + I + Γ model (a General Time Reversible model with a gamma-shaped distribution of rates across sites) was chosen, so that conversion rates for each nucleotide pair were estimated, and no partitioning was assumed. Two independent runs were performed on each dataset. To guarantee analysis was performed on the stabilised likelihood, 40% of each total run was treated as burn-in.

Species tree estimates were performed with BEAST* 2 v2.4.4 (Bouckaert et al., 2014) and SVDQuartets (Chifman and Kubatko, 2014, 2015). The BEAST* package performs simultaneous gene tree and species tree inference under a combined coalescent and sequence evolution model, and for this analysis the most variable 31 markers were selected. The tMRCA was estimated using as prior the split between the ingroup (*Darevskia*) and the outgroup (*Iranolacerta*) estimated elsewhere (10 Myrs: Mendes et al., 2016). To improve the convergence of the runs, this split was calibrated ignoring uncertainty. Run convergence for BEAST* 2 was evaluated by ESS values above 200 for all model parameters and posterior tree samples were combined to produce maximum clade credibility species trees with the software Tree Annotator v2.4.4 (Bouckaert et al., 2014), with node heights set to mean age estimates. To confirm the species tree topology, SVDQuartets (Chifman and Kubatko, 2014, 2015) was used with the most variable 300 markers concatenated and run together. This is not a concatenated species tree inference, instead the method calculates the unrooted topology for quartets of taxa for each value of a matrix of site patterns (the matrix of concatenated SNPs), and later infers the species phylogeny using quartets methods (Chifman and Kubatko, 2015). The primary concordance tree calculated by Bucky is also a reliable estimate of the species tree under the coalescent model (Degnan et al., 2009), and it was also used for comparison.

Species structure and population analyses

To identify any misclassified individuals, a discriminant analysis of principal components (DAPC) using the ADEGENET package (Jombart et al., 2010) was used on the SNP data set. The optimum number of clusters was chosen with the find.clusters() option, by comparing the different clustering solutions using a Bayesian Information Criterion (BIC).

Genetic composition was further investigated with STRUCTURE (Pritchard et al., 2000), which also allowed us to estimate the level of admixture for each species. This software applies a Bayesian clustering algorithm to identify subpopulations in Hardy-Weinberg equilibrium, assigning individuals to different clusters (or populations) and estimating the population allele frequencies. A random SNP per marker was selected and ten runs were performed, each with a different combination of randomly selected SNPs. Values from the different runs were combined with Clumpp (Jakobsson and Rosenberg, 2007). The best *K* was selected using a combination of methods: Delta *K* (Evanno et al., 2005) and the ADEGENET best K approximation calculated with the BIC.

Tests of introgression

Bucky (Larget et al., 2010) analysis was performed on a set of 300 loci. Given the low variability of some markers, and thus low phylogenetic information, only the most variable markers out of the 616 re-captured were selected for the Bucky run. This software takes as input unique trees from the tree distribution calculated for the Bayesian phylogenetic inference for each locus in MrBayes, as detailed above. It examines the tree distribution for each marker used and, with a non-parametric Dirichlet process, clusters markers into groups that share the same topology. The Dirichlet process uses an ***a priori*** level of discordance α, which combines the information in the sequences of those genes that are inferred to be congruent (Chung and Ane, 2011). The primary concordance tree is constructed based on the inferred clades for the largest proportion of loci analysed, including the highest concordance factors (CFs) for each clade represented. This value corresponds to the proportion of genes that truly have the clade in their trees and is calculated not only for the primary concordance tree, but also for the clades which are found in the distribution of trees but are not represented in the concordance tree. If the CF is high enough, we can assume the discordance which lead to the high proportion of this clade is relevant and these taxa present signs of incomplete lineage sorting (ILS), introgressive hybridization or horizontal gene transfer (HGT). Since no assumptions are made by this software for the reasons of discordance, it should be used only to detect it and downstream comparative analyses should be undertaken in order to understand the causes of discordance.

Patterson's *D*-statistics (Durand et al., 2011) were used for formal tests of whether the pattern of shared variation between species provided by Bucky is better explained by gene flow than incomplete lineage sorting. These tests, also called ABBA/BABA tests, determine whether the number of derived alleles (B) shared by reciprocally monophyletic taxa is greater than expected under the null model of incomplete lineage sorting (ILS). Ancestral (A) and derived (B) alleles are identified by comparison with an outgroup taxon. The numbers of ABBA and BABA patterns are calculated in a group of four clades, two sister clades (P1 and P2), a putative admixed clade (P3) and the outgroup (O). If no differential gene flow has happened between one of the sister taxa (P1, P2) and the “admixed” taxon (P3) the numbers of ABBA and BABA patterns are expected to be the same. *D*-statistics were estimated with the evobiR package in R (https://github.com/coleoguy/evobir) which takes as input a fasta file with the concatenated sequences of each of the taxa (P1, P2, P3

and O) without any missing data. Different combinations of taxa were used to test evidence of introgression from previous analyses and, for each tested combination, individuals of each species were randomly arranged so that for each combination of taxa *D*-statistics were calculated for multiple sets of individuals. These combinations were made such that the *BEAST2 species tree relationships were respected. For each combination of taxa (hereafter referenced as a test), 150 random combinations of individuals were run and *D*-statistics estimated for each. Significant deviations from zero of the distribution of the estimated *D*-statistics per test were calculated with Student's t-test and a Benjamini-Hochberg correction (Benjamini and Hochberg, 1995) for multiple comparisons was used for the *p*-values. In the four-taxon tests a significant deviation from 0 of the *D*-statistic distribution indicated introgression between P1 and P3 (if the mean was < 0) or between P2 and P3 (if the mean was > 0), compared with the expectation from ILS alone.

In addition to the ABBA/BABA tests, formal tests with three-population (f_3) and four-population (f_4) statistics (Reich et al., 2009) were used to estimate admixture proportions in groups of populations (or taxa), testing for introgression. f_3 *and* f_4 statistics estimate correlations between allele frequencies of three (f_3) or four (f_4) populations joined by an unrooted tree. If the populations tested are unadmixed, the genetic drift (measured by the allele frequency as a proxy) should be uncorrelated. If correlation between populations is found, this is indicative of gene flow. The f_3 statistic (X; Y, W) tests for admixture between a test population X and two reference populations, Y and W. In case of no admixture, its expected value is positive, and negative in case of admixture between X and Y, X and W, or both X and Y/W. On the other hand, the f_4 statistic tests the correlation between allele frequency differences in an unrooted tree (A,B;C,D). It is equal to zero if there are no correlations, and no admixture across the tree branches, positive if admixture occurs between A and C, between B and D, or both, and negative if admixture occurs between A and D, B and C, or both. The statistical significance for both of these tests was assessed using a Z-score (f_3 *or* f_4 divided by its standard deviation), and a threshold Z-score of 3, which corresponds to a p-value of 0.01, was used. *threepop* and *fourpop* programs in the TreeMix package (v 1.13; (Pickrell and Pritchard, 2012)) were used to calculate f_3 and f_4 statistics, respectively, as well as Z-scores to assess the significance of the test values for all possible combinations of three and four taxa within *Darevskia.* These Z-scores were then transformed into P-values and significance was confirmed after a Benjamini-

Hochberg correction as described above.

To visualise the extent, directionality and time of gene flow the graph-based method developed in TreeMix was used. This approach uses allele frequencies at genome-wide polymorphisms and a Gaussian approximation of the genetic drift among populations to estimate population graphs with 0-21 migration/admixture events connecting populations. The number of admixture events was estimated by comparing the change in log likelihood for each additional event. The proportion of the variance in population covariances explained by each population graph with different number of admixture events was calculated, and the best model estimated from the plot of the changes in the percentage of explained variance. Following this, we added 4 and 11 migration edges to the tree and observed the estimated admixture events. For TreeMix, f_3 and f_4 a total of 5823 SNPs was used, missing data excluded. Only ingroup species were tested here: *Iranolacerta* was not included.

RESULTS

Mapping and SNP calling

From the 574,341 bases in the transcriptome reference alignment used (Freitas et al, in prep: Chapter 4), a total of 429,078 sites for 75 individuals were recovered from the mapping of capture sequencing reads and SNP calling analyses, of which 11,738 were indels, 2,156 were MNPs and 408,082 were SNPs. On average, 226 bp were missing from each reference contig, and the amount of recovered sequence was between 25 and 2893 bp per target (reference contigs ranged from 550 to 3740bp: Freitas et al, in prep). Two individuals, one outgroup and one ingroup, had to be removed due to very low read coverage. After filtering and resolution of MNPs, 433,168 sites were left corresponding to 64,150 SNPs. From the total 627 contigs targeted, 616 were used in this work. All contigs analysed were variable in this dataset (except one: Locus_9712), ranging from 1 to 379 SNPs in total (0.12 to 24%). A total of 45,959 positions showed a minor allele count of 2 or higher, i.e. they were phylogenetically informative.

Phylogenetic inference

Species tree topologies calculated by SVDQuartets (Figure S5.1), *BEAST2 (Figure 5.1) and the population tree from Bucky (Figure 5.2) were concordant for nearly all clades. Test

runs with fewer positions in SVDQuartets (Figure S5.1) showed some inconsistency in the position for *D. parvula*, which grouped with the Rudis clade (*D. rudis*, *D. valentini* and *D. portschinskii*) instead of with *D. mixta*/*D. clarkorum* clade for those SVDQuartets analyses (data not shown). Also, in the SVDQuartets tree the *D. mixta*/*D. clarkorum* clade branched with the main *D. raddei* clade, while in the remaining analyses (*BEAST2 and Bucky) it branched with *D. parvula*. SVDQuartets and Bucky do not allow for clock rate assumptions, so the resulting tree branch lengths were not estimated.

In the *BEAST2 species tree the most basal split within *Darevskia* is between the Rudis clade and the remaining species, dated to have happened approximately ~5 [3.9, 6.1] My ago (posterior mode [95% credible interval]). The next split divided the Saxicola clade (*D. praticola*, *D. brauneri* and *D. saxicola*) from the Caucasica clade (*D. raddei*, *D. mixta*, *D. parvula*, *D. chlorogaster*, *D. steineri*, *D. defilippi, D. clarkorum*, *D. derjugini*), at ~4.5 [3.5, 5.5] My ago.

The species tree inference (Figure 5.1) also showed that species with different habitat requirements, saxicolous (rocky habitat) and ground-dwelling (forest or grass habitats), can be found in all major *Darevskia* clades.

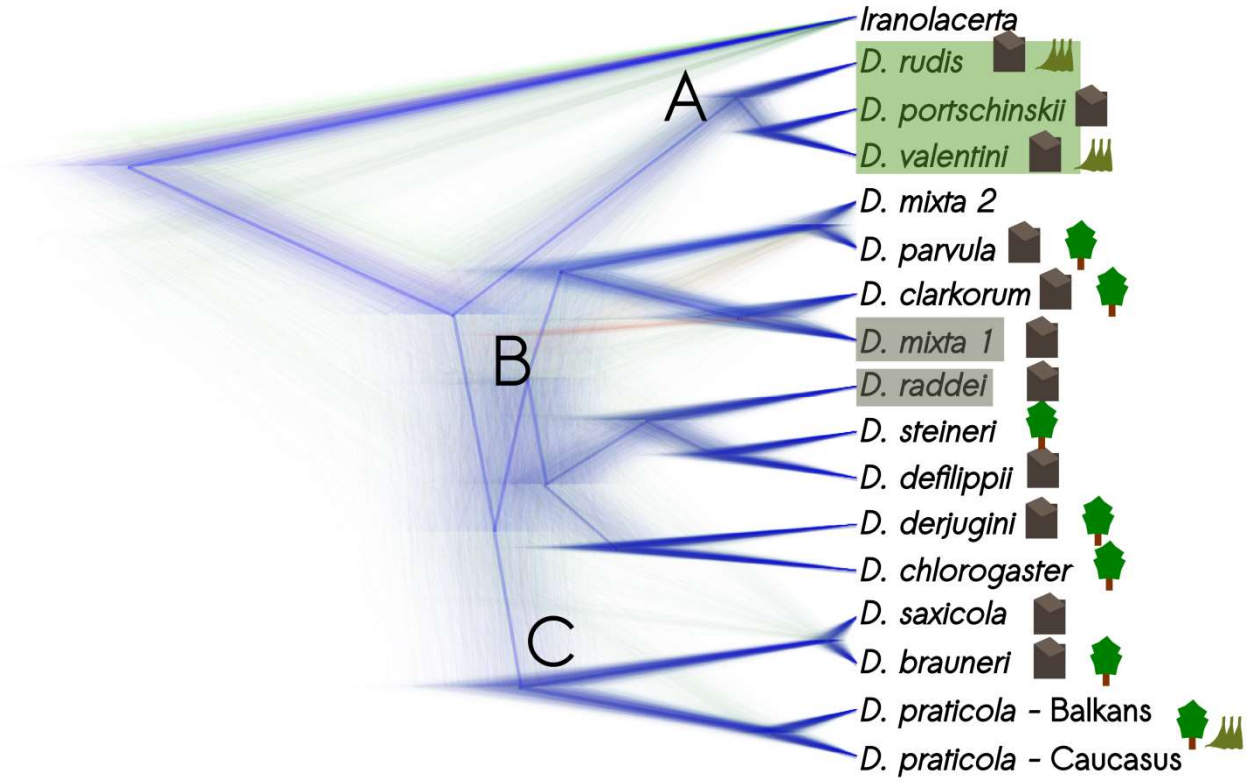


Figure 5.1 – Densitree representation of the *BEAST2 species tree inference for the sexual *Darevskia* species analysed in this study. Parthenogenetic parental species are highlighted, green the paternal species and grey the maternal species. The recovered clades are noted with letters, A the Rudis clade, B the Caucasica and C the Saxicola. Habitat requirements are identified for all species, saxicolous (grey block) and ground dwelling (tree for forest habitat, grass for grasslands). Branch colours correspond to the likelihood of the trees: blue is the most likely tree topology, red the second most likely topology and green the third most likely topology. Consensus tree is drawn as a single blue line above all the others.

Introgression tests

Bucky analysis showed alternative clades not represented in the primary concordance tree have higher concordance factor (CF) than the other combination of taxa, and thus reject the null hypothesis that the coalescent model is sufficient to explain the observed gene tree discordance. To explain this incongruence, alternative hypotheses, introgressive hybridization or ILS, have to be considered.

Evidence for incongruence (Figure 5.2) was found at different levels in the tree, not only between recently diverged taxa (e.g. *D. chlorogaster , D. raddei*), but also between major clade pairs (e.g. Rudis clade, *D. parvula*). No evidence for gene flow/ILS was found between any of the parental pairs known to have generated parthenogens.

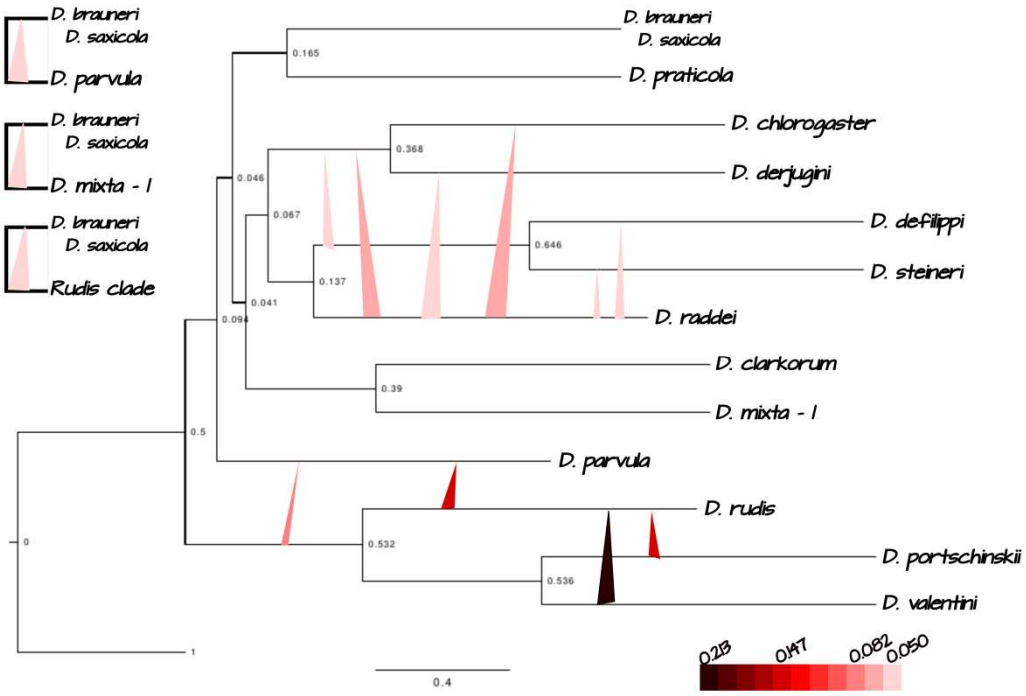


Figure 5.2 - Primary concordance tree inferred with Bucky. Insets correspond to significant clades not represented in the primary concordance tree, and colour correspondence to CF values is identified in the gradient.

D-statistics were estimated for several different combinations of species in order to test if Bucky gene tree discordance results are due to introgressive hybridization or ILS. From the 79 *D*-statistics tests, more than half were significantly different than zero (Table S5.1).
Regarding the f3 and *f4* formal tests, only the four-population values were significant. A significantly negative *f3* (X; Y, W) value indicates that the frequencies of alleles in population X tend to be intermediate between Y and W, which can only arise if population X resulted from a mixture of populations related to Y and W. However, if populations had enough time for drift to happen, or X had suffered bottlenecks after the admixture event, consistent with the complex demographic history of this genus, the *f3* statistic may not recover a signal of introgression (Decker et al., 2014; Kamdem et al., 2016). The *f4* formal tests provided additional evidences to widespread admixture in the genus, however because this test should be evaluated as admixture between either/or both species pairs tested, it was only used here as a confirmation of other admixture evidence (Figure S5.2).
D-statistics and *f4* formal tests showed that directional hybridization is widespread in the genus *Darevskia* (Figure 5.3). Even though most clades presented evidence for introgression, there was no evidence for directional hybridization between the parental pairs of the parthenogenetic hybrids, except *D*-statistics between *D. portschinskii* and *D. mixta* – 1. However, whenever the pair *D. portschinskii* and *D. mixta*-1 were analysed together with other species that have also presented evidence for gene flow with either one of these two, a significant test favoured gene flow between the third species and either *D. portschinskii* or *D. mixta* -1 (see tests 6, 70 and 76 as an example, in Table S5.1).

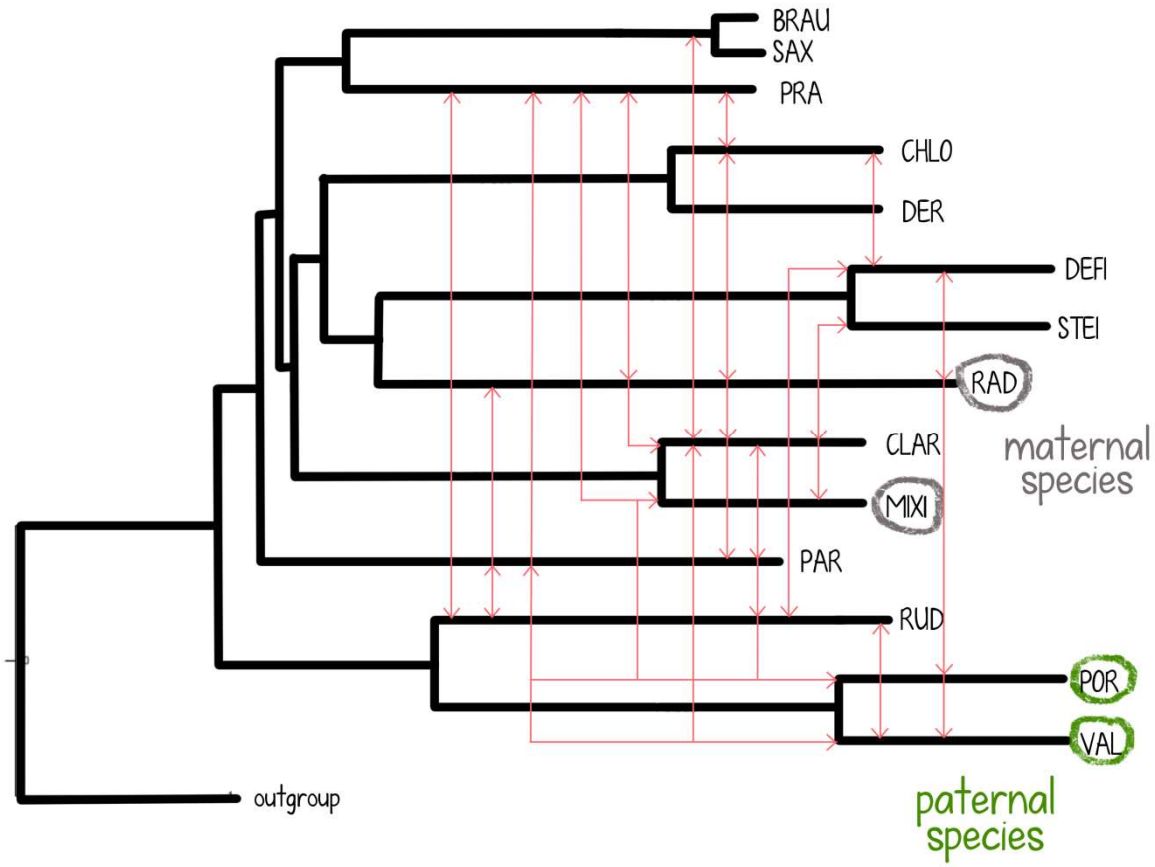


Figure 5.3 – Coloured vertical arrows depict ABBA/BABA significant relationships within *Darevskia* clades (for detailed information about all the significant and non-significant tests performed see Table S5.1). Species tree adapted from Bucky result.

