RESOLVING RETICULATE SPECIATION IN BISEXUAL AND PARTHENOGENETIC LIZARDS OF GENUS *DAREVSKIA* IN EAST ANATOLIA AND CAUCASUS

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

MERIÇ ERDOLU

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF DOCTOR OF PHILOSOPHY
IN
BIOLOGY

JANUARY 2023

Approval of the thesis:

RESOLVING RETICULATE SPECIATION IN BISEXUAL AND PARTHENOGENETIC LIZARDS OF GENUS *DAREVSKIA* IN EAST ANATOLIA AND CAUCASUS

submitted by **MERIÇ ERDOLU** in partial fulfillment of the requirements for the degree of **Doctor of Philosophy in Biology Department, Middle East Technical University** by,

Prof. Dr. Halil Kalıpçılar
Dean, Graduate School of **Natural and Applied Sciences** _____

Prof. Dr. Ayşe Gül Gözen
Head of Department, **Biology** _____

Prof. Dr. Mehmet Somel
Supervisor, **Biology, METU** _____

Assist. Prof. Dr. Alexey Yanchukov
Co-supervisor, **Biology, Zonguldak Bülent Ecevit University** _____

Examining Committee Members:

Prof. Dr. Cemal Can Bilgin
Biology, METU _____

Prof. Dr. Mehmet Somel
Biology, METU _____

Assist. Prof. Dr. İsmail Kudret Sağlam
Molecular Biology And Genetics, Koç University _____

Prof. Dr. Raşit Bilgin
Institute of Environmental Sciences, Boğaziçi University _____

Assist. Prof. Dr. Emrah Çoraman
Eurasia Ins. of Earth Sciences Ins., Istanbul Technical University _____

Date:20.01.2023

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Surname: Meriç Erdolu

Signature :

ABSTRACT

RESOLVING RETICULATE SPECIATION IN BISEXUAL AND PARTHENOGENETIC LIZARDS OF GENUS *DAREVSKIA* IN EAST ANATOLIA AND CAUCASUS

Erdolu, Meriç
Ph.D., Department of Biology
Supervisor: Prof. Dr. Mehmet Somel
Co-Supervisor: Assist. Prof. Dr. Alexey Yanchukov

January 2023, 137 pages

The genus *Darevskia* (rock lizards) from East Anatolia, Georgia and Armenia represents a rare biological system where both sexually reproducing and parthenogenetic forms are found together. In total, seven nominal parthenogenetic species (*D. unisexualis, D. rostombekovi, D. dahli, D. sapphirina, D. bendimahiensis, D. uzzelli*) are estimated to have formed several hundred thousand years ago via interspecific hybridization between four bisexual species, of which two (*D. mixta* and *D. raddei*) always acted as maternal and two (*D. valentini* and *D. portschinskii*) acted as paternal parents. The same parental pairs were able to produce different parthenogenetic species, each of which possesses distinct morphologies and ecological preferences. The *Darevskia* system can thus offer a unique insight into genetic mechanisms of the origin and maintenance of parthenogenesis in vertebrates since the process of hybridization has occurred in parallel. The dissertation is composed of four chapters.

In Chapter 1, we introduce the genus *Darevskia* and summarize the studies on this genus. Also, we give the objectives of the thesis. In Chapter 2, we describe in detail

the methods used in the study.

In Chapter 3, I present my contribution to a published study on the paternal origin and phylogenetic history of all seven parthenogenetic species. In this study, we analyzed genomic data of specimens from the bisexual species *Darevskia raddei* (n=25), *D. portschinskii* (n=9), *D. mixta* (n=12), *D. valentini* (n=18), *D. derjugini* (n=1), and *D. rudis* (n=6), as well as the parthenogenetic species *D. sapphirina* (n=8), *D. bendimahiensis* (n=10), *D. rostombekovi* (n=6), *D. dahli* (n=11), *D. uzzelli* (n=8), *D. unisexualis* (n=7), and *D. armeniaca* (n=13) species. With this data we aimed to identify those extant bisexual populations that are genetically closest to the actual paternal progenitors of the parthenogenetic species. We used 12,650 Z-chromosomal ddRAD-seq loci and several bioinformatics methods and tools for population genetics analysis. We found that the most likely paternal ancestor of each parthenogenetic species is also the geographically nearest one. Also, we found that hybridization seemed to have occurred during two time periods (0.83 Mya and 1.57 Mya).

In Chapter 4, we reveal the genetic details of the origin of two nominal parthenogenetic species *D. sapphirina* and *D. bendimahiensis*, endemic to Lake Van basin. The analysis was based on 1,381,322 autosomal ddRAD-seq loci. In line with the previous results based on 10 microsatellite markers as well as Z-linked ddRAD loci, the two nominal species were found to be indistinguishable genetically and should be revised into a single taxon. Their most likely paternal ancestor is the population of *D. valentini* from Çaldıran, located only 4 km away from the nearest population of *D. bendimahiensis*. In contrast, the most likely maternal ancestor of *D. sapphirina* and *D. bendimahiensis* is the geographically distant population of *D. raddei* from Doğubayazıt population in the Aras river basin, not the sympatric *D. raddei* population in the Lake Van basin.

Finally, we designed and employed several methods to estimate the number of hybridization events that led to the origin of *D. bendimahiensis* / *D. sapphirina*. The pattern of allele frequency correlation between the parental and parthenogenetic population, as well as the presence of multiple ddRAD loci with three distinct alleles, do not support the null hypothesis of a single hybridization event between just two individuals of the parental species. The paper draft summarising these fundings is

currently in preparation.

Keywords: Darevskia, rock lizrads, parthenogenesis, hybridization, reticulate speciation, RADseq, asexual vertebrates, clonal reproduction

ÖZ

DOĞU ANADOLU VE KAFKASLARDAKİ DAREVSKİA GENUSUNA AİT EŞEYLİ VE EŞEYSİZ (PARTENOGENETİK) ÜREYEN KERTENKELELERİN AĞSI TÜRLEŞME SÜRECİNİN ÇÖZÜMLENMESİ

Erdolu, Meriç
Doktora, Biyoloji Bölümü
Tez Yöneticisi: Prof. Dr. Mehmet Somel
Ortak Tez Yöneticisi: Dr. Öğr. Üyesi. Alexey Yanchukov

Ocak 2023 , 137 sayfa

Doğu Anadolu, Gürcistan ve Ermenistan'dan gelen *Darevskia* kertenkele (kaya kertenkeleleri olarak bilinir) cinsi, hem eşeyli üreyen hem de eşeysiz (partenogenetik) üreyen tek eşeyli türlerin bir arada bulunduğu nadir bir biyolojik sistemi temsil eder. Toplam yedi eşeyiz üreyen (partenogenetik) türün (*D. unisexualis, D. rostombekovi, D. dahli, D. sapphirina, D. bendimahiensis, D. uzzelli*) yüz binlerce yıl önce dört eşeyli üreyen *Darevskia* türünün (*D. mixta* ve *D. raddei* türlerinden sadece dişi bireyler) ve (*D. valentini* ve *D. portschinskii* türlerinden sadece erkek bireyler) hibridizasyonu sonucu ortaya çıkmışlardır. Aynı ebeveyn çiftleri, her biri farklı morfolojilere ve ekolojik tercihlere sahip olan farklı partenogenetik türler üretebilir ve şu anda yedi tür partenogenetik olarak tanımlanmıştır. *Darevskia* sistemi, hibridizasyon süreci birkaç tek eşeyli kardeş populasyonda paralel olarak meydana geldiğinden, ağsı türleşmenin genetik mekanizmalarına benzersiz bir bakış açısı sunabilir. Çalışma dört bölümden meydana gelmektedir.

Birinci bölümde *Darevskia* cinsi tanıtılıp onunla ilgili yapılan çalışmalar özetlenmektedir. İkinci bölümde ise çalışmada kullanılan metotlar detaylı olarak anlatılmaktadır.