TreeMix was used to construct a tree connecting the 15 taxa in order to describe and date admixture events. The variance explained by the TreeMix ML model without migration edges was high (98.9%, Figure S5.3), which means the main relationships between taxa can also be explained by a bifurcating tree (with no assumptions for gene flow). However, this value is in agreement with other studies that have used TreeMix to detect admixture between different taxa, even when different types of markers are used (e.g. RAD sequencing (Kamdem et al., 2016) and SNP data (Decker et al., 2014)). Most of the genetic affinities found by TreeMix (Figure 5.4) was consistent with the evidence for introgression found by both the formal tests and Bucky (comparison of all analyses of introgression in Figure S5.2). No evidence for introgression between the parthenogens'

parental pairs was found, even though there was an estimated introgression event *D. mixta* (maternal species) and the ancestral state of *D. valentini* and *D. portschinskii* (paternal species).

The TreeMix model also showed admixture events happened along the evolutionary time-scale of *Darevskia*, with some events happening more recently (such as the *D. raddei-D. chlorogaster* introgression) and others estimated to have happened before the split within its clades (such as the *D. praticola* – Caucasica clade). The *D.parvula* - Rudis clade migration edge falls after the split within the Rudis clade, and specifically between *D. portschinskii – D. parvula*. This analysis also showed, in concordance with all the others, a high number of introgressive events between the parthenogens' sexual parentals and other *Darevskia* species, but not within the parthenogens' parental pairs.
The values of the residual covariance between each species pair identify deviations the proposed model. In Figure 5.4, heat colours depict the residual covariance between each species pair and darker colours indicate taxa more closely related to each other than expected by the proposed model, which could mean that the bifurcating model together with the estimated migration events may still not fully explain the covariance among *Darevskia* species.

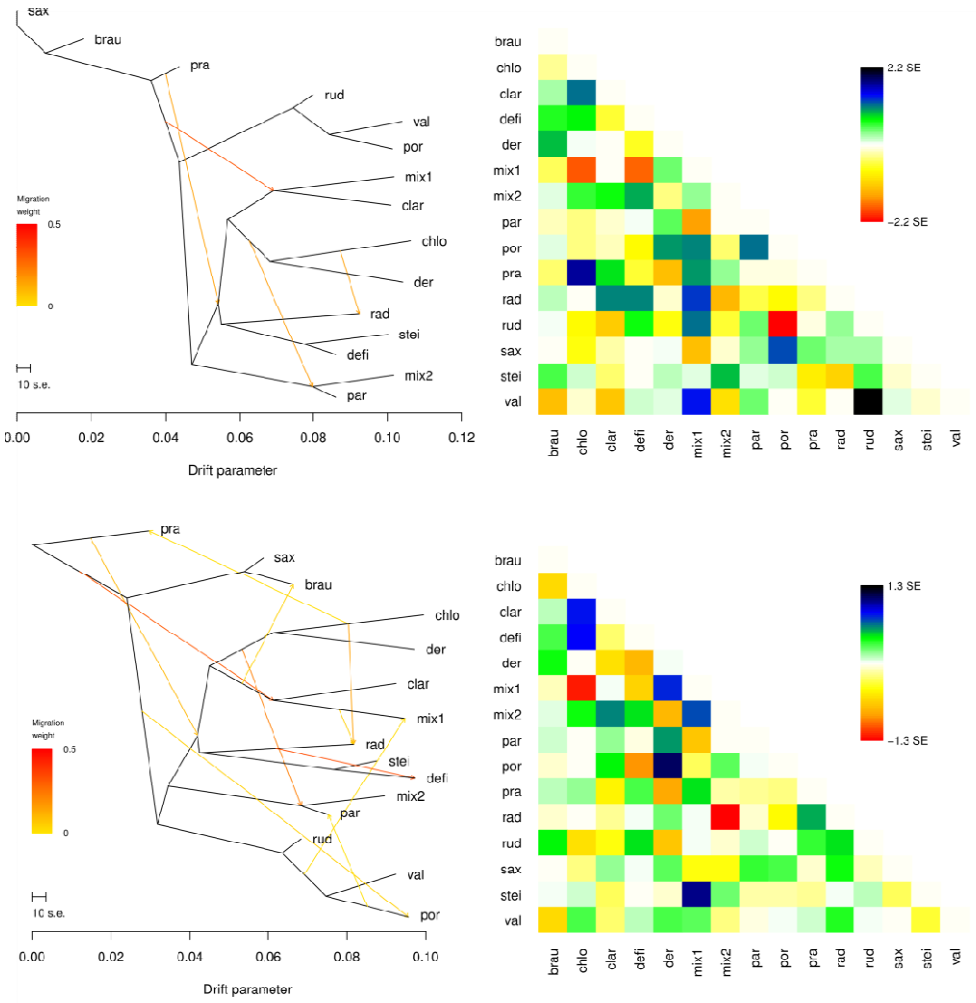


Figure 5.4 – TreeMix Maximum Likelihood (ML) trees estimated from the allelic frequency of 5823 SNPs. ML tree (left) and residual fit from the ML model (right) inferred with 4 migrations (top) and 11 migration edges (bottom). The arrows on the ML trees (left) indicate the directionality of gene flow migration edge and the colour of the edge reflect the intensity of admixture. The colours in the residual fit matrix represent the standard error depicted in each corresponding tree (light colours are close to 0, dark colours have higher standard error).

DISCUSSION

Recent studies on speciation and species divergence have not only challenged the conventional idea of the disruptive role hybridization has on divergence and speciation (Martin et al., 2015), but have also been showing how hybridization can accelerate speciation by adaptive introgression (Cronk and Yang, 2016; Pereira et al., 2014), or even originate new hybrid species (Alves et al., 2001; Trier et al., 2014). Enhanced with the advent of high-throughput sequencing techniques and the exponential increase of analytical power, which allowed for the analysis of cross genome markers in several individuals, introgressive hybridization has been acknowledged as a part of the evolutionary history of organisms (Abbott et al., 2016), with different parts of the genome presenting different permeability rates to gene flow.

In vertebrates, hybridization can frequently lead to the origin of parthenogenetic hybrid lineages, however it is not known whether this hybridization is a specific property of the parthenogens' parental species pair, or how hybrid asexuality affects gene flow among such groups. In this study we tested the rate and frequency of hybridization with gene flow in a group that suffered repeated secondary contact during divergence and that presents a reticulate evolutionary history (Freitas et al., 2016). Here we aimed to understand the role of hybrid asexuality on the general patterns of within group introgression: if hybrid asexuality is dependent on the genetic distance between the parental clades, then we expect gene flow only between young diverged clades, or ancient gene flow between the parthenogens' parental species pairs (until the origin event); on the other hand, if hybrid asexuality depends on the specificity of the species pairs, then gene flow will be found widespread except between the parental species pairs.

Evolutionary history of *Darevskia*

Taxonomic groups with highly dynamic species ranges, such as species inhabiting regions strongly affected by climate-induced geographic range shifts during the Pleistocene, are more prone to secondary contact and intermittent gene flow, because behavioural reproductive barriers are not likely to evolve while in allopatry (Christe et al., 2016). The Early Pliocene was marked by great orogenic and climatic changes in the Caucasus mountain ranges, leading to a regression of sea water and the formation of the modern outlines of the Black and Caspian Seas (Popov et al., 2006). This was induced by high

volcanic activity in the region, continuing until the Mid Pliocene (Tarkhnishvili et al., 2013). The first split within *Darevskia* was dated to ~5 M yrs (beginning of Pliocene), coincident with these strong changes. Speciation events in *Darevskia* seemed to have ranged between ~5 (beginning of Pliocene) to 0.3 (Pleistocene) My ago. However, given the lack of uncertainty associated to the calibration used for the species tree construction, interpretation of the age of the splits within *Darevskia* should be taken cautiously.

When speciation events happen in a few generations time scale, as in situations with abrupt climate changes, divergent lineages can retain shared ancient polymorphisms (and/or if they had large effective population size) (Takahashi et al., 2001). This incomplete lineage sorting (ILS) can lead to incongruence in the gene trees, which can mislead the detection of introgression hybridization. Because of that, a multifaceted approach has to be taken in order to distinguish the different degrees of incongruence that we can find in a group divergence continuum.

The present species tree matches the genus phylogeny previously proposed (Murphy et al., 2000), except for the position of *D. parvula* and *D. mixta*. In the available phylogeny produced with mitochondrial DNA (Cyt-*b* (Murphy et al., 2000) and ND4 (Freitas et al, in prep) and allozymes (Fu et al., 1997), *D. parvula* is placed within the Rudis clade, while in our study *D. parvula* is closer to the species of the Caucasica clade, even though early diverging. However, our analyses showed high degrees of introgression, identified by Bucky (Figure 5.2), and confirmed by ABBA/BABA tests (Figure 5.3), f_4 (Figure S5.2) and TreeMix (Figure 5.4), between *D. parvula* and the Rudis clade, as well as between *D. parvula* and the Caucasica clade, which could have lead to the incongruence in the position of this species in the different studies. The same happens with *D. mixta,* which instead of branching with *D. raddei* (Murphy et al., 1996) in our study is congruently found together with *D. clarkorum.* Like with *D. parvula, D mixta* also showed evidences for frequent hybridization with gene flow with other sexual species, specifically *D. raddei* (Figure 5.4), which could explain the incongruence found.

The species tree analysis also showed that the niche change from saxicolous habitat ground-dwelling environments happened more than once in *Darevskia* evolution, since neither of the habitat requirements was restricted to all species in each clade. The species *D. praticola*, *D. derjugini*, *D. steineri, D. clarkorum* and *D. chlorogaster* are ground-dwelling, while *D. raddei*, *D. defilippii*, *D. portschinskii* are saxicolous (Ahmadzadeh et al., 2013; Tarkhnishvili et al., 2013). But despite such habitat differences potentially limiting the

possibility for hybridization and gene flow between species with different habitat requirements, our results do not show any evidences that species with similar behaviours share higher levels of admixture between them. So, given these species are unlikely to have overlapped recently, either *Darevskia* species can rapidly change niche and adapt to new habitats or niche changes happened recently in *Darevskia* evolution and allopatric isolation still has not made a clear effect on the admixture.

Hybridization and gene flow

The different types of introgression analyses performed here behave differently regarding incomplete lineage sorting (ILS). While Bucky does not distinguish between ILS and gene flow (Larget et al., 2010), the formal tests (*D* and *f4* statistics*)* used are not expected to be affected by ILS, and if these tests are significant, gene flow must be the only responsible actor (Durand et al., 2011). Also, these formal tests are not sensitive to recombination, since they can detect gene flow even after recombination has partially diluted the introgressed genome. Bucky, on the other hand, since it is based on gene tree analyses, can be affected more intensively by recombination. If recombination breaks the introgressed sequence analysed, clades may not coalesce, because recombination breaks the sequence and also dilutes the signal of the introgression. Following this reasoning, while *D*-statistics and *f4* formal tests can detect both old and recent hybridization with gene flow events, Bucky can only detect gene flow events that have not been diluted by recombination, and thus, are more recent. To complement these analyses, TreeMix was used not only to confirm the patterns of introgression found, but also as a way to date the historical splits and admixture within *Darevskia*.

Together, all analyses show a high frequency of gene flow between all major clades, that have happened not only between recently diverged taxa, but also between more anciently diverged clades. TreeMix results, and Bucky vs formal tests, showed admixture events have happened along the divergence continuum during *Darevskia* evolution, even though the precise date and amount of introgressed alleles could not be measured with this dataset.

This widespread evidence of introgressive hybridization in *Darevskia* confirms the expectation for episodic secondary contact throughout evolution of this group. Even though *Darevskia* species occupy different habitats, they are distributed in very

heterogeneous landscapes, which could allow for species with different habitat uses to hybridize. Moreover, the absence of introgressive hybridization between some species pairs could be only due to biogeographic restrictions, since some species ranges do not overlap, rather than to genetic constraints. Evidence for introgressive hybridization between diverging clades has been found in several systems (Mavárez et al., 2006; Pereira et al., 2014; Stankowski and Streisfeld, 2015), however this is the first study to find such widespread patterns of deep branch gene flow in a vertebrate group with known recurrent hybridization, past and present.

Gene flow, asexuality and reproductive barriers in diverging clades

Despite the high number of introgressive events detected among *Darevskia,* no evidence of hybridization with gene flow was found between the parental species pairs of the asexual hybrids, *D. raddei* and *D. mixta* (maternal species) and *D. portschinskii* and *D. rudis/D. valentini* (paternal species), in Bucky, *D*-statistics, *f4* or TreeMix (with one exception, see Results). Contrary to other sexual species that might have never had the change to geographically overlap, and thus hybridize, the existence of the asexual hybrids is in itself an evidence for secondary contact and hybridization between these sexual species pairs. Thus, the lack of gene flow was prevented by reproductive barriers between the species pairs, and can be interpreted in the light of the Balance and Phylogenetic constraint hypotheses. When these taxa hybridized in the past, they originated a hybrid which was capable of reproducing asexually, but not sexually viable. Thus, it could not have acted as a vector of gene flow and represented the reproductive isolation between both clades (Janko et al., 2016). If the origin of parthenogenetic hybrids in vertebrates was due to the phylogenetic distance between parentals, we would find evidence of gene flow only between pairs of lineages younger than those that generated the parthenogens. However, we find gene flow not only between young lineages, but also deep branch gene flow, recent and past. On the other hand, if only some specific species pairs can produce parthenogens, we would find widespread gene flow in the genus, except between the parental species pairs before the origin of the parthenogens. Our study showed that *t*he hybrid parthenogenetic lineages are the only record of the past hybridization between the parthenogens parental species, and no evidence of gene flow (before or after that hybridization event) was found. Extant parthenogenetic *Darevskia* are expected to be relatively recent (~100 Kyrs, Freitas et al, 2016). Thus, the absence of any gene flow

between these parentals is most likely explained by the fact that every time they contacted they originated a hybrid that was not capable of sexually reproducing with any of the parentals. Even though both hypotheses are not mutually exclusive, the Phylogenetic Constraint Hypothesis better explains the origin of parthenogenetic hybrids and the patterns of gene flow found in this genus.

The expectations of the Phylogenetic Constraint Hypothesis, and the evidence provided by this genus, are that parthenogenetic hybrids can only be originated under rather special conditions. Given this, the scarcity of asexual lineages could be partially because of their low origination rate (Janko et al., 2008). Sexual reproduction is widespread in nature, and its higher potential for adaptation when in comparison to asexual reproductive species is generally used to explain its prevalence (Luijckx et al., 2017; McDonald et al., 2016). However, the reproductive potential asexual individuals present (Maynard Smith, 1971), the fact it has been shown that some lineages are known to have been reproducing solely asexually for millions of years (Judson and Normark, 1996), and their potential to occupy more diverse habitats than sexual relatives (Fontcuberta García-Cuenca et al., 2016) are an indication that a complementary explanation for the scarcity of aseexual lineages is needed. If the conditions that lead to the origin of asexual lineages are so particular that only rarely can happen in nature, more than their low potential to adaptation the rarity of asexual lineages could simply be due to low origination rates.

In this study we found extensive evidence of episodic introgression between the different clades in *Darevskia*, both anciently and recently diverged, with the exception of the species pairs that originated the parthenogenetic hybrids. Regardless of the reproduction type, this study showed the depth of hybridization and gene flow during taxa divergence in *Darevskia*, contributing to the rising number of evidences that species divergence and gene flow are not two exclusive events (Mavárez et al., 2006; Pereira et al., 2014; Stankowski and Streisfeld, 2015). The low evidence for introgression within the parental pairs of the parthenogens suggests these species, if they were in contact more than once in the past, could only have originated hybrids that were not capable of sexually reproducing with the parentals, and thus not being able to vector gene flow between them. This study highlights the evidence for hybrid asexuality acting as a reproductive isolation barrier that arises before other pre-zygotic mechanisms, and the specificity of the conditions that surround the origin of hybrid asexuality in vertebrates, contributing to its low

frequency in nature.

REFERENCES

Abbott, R., Albach, D., Ansell, S., Arntzen, J.W., Baird, S.J.E., Bierne, N., Boughman, J., Brelsford, A., Buerkle, C.A., Buggs, R., et al. (2013). Hybridization and speciation. J. Evol. Biol. *26*, 229–246.

Abbott, R.J., Barton, N.H., and Good, J.M. (2016). Genomics of hybridization and its evolutionary consequences. Mol. Ecol. n/a – n/a.

Ahmadzadeh, F., Flecks, M., Carretero, M.A., Mozaffari, O., Böhme, W., Harris, D.J., Freitas, S., and Rödder, D. (2013). Cryptic Speciation Patterns in Iranian Rock Lizards Uncovered by Integrative Taxonomy. PLOS ONE *8*, e80563.

Alves, M.J., Coelho, M.M., and Collares-Pereira, M.J. (2001). Evolution in action through hybridisation and polyploidy in an Iberian freshwater fish: a genetic review. Genetica *111*, 375–385.

Avise, J.C. (2008). Clonality: the genetics, ecology, and evolution of sexual abstinence in vertebrate animals (Oxford ; New York: Oxford University Press).

Benjamini, Y., and Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. J. R. Stat. Soc. Ser. B Methodol. *57*, 289–300.

Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C.-H., Xie, D., Suchard, M.A., Rambaut, A., and Drummond, A.J. (2014). BEAST 2: A Software Platform for Bayesian Evolutionary Analysis. PLOS Comput. Biol. *10*, e1003537.

Chifman, J., and Kubatko, L. (2014). Quartet Inference from SNP Data Under the Coalescent Model. Bioinformatics *30*, 3317–3324.

Chifman, J., and Kubatko, L. (2015). Identifiability of the unrooted species tree topology under the coalescent model with time-reversible substitution processes, site-specific rate variation, and invariable sites. J. Theor. Biol. *374*, 35–47.

Christe, C., Stölting, K.N., Paris, M., Fraïsse, C., Bierne, N., and Lexer, C. (2016). Adaptive evolution and segregating load contribute to the genomic landscape of divergence in two tree species connected by episodic gene flow. Mol. Ecol.

Chung, Y., and Ane, C. (2011). Comparing Two Bayesian Methods for Gene Tree/Species Tree Reconstruction: Simulations with Incomplete Lineage Sorting and Horizontal Gene Transfer. Syst. Biol. *60*, 261–275.

Cronk, Q., and Yang, J.Y. (2016). Hybridization between pollination syndromes as an ecological and evolutionary resource. Mol. Ecol. *25*, 5827–5829.

Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T., et al. (2011). The variant call format and VCFtools. Bioinformatics *27*, 2156–2158.

Danielyan, F., Arakelyan, M., and Stepanyan, I. (2008). Hybrids of Darevskia valentini, D. armeniaca and D. unisexualis from a sympatric population in Armenia. Amphib.-Reptil. *29*, 487–504.

Darevsky, I. (1967). Rock lizards of the Caucasus: systematics, ecology and phylogenesis of the polymorphic groups of Caucasian rock lizards of the subgenus Archaeolacerta. Leningrad. Nauka.

Decker, J.E., McKay, S.D., Rolf, M.M., Kim, J., Molina Alcalá, A., Sonstegard, T.S., Hanotte, O., Götherström, A., Seabury, C.M., Praharani, L., et al. (2014). Worldwide Patterns of Ancestry, Divergence, and Admixture in Domesticated Cattle. PLoS Genet. *10*, e1004254.

Degnan, J.H., DeGiorgio, M., Bryant, D., and Rosenberg, N.A. (2009). Properties of Consensus Methods for Inferring Species Trees from Gene Trees. Syst. Biol. syp008.

Durand, E.Y., Patterson, N., Reich, D., and Slatkin, M. (2011). Testing for Ancient Admixture between Closely Related Populations. Mol. Biol. Evol. *28*, 2239–2252.

Evanno, G., Regnaut, S., and Goudet, J. (2005). Detecting the number of clusters of individuals using the software structure: a simulation study. Mol. Ecol. *14*, 2611–2620.

Fontcuberta García-Cuenca, A., Dumas, Z., and Schwander, T. (2016). Extreme genetic diversity in asexual grass thrips populations. J. Evol. Biol. *29*, 887–899.

Freitas, S., Rocha, S., Campos, J., Ahmadzadeh, F., Corti, C., Sillero, N., Ilgaz, Ç., Kumlutaş, Y., Arakelyan, M., Harris, D.J., et al. (2016). Parthenogenesis through the ice ages: A biogeographic analysis of Caucasian rock lizards (genus Darevskia). Mol. Phylogenet. Evol. *102*, 117–127.

Fu, J., Murphy, R.W., and Darevsky, I.S. (1997). Toward the phylogeny of caucasian rock lizards: implications from mitochondrial DNA gene sequences (Reptilia: Lacertidae). Zool. J. Linn. Soc. *120*, 463–477.

Garrison, E., and Marth, G. (2012). Haplotype-based variant detection from short-read sequencing. ArXiv Prepr. ArXiv12073907.

Huelsenbeck, J.P., and Ronquist, F. (2001). MRBAYES: Bayesian Inference of phylogenetic trees. Bioinformatics *17*, 754–755.

Jakobsson, M., and Rosenberg, N.A. (2007). CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. Bioinformatics *23*, 1801–1806.

Janko, K., Drozd, P., Flegr, J., and Pannell, J.R. (2008). Clonal Turnover Versus Clonal Decay: A Null Model for Observed Patterns of Asexual Longevity, Diversity and Distribution. Evolution *62*, 1264–1270.

Janko, K., Pačes, J., Wilkinson-Herbots, H., Costa, R.J., Röslein, J., Drozd, P., Iakovenko, N., Rídl, J., Kočí, J., Reifová, R., et al. (2016). Hybrid asexuality as a primary reproductive barrier: on the interconnection between asexuality and speciation.

Jombart, T., Devillard, S., and Balloux, F. (2010). Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. BMC Genet. *11*, 94.

Judson, O.P., and Normark, B.B. (1996). Ancient asexual scandals. Trends Ecol. Evol. *11*, 41–46.

Kamdem, C., Fouet, C., Gamez, S., and White, B. (2016). Genomic signatures of introgression at late stages of speciation in a malaria mosquito. bioRxiv.

Larget, B.R., Kotha, S.K., Dewey, C.N., and Ané, C. (2010). BUCKy: Gene Tree / Species Tree Reconciliation with Bayesian Concordance Analysis. Bioinformatics btq539.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and Subgroup, 1000 Genome Project Data Processing (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078–2079.

Luijckx, P., Ho, E.K.H., Gasim, M., Chen, S., Stanic, A., Yanchus, C., Kim, Y.S., and Agrawal, A.F. (2017). Higher rates of sex evolve during adaptation to more complex environments. Proc. Natl. Acad. Sci. 201604072.

Lunter, G., and Goodson, M. (2011). Stampy: A statistical algorithm for sensitive and fast mapping of Illumina sequence reads. Genome Res. *21*, 936–939.

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal *17*, pp. 10–12.

Mavárez, J., Salazar, C.A., Bermingham, E., Salcedo, C., Jiggins, C.D., and Linares, M. (2006). Speciation by hybridization in Heliconius butterflies. Nature *441*, 868–871.

Maynard Smith, J. (1971). The origin and the maintenance of sex. In Group Selection, (Chicago: Aldine-Atherton), pp. 163–175.

McDonald, M.J., Rice, D.P., and Desai, M.M. (2016). Sex speeds adaptation by altering the dynamics of molecular evolution. Nature *531*, 233–236.

Mendes, J., Harris, D.J., Carranza, S., and 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. Mol. Phylogenet. Evol. *100*, 254–267.

Moritz, C., Brown, W.M., Densmore, L.D., Vyas, S., Donnellan, S., Adams, M., and Baverstock, P. (1989). Genetic diversity and the dynamics of hybrid parthenogenesis in Cnemidophorus (Teiidae) and Heteronotia (Gekkonidae). In Evolution and Ecology of Unisexual Vertebrates, (Albany),.

Murphy, R. (2000). A fine line between sex and unisexuality: the phylogenetic constraints on parthenogenesis in lacertid lizards. Zool. J. Linn. Soc. *130*, 527–549.

Murphy, R.W., Darevsky, I.S., MacCulloch, R.D., Fu, J., and Kupriyanova, L.A. (1996). Evolution of the bisexual species of Caucasian rock lizards: a phylogenetic evaluation of

allozyme data. Russ. J. Herpetol. *3*, 18–3.

Pereira, R.J., Barreto, F.S., and Burton, R.S. (2014). Ecological Novelty by Hybridization: Experimental Evidence for Increased Thermal Tolerance by Transgressive Segregation in Tigriopus Californicus. Evolution *68*, 204–215.

Pickrell, J.K., and Pritchard, J.K. (2012). Inference of Population Splits and Mixtures from Genome-Wide Allele Frequency Data. PLoS Genet *8*, e1002967.

Popov, S.V., Shcherba, I.G., Ilyina, L.B., Nevesskaya, L.A., Paramonova, N.P., Khondkarian, S.O., and Magyar, I. (2006). Late Miocene to Pliocene palaeogeography of the Paratethys and its relation to the Mediterranean. Palaeogeogr. Palaeoclimatol. Palaeoecol. *238*, 91–106.

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

Reich, D., Thangaraj, K., Patterson, N., Price, A.L., and Singh, L. (2009). Reconstructing Indian population history. Nature *461*, 489–494.

Ronquist, F., and Huelsenbeck, J.P. (2003). MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics *19*, 1572–1574.

Sambrook, J., and Russell, D.W. (2001). Molecular Cloning: A Laboratory Manual (CSHL Press).

Soltis, P.S. (2013). Hybridization, speciation and novelty. J. Evol. Biol. *26*, 291–293.

Stankowski, S., and Streisfeld, M.A. (2015). Introgressive hybridization facilitates adaptive divergence in a recent radiation of monkeyflowers. Proc R Soc B *282*, 20151666.

Takahashi, K., Terai, Y., Nishida, M., and Okada, N. (2001). Phylogenetic Relationships and Ancient Incomplete Lineage Sorting Among Cichlid Fishes in Lake Tanganyika as Revealed by Analysis of the Insertion of Retroposons. Mol. Biol. Evol. *18*, 2057–2066.

Tarkhnishvili, D., Murtskhvaladze, M., and Gavashelishvili, A. (2013). Speciation in Caucasian lizards: climatic dissimilarity of the habitats is more important than isolation time. Biol. J. Linn. Soc. *109*, 876–892.

Trier, C.N., Hermansen, J.S., Sætre, G.-P., and Bailey, R.I. (2014). Evidence for Mito-Nuclear and Sex-Linked Reproductive Barriers between the Hybrid Italian Sparrow and Its Parent Species. PLoS Genet. *10*, e1004075.

CHAPTER 6
GENERAL DISCUSSION

Figure 6.0 - a very high and foggy mountain (Armenia).

Evolutionary History of *Darevskia:* new insights

Darevskia lizards are a group of rock lizards found throughout the Caucasus and adjacent regions, and the Balkans. The Caucasus mountain range is marked by highly irregular topology, with two mountain massifs, Greater and Lesser Caucasus, and several plateaus and planes. This region has been strongly affected by climate-induced changes during time, and species inhabiting the Caucasus consequently present highly dynamic species ranges with several geographic range shifts throughout geological times. Extreme orogenic changes in the Caucasus mountain ranges happened during the Early Pliocene, leading to a regression of sea level and the formation of the modern outlines of the Black and Caspian Seas (Popov et al., 2006). This was induced by high volcanic activity in the region, continuing until the Mid Pliocene (Tarkhnishvilli, 2014). In the Pleistocene cyclical ice-ages, organisms were forced to contract during colder periods, and were able to expand when the temperatures rose, opening the possibility of constant geographic range shifts, secondary contact and possibly prone to hybridization and gene flow, since behavioural barriers are not likely to evolve while in allopatry (e.g. Christe et al., 2016).

Most *Darevskia* species are found in the Caucasus and adjacent regions, except for one, *D. praticola* sensu lato*,* which has a disjoint distribution range, found in the Caucasus region but is also the only *Darevskia* representative in Europe, more specifically in the Balkan region. Its disjoint distribution is another evidence of the intense effects climate-induced changes had over species ranges, and more specifically on *Darevskia*. Our results show this separation likely has happened during the Pliocene-Pleistocene transition, when the Balkans and Asia Minor became connected (Kerey et al., 2004). During subsequent Pleistocene glacial cycles the sea level fluctuated and the Black Sea connection to the Mediterranean Sea (via Marmara Sea) was frequently discontinued (or at least significantly restrained) (Popov et al., 2006). Only during the late Pleistocene the forests could expand again (Willis et al., 1999), after the aridification of the climate during the interglacials (Esin et al., 2014). These events were responsible for the vicariance in *D. praticola,* and similar paleoclimatic scenarios have already been proposed for lineage diversification in other reptile and amphibian species in this region (Ahmadzadeh et al., 2013a, 2013b; Wielstra et al., 2013).

In the phylogenetic reconstruction of the genus 14 species were included. This analyses confirmed the three main phylogenetic clades in *Darevskia* previously described (Murphy et al., 2000), despite more species being included here. Only one species presented a

different position in the phylogenetic arrangement, *D. parvula*, which instead of grouping in the Rudis clade, was closer to the Caucasica clade. This incongruence was likely due to the high levels on introgression, and possibly incomplete lineages sorting(ILS), found between this species and the species from the rudis clade, and others from the Caucasica clade. In rapid speciation events, and/or when populations with large effective population size (N_e) diverge,the segregation of alleles may not happen simultaneously, leading to gene trees to present different evolutionary histories for the same system. Because here we are using multiple markers, we can provide a more accurate estimate of phylogeny despite incomplete lineage sorting, or even other effects that could mask the phylogeny such as selection or genetic drift (Maddison et al., 2006).

In relation to their habitat requirements, *Darevskia* species can be saxicolous or ground dwelling. Even though the ancestral state could not be recovered, phylogenetic inference also shows that niche shift from saxicolous to ground-dwelling happened more than once throughout *Darevskia*'s evolution, since neither of the habitat requirements is restricted to all species of any given clade. Even though habitat differences could limit the potential for interspecific hybridization, the results reported here do not show any evidence that species with similar behaviours share higher levels of admixture between them. Examples of ground-dwelling species (eg *D. chlorogaster*) and saxicolous (*D. raddei*) (Ahmadzadeh et al., 2013) that are not likely to overlap in range in the present but still show high levels of introgression between them, suggest that *Darevskia* sexual species can rapidly change niche and adapt to new habitats, or niche changes happened recently in *Darevskia* evolution and allopatric isolation still has not made a clear effect on the admixture.

Finally, parthenogenetic species are not frequently in sympatry with their sexual parentals, or with any other sexual species, despite the similarity on niche requirements between parthenogens and their parentals. At a geographic scale, the suitable area of *D. unisexualis* predicted by ecological niche modelling widely overlaps that of *D. raddei* sensu lato. This niche overlap (Figure 6: Chapter 2 – Freitas et al, 2016), together with the wide distribution range of *D. unisexualis* (despite its recent origin) and the evidences for population expansion, suggest this parthenogen may even outcompete its maternal species within its range. At a local scale, this has been already demonstrated for another parthenogenetic species, *D. dahli*, which has been shown to outcompete its sexual maternal species, *D. mixta* (Tarkhnishvili et al., 2010).

Origin of asexuality in *Darevskia*

Parthenogenesis is a rare reproductive mode that, despite being found in most animal groups, is observed in less than 0.1% of known vertebrate species (Vrijenhoek, 1989). Given the low number of species, the switch from sexual to parthenogenetic reproduction is expected to happen only rarely. Most sexual-parthenogenetic complexes in vertebrates show a multiple and polyphyletic origin of parthenogenetic lineages (Crease et al., 1989; Grismer et al., 2014; Simon et al., 2003), where parthenogenesis has evolved more than once. In non-vertebrates asexuality is generally originated from spontaneous thelytokous parthenogenesis (Bullini, 1994), but most parthenogenetic vertebrates are originated by interspecific hybridization (Kearney et al., 2009).

The reassessment of the parentage inferences (Chapter 3: Freitas et al, in prep) confirms the hybrid genetic profile of parthenogenetic *Darevskia* and that they originate from the directional hybridization between specific sexual species pairs. Only the hybridization between these sexuals resulted in successful parthenogenetic hybrids, which occurred multiple times during a single temporal event, possibly ~100,000 yrs ago, during the Pleistocene (Chapter 2 - Freitas et al., 2016; Chapter 3 – Freitas et al, in prep). The cyclical Pleistocene ice ages and subsequent expansion-contraction of organisms in a very heterogeneous habitat allowed for the secondary contact of sexual *Darevskia* lineages in incomplete stages of reproductive isolation (Vrijenhoek, 1989). This paved the way for repeated hybridization events in separate geographical areas that eventually originated hybrids that could not reproduce sexually, but instead were able to reproduce parthenogenetically.

The maintenance of asexual species depends on the balance between the generation of asexual lineages from sexual progenitors and their loss. Many evolutionary occurrences can be involved in the extinction of asexual lineages, such as mutation accumulation (Haigh, 1978) or failure to adapt (Lively, 2010), but recent studies have also shown asexual lineage loss can happen simply by neutral processes (Janko, 2014; Schwander and Crespi, 2009), contradicting the general belief that parthenogenetic lineages are short lived only due to the constraints of asexuality.

The study of the origin of different parthenogenetic vertebrates has motivated the construction of two hypothesis addressing how hybridization can originate asexual reproduction in this group. The Balance Hypothesis, impelled by the study of *Aspidoscelis* (previously *Cnemidophorus*) and *Heteronotia* systems, suggests that parthenogenetic

vertebrates mostly arise by the hybridization of two sexual species divergent enough to disrupt meiosis in the hybrids, yet not so divergent as to seriously compromise hybrid viability or fertility (Kearney et al., 2009). The Phylogenetic Constraint Hypothesis, impelled by the study of *Darevskia,* suggests parthenogenetic lineages are originated by the hybridization between specific pairs of sexual species, which likely present genetic peculiarities that allow them to interbreed and produce hybrids capable of reproducing parthenogenetically (Avise, 2008).

The results reported here show that only a few sexual species pairs were responsible for the origin of the parthenogens, that some species acted always as the maternal species and others always as the paternal, that the phylogenetic distances between each pair are included in a wide range interval, and that introgression is found widespread across the genus but absent between the parental pairs. This evidence is more congruent with the Phylogenetic Constraint Hypothesis. Even though other parthenogenetic hybrid vertebrate models also suggest specificity in the sexual parental pairs is more relevant than the phylogenetic distance between them (Grismer et al., 2014; Janko et al., 2003), the Balance Hypothesis is still the most widely accepted explanation for the origin of vertebrate asexuality (Kearney et al., 2009).

Until the discovery of *Darevskia* parthenogenetic species almost 60 years ago, parthenogenesis was thought to be restricted to non-vertebrate species. This strongly impacted our interpretation of asexuality in nature and the notions of the evolution of sexual reproduction. However, we are still discussing and trying to understand how asexuality originates in vertebrates, its implications and consequences. The fact that the Balance Hypothesis is the most common explanation on the origin of parthenogenesis, despite the evidence of other forces that could be more influential than phylogenetic distance alone, leads scientists to focus on the wrong questions and miss the point to study. It is important to collect more evidence in other parthenogenetic vertebrates and test whether the *Darevskia* system provides a good model that can be used to understand the origin of asexuality in vertebrates.

Methodology: advantages and caveats

The recent appearance of next generation sequencing (NGS) techniques allowed for the generation of a very large amount of data, with considerable low initial economic

investment (Ekblom and Galindo, 2011). The work presented in this thesis is representative of the intense transformation evolutionary genetic analysis methods have suffered in the last couple of decades. From Sanger sequencing and microsatellites genotyping, to NGS methods including RNA seq and capture sequence analysis, the advantages and caveats for each type of analyses were tested.

Sanger sequencing and microsatellite genotyping generate markers in a highly targeted and straightforward way. These methods have been widely used in phylogenetic and population genetic studies, are easy to apply cross species and use in comparative studies (due to the high amount of data already produced) and have been tested extensively. These type of markers can still prove to be a cost-effective and quick method, as demonstrated by the effectiveness of ploidy detection using cross-species microsatellites (Chapter 3: Freitas et al, in prep), but also for general surveys of many individuals (Mira et al., 2017), species assignment and discrimination (Harrison et al., 2014), just to state some examples. However, the analyses of only a restricted number of markers can also depict a very biased account of the evolutionary history of the model studied.

NGS techniques allow the study of several individuals and markers cross-genome in a fast and cost-effective way (Carstens et al., 2012; Metzker, 2010). However, with many NGS methods there is less control over which regions of the genome are sequenced, and if those regions are homologous across the individuals studied and also phylogenetically informative (Carstens et al., 2012). Considering our goal was to analyse interspecific relationships recurring to phylogenetic inferences, this was a strong caveat that had to be taken into account.

To overcome this problem, RNA sequencing and transcriptome de novo assembly was used to look for several markers that are long, phylogenetically informative and conserved enough that would be able to not only be used to sequence several different regions of the genome, but also that could be used across the species of the genus (and also in the outgroup) so that phylogenetic analyses could be performed downstream.

This approach proved to be successful, and provided new insights into the evolutionary history of *Darevskia*, such as the high levels of interspecific gene flow found, that would have been more difficult to achieve if still analysing one locus per individual at a time.

Gene flow between sexual clades

In vertebrates, and more specifically in *Darevskia,* interspecific hybridization can originate

new hybrid parthenogenetic species. However, how the hybrid origin of parthenogenesis fits the overall history of hybridization within the group it is still not understood. If sexual species are known to have hybridised in the past, originating the parthenogenetic vertebrates, they are also likely to have hybridised not only to the extent of originating new parthenogenetic species, but possibly allowing for fertile sexual hybrids to vector gene flow between different clades.

Tests for introgression within the genus show a high degree of gene flow between all major clades. Due to the extensive hybridization found at present, either between parthenogenetic and sexual species (Chapter 3: Freitas et al, in prep), or between different sexual species (Arakelyan and Danielyan, 2011), some recent gene flow could be expected. Results presented here show recent gene flow not only between closely related species, but also deep branch gene flow, both recent and in the past between distantly related clades. This introgression between taxa with deep divergence, at least in some examples such as the case of *D. parvula* and the rudis clade species, happened after a period of divergence.

Several evidences of hybridization with gene flow between diverging clades have been found in other systems (Cronk and Yang, 2016; Pereira et al., 2014; Stankowski and Streisfeld, 2015). However, such widespread patterns where introgression has happened between all major clades, and in a very high number of species pairs, either recently or anciently diverged, are not frequently found. In groups with sexual-parthenogenetic reproducing species, pre- or post-zygotic reproductive barriers could be more difficult to attain and hybrid asexuality can, thus, effectively create reproductive barriers between diverging sexual taxa, that were likely faced with recurrent secondary contact.

Models such as *Darevskia* can help to understand not only the significance of sex and asexual reproduction, but also how species can escape the evolutionary constraints of sexual reproduction and originate organisms that reproduce asexually.

Final remarks

Parthenogenetic species have some advantages over sexual species, they can reproduce twice as fast as the sexual species (Maynard Smith, 1978) and are not affected by other costs of sex (Lehtonen et al., 2012). In the short term, this may provide an advantage when in sympatry with sexual species (Burke et al., 2015; Tarkhnishvili et al., 2010). However, sexual reproduction has been shown to be more successful in changing

environments and is known as a driver of evolution and speciation. In an ever-changing environment, with several different pressures, asexual species which lack the recombination benefits of sexual reproduction are expected to be at a disadvantage when in competition with sexual species.

If asexuality is so disadvantageous in relation to sexual reproduction, we would expect sexual species were positively selected and asexual reproduction extinct from Nature. Even if sexual vertebrates outnumber parthenogenetic ones, asexuality in vertebrates is still found recurrently in many groups. Despite their relative abundance and phylogenetic distribution, we have very reduced information about the mechanisms that are relevant in the origin and evolution and parthenogenetic vertebrates.Given the lack of recombination and higher reproduction rate, asexual species are likely to show different patterns of genomic evolution than sexuals. The mutation rate, transposable elements (TEs) propagation and mutation accumulation will not act the same way in a genome that can recombine and lose alleles and allelic combinations in one generation, to other that will pass on the same allelic combination as a bloc to their descendants.

Future research in *Darevskia* should focus on the study of the differential evolution between sexual and asexual genomes. To do that, capture sequence data on the same markers already sequenced on the sexual species should be phased and mutation rate estimated. Several theoretical studies have shown a increase of the substitution rate in asexual organisms, when comparing to their sexual relatives, in the presence of clonal interference. Some studies have already provided indirect empirical evidences that substitution rate is faster in asexuals in relation to their sexual relatives, showing longer branch length in phylogenetic inferences (Tucker et al., 2013) or even higher number of mutation accumulation after the same number of generations in a changing environment (McDonald et al., 2016). However, no empirical study addressed so far the rate of substitution directly of sexual and asexual relatives, evolving in the same conditions.

Studying asexual species in the context of their sexual relatives can help to understand the effect of asexuality (or the absence of sex) on the genome. *Darevskia* lizards provide a promising model for the study of the evolution of asexuality and why sexual reproduction is so widespread in the tree of life.