Üçüncü bölümde, eşeysiz üreyen yedi hibrit *Darevskia* türünün olası atasal baba populasyonlarıyla ilgili çalışmaya katkılarımı sunuyorum. Bu çalışmada eşeyli üreyen, *D. raddei* (n=25), *D. portschinskii* (n=9), *D. mixta* (n=12), *D. valentini* (n=18), *D. derjugini* (n=1), *D. rudis* (n=6) ve partenogenetik, *D. sapphirina* (n=8), *D. bendimahiensis* (n=10), *D. rostombekovi* (n=6), *D. dahli* (n =11), *D. uzzelli* (n=8), *D. unisexualis* (n=7), *D. armeniaca* (n=13) örneklerine ait genetik veriyi analiz ettik. Bu veriyle partenogenetik türlerin en olası baba atalarını bulmayı amaçladık. 12,650 ddRAD-seq lokusu ve birçok biyoinformatik metot ve araç kullanarak populasyon genetiği analizlerini gerçekleştirdik. Sonuç olarak tüm partenogenetik türlerin baba atalarının coğrafi olarak en yakın aday baba popülasyonu olduğunu bulduk. Ayrıca hibritleşmenin iki zaman periyodu arasında gerçekleştiğini bulduk (0.83 My - 1.57 My) .

Dördüncü bölümde, endemik *D. sapphirina* ve *D. bendimahiensis* hibrit türlerinin Van Gölü havzasındaki en olası atasal ebeveyn populasyonları ve geçmişlerindeki hibritleşme sayısıyla ilgili analizleri ve sonuçları sunuyorum. Populasyon genetiği analizlerini gerçekleştirmek için 1,381,322 otozomal ddRAD-seq lokuslarını kullandık. 10 mikrosatelit ve ddRAD Z kromozomu belirteçlerinin sonuçlarıyla birlikte iki nominal türün genetik olarak ayırt edilemediklerini ve tek bir takson olarak revize edilmesi gerektiğini bulduk. Atasal baba populasyon sonuçlarının aksine anne atalarının Van Gölü havzasındaki en yakın aday anne populasyonu olmadığını, bunun yerine Aras nehri havzasındaki *D. raddei*'nin Doğubayazıt populasyonunun muhtemel en yakın anne popülasyonu olduğunu bulduk fakat en olası atasal baba populasyonunun coğrafi olarak da en yakın aday baba populasyonu olan *D. valentini*'nin Çaldıran populasyonu olduğunu bulduk.

Son olarak, *D. bendimahiensis*/*D. sapphirina* gruplarının geçmişlerindeki hibridizasyon olaylarının sayısını tahmin etmek için birçok biyoinformatik yöntem kullandık. *D. bendimahiensis* ve *D. sapphirina* populasyonlarındaki çoklu haplotipik lokuslara karşılık gelen allellerin frekanslarının korelasyonları ve fazla farklılaşmış allellerin varlığı, iki ebeveyn birey arasında tek bir hibritleşmenin gerçekleştiğini söyleyen boş hipotezi desteklemiyor. Bu çalışmanın sonuçlarını özetleyen makale şu an hazırlan-

maktadır.

Anahtar Kelimeler: Darevskia, kaya kertenkeleleri, partenogenez, hibritleşme, ağsı türleşme, RADseq, eşeysiz üreyen omurgalılar, klonal üreme

Dedicated to My Beloved Laborer Father and Mother

ACKNOWLEDGMENTS

Firstly, I would like to express my sincere gratitude to my advisors Assist. Prof. Dr. Alexey Yanchukov and Prof. Dr. Mehmet Somel, who patiently helped and supported me while preparing this thesis. I would never have finished this study without their guidance. Moreover, if one of the most important things for a professor is to influence a student's perspective on science and life, my professors have greatly influenced and contributed to my views.

I would like to express my deepest appreciation and gratitude to Prof. Dr. C. Can Bilgin and Assist. Prof. Dr. İsmail K. Sağlam, who were on my thesis monitoring committee (TİK), for their insightful discussions, contributions, and valuable time.

I would like to thank all my lab mates from CompEvo for their contributions to the study in the discussions at the lab meetings. Moreover, special thanks to N. Göksun Özhan, who helped with the lizard silhouette used in Figure 4.3, and to Z. Gözde Turan, who evaluated all the figures in the article presented in Chapter 4 and made precious suggestions for aesthetic correction, and to Melih Yıldız, who reviewed some parts of the article and gave very important opinions for the study presented in Chapter 4. I would also like to thank these people for their beautiful and sincere friendships.

Many thanks to my friends and colleagues Ortaç Çetintaş and Halil Mert Solak from Bülent Ecevit University Evolutionary Biology Lab who assisted with the handling of samples in the study presented in Chapter 3.

Lastly, I would like to express my sincere thanks and appreciation to my mother Işık Erdolu, and to my father Mahmut Erdolu, they patiently answered all my questions when I was little and reinforced my interest in science and philosophy, and thanks a lot to them and my beloved sister M. Dicle Erdoğan and my brother-in-law Şahan Erdoğan for their support and motivation during the study. I would not accomplish this work without them.

This study is supported by a joint scientific grant program of the Scientific and Technical Research Council of Turkey (TÜBİTAK grant no. 216Z189 to Alexey Yanchukov) and LEPL Shota Rustaveli National Science Foundation (SRNSF grant No04/41 to David Tarkhnishvili).

In this study, the Turkish National e-Science e-Infrastructure (TRUBA) servers were also used during data analyses.

TABLE OF CONTENTS

LIST OF TABLES

TABLES

LIST OF FIGURES

FIGURES

LIST OF ABBREVIATIONS

DNA	Deoxyribonucleic Acid
RE	Restriction enzyme
ML	Maximum Likelihood
mtDNA	Mitochondrial DNA
nDNA	Nuclear DNA
ddRAD-seq	Double digest restriction-site associated DNA
N	Population size
Ne	Effective population size
n	Sample size
Myr	Million Years
kyr	Thousand Years
Mya	Million Years Ago
stdev	Standard deviation
MCMC	Markov chain Monte Carlo
MALO	Multi-Allelic Locus
bp	Base pair
PCR	Polymerase Chain Reaction
SNP	Single Nucleotide Polymorphism
SNV	Single Nucleotide Variation
Z-linked	Related with chromosome Z
NJ	Neighbor joining

CHAPTER 1

GENERAL INTRODUCTION

1.1 Asexuality in Vertebrates

Vertebrates are generally bisexual, and asexual reproduction is more frequently observed among invertebrate animals than vertebrates. However, there are still numerous asexual vertebrate species in nature. From fish to amniotes, asexual reproduction can be seen in many vertebrate species in nature (Laskowski et al., 2019; Avise, 2015).

These asexual vertebrate species reproduce in three modes: gynogenesis, hybridogenesis, and parthenogenesis (Neaves & Baumann, 2011). In gynogenesis, females produce a diploid or polyploid egg, but in order to start the embryogenesis, this egg cell must interact with a sperm cell from a male of a genetically related sexual species. Here there is no genetic contribution from sperm to the offspring. Gynogenesis is seen in some fish and salamander species. In hybridogenesis, females produce a haploid egg and this egg must be fertilized by a sperm from another related species so as to produce an offspring, but even though this offspring expresses all maternal and paternal chromosomes in its somatic cells when it produces egg cells, paternal chromosomes are not included in the oogenesis process. Hence, the haploid eggs involve only the maternal chromosome complement (Neaves & Baumann, 2011; Avise, 2015; Fujita et al., 2020). Because the offspring are semi-clonal, this reproduction type is called hybridogenesis from "hybrid" and "genesis" meaning birth, formation, or existence, altogether meaning hybrid birth in a literary sense. Hybridogenesis is generally seen in some frogs, salamanders, and fish species.