REFERENCES

Ahmadzadeh, F., Flecks, M., Rödder, D., Böhme, W., Ilgaz, Ç., Harris, D.J., Engler, J.O., Üzüm, N., and Carretero, M.A. (2013a). Multiple dispersal out of Anatolia: biogeography and evolution of oriental green lizards. Biol. J. Linn. Soc. *110*, 398–408.

Ahmadzadeh, F., Flecks, M., Carretero, M.A., Mozaffari, O., Böhme, W., Harris, D.J., Freitas, S., and Rödder, D. (2013b). Cryptic Speciation Patterns in Iranian Rock Lizards Uncovered by Integrative Taxonomy. PLOS ONE *8*, e80563.

Arakelyan, M., and Danielyan, F. (2011). Syntopic populations of Darevskia biparental species in Armenia and Nagrono-Karabakh. (Luxembourg and Germany).

Avise, J.C. (2008). Clonality: the genetics, ecology, and evolution of sexual abstinence in vertebrate animals (Oxford ; New York: Oxford University Press).

Bullini, L. (1994). Origin and evolution of animal hybrid species. Trends Ecol. Evol. *9*, 422–426.

Burke, N.W., Crean, A.J., and Bonduriansky, R. (2015). The role of sexual conflict in the evolution of facultative parthenogenesis: a study on the spiny leaf stick insect. Anim. Behav. *101*, 117–127.

Carstens, B., Lemmon, A.R., and Lemmon, E.M. (2012). The Promises and Pitfalls of Next-Generation Sequencing Data in Phylogeography. Syst. Biol. *61*, 713–715.

Christe, C., Stölting, K.N., Paris, M., Fraïsse, C., Bierne, N., and Lexer, C. (2016). Adaptive evolution and segregating load contribute to the genomic landscape of divergence in two tree species connected by episodic gene flow. Mol. Ecol.

Crease, T.J., Stanton, D.J., and Hebert, P.D.N. (1989). Polyphyletic Origins of Asexuality in Daphnia pulex. II. Mitochondrial-DNA Variation. Evolution *43*, 1016–1026.

Cronk, Q., and Yang, J.Y. (2016). Hybridization between pollination syndromes as an ecological and evolutionary resource. Mol. Ecol. *25*, 5827–5829.

Ekblom, R., and Galindo, J. (2011). Applications of next generation sequencing in

molecular ecology of non-model organisms. Heredity *107*, 1–15.

Freitas, S., Rocha, S., Campos, J., Ahmadzadeh, F., Corti, C., Sillero, N., Ilgaz, Ç., Kumlutaş, Y., Arakelyan, M., Harris, D.J., et al. (2016). Parthenogenesis through the ice ages: A biogeographic analysis of Caucasian rock lizards (genus Darevskia). Mol. Phylogenet. Evol. *102*, 117–127.

Grismer, J.L., Bauer, A.M., Grismer, L.L., Thirakhupt, K., Aowphol, A., Oaks, J.R., Wood, P.L., Onn, C.K., Thy, N., Cota, M., et al. (2014). Multiple origins of parthenogenesis, and a revised species phylogeny for the Southeast Asian butterfly lizards, Leiolepis. Biol. J. Linn. Soc. n/a – n/a.

Haigh, J. (1978). The accumulation of deleterious genes in a population—Muller's Ratchet. Theor. Popul. Biol. *14*, 251–267.

Harrison, H.B., Feldheim, K.A., Jones, G.P., Ma, K., Mansour, H., Perumal, S., Williamson, D.H., and Berumen, M.L. (2014). Validation of microsatellite multiplexes for parentage analysis and species discrimination in two hybridizing species of coral reef fish (Plectropomus spp., Serranidae). Ecol. Evol. *4*, 2046–2057.

Janko, K., Kotlík, P., and Ráb, P. (2003). Evolutionary history of asexual hybrid loaches (Cobitis: Teleostei) inferred from phylogenetic analysis of mitochondrial DNA variation. J. Evol. Biol. *16*, 1280–1287.

Janko, K. (2014). Let Us Not Be Unfair to Asexuals: Their Ephemerality May Be Explained by Neutral Models Without Invoking Any Evolutionary Constraints of Asexuality. Evolution *68*, 569–576.

Kearney, M., Fujita, M.K., and Ridenour, J. (2009). Lost Sex in the Reptiles: Constraints and Correlations. In Lost Sex, I. Schön, K. Martens, and P. Dijk, eds. (Springer Netherlands), pp. 447–474.

Lehtonen, J., Jennions, M.D., and Kokko, H. (2012). The many costs of sex. Trends Ecol. Evol. *27*, 172–178.

Lively, C.M. (2010). A Review of Red Queen Models for the Persistence of Obligate Sexual Reproduction. J. Hered. *101*, S13–S20.

Maddison, W.P., Knowles, L.L., and Collins, T. (2006). Inferring Phylogeny Despite Incomplete Lineage Sorting. Syst. Biol. *55*, 21–30.

Maynard Smith, J. (1978). The Evolution of Sex (New York: Cambridge University Press).

McDonald, M.J., Rice, D.P., and Desai, M.M. (2016). Sex speeds adaptation by altering the dynamics of molecular evolution. Nature *531*, 233–236.

Metzker, M.L. (2010). Sequencing technologies — the next generation. Nat. Rev. Genet. *11*, 31–46.

Mira, Ó., Sánchez-Prieto, C.B., Dawson, D.A., Burke, T., Tinaut, A., and Martínez, J.G. (2017). Parnassius apollo nevadensis: identification of recent population structure and source–sink dynamics. Conserv. Genet.

Murphy, R.W., Fu, J., Macculloch, R.D., Darevsky, I.S., and Kupriyanova, L.A. (2000). A fine line between sex and unisexuality: the phylogenetic constraints on parthenogenesis in lacertid lizards. Zool. J. Linn. Soc. *130*, 527–549.

Pereira, R.J., Barreto, F.S., and Burton, R.S. (2014). Ecological Novelty by Hybridization: Experimental Evidence for Increased Thermal Tolerance by Transgressive Segregation in Tigriopus Californicus. Evolution *68*, 204–215.

Popov, S.V., Shcherba, I.G., Ilyina, L.B., Nevesskaya, L.A., Paramonova, N.P., Khondkarian, S.O., and Magyar, I. (2006). Late Miocene to Pliocene palaeogeography of the Paratethys and its relation to the Mediterranean. Palaeogeogr. Palaeoclimatol. Palaeoecol. *238*, 91–106.

Schwander, T., and Crespi, B.J. (2009). Multiple direct transitions from sexual reproduction to apomictic parthenogenesis in Timema stick insects. Evol. Int. J. Org. Evol. *63*, 84–103.

Simon, J.-C., Delmotte, F., Rispe, C., and Crease, T. (2003). Phylogenetic relationships between parthenogens and their sexual relatives: the possible routes to parthenogenesis in animals. Biol. J. Linn. Soc. *79*, 151–163.

Stankowski, S., and Streisfeld, M.A. (2015). Introgressive hybridization facilitates adaptive divergence in a recent radiation of monkeyflowers. Proc R Soc B *282*, 20151666.

Tarkhnishvilli, D. (2014). Historical Biogeography of the Caucasus. Nova Science. Publishers. New York. 234 pp.

Tarkhnishvili, D., Gavashelishvili, A., Avaliani, A., Murtskhvaladze, M., and Mumladze, L. (2010). Unisexual rock lizard might be outcompeting its bisexual progenitors in the Caucasus. Biol. J. Linn. Soc. *101*, 447–460.

Tucker, A.E., Ackerman, M.S., Eads, B.D., Xu, S., and Lynch, M. (2013). Population-genomic insights into the evolutionary origin and fate of obligately asexual Daphnia pulex. Proc. Natl. Acad. Sci. *110*, 15740–15745.

Vrijenhoek, R.C. (1989). Genetic and ecological constraints on the origins and establishment of unisexual vertebrates. Evol. Ecol. Unisexual Vertebr. *466*, 24–31.

Wielstra, B., Crnobrnja-Isailović, J., Litvinchuk, S.N., Reijnen, B.T., Skidmore, A.K., Sotiropoulos, K., Toxopeus, A.G., Tzankov, N., Vukov, T., and Arntzen, J.W. (2013). Tracing glacial refugia of Triturus newts based on mitochondrial DNA phylogeography and species distribution modeling. Front. Zool. *10*, 13.

APPENDIX

CHAPTER 2.2 – SUPPLEMENTARY INFORMATION

Table 1: Individuals included in this study: location, accession numbers, latitude/longitude, haplotypes for each marker analysed – mtDNA (concatenated Cyt-*b* and ND4 sequences), MC1R and C-mos).

lineage mtDNA	Haplotype mtDNA	Haplotype MC1R	Haplotype CMOS	*SPECIES*	SUBSPECIES	COUNTRY	LATITUDE	LONGITUDE	code
1,1	15	10,10	1,1	*D. raddei*	*nairensis*	Armenia	40,183	44,496	10335
1,1	18	3,10	1,1	*D. raddei*	*nairensis*	Armenia	40,183	44,496	10342
1,1	19	8,10	1,1	*D. raddei*	*nairensis*	Armenia	40,183	44,496	10343
1,1	94	3,10	NA	*D. raddei*	*nairensis*	Armenia	40,183	44,496	9926
1,1	15	NA	1,1	*D. raddei*	*nairensis*	Armenia	40,183	44,496	10370
1,1	46	11,8	1,1	*D. raddei*	*nairensis*	Armenia	40,438	45,101	9819
1,1	90	11,3	1,1	*D. raddei*	*nairensis*	Armenia	40,438	45,101	631
1,1	90	8,8	NA	*D. raddei*	*nairensis*	Armenia	40,438	45,101	9802
1,1	96	3,2	1,1	*D. raddei*	*nairensis*	Armenia	40,438	45,101	626
1,1	97	11,3	1,1	*D. raddei*	*nairensis*	Armenia	40,438	45,101	638
1,1	4	1,2	1,1	*D. raddei*	*nairensis*	Armenia	40,462	45,082	9907
1,1	4	11,8	1,1	*D. raddei*	*nairensis*	Armenia	40,462	45,083	625
1,1	4	11,8	NA	*D. raddei*	*nairensis*	Armenia	40,462	45,083	4513
1,1	4	8,8	1,1	*D. raddei*	*nairensis*	Armenia	40,462	45,083	4531
1,1	45	1,19	1,1	*D. raddei*	*nairensis*	Armenia	40,462	45,083	4370
1,1	46	3,2	1,1	*D. raddei*	*nairensis*	Armenia	40,462	45,083	4371
1,1	46	3,2	1,1	*D. raddei*	*nairensis*	Armenia	40,462	45,083	4494
1,1	46	3,2	1,1	*D. raddei*	*nairensis*	Armenia	40,462	45,083	4495
1,1	4	1,2	1,1	*D. raddei*	*nairensis*	Armenia	40,510	44,952	10016
1,1	4	3,15	NA	*D. raddei*	*nairensis*	Armenia	40,514	44,936	4368
1,1	4	1,15	1,1	*D. raddei*	*nairensis*	Armenia	40,514	44,936	4369
1,1	5	7,8	1,1	*D. raddei*	*nairensis*	Armenia	40,609	44,609	10114
1,1	92	3,3	1,1	*D. raddei*	*raddei*	Armenia	40,730	44,997	644
1,1	53	3,3	1,1	*D. raddei*	*raddei*	Armenia	40,736	45,019	4978
1,1	5	3,3	1,1	*D. raddei*	*raddei*	Armenia	40,739	45,020	10019
1,1	13	3,3	1,1	*D. raddei*	*raddei*	Armenia	40,739	45,020	10291
1,1	13	3,3	1,1	*D. raddei*	*raddei*	Armenia	40,739	45,020	10492

(Table 1 cont.)

1,1	21	11,3	1,1	*D. raddei*	*raddei*	Armenia	40,739	45,020	10569
1,1	5	3,15	1,1	*D. raddei*	*raddei*	Armenia	40,740	45,022	4395
1,1	13	3,3	NA	*D. raddei*	*raddei*	Armenia	40,740	45,022	4359
1,1	48	3,3	1,1	*D. raddei*	*raddei*	Armenia	40,740	45,022	4396
1,1	49	3,3	1,1	*D. raddei*	*raddei*	Armenia	40,740	45,022	4397
1,1	51	11,3	NA	*D. raddei*	*raddei*	Armenia	40,740	45,022	4560
1,1	99	NA	NA	*D. raddei*	*raddei*	Armenia	40,740	45,022	4360
1,1	5	3,3	NA	*D. raddei*	*raddei*	Armenia	40,788	45,034	12653
1,1	110	3,3	1,1	*D. raddei*	*raddei*	Armenia	41,017	45,137	654
1,1	110	3,3	1,1	*D. raddei*	*raddei*	Armenia	41,017	45,137	4361
1,1	93	3,3	1,1	*D. raddei*	*hybrid Portschinskii*	Armenia	41,022	45,053	650
1,2	29	11,11	NA	*D. raddei*	*nairensis*	Armenia	40,383	44,714	12438
1,2	36	7,15	NA	*D. raddei*	*nairensis*	Armenia	40,383	44,714	12868
1,2	33	11,14	NA	*D. raddei*	*nairensis*	Georgia	41,190	43,750	12805
1,2	34	11,11	NA	*D. raddei*	*nairensis*	Georgia	41,190	43,750	12806
1,2	32	14,14	NA	*D. raddei*	*nairensis*	Georgia	41,458	43,460	12810
1,2	32	11,11	NA	*D. raddei*	*nairensis*	Georgia	41,467	43,436	12836
1,2	32	11,11	NA	*D. raddei*	*nairensis*	Georgia	41,467	43,439	12841
1,2	32	11,11	NA	*D. raddei*	*nairensis*	Georgia	41,480	43,286	12660
1,2	103	6,3	1,1	*D. raddei*	*raddei*	Armenia	39,973	44,951	4341
1,2	103	37,37	1,26	*D. raddei*	*raddei*	Armenia	39,973	44,951	4342
1,2	106	3,3	NA	*D. raddei*	*raddei*	Armenia	40,021	44,986	670
1,2	36	3,3	1,26	*D. raddei*	*raddei*	Armenia	40,113	44,730	4343
1,2	36	3,3	1,26	*D. raddei*	*raddei*	Armenia	40,113	44,730	4344
1,2	32	11,11	1,1	*D. raddei*	*raddei*	Georgia	41,439	43,312	4985
1,2	32	11,11	1,1	*D. raddei*	*raddei*	Georgia	41,439	43,312	4986
1,2	55	11,11	1,1	*D. raddei*	*raddei*	Georgia	41,439	43,312	4987
1,2	55	11,11	1,1	*D. raddei*	*raddei*	Georgia	41,439	43,312	4988
1,2	55	11,11	1,1	*D. raddei*	*raddei*	Georgia	41,439	43,312	4989
2	17	11,1	1,1	*D. raddei*	*nairensis*	Armenia	40,300	43,775	10338

(Table 1 cont.)

2	17	1,1	NA	*D. raddei*	*nairensis*	Armenia	40,300	43,775 10341
2	16	11,12	1,1	*D. raddei*	*nairensis*	Armenia	40,358	43,786 10337
2	30	1,1	NA	*D. raddei*	*nairensis*	Armenia	40,371	44,268 12439
2	31	11,8	NA	*D. raddei*	*nairensis*	Armenia	40,371	44,268 12532
2	14	11,1	1,1	*D. raddei*	*nairensis*	Armenia	40,388	44,227 9943
2	14	1,1	1,1	*D. raddei*	*nairensis*	Armenia	40,388	44,227 9953
2	95	1,1	1,1	*D. raddei*	*nairensis*	Armenia	40,388	44,227 9971
2	14	1,1	1,1	*D. raddei*	*nairensis*	Armenia	40,388	44,227 10303
2	14	1,1	1,1	*D. raddei*	*nairensis*	Armenia	40,388	44,227 10307
2	14	1,1	1,1	*D. raddei*	*nairensis*	Armenia	40,388	44,227 10310
3	70	NA	1,1	*D. raddei*	*vanensis*	Iran	38,106	44,562 5963
3	69	27,26	1,1	*D. raddei*	*vanensis*	Iran	38,106	44,562 5961
3	59	NA	1,1	*D. raddei*	*vanensis*	Turkey	38,429	43,411 5405
3	64	27,27	1,1	*D. raddei*	*vanensis*	Turkey	38,429	43,411 5576
3	64	27,26	NA	*D. raddei*	*vanensis*	Turkey	38,429	43,411 5667
3	63	27,27	1,1	*D. raddei*	*vanensis*	Turkey	38,661	43,836 5566
3	65	27,27	NA	*D. raddei*	*vanensis*	Turkey	38,661	43,836 5607
3	86	27,27	1,1	*D. raddei*	*vanensis*	Iran	38,721	44,613 6085
3	87	27,27	1,1	*D. raddei*	*vanensis*	Iran	38,721	44,613 6086
3	85	27,27	1,1	*D. raddei*	*vanensis*	Iran	38,964	44,682 6072
3	85	27,27	1,1	*D. raddei*	*vanensis*	Iran	38,964	44,682 6073
3	85	27,27	1,1	*D. raddei*	*vanensis*	Iran	38,964	44,682 6074
3	85	27,27	NA	*D. raddei*	*vanensis*	Iran	38,964	44,682 6075
3	60	26,26	1,1	*D. raddei*	*vanensis*	Turkey	39,057	43,757 5406
3	61	26,26	1,19	*D. raddei*	*vanensis*	Turkey	39,057	43,757 5498
3	62	26,26	1,19	*D. raddei*	*vanensis*	Turkey	39,057	43,757 5507
4,1	20	NA	NA	*D. portschinskii*	*hybrid Rostombekovi*	Armenia	40,757	44,804 10363
4,1	72	6,29	1,1	*D. raddei*	*raddei*	Iran	38,424	46,416 5978
4,1	72	29,30	1,1	*D. raddei*	*raddei*	Iran	38,424	46,416 5979
4,1	72	30,30	1,1	*D. raddei*	*raddei*	Iran	38,424	46,416 5980

(Table 1 cont.)

4,1	81	30,30	1,1	*D. raddei*	*raddei*	Iran	38,479	46,475 6052
4,1	81	6,6	1,1	*D. raddei*	*raddei*	Iran	38,479	46,475 6053
4,1	81	13,6	1,1	*D. raddei*	*raddei*	Iran	38,479	46,475 6054
4,1	109	11,6	1,1	*D. raddei*	*raddei*	Armenia	38,927	46,438 4351
4,1	111	6,6	1,1	*D. raddei*	*raddei*	Armenia	39,102	46,489 4353
4,1	112	13,6	1,1	*D. raddei*	*raddei*	Armenia	39,102	46,489 4354
4,1	100	6,6	1,1	*D. raddei*	*raddei*	Armenia	39,191	46,378 4349
4,1	101	6,6	1,1	*D. raddei*	*raddei*	Armenia	39,191	46,378 4350
4,1	27	6,6	NA	*D. raddei*	*raddei*	Armenia	39,211	46,379 12390
4,1	28	6,6	NA	*D. raddei*	*raddei*	Armenia	39,211	46,379 12400
4,1	39	6,9	NA	*D. raddei*	*raddei*	Armenia	39,338	46,452 12935
4,1	40	6,6	NA	*D. raddei*	*raddei*	Armenia	39,338	46,452 12936
4,1	116	6,9	1,1	*D. raddei*	*raddei*	Armenia	39,380	46,250 4357
4,1	117	6,39	1,1	*D. raddei*	*raddei*	Armenia	39,380	46,250 4358
4,1	41	6,6	NA	*D. raddei*	*raddei*	Armenia	39,421	46,374 12947
4,1	35	6,6	NA	*D. raddei*	*raddei*	Armenia	39,445	46,158 12863
4,1	25	6,6	NA	*D. raddei*	*raddei*	Armenia	39,446	46,159 12353
4,1	24	6,6	NA	*D. raddei*	*raddei*	Armenia	39,496	46,121 12331
4,1	24	6,9	NA	*D. raddei*	*raddei*	Armenia	39,496	46,121 12341
4,1	11	6,9	1,1	*D. raddei*	*raddei*	Armenia	39,498	46,348 10197
4,1	24	6,6	1,1	*D. raddei*	*raddei*	Armenia	39,560	46,004 4356
4,1	112	6,6	1,1	*D. raddei*	*raddei*	Armenia	39,560	46,004 4355
4,1	108	6,3	1,1	*D. raddei*	*raddei*	Armenia	39,707	45,208 4346
4,1	114	6,6	1,29	*D. raddei*	*raddei*	Armenia	39,707	45,208 4348
4,1	107	NA	NA	*D. raddei*	*raddei*	Armenia	39,707	45,208 4345
4,1	113	NA	1,29	*D. raddei*	*raddei*	Armenia	39,707	45,208 4347
4,1	38	13,3	NA	*D. raddei*	*raddei*	Armenia	39,837	45,670 12917
4,1	38	13,6	NA	*D. raddei*	*raddei*	Armenia	39,837	45,670 12918
4,1	38	13,13	NA	*D. raddei*	*raddei*	Armenia	39,838	45,661 12913
4,1	38	6,3	NA	*D. raddei*	*raddei*	Armenia	39,838	45,661 12922

(Table 1 cont.)

4,1	26	13,6	NA	*D. raddei*	*raddei*	Armenia	39,843	45,295 12361
4,1	9	6,6	NA	*D. raddei*	*raddei*	NKR	40,043	46,518 10126
4,1	9	6,6	1,1	*D. raddei*	*raddei*	NKR	40,043	46,518 10199
4,1	9	6,6	1,1	*D. raddei*	*raddei*	NKR	40,048	46,237 4366
4,1	120	6,6	1,1	*D. raddei*	*raddei*	NKR	40,048	46,237 4367
4,1	9	6,6	NA	*D. raddei*	*raddei*	NKR	40,048	46,233 10229
4,1	12	6,6	1,1	*D. raddei*	*raddei*	NKR	40,048	46,233 10209
4,1	9	6,6	1,1	*D. raddei*	*raddei*	NKR	40,051	46,232 10202
4,1	12	6,6	1,1	*D. raddei*	*raddei*	NKR	40,054	46,238 10262
4,1	98	NA	1,1	*D. raddei*	*raddei*	NKR	40,058	46,530 4365
4,1	9	6,6	1,1	*D. raddei*	*raddei*	NKR	40,058	46,530 4364
4,1	8	6,6	NA	*D. raddei*	*raddei*	NKR	40,090	46,170 10110
4,1	10	6,6	1,1	*D. raddei*	*raddei*	NKR	40,090	46,170 10141
4,1	9	6,6	1,25	*D. raddei*	*raddei*	NKR	40,123	46,418 604
4,1	9	6,36	1,1	*D. raddei*	*raddei*	NKR	40,123	46,418 605
4,1	9	6,6	1,1	*D. raddei*	*raddei*	NKR	40,140	46,520 4362
4,1	47	6,6	1,24	*D. raddei*	*raddei*	NKR	40,140	46,520 632
4,1	89	6,6	1,24	*D. raddei*	*raddei*	NKR	40,140	46,520 630
4,1	91	6,6	1,1	*D. raddei*	*raddei*	NKR	40,140	46,520 635
4,1	118	6,6	1,1	*D. raddei*	*raddei*	NKR	40,140	46,520 4363
4,1	9	6,20	1,1	*D. raddei*	*raddei*	NKR	40,252	46,640 639
4,1	9	6,6	1,1	*D. raddei*	*raddei*	NKR	40,306	46,697 598
4,1	47	6,20	1,1	*D. raddei*	*raddei*	NKR	40,306	46,697 4372
4,1	98	6,6	1,1	*D. raddei*	*raddei*	NKR	40,306	46,697 597
4,2	42	16,17	NA	*D. raddei*	*raddei*	Iran	37,161	49,022 13792
4,2	43	NA	NA	*D. raddei*	*raddei*	Iran	37,161	49,022 13793
4,2	75	16,17	1,1	*D. raddei*	*raddei*	Iran	37,602	48,689 5996
4,2	75	32,32	1,1	*D. raddei*	*raddei*	Iran	37,602	48,689 5997
4,2	75	NA	1,1	*D. raddei*	*raddei*	Iran	37,602	48,689 5995
4,2	88	33,17	1,1	*D. raddei*	*raddei*	Iran	37,672	48,509 6094

(Table 1 cont.)

	4,2	76	33,33	1,1	*D. raddei*	*raddei*	Iran	38,088	48,468 6012
	5	74	31,3	1,1	*D. raddei*	*raddei*	Iran	38,842	47,001 5988
	5	73	NA	1,1	*D. raddei*	*raddei*	Iran	38,842	47,001 5987
	5	71	NA	1,1	*D. raddei*	*raddei*	Iran	38,863	46,824 5972
	5	71	11,6	1,1	*D. raddei*	*raddei*	Iran	38,863	46,824 5971
	5	71	13,6	1,1	*D. raddei*	*raddei*	Iran	38,863	46,824 5973
	5	82	3,3	1,1	*D. raddei*	*raddei*	Iran	38,868	46,968 6055
	5	83	3,3	1,1	*D. raddei*	*raddei*	Iran	38,868	46,968 6056
	5	84	3,3	1,22	*D. raddei*	*raddei*	Iran	38,868	46,968 6057
	6	44	13,18	NA	*D. raddei*	*raddei*	Iran	38,277	47,698 13796
	6	67	13,16	1,21	*D. raddei*	*raddei*	Iran	38,539	47,666 5950
	6	54	NA	1,5	*D. raddei*	*raddei*	Iran	38,757	47,522 4982
	6	77	13,34	1,1	*D. raddei*	*raddei*	Iran	38,845	47,638 6032
	6	78	13,3	1,22	*D. raddei*	*raddei*	Iran	38,845	47,638 6034
	6	79	11,11	NA	*D. raddei*	*raddei*	Iran	38,851	47,663 6037
	6	80	11,35	1,1	*D. raddei*	*raddei*	Iran	38,851	47,663 6036
	6	79	NA	1,23	*D. raddei*	*raddei*	Iran	38,851	47,663 6035
	68	68	28,28	22,22	*D. raddei*	*raddei*	Iran	37,780	46,409 5951
	68	68	28,28	22,22	*D. raddei*	*raddei*	Iran	37,780	46,409 5952
	68	68	28,28	22,22	*D. raddei*	*raddei*	Iran	37,780	46,409 5953
2.uni		6	1,4	1,2	*D. unisexualis*	NA	Turkey	39,887	42,360 5499
2.uni		6	1,4	1,2	*D. unisexualis*	NA	Turkey	39,887	42,360 5548
2.uni		6	1,4	1,2	*D. unisexualis*	NA	Turkey	39,887	42,360 5809
2.uni		6	1,4	1,2	*D. unisexualis*	NA	Turkey	39,887	42,360 5823
2.uni		6	NA	1,2	*D. unisexualis*	NA	Turkey	39,887	42,360 5544
2.uni		22	1,4	1,2	*D. unisexualis*	NA	Armenia	40,154	45,496 10577
2.uni		6	1,4	NA	*D. unisexualis*	NA	Armenia	40,180	45,631 12362
2.uni		6	1,4	1,2	*D. unisexualis*	NA	Armenia	40,388	45,209 10064
2.uni		6	1,4	1,2	*D. unisexualis*	NA	Armenia	40,388	45,209 10454
2.uni		6	1,4	1,2	*D. unisexualis*	NA	Armenia	40,462	45,082 9915

(Table 1 cont.)

2.uni	23	1,4	1,2	*D. unisexualis*	NA	Armenia	40,462	45,082 10591
2.uni	6	1,4	1,2	*D. unisexualis*	NA	Armenia	40,462	45,083 4497
2.uni	23	1,4	1,2	*D. unisexualis*	NA	Armenia	40,462	45,083 4499
2.uni	50	1,4	1,2	*D. unisexualis*	NA	Armenia	40,462	45,083 4498
2.uni	6	1,4	NA	*D. unisexualis*	NA	Armenia	40,495	45,355 12376
2.uni	37	1,4	NA	*D. unisexualis*	NA	Armenia	40,504	44,748 12890
2.uni	6	1,4	1,2	*D. unisexualis*	NA	Armenia	40,540	44,373 4461
2.uni	6	1,4	1,2	*D. unisexualis*	NA	Armenia	40,540	44,373 4463
2.uni	6	1,4	1,2	*D. unisexualis*	NA	Armenia	40,540	44,373 4465
2.uni	23	1,4	1,2	*D. unisexualis*	NA	Armenia	40,540	44,373 4458
2.uni	6	1,4	1,2	*D. unisexualis*	NA	Armenia	40,564	45,010 4971
2.uni	6	1,4	1,2	*D. unisexualis*	NA	Armenia	40,564	45,010 4972
2.uni	52	1,21	NA	*D. unisexualis*	NA	Armenia	40,564	45,010 4970
2.uni	6	NA	1,2	*D. unisexualis*	NA	Armenia	40,619	44,564 659
2.uni	6	NA	1,2	*D. unisexualis*	NA	Armenia	40,619	44,564 668
2.uni	6	1,4	1,2	*D. unisexualis*	NA	Armenia	40,823	44,289 4629
2.uni	6	1,4	1,2	*D. unisexualis*	NA	Armenia	40,848	44,252 9947
2.uni	6	1,4	1,2	*D. unisexualis*	NA	Armenia	40,853	43,948 10301
2.uni	6	1,4	1,2	*D. unisexualis*	NA	Armenia	40,853	43,948 10305
2.uni	6	1,4	1,2	*D. unisexualis*	NA	Armenia	40,853	43,948 10308
2.uni	6	1,4	1,2	*D. unisexualis*	NA	petshop	NA	NA 4990
2.uni	6	1,4	1,2	*D. unisexualis*	NA	petshop	NA	NA 4991
2.uzz	57	1,25	1,2	*D. uzzeli*	NA	Turkey	39,887	42,360 5268
2.uzz	58	1,25	1,2	*D. uzzeli*	NA	Turkey	39,887	42,360 5273
2.uzz	58	1,25	1,2	*D. uzzeli*	NA	Turkey	39,887	42,360 5292
2.uzz	58	1,25	1,2	*D. uzzeli*	NA	Turkey	39,887	42,360 5399
2.uzz	58	1,25	1,2	*D. uzzeli*	NA	Turkey	39,887	42,360 5496
3.bendi	66	27,4	1,2	*D. bendimahiensis*	NA	Turkey	39,057	43,757 5623
3.bendi	66	27,4	1,2	*D. bendimahiensis*	NA	Turkey	39,057	43,757 5764
3.bendi	66	27,4	1,2	*D. bendimahiensis*	NA	Turkey	39,057	43,757 5777

(Table 1 cont.)

D.portschinskii		7	5,5	3,3	*D. portschinskii*	NA	Armenia	40,739	45,020	10071
D.portschinskii		7	5,5	3,3	*D. portschinskii*	NA	Armenia	40,740	45,022	4635
D.portschinskii		115	NA	3,3	*D. portschinskii*	NA	Armenia	40,996	44,389	4588
D.portschinskii		115	5,5	NA	*D. portschinskii*	NA	Armenia	40,996	44,389	4592
D.rudis		56	23,24	6,7	*D. rudis*	NA	Turkey	41,337	32,588	5250
D.rudis		56	22,22	6,8	*D. rudis*	NA	Turkey	41,337	32,588	5255
D.valentini		104	25,38	2,28	*D. valentini*	NA	Armenia	40,514	44,936	4547
D.valentini		105	25,38	2,2	*D. valentini*	NA	Armenia	40,514	44,936	4551
D.valentini		102	25,25	2,2	*D. valentini*	NA	Armenia	40,540	44,373	4455
D.valentini		119	25,25	2,2	*D. valentini*	NA	Armenia	40,655	44,581	4488
D.valentini		119	25,25	2,2	*D. valentini*	NA	Armenia	40,655	44,581	4489
NA	NA	5,5		3,3	*D. portschinskii*	NA	Nagorno-Karabakh Republic	40,048	46,237	4638
NA	NA	NA		3,3	*D. portschinskii*	NA	Nagorno-Karabakh Republic	40,048	46,237	4636
NA	NA	NA		3,3	*D. portschinskii*	NA	Nagorno-Karabakh Republic	40,048	46,237	4637
NA	NA	NA		3,3	*D. portschinskii*	NA	Nagorno-Karabakh Republic	40,048	46,237	4639
NA	NA	NA		1,1	*D. raddei*	*raddei*	Iran	38,088	48,468	6013
NA	NA	NA		1,1	*D. raddei*	*raddei*	Iran	38,106	44,562	5962
NA	NA	NA		1,21	*D. raddei*	*raddei*	Iran	38,539	47,666	5949
NA	NA	NA		1,1	*D. raddei*	*raddei*	Turkey	**38,661**	**43,836**	5506
NA	NA	NA		1,1	*D. raddei*	*raddei*	Iran	38,842	47,001	5986
NA	NA	NA		22,22	*D. raddei*	*raddei*	Iran	38,845	47,638	6033
NA	NA	NA		1,1	*D. raddei*	*raddei*	Armenia	38,927	46,438	4352
NA	NA	3,37		1,1	*D. raddei*	*raddei*	Armenia	40,021	44,986	669
NA	NA	25,25		NA	*D. valentini*	NA	Armenia	40,540	44,373	4472
outgroup		1	NA	NA	*Algyroides sp.*	NA	NA	NA	NA	Algyroides_fitzingeri
outgroup		2	NA	NA	*Archaeolacerta sp.*	NA	NA	NA	NA	Archaeolacerta_bedriagae
outgroup		3	NA	NA	*Iranolacerta sp.*	NA	NA	NA	NA	Iranolacerta

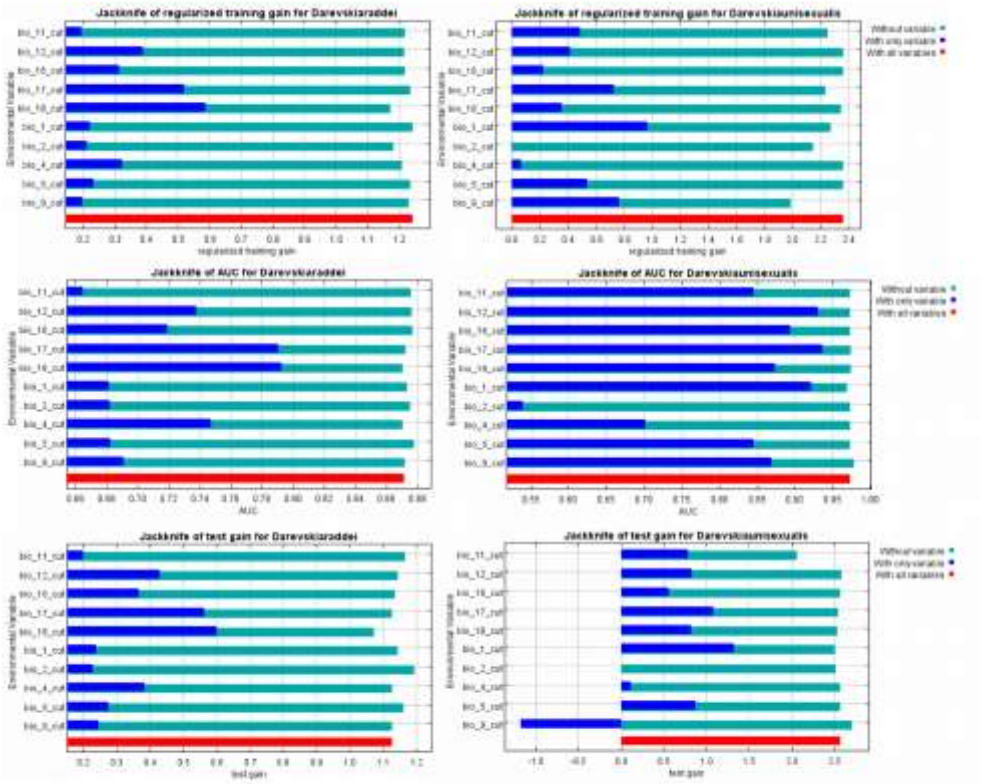
Table 2 – Forward and Reverse primer sequences for all markers used in the study (Cyt-*b*, ND4, MC1R, Cmos).

	primers		
	Forward	Reverse	length
cytb	Glu-DG: 5'-TGACTTGAARAACCAYCGTTG-3'	PeiL: 5'-TCTTCTACTGGTTGTCCTCCGATTCA-3'	898
ND4	ND4: 5'-CACCTATGACTACCAAAAGCTCATGTAGAAGC-3'	LEU: 5'-CATTACTTTTACTTGGAATTTGCACCA-3'	790
MC1R	MC1R_F: 5'-GGCNGCCATYGTCAAGAACCGGAACC-3'	MC1R_R: 5'-CTCCGRAAGGCRTAAATGATGGGGTCCAC-3'	659
Cmos	LSC1: 5'-CTCTGGKGGCTTTGGKKCTGTSTACAAGG-3'	LSC2: 5'-GGTGATGGCAAANGAGTAGATGTCTGC-3'	535

Table 3 – Percent contribution and permutation importance of each variable used for the ecological models constructed for both each species.

	Percent contribution		*Permutation importance*		
variable code	*D. raddei*	*D. unisexualis*	*D. raddei*	*D. unisexualis*	Variable
Bio 1	4	23	4	23	Annual Mean Temperature
Bio 2	9	6	10	6	Mean Diurnal Range (Mean of monthly (max temp - min temp)
Bio 4	10	0	10	0	Temperature Seasonality (standard deviation *100)
Bio 5	2	0	2	0	Max Temperature of Warmest Month
Bio 9	3	37	3	37	Mean Temperature of Driest Quarter
Bio 11	9	9	9	9	Mean Temperature of Coldest Quarter
Bio 12	7	0	7	0	Annual Precipitation
Bio 16	1	0	1	0	Precipitation of Wettest Quarter
Bio 17	10	14	10	14	Precipitation of Driest Quarter
Bio 18	45	11	45	11	Precipitation of Warmest Quarter

Figure 1 – Graphics of the contribution (jacknife of regularised training gain, AUC and test gain) of each environmental variable on the *D. raddei* and *D. unisexualis* ecological niche models.



CHAPTER 3 – SUPPLEMENTARY INFORMATION

Table 3.1: Samples information of the genotyped individuals: Sample identification (ID), species code (code), species name, sympatric locality, ploidy number, country of origin, latitude and longitude.

ID	code	*species*	Sympatric locality	Ploidy	Country	Latitude	Longitude
12300	arm	*D. armeniaca*		2N	Georgia	41.39	43.47
5398	arm	*D. armeniaca*		2N	Turkey	41.23	42.45
12257	arm	*D. armeniaca*		2N	Armenia	41.07	43.66
6267	arm	*D. armeniaca*		2N	NA	40.62	44.56
6271	arm	*D. armeniaca*		2N	NA	40.62	44.56
6272	arm	*D. armeniaca*		2N	NA	40.62	44.56
10431	arm	*D. armeniaca*		2N	Armenia	40.94	44.48
10367	arm	*D. armeniaca*		2N	Armenia	40.76	44.80
10453	arm	*D. armeniaca*		2N	Armenia	40.76	44.80
4609	arm	*D. armeniaca*		2N	Armenia	40.89	44.20
4610	arm	*D. armeniaca*		2N	Armenia	40.89	44.20
4613	arm	*D. armeniaca*		2N	Armenia	40.89	44.20
10177	arm	*D. armeniaca*		2N	Armenia	40.95	44.47
10410	arm	*D. armeniaca*		2N	Armenia	40.62	44.56
12887	arm	*D. armeniaca*		2N	Armenia	40.50	44.75
10316	arm	*D. armeniaca*		2N	Armenia	40.85	43.95
10321	arm	*D. armeniaca*		2N	Armenia	40.85	43.95
10334	arm	*D. armeniaca*		2N	Armenia	40.85	43.95
9820	arm	*D. armeniaca*		2N	Armenia	NA	NA
6268	arm	*D. armeniaca*		2N	NA	40.82	44.47
9958	arm	*D. armeniaca*		2N	Armenia	40.74	44.82
10406	arm	*D. armeniaca*		2N	Armenia	41.17	44.32
10459	arm	*D. armeniaca*		2N	Armenia	41.17	44.32
10445	arm	*D. armeniaca*		2N	Armenia	41.01	44.38
10450	arm	*D. armeniaca*		2N	Armenia	41.17	44.32
4490	arm	*D. armeniaca*		2N	Armenia	40.65	44.58
4491	arm	*D. armeniaca*		2N	Armenia	40.65	44.58
12766	dah	*D. dahli*		2N	Georgia	41.81	43.44
12180	k_arm	*D. armeniaca*	Kuchak	2N	Armenia	NA	NA
12181	k_arm	*D. armeniaca*	Kuchak	2N	Armenia	NA	NA
12185	k_arm	*D. armeniaca*	Kuchak	2N	Armenia	NA	NA
12186	k_arm	*D. armeniaca*	Kuchak	2N	Armenia	NA	NA
12189	k_arm	*D. armeniaca*	Kuchak	2N	Armenia	NA	NA

12190	k_arm	*D. armeniaca*	Kuchak	2N	Armenia	NA	NA
12228	k_arm	*D. armeniaca*	Kuchak	2N	Armenia	NA	NA
12233	k_arm	*D. armeniaca*	Kuchak	2N	Armenia	44.39	40.53
12238	k_arm	*D. armeniaca*	Kuchak	2N	Armenia	44.39	40.53
12239	k_arm	*D. armeniaca*	Kuchak	2N	Armenia	44.39	40.53
12243	k_arm	*D. armeniaca*	Kuchak	2N	Armenia	44.39	40.53
12246	k_arm	*D. armeniaca*	Kuchak	2N	Armenia	44.39	40.53
12248	k_arm	*D. armeniaca*	Kuchak	2N	Armenia	44.39	40.53
12249	k_arm	*D. armeniaca*	Kuchak	2N	Armenia	44.39	40.53
12253	k_arm	*D. armeniaca*	Kuchak	2N	Armenia	44.39	40.53
12258	k_arm	*D. armeniaca*	Kuchak	2N	Armenia	44.39	40.53
12259	k_arm	*D. armeniaca*	Kuchak	2N	Armenia	44.39	40.53
12268	k_arm	*D. armeniaca*	Kuchak	2N	Armenia	44.39	40.53
12269	k_arm	*D. armeniaca*	Kuchak	2N	Armenia	44.39	40.53
12271	k_arm	*D. armeniaca*	Kuchak	2N	Armenia	44.39	40.53
12273	k_arm	*D. armeniaca*	Kuchak	2N	Armenia	44.39	40.53
12274	k_arm	*D. armeniaca*	Kuchak	2N	Armenia	44.39	40.53
12278	k_arm	*D. armeniaca*	Kuchak	2N	Armenia	44.39	40.53
12279	k_arm	*D. armeniaca*	Kuchak	2N	Armenia	44.39	40.53
12288	k_arm	*D. armeniaca*	Kuchak	2N	Armenia	44.39	40.53
12289	k_arm	*D. armeniaca*	Kuchak	2N	Armenia	44.39	40.53
12293	k_arm	*D. armeniaca*	Kuchak	2N	Armenia	44.39	40.53
12298	k_arm	*D. armeniaca*	Kuchak	2N	Armenia	44.39	40.53
12303	k_arm	*D. armeniaca*	Kuchak	2N	Armenia	44.39	40.53
12308	k_arm	*D. armeniaca*	Kuchak	2N	Armenia	44.39	40.53
12309	k_arm	*D. armeniaca*	Kuchak	2N	Armenia	44.39	40.53
12419	k_arm	*D. armeniaca*	Kuchak	2N	Armenia	44.39	40.53
12424	k_arm	*D. armeniaca*	Kuchak	2N	Armenia	44.39	40.53
12435	k_arm	*D. armeniaca*	Kuchak	2N	Armenia	44.39	40.53
12444	k_arm	*D. armeniaca*	Kuchak	2N	Armenia	44.39	40.53
12445	k_arm	*D. armeniaca*	Kuchak	2N	Armenia	44.39	40.53
12448	k_arm	*D. armeniaca*	Kuchak	2N	Armenia	44.39	40.53
12457	k_arm	*D. armeniaca*	Kuchak	2N	Armenia	44.39	40.53
12458	k_arm	*D. armeniaca*	Kuchak	2N	Armenia	44.39	40.53

(table 3.1 cont.)