In the parthenogenetic (from the Greek, "parthenos", virgin + "genesis", birth, formation, existence, meaning "virgin birth") reproduction mode, female individuals only

produce and lay diploid or polyploid eggs and after hatching, new viable offspring, which are genetically clones of their mother, begin their life. This type of reproduction is also called "true parthenogenesis" and does not include any male contribution (Neaves & Baumann, 2011; Avise, 2015). It is seen in some reptile species such as lizards, and snakes; (Kearney et al., 2009) in fish; (Feldheim et al., 2010) and in some bird species (Schut et al., 2008).

Parthenogenesis is observed in two types: automictic and apomictic parthenogenesis. In automictic parthenogenesis, the egg cell fuses with one of the polar bodies produced by regular canonical meiosis (Kooi & Schwander, 2015), so this kind of parthenogenesis is called automictic (meaning "auto" - "mixing"), whereas, in apomictic reproduction, there is no genetic recombination ("apo" - "mixing" means "away from" + "mixing") and the offspring are clones of their mother (Hales et al., 2002). In addition, parthenogenetic reproduction can be also observed occasionally in organisms that reproduce sexually such as sharks (Chapman et al., 2007), Komodo dragons (Watts et al., 2006), and snakes (Booth et al., 2011). This kind of parthenogenesis is called facultative parthenogenesis (Lampert, 2008; Kooi & Schwander, 2015) and it is seen only in captive specimens that had been kept in isolation from male individuals for a very long time. On the other hand, if females from all-female lineages cannot reproduce sexually (even if they mate with males from sexual lineages), this parthenogenetic reproduction is called obligate parthenogenesis (Kooi & Schwander, 2015) and it is seen in some reptile families such as Gekkonidae, Lacertidae (Kearney et al., 2009), Typhlopidae (Booth & Schuet, 2016).

1.2 Origin of Clonal Vertebrate Populations

Genus *Darevskia* is one of the vertebrate genera that have parthenogenetic species. Two hypotheses have been put forward to explain the appearance of parthenogenesis based on the genus *Darevskia*, and with the first of them in 1985, hybridization was accepted as a principal mechanism of parthenogenesis in vertebrates (Darevsky, 1985). Beyond hybridization, these two hypotheses, the "Balancing Hypothesis" and the "Phylogenetic Constraint Hypothesis" (Murphy et al., 2000), provide two possible explanations for the evolution of true parthenogenetic vertebrates.

The genetic distance between two species can contribute to reproductive isolation through the presence of incompatible genes that prevent fertilization or the development of viable and fertile offspring (Martin and Mendelson, 2018; Presgraves and Meiklejohn, 2021). The Balance Hypothesis argued by Moritz et al. (1989) proposes that when two different species mate, their genetic distance should allow the offspring to produce viable and fertile offspring, but also this genetic interaction should not allow for normal meiosis. And the offspring should produce unreduced egg cells for a parthenogenetic lineage. The Phylogenetic Constraint Hypothesis proposed by Darevsky, Kupriyanova, and Uzzell (1985) is based on specific genetic factors coming from phylogenetic differences between two different bisexual species, and thanks to these specific factors, these two species are able to produce parthenogenetic offspring. While the former hypothesis suggests a general role for specific genetic factors in the two parental species involved in creating a parthenogenetic species, the latter hypothesis concentrates on the genetic distance between the two parental species.

1.2.1 Advantages and Disadvantages of Clonal Vertebrate Populations in Nature

Sex is a costly event because individuals have to expend time and energy to find a suitable partner for courtship, copulation may be interrupted, at least one of the partners may die, and the process may end without finishing ejaculation and fertilization (Meirmans, 2012). Furthermore, sexual females have a "two-fold cost of sex" (also called two-fold cost of males) in sexual species because they must invest $\sim$50% of their resources into male offspring, which do not carry the potential of making themselves offspring (Maynard Smith, 1971). Thus only female individuals can contribute directly to the growth rate of a population. And a sexual population is disadvantaged against parthenogenetic conspecifics in terms of growth rate. Moreover, segregation generates 50% homozygotes in meiosis. And if there is overdominance, sexual reproduction may adversely affect that population (Meirmans, 2012; Gibson et al., 2017).