12459	k_arm	*D. armeniaca*	Kuchak	2N	Armenia	44.39	40.53
12466	k_arm	*D. armeniaca*	Kuchak	2N	Armenia	44.39	40.53
12519	k_arm	*D. armeniaca*	Kuchak	2N	Armenia	44.39	40.53
12531	k_arm	*D. armeniaca*	Kuchak	2N	Armenia	44.39	40.53
12533	k_arm	*D. armeniaca*	Kuchak	2N	Armenia	44.39	40.53
12543	k_arm	*D. armeniaca*	Kuchak	2N	Armenia	44.39	40.53
12544	k_arm	*D. armeniaca*	Kuchak	2N	Armenia	44.39	40.53
12548	k_arm	*D. armeniaca*	Kuchak	2N	Armenia	44.39	40.53
12558	k_arm	*D. armeniaca*	Kuchak	2N	Armenia	44.39	40.53
12560	k_arm	*D. armeniaca*	Kuchak	2N	Armenia	44.39	40.53
12563	k_arm	*D. armeniaca*	Kuchak	2N	Armenia	44.39	40.53
12566	k_arm	*D. armeniaca*	Kuchak	2N	Armenia	44.39	40.53
12567	k_arm	*D. armeniaca*	Kuchak	2N	Armenia	44.39	40.53
12571	k_arm	*D. armeniaca*	Kuchak	2N	Armenia	44.39	40.53
12575	k_arm	*D. armeniaca*	Kuchak	2N	Armenia	44.39	40.53
12580	k_arm	*D. armeniaca*	Kuchak	2N	Armenia	44.39	40.53
12581	k_arm	*D. armeniaca*	Kuchak	2N	Armenia	44.39	40.53
12584	k_arm	*D. armeniaca*	Kuchak	2N	Armenia	44.39	40.53
12585	k_arm	*D. armeniaca*	Kuchak	2N	Armenia	44.39	40.53
12586	k_arm	*D. armeniaca*	Kuchak	2N	Armenia	44.39	40.53
12592	k_arm	*D. armeniaca*	Kuchak	2N	Armenia	44.39	40.53
12595	k_arm	*D. armeniaca*	Kuchak	2N	Armenia	44.39	40.53
12596	k_arm	*D. armeniaca*	Kuchak	2N	Armenia	44.39	40.53
12600	k_arm	*D. armeniaca*	Kuchak	2N	Armenia	44.39	40.53
12601	k_arm	*D. armeniaca*	Kuchak	2N	Armenia	44.39	40.53
12602	k_arm	*D. armeniaca*	Kuchak	2N	Armenia	44.39	40.53
12604	k_arm	*D. armeniaca*	Kuchak	2N	Armenia	44.39	40.53
12605	k_arm	*D. armeniaca*	Kuchak	2N	Armenia	44.39	40.53
12607	k_arm	*D. armeniaca*	Kuchak	2N	Armenia	44.39	40.53
12608	k_arm	*D. armeniaca*	Kuchak	2N	Armenia	44.39	40.53
12609	k_arm	*D. armeniaca*	Kuchak	2N	Armenia	44.39	40.53
12617	k_arm	*D. armeniaca*	Kuchak	2N	Armenia	44.39	40.53
12618	k_arm	*D. armeniaca*	Kuchak	2N	Armenia	44.39	40.53
12620	k_arm	*D. armeniaca*	Kuchak	2N	Armenia	44.39	40.53

(table 3.1 cont.)

12621	k_arm	*D. armeniaca*	Kuchak	2N	Armenia	44.39	40.53
12627	k_arm	*D. armeniaca*	Kuchak	2N	Armenia	44.39	40.53
12630	k_arm	*D. armeniaca*	Kuchak	2N	Armenia	44.39	40.53
12632	k_arm	*D. armeniaca*	Kuchak	2N	Armenia	44.39	40.53
12639	k_arm	*D. armeniaca*	Kuchak	2N	Armenia	44.39	40.53
12644	k_arm	*D. armeniaca*	Kuchak	2N	Armenia	44.39	40.53
12650	k_arm	*D. armeniaca*	Kuchak	2N	Armenia	44.39	40.53
12657	k_arm	*D. armeniaca*	Kuchak	2N	Armenia	44.39	40.53
12739	k_arm	*D. armeniaca*	Kuchak	2N	Armenia	44.39	40.53
12803	k_arm	*D. armeniaca*	Kuchak	2N	Armenia	44.39	40.53
12850	k_arm	*D. armeniaca*	Kuchak	2N	Armenia	44.39	40.53
12853	k_arm	*D. armeniaca*	Kuchak	2N	Armenia	44.39	40.53
12956	k_arm	*D. armeniaca*	Kuchak	2N	Armenia	44.39	40.53
12957	k_arm	*D. armeniaca*	Kuchak	2N	Armenia	44.39	40.53
12965	k_arm	*D. armeniaca*	Kuchak	2N	Armenia	44.39	40.53
12969	k_arm	*D. armeniaca*	Kuchak	2N	Armenia	44.39	40.53
12976	k_arm	*D. armeniaca*	Kuchak	2N	Armenia	44.39	40.53
12985	k_arm	*D. armeniaca*	Kuchak	2N	Armenia	44.39	40.53
12995	k_arm	*D. armeniaca*	Kuchak	2N	Armenia	44.39	40.53
12997	k_arm	*D. armeniaca*	Kuchak	2N	Armenia	44.39	40.53
12999	k_arm	*D. armeniaca*	Kuchak	2N	Armenia	44.39	40.53
13005	k_arm	*D. armeniaca*	Kuchak	2N	Armenia	44.39	40.53
13008	k_arm	*D. armeniaca*	Kuchak	2N	Armenia	44.39	40.53
13018	k_arm	*D. armeniaca*	Kuchak	2N	Armenia	44.39	40.53
13029	k_arm	*D. armeniaca*	Kuchak	2N	Armenia	44.39	40.53
13038	k_arm	*D. armeniaca*	Kuchak	2N	Armenia	44.39	40.53
13039	k_arm	*D. armeniaca*	Kuchak	2N	Armenia	44.39	40.53
13040	k_arm	*D. armeniaca*	Kuchak	2N	Armenia	44.39	40.53
13050	k_arm	*D. armeniaca*	Kuchak	2N	Armenia	44.38	40.53
4459	k_axv	*D. armeniaca x D. valentini*	Kuchak	3N	Armenia	40.54	44.37
12176	k_axv	*D. armeniaca x D. valentini*	Kuchak	4N	NA	NA	NA
12978	k_axv	*D. armeniaca x D. valentini*	Kuchak	3N	Armenia	44.39	40.53
4458	k_uni	*D. unisexualis*	Kuchak	2N	Armenia	40.54	44.37
4461	k_uni	*D. unisexualis*	Kuchak	2N	Armenia	40.54	44.37

(table 3.1 cont.)

12247	k_uni	*D. unisexualis*	Kuchak	2N	Armenia	44.39	40.53
12591	k_uni	*D. unisexualis*	Kuchak	2N	Armenia	44.39	40.53
12652	k_uni	*D. unisexualis*	Kuchak	2N	Armenia	44.39	40.53
4473	k_uxv	*D. unisexualis x D. valentini*	Kuchak	3N	Armenia	40.54	44.37
4475	k_uxv	*D. unisexualis x D. valentini*	Kuchak	3N	Armenia	40.54	44.37
12182	k_uxv	*D. unisexualis x D. valentini*	Kuchak	3N	NA	NA	NA
12237	k_uxv	*D. unisexualis x D. valentini*	Kuchak	3N	Armenia	44.39	40.53
12428	k_uxv	*D. unisexualis x D. valentini*	Kuchak	3N	Armenia	44.39	40.53
12454	k_uxv	*D. unisexualis x D. valentini*	Kuchak	3N	Armenia	44.39	40.53
12523	k_uxv	*D. unisexualis x D. valentini*	Kuchak	3N	Armenia	44.39	40.53
12524	k_uxv	*D. unisexualis x D. valentini*	Kuchak	3N	Armenia	44.39	40.53
12555	k_uxv	*D. unisexualis x D. valentini*	Kuchak	3N	Armenia	44.39	40.53
12572	k_uxv	*D. unisexualis x D. valentini*	Kuchak	3N	Armenia	44.39	40.53
12574	k_uxv	*D. unisexualis x D. valentini*	Kuchak	3N	Armenia	44.39	40.53
12612	k_uxv	*D. unisexualis x D. valentini*	Kuchak	3N	Armenia	44.39	40.53
12622	k_uxv	*D. unisexualis x D. valentini*	Kuchak	3N	Armenia	44.39	40.53
12626	k_uxv	*D. unisexualis x D. valentini*	Kuchak	3N	Armenia	44.39	40.53
12628	k_uxv	*D. unisexualis x D. valentini*	Kuchak	3N	Armenia	44.39	40.53
12631	k_uxv	*D. unisexualis x D. valentini*	Kuchak	3N	Armenia	44.39	40.53
12636	k_uxv	*D. unisexualis x D. valentini*	Kuchak	3N	Armenia	44.39	40.53
12637	k_uxv	*D. unisexualis x D. valentini*	Kuchak	3N	Armenia	44.39	40.53
12651	k_uxv	*D. unisexualis x D. valentini*	Kuchak	3N	Armenia	44.39	40.53
13019	k_uxv	*D. unisexualis x D. valentini*	Kuchak	3N	Armenia	44.39	40.53
13020	k_uxv	*D. unisexualis x D. valentini*	Kuchak	3N	Armenia	44.39	40.53
4480	k_val	*D. valentini*	Kuchak	2N	Armenia	40.54	44.37
12175	k_val	*D. valentini*	Kuchak	2N	NA	NA	NA
12229	k_val	*D. valentini*	Kuchak	2N	NA	NA	NA
12234	k_val	*D. valentini*	Kuchak	2N	Armenia	44.39	40.53
12256	k_val	*D. valentini*	Kuchak	2N	Armenia	44.39	40.53
12299	k_val	*D. valentini*	Kuchak	2N	Armenia	44.39	40.53
12437	k_val	*D. valentini*	Kuchak	2N	Armenia	44.39	40.53
12449	k_val	*D. valentini*	Kuchak	2N	Armenia	44.39	40.53
12464	k_val	*D. valentini*	Kuchak	2N	Armenia	44.39	40.53
12468	k_val	*D. valentini*	Kuchak	2N	Armenia	44.39	40.53

(table 3.1 cont.)

12554	k_val	*D. valentini*	Kuchak	2N	Armenia	44.39	40.53
12564	k_val	*D. valentini*	Kuchak	2N	Armenia	44.39	40.53
12570	k_val	*D. valentini*	Kuchak	2N	Armenia	44.39	40.53
12594	k_val	*D. valentini*	Kuchak	2N	Armenia	44.39	40.53
12615	k_val	*D. valentini*	Kuchak	2N	Armenia	44.39	40.53
12624	k_val	*D. valentini*	Kuchak	2N	Armenia	44.39	40.53
12641	k_val	*D. valentini*	Kuchak	2N	Armenia	44.39	40.53
12642	k_val	*D. valentini*	Kuchak	2N	Armenia	44.38	40.53
12958	k_val	*D. valentini*	Kuchak	2N	Armenia	44.39	40.53
12959	k_val	*D. valentini*	Kuchak	2N	Armenia	44.39	40.53
12966	k_val	*D. valentini*	Kuchak	2N	Armenia	44.39	40.53
12968	k_val	*D. valentini*	Kuchak	2N	Armenia	44.39	40.53
12988	k_val	*D. valentini*	Kuchak	2N	Armenia	44.39	40.53
12989	k_val	*D. valentini*	Kuchak	2N	Armenia	44.39	40.53
13009	k_val	*D. valentini*	Kuchak	2N	Armenia	44.39	40.53
12554b	k_val	*D. valentini*	Kuchak	2N	#N/A	#N/A	#N/A
4530	Lch_nai	*D. raddei*	Lchap	2N	Armenia	40.46	45.08
4512	Lch_nai-1	*D. raddei*	Lchap	2N	Armenia	40.46	45.08
4513	Lch_nai-1	*D. raddei*	Lchap	2N	Armenia	40.46	45.08
4531	Lch_nai-1	*D. raddei*	Lchap	2N	Armenia	40.46	45.08
4497	Lch_uni	*D. unisexualis*	Lchap	2N	Armenia	40.46	45.08
4498	Lch_uni	*D. unisexualis*	Lchap	2N	Armenia	40.46	45.08
4499	Lch_uni	*D. unisexualis*	Lchap	2N	Armenia	40.46	45.08
4500	Lch_uni	*D. unisexualis*	Lchap	2N	Armenia	40.46	45.08
4501	Lch_uni	*D. unisexualis*	Lchap	2N	Armenia	40.46	45.08
4511	Lch_uni	*D. unisexualis*	Lchap	2N	Armenia	40.46	45.08
4532	Lsh_nai	*D. raddei*	Lchaschen	2N	Armenia	40.51	44.94
4535	Lsh_nai	*D. raddei*	Lchaschen	2N	Armenia	40.51	44.94
4539	Lsh_nai	*D. raddei*	Lchaschen	2N	Armenia	40.51	44.94
4533	Lsh_nai-1	*D. raddei*	Lchaschen	2N	Armenia	40.51	44.94
4534	Lsh_nai-1	*D. raddei*	Lchaschen	2N	Armenia	40.51	44.94
4538	Lsh_nai-1	*D. raddei*	Lchaschen	2N	Armenia	40.51	44.94
4536	Lsh_val	*D. valentini*	Lchaschen	2N	Armenia	40.51	44.94
4547	Lsh_val	*D. valentini*	Lchaschen	2N	Armenia	40.51	44.94

(table 3.1 cont.)

4537	Lsh_vxa	*D. armeniaca x D. valentini*	Lchaschen	2N	Armenia	40.51	44.94
9829	mix	*D. mixta*		2N	Georgia	42.31	42.69
9831	mix	*D. mixta*		2N	Georgia	42.31	42.69
9849	mix	*D. mixta*		2N	Georgia	42.31	42.69
10281	mix	*D. mixta*		2N	Georgia	42.31	42.69
9804	mix	*D. mixta*		2N	Georgia	NA	NA
9816	mix	*D. mixta*		2N	Georgia	NA	NA
9823	mix	*D. mixta*		2N	Georgia	NA	NA
10290	mix	*D. mixta*		2N	Georgia	NA	NA
10292	mix	*D. mixta*		2N	Georgia	NA	NA
15737	mix2	*D. mixta - 2*		2N	Turkey	40.72	40.60
15738	mix2	*D. mixta - 2*		2N	Turkey	40.72	40.60
15739	mix2	*D. mixta - 2*		2N	Turkey	40.72	40.60
15740	mix2	*D. mixta - 2*		2N	Turkey	40.72	40.60
15741	mix2	*D. mixta - 2*		2N	Turkey	40.72	40.60
15742	mix2	*D. mixta - 2*		2N	Turkey	40.72	40.60
15746	mix2	*D. mixta - 2*		2N	Turkey	40.72	40.60
676	por	*D. portschinksii*		2N	NKR	40.05	46.24
10363	PoxRo	*D. portschinksii x D. rostombekovi*		2N	Armenia	40.76	44.80
13285	pra	*D. praticola*		2N	Serbia	44.68	20.55
12760	rad	*D. raddei*		2N	Armenia	40.39	44.23
12975	rad	*D. raddei*		2N	Armenia	40.37	44.27
5566	rad	*D. raddei*		2N	Turkey	38.66	43.84
12934	rad	*D. raddei*		2N	Armenia	39.34	46.45
10180	rad	*D. raddei*		2N	NKR	40.04	46.52
6080	rad	*D. raddei*		2N	Iran	38.96	44.68
6081	rad	*D. raddei*		2N	Iran	38.96	44.68
15715	rad	*D. raddei*		2N	Turkey	39.43	43.99
10104	rad	*D. raddei*		2N	Armenia	39.50	46.35
5405	rad	*D. raddei*		2N	Turkey	38.43	43.41
10313	rad	*D. raddei*		2N	Armenia	40.85	43.95
10339	rad	*D. raddei*		2N	Armenia	40.85	43.95
4342	rad	*D. raddei*		2N	Armenia	39.97	44.95
6082	rad	*D. raddei*		2N	Iran	38.72	44.61

(table 3.1 cont.)

6083	rad	*D. raddei*		2N	Iran	38.72	44.61
5953	rad	*D. raddei*		2N	Iran	37.78	46.41
4982	rad	*D. raddei*		2N	Iran	38.76	47.52
5507	rad	*D. raddei*		2N	Turkey	39.06	43.76
12432	rad	*D. raddei*		2N	Armenia	44.39	40.53
12660	rad	*D. raddei*		2N	Armenia	44.39	40.53
12846	rad	*D. raddei*		2N	Armenia	44.39	40.53
572	rad	*D. raddei*		2N	Armenia	38.93	46.44
643	rad	*D. raddei*		2N	Armenia	38.93	46.44
4345	rad	*D. raddei*		2N	Armenia	39.71	45.21
DB4357	rad	*D. raddei*		2N	NA	NA	NA
5962	rad	*D. raddei*		2N	Iran	38.11	44.56
4348	rad	*D. raddei*		2N	Armenia	39.71	45.21
654	rad	*D. raddei*		2N	Armenia	41.02	45.14
4379	rad	*D. raddei*		2N	Armenia	39.10	46.49
635	rad	*D. raddei*		2N	NKR	40.14	46.52
4363	rad	*D. raddei*		2N	NKR	40.14	46.52
12351	rad	*D. raddei*		2N	Armenia	39.50	46.12
13027	rad-1	*D. raddei*		2N	Armenia	40.37	44.27
10383	rad-1	*D. raddei*		2N	Armenia	40.76	44.80
10385	rad-1	*D. raddei*		2N	Armenia	40.76	44.80
10389	rad-1	*D. raddei*		2N	Armenia	40.76	44.80
10391	rad-1	*D. raddei*		2N	Armenia	40.76	44.80
10404	rad-1	*D. raddei*		2N	Armenia	40.76	44.80
13028	rad-1	*D. raddei*		2N	Armenia	40.38	44.71
4373	rad-1	*D. raddei*		2N	Armenia	40.11	44.73
10314	rad-1	*D. raddei*		2N	Armenia	40.85	43.95
4492	rad-1	*D. raddei*	Lchap	2N	Armenia	40.46	45.08
4493	rad-1	*D. raddei*	Lchap	2N	Armenia	40.46	45.08
12759	rad-1	*D. raddei*		2N	Armenia	44.39	40.53
12882	rad-1	*D. raddei*		2N	Armenia	44.39	40.53
12883	rad-1	*D. raddei*		2N	Armenia	44.39	40.53
12967	rad-1	*D. raddei*		2N	Armenia	44.39	40.53
12391	rad-1	*D. raddei*		2N	Armenia	39.84	45.29

(table 3.1 cont.)

11447	rad-1	*D. raddei*		2N	Armenia	40.12	44.73
9838	rad-1	*D. raddei*		2N	Georgia	42.31	42.69
9874	raxpo	*D. raddei x D. portschinskii*		2N	Armenia	40.74	45.02
9805	rud	*D. rudis*		2N	Georgia	41.62	43.10
9822	rud	*D. rudis*		2N	Georgia	41.62	43.10
9824	rud	*D. rudis*		2N	Georgia	41.62	43.10
9844	rud	*D. rudis*		2N	Georgia	41.62	43.10
12260	rud	*D. rudis*		2N	Georgia	41.39	43.47
5279	rud	*D. rudis*		2N	Turkey	41.72	32.36
5397	rud	*D. rudis*		2N	Turkey	41.72	32.36
5410	rud	*D. rudis*		2N	Turkey	40.70	40.69
5277	rud	*D. rudis*		2N	Turkey	41.34	32.71
7839	rud	*D. rudis*		2N	Georgia	41.73	42.84
7849	rud	*D. rudis*		2N	Georgia	41.73	42.84
7859	rud	*D. rudis*		2N	Georgia	41.79	42.85
5276	rud	*D. rudis*		2N	Turkey	41.04	42.12
5395	rud	*D. rudis*		2N	Turkey	41.76	42.21
5412	rud	*D. rudis*		2N	Turkey	41.76	42.21
DB7859	rud	*D. rudis*		2N	Armenia	NA	NA
12798	rud	*D. rudis*		2N	Georgia	41.61	43.09
12802	rud	*D. rudis*		2N	Georgia	41.61	43.09
DB12811	rud	*D. rudis*		2N	Georgia	41.61	43.09
9910	S_axv	*D. armeniaca x D. valentini*	Sotk	4N	Armenia	40.53	44.39
9913	S_axv	*D. armeniaca x D. valentini*	Sotk	3N	Armenia	40.53	44.39
9944	S_axv	*D. armeniaca x D. valentini*	Sotk	3N	Armenia	40.53	44.39
9875	S_axv	*D. armeniaca x D. valentini*	Sotk	3N	Armenia	40.22	46.00
9877	S_axv	*D. armeniaca x D. valentini*	Sotk	3N	Armenia	40.22	46.00
9878	S_axv	*D. armeniaca x D. valentini*	Sotk	3N	Armenia	40.22	46.00
9881	S_axv	*D. armeniaca x D. valentini*	Sotk	3N	Armenia	40.22	46.00
9885	S_axv	*D. armeniaca x D. valentini*	Sotk	3N	Armenia	40.22	46.00
9950	S_axv	*D. armeniaca x D. valentini*	Sotk	3N	Armenia	40.22	46.00
10014	S_axv	*D. armeniaca x D. valentini*	Sotk	3N	Armenia	40.22	46.00
10034	S_axv	*D. armeniaca x D. valentini*	Sotk	3N	Armenia	40.22	46.00
10048	S_axv	*D. armeniaca x D. valentini*	Sotk	3N	Armenia	40.22	46.00

(table 3.1 cont.)

10050	S_axv	*D. armeniaca x D. valentini*	Sotk	3N	Armenia	40.22	46.00
10135	S_axv	*D. armeniaca x D. valentini*	Sotk	3N	Armenia	40.22	46.00
9968	S_val	*D. valentini*	Sotk	2N	Armenia	40.53	44.39
9884	S_val	*D. valentini*	Sotk	2N	Armenia	40.22	46.00
9893	S_val-1	*D. valentini*	Sotk	2N	Armenia	40.22	46.00
9894	S_val-1	*D. valentini*	Sotk	2N	Armenia	40.22	46.00
10012	S_val-1	*D. valentini*	Sotk	2N	Armenia	40.22	46.00
5499	uni	*D. unisexualis*		2N	Turkey	39.89	42.36
15708	uni	*D. unisexualis*		2N	Turkey	39.43	43.99
15710	uni	*D. unisexualis*		2N	Turkey	39.43	43.99
15713	uni	*D. unisexualis*		2N	Turkey	39.43	43.99
15716	uni	*D. unisexualis*		2N	Turkey	39.43	43.99
659	uni	*D. unisexualis*		2N	Armenia	40.62	44.56
668	uni	*D. unisexualis*		2N	Armenia	40.62	44.56
16372	uni	*D. unisexualis*		2N	Turkey	39.96	42.29
16376	uni	*D. unisexualis*		2N	Turkey	39.96	42.29
16379	uni	*D. unisexualis*		2N	Turkey	39.96	42.29
10301	uni	*D. unisexualis*		2N	Armenia	40.85	43.95
10305	uni	*D. unisexualis*		2N	Armenia	40.85	43.95
10308	uni	*D. unisexualis*		2N	Armenia	40.85	43.95
4510	uni	*D. unisexualis*	Lchap	2N	Armenia	40.46	45.08
10577	uni	*D. unisexualis*		2N	Armenia	40.15	45.50
10064	uni	*D. unisexualis*		2N	Armenia	40.39	45.21
10454	uni	*D. unisexualis*		2N	Armenia	40.39	45.21
4970	uni	*D. unisexualis*		2N	Armenia	40.56	45.01
4971	uni	*D. unisexualis*		2N	Armenia	40.56	45.01
4629	uni	*D. unisexualis*		2N	Armenia	40.82	44.29
9947	uni	*D. unisexualis*		2N	Armenia	40.85	44.25
4685	uzz	*D. uzzelli*		2N	Turkey	39.94	42.30
4686	uzz	*D. uzzelli*		2N	Turkey	39.94	42.30
5399	uzz	*D. uzzelli*		2N	Turkey	39.89	42.36
5496	uzz	*D. uzzelli*		2N	Turkey	39.89	42.36
5656	uzz	*D. uzzelli*		2N	Turkey	39.53	42.21
5686	uzz	*D. uzzelli*		2N	Turkey	39.53	42.21

(table 3.1 cont.)

5758	uzz	*D. uzzelli*		2N	Turkey	39.53	42.21
5767	uzz	*D. uzzelli*		2N	Turkey	39.53	42.21
5802	uzz	*D. uzzelli*		2N	Turkey	39.53	42.21
16381	uzz	*D. uzzelli*		2N	Turkey	39.96	42.29
16382	uzz	*D. uzzelli*		2N	Turkey	39.96	42.29
16383	uzz	*D. uzzelli*		2N	Turkey	39.96	42.29
16384	uzz	*D. uzzelli*		2N	Turkey	39.96	42.29
16394	uzz	*D. uzzelli*		2N	Turkey	39.80	42.49
16398	uzz	*D. uzzelli*		2N	Turkey	39.80	42.49
16399	uzz	*D. uzzelli*		2N	Turkey	39.80	42.49
16400	uzz	*D. uzzelli*		2N	Turkey	39.80	42.49
16401	uzz	*D. uzzelli*		2N	Turkey	39.80	42.49
16406	uzz	*D. uzzelli*		2N	Turkey	39.80	42.49
16413	uzz	*D. uzzelli*		2N	Turkey	39.80	42.49
16425	uzz	*D. uzzelli*		2N	Turkey	39.80	42.49
12310	val	*D. valentini*		2N	Armenia	NA	NA
12562	val	*D. valentini*		2N	Armenia	39.84	45.65
12573	val	*D. valentini*		2N	Armenia	39.84	45.65
12623	val	*D. valentini*		2N	Armenia	39.84	45.65
12727	val	*D. valentini*		2N	Georgia	41.16	43.79
DB4547	val	*D. valentini*	Lchaschen	2N	Armenia	40.51	44.94
DB4551	val	*D. valentini*	Lchaschen	2N	Armenia	40.51	44.94
12360	val	*D. valentini*		2N	Armenia	NA	NA
12948	val	*D. valentini*		2N	Armenia	NA	NA
658	val	*D. valentini*		2N	Armenia	NA	NA
660	val	*D. valentini*		2N	Armenia	NA	NA
4489	val	*D. valentini*		2N	Armenia	40.65	44.58
12242	val-1	*D. valentini*		2N	Armenia	41.07	43.75
12973	val-1	*D. valentini*		2N	Georgia	41.34	43.49
12232	val-1	*D. valentini*		2N	Georgia	41.16	43.79
12411	val-1	*D. valentini*		2N	Georgia	41.16	43.79
12414	val-1	*D. valentini*		2N	Georgia	41.16	43.79
12729	val-1	*D. valentini*		2N	Georgia	41.16	43.79
12730	val-1	*D. valentini*		2N	Georgia	41.16	43.79

(table 3.1 cont.)

Table 3.2: Mulitplex PCR details.

marker	repeated individuals	random repeats	% of repeats (repeated inds + random repeats / total genotypes)	samples	Tm	total genotypes
C113	16	128	37.50%	384		363
Ph128	31	128	41.41%	384		329
Pb55	15	128	37.24%	384		281
Ph39	4	128	34.38%	384		142
P054	17	128	37.76%	384		371
P011	18	128	38.02%	384		358
Du323	14	128	36.98%	384		335
D119	17	128	37.76%	384		363
Ph124	14	128	36.98%	384		282
C118	15	128	37.24%	384		286
Ph170	14	128	36.98%	384		155
Lv-4-72	8	128	35.42%	384		137

Table 3.3: Private alleles for each diploid species.

Summary of Private Alleles by Population

No. Loci	10
No. Samples	317
No. Pops.	18

Pop	Locus	Allele	Freq	Locality (if sympatric)
D. unisexualis	D119.	262	0.024	
	P011	243	0.119	
	P011	247	0.048	
	P011	251	0.048	
D. uzzelli	C113.	258	0.533	
	Du323.	212	0.036	
D. armeniaca	D119.	322	0.022	
	P054	99	0.022	
	Ph124	316	0.022	
D. mixta	D119.	202	0.100	
	D119.	216	0.050	
	D119.	276	0.050	
	D119.	298	0.100	
	D119.	318	0.100	
	Ph124	286	0.050	
D. mixta 2	D119.	214	0.250	
	P054	77	0.333	
	P054	81	0.083	
D. raddei nairensis	P011	199	0.075	
D. raddei raddei	C118.	145	0.048	
	D119.	170	0.043	
	D119.	188	0.022	
	Du323.	186	0.023	
	P054	129	0.023	
	Ph170	225	0.022	
	D119.	190	0.036	

D. rudis	D119.	196	0.071	
	D119.	198	0.071	
	D119.	200	0.071	
	D119.	206	0.036	
	Du323.	188	0.077	
	Ph124	292	0.143	
	Ph124	295	0.036	
	Ph124	298	0.179	
D. valentini	C113.	246	0.022	
	C113.	254	0.022	
	D119.	258	0.065	
	Pb55	239	0.024	
D. valentini	D119.	302	0.019	
	Du323.	216	0.038	
	Ph128	250	0.025	
	Ph170	273	0.068	
D. armeniaca	C113.	228	0.005	Kuchak
	C113.	236	0.005	
	P011	189	0.010	
	P011	191	0.005	
	Ph128	228	0.005	
D. unisexualis	C113.	244	0.300	
D. valentini	Du323.	178	0.100	Sotk
	P011	239	0.100	

(Table 3.3 cont.)

Table 3.4: Private alleles for each polyploid hybrid grouping

Population	Population	Allele present	marker	allele	individuals
Sotk	*D. armeniaca x D. valentini*	Only in Sotk hybrids	Ph128	246	9913
					9944
			Ph39	169	10050
		only in *D. valentini* whole range and Kuchak	Du323	194	9913
			Ph170	246	9913
			P011	225	9913
		D. valentini whole range	D119	242	10048
		D. valentini Kuchak	Lv-4-72	150	9944
Kuchak	D. unisexualis x D. valentini	only in *D. unisexualis x D. valentini* Kuchak hybrids	C113	258	12574
			P011	261	12182
			Lv-4-72	132	12454
		D. valentini whole range	P011	223	12651
			Ph39	115	12454
		D. unisexualis whole range	P011	245	4473
					12237
					12523
					12612
					12631
					12651
		D. valentini whole range and *D. armeniaca* Kuchak (90%)	Lv-4-72	166	12555

Figure S3.1 - Discriminant analysis of principal components (DAPC) of the sexual (*D. raddei, D. mixta, D. valentini* and *D. rudis*) and parthenogenetic (*D. armeniaca, D. unisexualis* and *D. uzzelli*) species. Only diploid individuals are used in this analysis. Colours used correspond to the same coding as in the other figures.

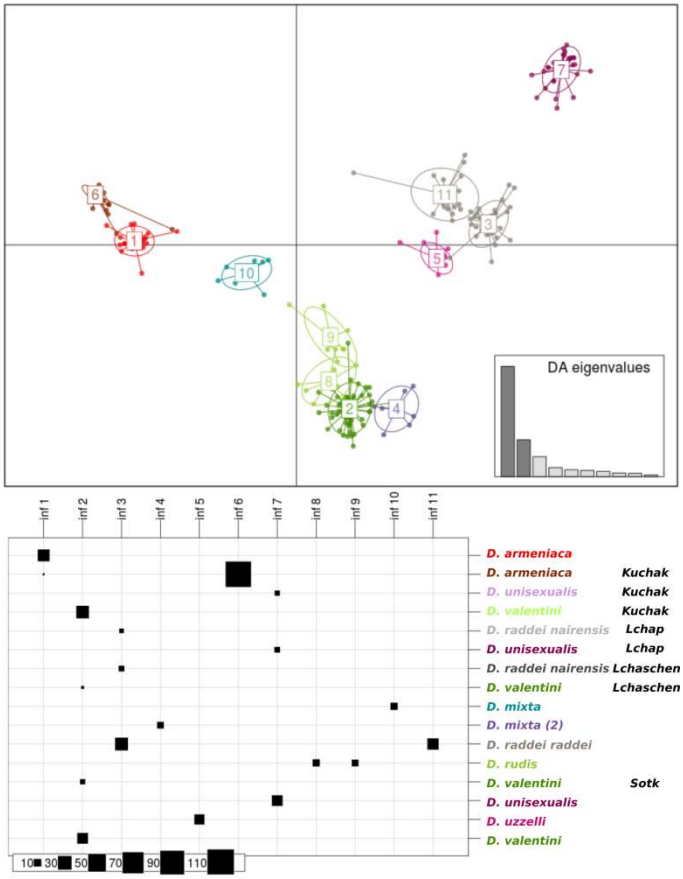


Figure S3.2 – Neighbour-joining tree of the F_{ST} pairwise distances.

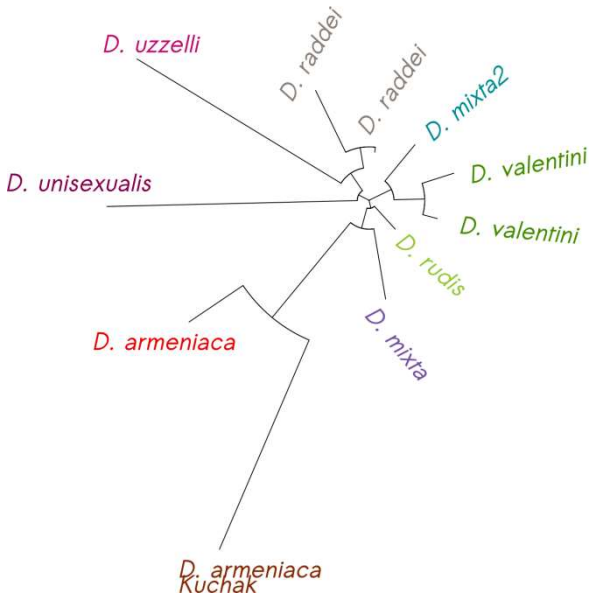


Figure S3.3 – Isolation by distance calculated for the sexual species *D.valentini.*

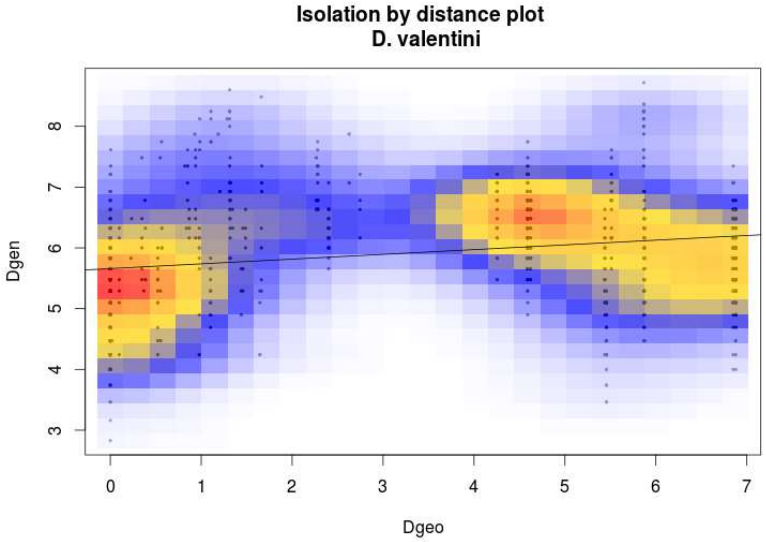
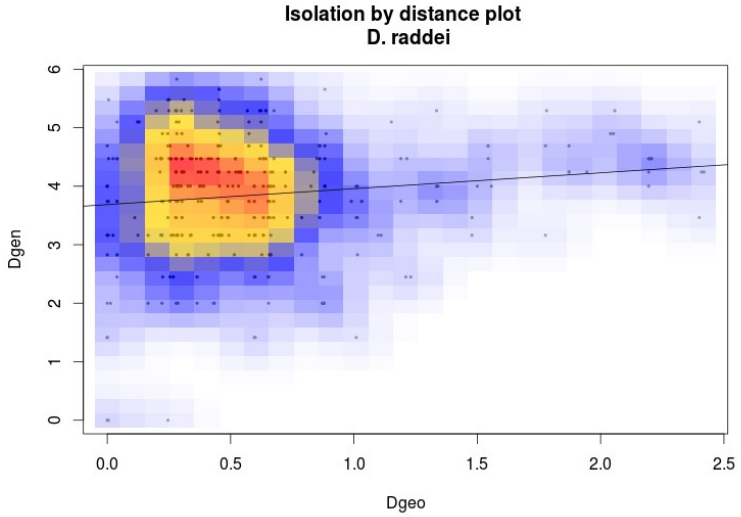


Figure S3.4 – Isolation by distance calculated for the sexual species *D. raddei.*



CHAPTER 4 – SUPPLEMENTARY INFORMATION

Table 4.1: Flagstat mapping results for all tissue reads against the *D. valentini* – tail reference (VT) and *D. unisexualis* – brain/ovaries reference (UBUO).

valtail		against VT		against UBUO
mapped_read	total	78691840		78691840
	duplicate	0		0
	mapped	63819207	(81.10%:-nan	42472619
	paired in sequencing	78691840		78691840
	read1	39345920		39345920
	read2	39345920		39345920
	properly paired	47831750	(60.78%:-nan	35436424
	with itself and mate mapped	62150926		38970154
	singletons	1668281	(2.12%:-nan%	3502465
	with mate mapped to a different chr	12628986		1900598
	with mate mapped to a different chr (mapQ>=5)	5695640		1400322
unitail				
mapped_read	total	68223786		68223786
	duplicate	0		0
	mapped	48857880	(71.61%:-nan	35132790
	paired in sequencing	68223786		68223786
	read1	34111893		34111893
	read2	34111893		34111893
	properly paired	42921368	(62.91%:-nan	30060586
	with itself and mate mapped	47469432		32957546
	singletons	1388448	(2.04%:-nan%	2175244
	with mate mapped to a different chr	3278990		1691734
	with mate mapped to a different chr (mapQ>=5)	1817674		1069842
unibra				
mapped_read	total	63666608		63666608
	duplicate	0		0
	mapped	41599454	(65.34%:-nan	33719298
	paired in sequencing	63666608		63666608
	read1	31833304		31833304
	read2	31833304		31833304
	properly paired	37722348	(59.25%:-nan	28921298
	with itself and mate mapped	40366830		32022752
	singletons	1232624	(1.94%:-nan%	1696546
	with mate mapped to a different chr	1447650		2014580
	with mate mapped to a different chr (mapQ>=5)	885262		1180912
uniova				
mapped_read	total	72444128		72444128
	duplicate	0		0
	mapped	50419497	(69.60%:-nan	36653378
	paired in sequencing	72444128		72444128
	read1	36222064		36222064
	read2	36222064		36222064
	properly paired	46447376	(64.11%:-nan	31820668
	with itself and mate mapped	49210930		34850752
	singletons	1208567	(1.67%:-nan%	1802626
	with mate mapped to a different chr	961250		1413538
	with mate mapped to a different chr (mapQ>=5)	670124		876496

CHAPTER 5 – SUPPLEMENTARY INFORMATION

Table 5.1: Species information for all individuals included in the analyses.

pop_name	*species*	code	Reproduction mode
brau	*D. brauneri*	Dbrau6299	sexual
brau	*D. brauneri*	Dsax6276	sexual
chlo	*D. chlorogaster*	Dchlo6132	sexual
chlo	*D. chlorogaster*	Dchlo6146	sexual
clar	*D. clarkorum*	Dclar5440	sexual
clar	*D. clarkorum*	Dclar5818	sexual
defi	*D. defilippii*	Ddefi6200	sexual
defi	*D. defilippii*	Ddefi6208	sexual
defi	*D. defilippii*	Drad6093	sexual
der	*D. derjugini*	Dder16266	sexual
mix1	*D. mixta*	Dmix10290	sexual
mix1	*D. mixta*	Dmix9816	sexual
mix1	*D. mixta*	Dmix9829	sexual
mix1	*D. mixta*	Dmix9849	sexual
mix2	*D. mixta - 2*	Dmix15741	sexual
mix2	*D. mixta - 2*	Dmix16249	sexual
par	*D. parvula*	Dpar12738	sexual
par	*D. parvula*	Dpar16237	sexual
par	*D. parvula*	Drud16387	sexual
por1	*D. portschinskii*	Dpor10409	sexual
por1	*D. portschinskii*	Dpor12777	sexual
por2	*D. portschinskii*	Dpor10250	sexual
por2	*D. portschinskii*	Dpor6280	sexual
por2	*D. portschinskii*	Dpor676	sexual
praB	*D. praticola - Balkans*	Dpra13285	sexual
praB	*D. praticola - Balkans*	Dpra16331	sexual
praC	*D. praticola - Caucasus*	Dpra10360	sexual
praC	*D. praticola - Caucasus*	Dpra14664	sexual
praC	*D. praticola - Caucasus*	Dpra14665	sexual
rad1	*D. raddei*	Dnai10303	sexual
rad1	*D. raddei*	Dnai12759	sexual
rad1	*D. raddei*	Dpor12447	sexual
rad1	*D. raddei*	Drad10291	sexual
rad1	*D. raddei*	Drad10338	sexual
rad1	*D. raddei*	Drad11447	sexual
rad1	*D. raddei*	Drad12882	sexual
rad1	*D. raddei*	Drad12934	sexual
rad1	*D. raddei*	Drad15728	sexual
rad1	*D. raddei*	Drad4493	sexual
rad1	*D. raddei*	Drad4982	sexual
rad2	*D. raddei*	Drad15715	sexual
rad2	*D. raddei*	Drad5405	sexual
rad2	*D. raddei*	Drad6281	sexual
rud1	*D. rudis*	Drud7859	sexual
rud1	*D. rudis*	Drud9805	sexual
rud1	*D. rudis*	Dval12260	sexual
rud2	*D. rudis*	Drud5397	sexual
rud2	*D. rudis*	Drud5410	sexual
rud2	*D. rudis*	Drud5512	sexual
rud2	*D. rudis*	Dval16280	sexual
sax	*D. saxicola*	Dder6287	sexual
sax	*D. saxicola*	Dsax6286	sexual

(Table 5.1 cont.)

stei	*D. steineri*	Dstei6111	sexual
stei	*D. steineri*	Dstei6112	sexual
stei	*D. steineri*	Dstei6115	sexual
val	*D. valentini*	Dval10012	sexual
val	*D. valentini*	Dval10048	sexual
val	*D. valentini*	Dval10050	sexual
val	*D. valentini*	Dval12229	sexual
val	*D. valentini*	Dval12310	sexual
val	*D. valentini*	Dval12360	sexual
val	*D. valentini*	Dval12464	sexual
val	*D. valentini*	Dval12562	sexual
val	*D. valentini*	Dval12641	sexual
val	*D. valentini*	Dval13009	sexual
val	*D. valentini*	Dval4536	sexual
val	*D. valentini*	Dval658	sexual
val	*D. valentini*	Dval9893	sexual
val	*D. valentini*	Dval9894	sexual

Figure S5.2 – Summary of the introgression hybridization events detected with the different analyses used in this study.

	brau	chlo	clar	def i	der	mix1	mix2	par	por	pra	rad	rud	sax	stei	val	out
brau																
chlo																
clar	2	**2, 4(k)**														
def i	4(a,b)	2														
der																
mix1	4(c,d)															
mix2																
par		2	2													
por	4(e)	4(f)	2, 4(g)	2		2, 4(h)	4(i)	**2, 3, 4(j)**								
pra		**2, 3**	2			2			2							
rad		**1, 2, 3, 4(a,l)**		**1, 2, 3**	**1, 4(b)**	3		**2, 4(c)**		**2, 4(o)**						
rud	4(m)	4(n)	2	2				**1, 2**	1	2	2					
sax									4(o,p)							
stei	4(k,l)		2			2		4(d)		4(p)	1					
val			2	2				2	4(m,n)			**1, 2, 4(e-j)**				
out																

BUCKY	1
ABBA/BABA	2
TM-MIGRATIC	3
F4	4

#	P1	P2	P3	O	Signal of *D*-statistic	significant?	significant tests (yellow)
1	POR	VAL	RAD	IRANO		NO	1
2	RUD	VAL	RAD	IRANO	<	YES	2
3	POR	RUD	RAD	IRANO	>	YES	3
4	RAD	MIX1	VAL	IRANO		NO	4
5	RAD	MIX1	RUD	IRANO		NO	5
6	RAD	MIX1	POR	IRANO	>	YES	6
47	RAD	DEFI	CHLO	IRANO		NO	47
48	RAD	STEI	CHLO	IRANO	<	YES	48
49	PAR	CHLO	RAD	IRANO	>	YES	49
50	MIX2	CHLO	RAD	IRANO	>	YES	50
51	MIX1	CHLO	RAD	IRANO	>	YES	51
52	CLAR	CHLO	RAD	IRANO	>	YES	52
59	RAD	CLAR	VAL	IRANO	>	YES	59
60	RAD	CLAR	RUD	IRANO	>	YES	60
61	RAD	CLAR	POR	IRANO	>	YES	61
62	RAD	PRA	VAL	IRANO		NO	62
63	RAD	PRA	RUD	IRANO	>	YES	63
64	RAD	PRA	POR	IRANO	>	YES	64
65	MIX1	CLAR	VAL	IRANO	>	YES	65
66	MIX1	CLAR	RUD	IRANO	>	YES	66
67	MIX1	CLAR	POR	IRANO	>	YES	67
68	MIX1	PRA	VAL	IRANO		NO	68
69	MIX1	PRA	RUD	IRANO		NO	69
70	MIX1	PRA	POR	IRANO	>	YES	70
71	CLAR	CHLO	DEFI	IRANO		NO	71
72	MIX1	CLAR	CHLO	IRANO	>	YES	72
73	RAD	CHLO	PRA	IRANO	<	YES	73
74	PARM	MIX1	VAL	IRANO	<	YES	74
75	PARM	MIX1	RUD	IRANO		NO	75
76	PARM	MIX1	POR	IRANO	<	YES	76
77	PARM	CLAR	VAL	IRANO		NO	77
78	PARM	CLAR	RUD	IRANO		NO	78
79	PARM	CLAR	POR	IRANO		NO	79
80	PARM	PRA	VAL	IRANO	<	YES	80
81	PARM	PRA	RUD	IRANO	<	YES	81
82	PARM	PRA	POR	IRANO		NO	82
83	PARM	DEFI	VAL	IRANO		NO	83
84	PARM	DEFI	RUD	IRANO	>	YES	84
85	PARM	DEFI	POR	IRANO		NO	85
86	PARM	STEI	VAL	IRANO	<	YES	86
87	PARM	STEI	RUD	IRANO	<	YES	87
88	PARM	STEI	POR	IRANO		NO	88
89	PARM	RAD	VAL	IRANO	<	YES	89
90	PARM	RAD	RUD	IRANO	<	YES	90
91	PARM	RAD	POR	IRANO	<	YES	91
92	PARM	CHLO	VAL	IRANO	<	YES	92
93	PARM	CHLO	RUD	IRANO	<	YES	93
94	PARM	CHLO	POR	IRANO	<	YES	94
95	PARM	BRAUS	VAL	IRANO	<	YES	95
96	PARM	BRAU	RUD	IRANO	<	YES	96
97	PARM	BRAU	POR	IRANO	<	YES	97
98	DEFI	STEI	VAL	IRANO	<	YES	98
99	DEFI	STEI	RUD	IRANO	<	YES	99
100	DEFI	STEI	POR	IRANO	<	YES	100
101	DEFI	STEI	RAD	IRANO	<	YES	101
102	PARM	MIX1	CHLO	IRANO		NO	102
103	PARM	MIX1	RAD	IRANO		NO	103
104	PARM	MIX1	DEFI	IRANO		NO	104
105	PARM	MIX1	STEI	IRANO	>	YES	105
106	PARM	MIX1	BRAUS	IRANO		NO	106
107	PARM	MIX1	PRA	IRANO	>	YES	107
108	PARM	CLAR	CHLO	IRANO	>	YES	108
109	PARM	CLAR	RAD	IRANO		NO	109
110	PARM	CLAR	DEFI	IRANO		NO	110
111	PARM	CLAR	STEI	IRANO	>	YES	111
112	PARM	CLAR	BRAUS	IRANO	>	YES	112
113	PARM	CLAR	PRA	IRANO	>	YES	113
114	DEFI	STEI	CHLO	IRANO	<	YES	114
115	RAD	DEFI	CHLO	VALg	<	YES	115
116	RAD	STEI	CHLO	VALg	<	YES	116
117	PARM	CHLO	RAD	VALg	>	YES	117
118	CLAR	CHLO	DEFI	VALg	>	YES	118
119	MIX1	CLAR	CHLO	VALg	>	YES	119
120	RAD	CHLO	PRA	VALg	>	YES	120
121	VAL	POR	RUD	IRANO	<	YES	121
122	CLAR	MIX1	PARM	IRANO	<	YES	122
123	RAD	MIX1	PARM	IRANO	<	YES	123
124	CHLO	MIX1	PARM	IRANO	<	YES	124

The largest P value that has P<(i/m)Q is significant, and all of the P values smaller than it are also significant, even the ones that aren't less than their Benjamini-Hochberg critical value.

I = rank
m = total number of tests
Q = rate of false discovery

D < 0 P1 IS CLOSER THAN P2
D > 0 P2 IS CLOSER THAN P1

Table S5.1 – (previous page) Combinations of species (tests) used to estimate introgressive hybridization with *D*-statistics.

Figure S5.1 – SVDQuartets species tree inference calculated with 300 markers.

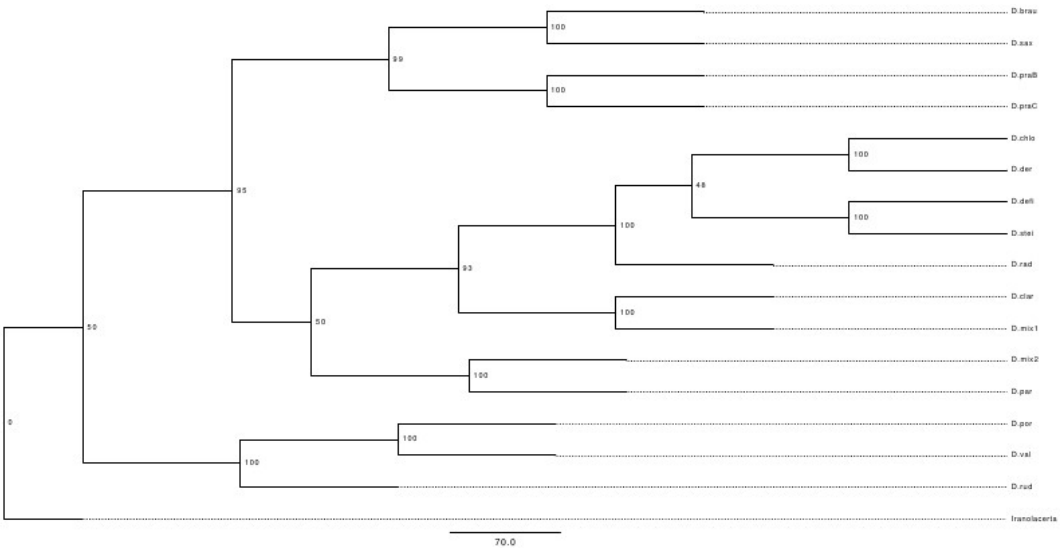


Figure S5.2 – Distribution of *D*-statistics values calculated for each test.

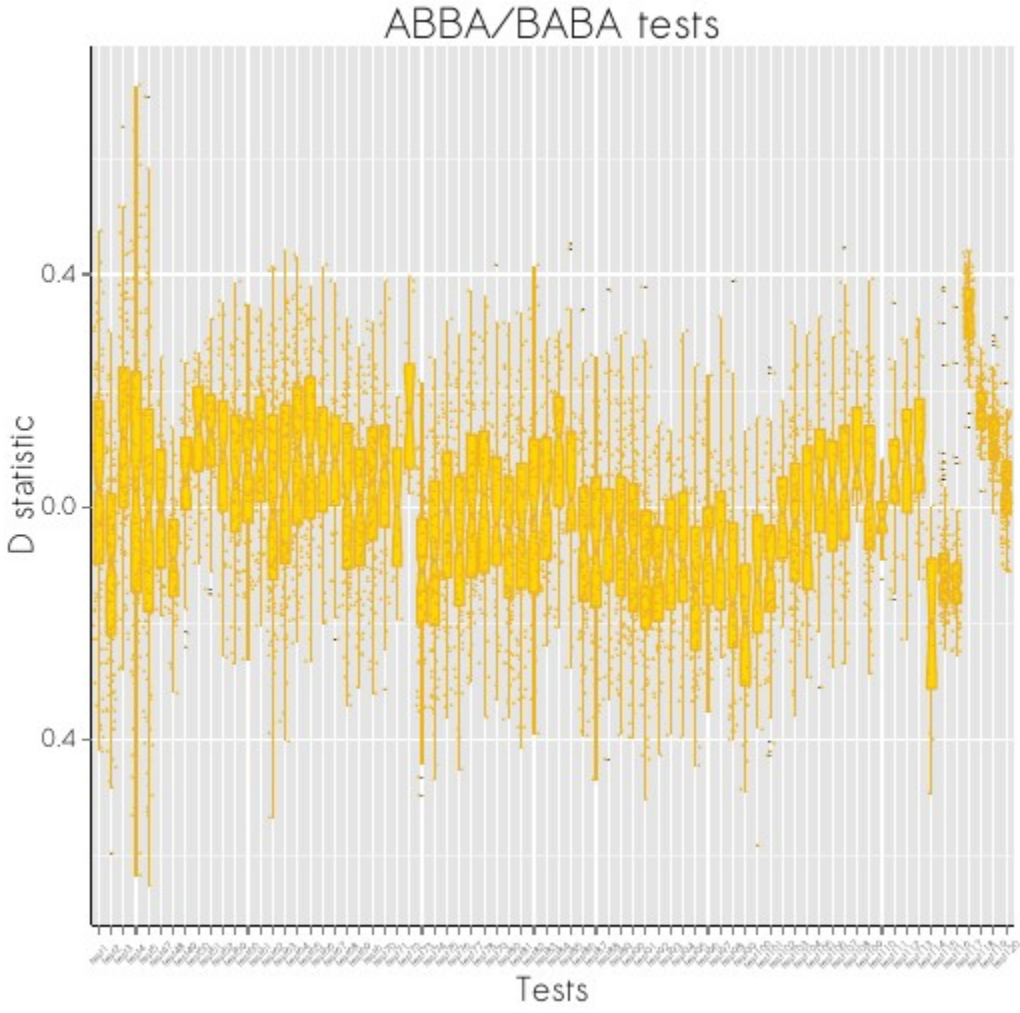


Figure S5.3 - The fraction of variance in relatedness between populations accounted for by phylogenetic models with 0 through 21 migration edges. The fraction of variance in the sample covariance matrix () accounted for by the model covariance matrix (). (Pickrell and Pritchard, 2012) showed that the fraction began to asymptote at 0.998 when the models accurately depicted relationships between simulated populations. We also observed this asymptote near 0.998 in our empirical analysis, leading us to conclude that the relationships between the Darevskia sexual taxa were accurately described by a phylogenetic network with 11 migration edges.

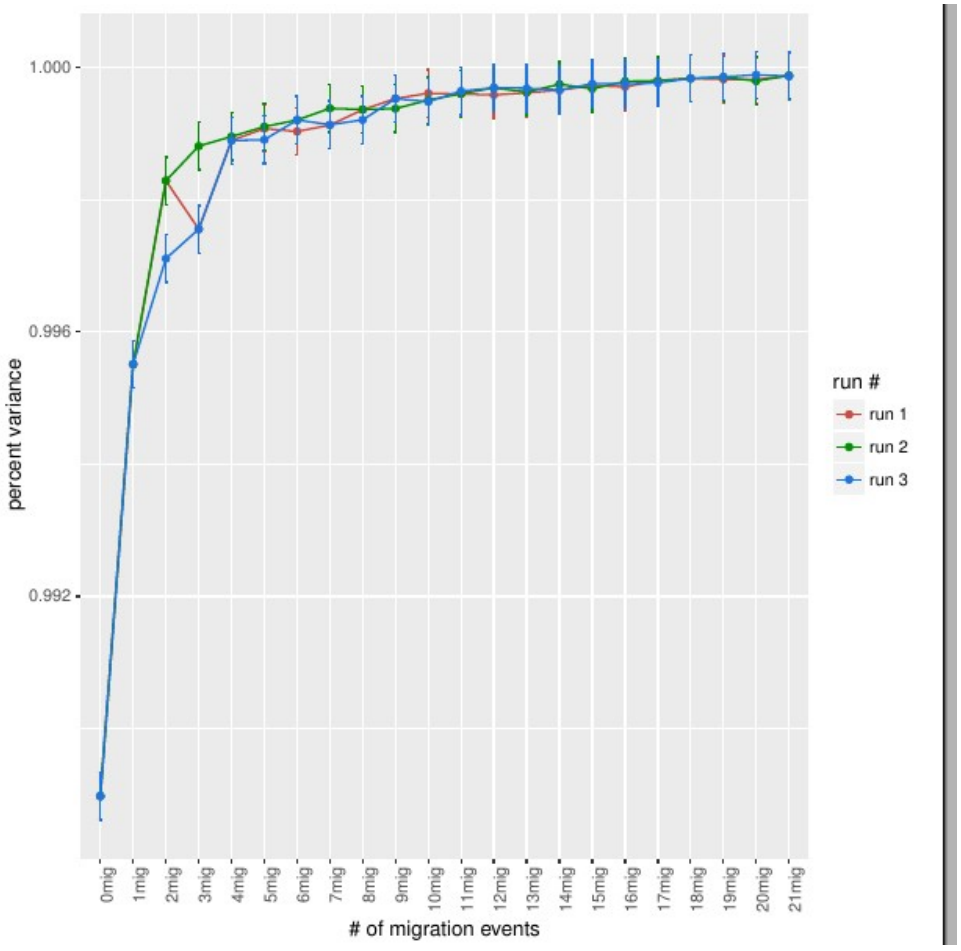
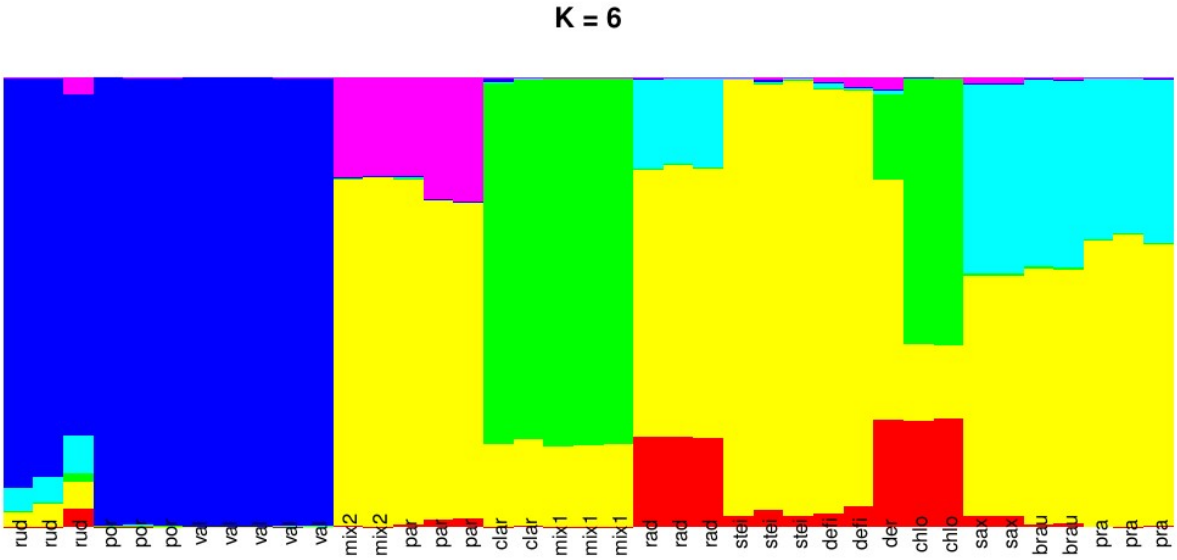


Figure S5.4 – STRUCTURE analysis using the optimum *K* value of 6. All species were included in this analysis, but only 3 individuals was used per species.



FCUP
Why Sex? *Darevskia* answers.