Parthenogenetic populations as other asexual lineages do not have to bear the costs of sex. Nevertheless, these populations are also vulnerable to negative consequences of asexual reproduction, as described in the Red Queen (Van Vallen, 1973), Muller's

Ratchet (Muller, 1932), and Kondrashov's Hatchet (Kondrashov, 1988; Rice and Friberg, 2009) hypotheses. In the Red Queen Hypothesis, Leigh Van Vallen proposed that a species that cannot evolve at a rate matching those of co-evolving adversaries will be outcompeted. Because sexual reproduction can generate new combinations of alleles from different individuals, species can evolve faster in response to their natural adversaries, such as predators, parasites, or competitors. Thus, clonal reproduction may be a disadvantage in the face of rapidly evolving adversaries. On the other hand, Muller(1932) argued that because deleterious mutations accumulate over generations, species that lack recombination may become extinct over time.

Bacteria can overcome issues related to Muller's ratchet by horizontal gene transfer from other bacteria (Takeuchi et al., 2013), or through very large effective population sizes. Likewise, bisexual species can deal with issues connected with Muller's ratchet by creating new allele combinations. In contrast, parthenogenic species should be vulnerable to Muller's ratchet and thus have limited long-term success. Meanwhile, Alexey Kondrashov argues that through sexual reproduction, mildly deleterious mutations accumulated in the population are dispersed among individuals in such a way that they do not cause the death of the individual due to the synergistic effect of the mildly deleterious mutations, so the fitness of the individuals is not reduced even if slightly harmful mutations accumulate in the gene pool. This hypothesis is called "Kondrashov's hatchet" and according to this idea, lack of sexual reproduction can decrease the fitness of the population as a long-term effect (Kondrashov, 1988; Rice and Friberg, 2009).

Nonetheless, clonal vertebrates may have high genetic diversity in some way. For instance, the Amazon molly, *Poecilia formosa*, is a gynogenetic fish species that results from interspecific hybridization living in freshwaters in northeastern Mexico and southern U.S.A., and its populations have 10-fold higher heterozygosity than in its sexual progenitor species (Warren et al., 2018). In this way, they may mitigate the effects of Red Queen. Also, high genetic diversity may decrease the effects of Muller's ratchet. This is because when there is a greater variety of genes in a population, it increases the chances that beneficial mutations will occur and counteract any deleterious ones.

In clonal females, during oogenesis, some parts of meiosis I occur differently from normal meiosis in sexual females, so events producing new genetic combinations, such as recombination and independent assortment, do not occur (Stenberg and Saura, 2009). Therefore, egg cells carry completely the same genetic material as the mother. For example, Lutes et al. (2010) report that heterozygosity in parthenogenetic species of genus *Aspidoscelis* is preserved thanks to a sister chromosome pairing during meiosis I. Because sister chromosomes are matched with sister chromosomes instead of homologous chromosomes, any exchange is done between sister chromosomes, and thanks to this, total heterozygosity is not changed at the end of the process. And *Aspidoscelis* lizards could be a model for *Darevskia*.

1.3 Genus *Darevskia*

Genus *Darevskia* from the family Lacertidae was named by Oscar Arribas in 1997. Beforehand, *Darevskia* species were included in the genus *Lacerta* (Arribas, 1999) from the same family. The genus *Darevskia* involves both bisexual species and parthenogenetic species that have arisen by interspecific hybridization of these bisexual species (Darevsky, 1966; Darevsky, 1967; Murphy, 2000; Arnold, 2007). One individual from the *D. sapphirina* species appears in Figure1.1. This individual was captured during our field study in the Lake Van region in 2018 when we also gathered specimens from numerous *D. bendimahiensis, D. valentini* and *D. raddei* subpopulations in the region.

Figure 1.1: A captured *Darevskia sapphirina* lizard during fieldwork in the Lake Van basin.

True parthenogenesis in vertebrates was first described >50 years ago with parthenogenetic *Darevskia* species by Ilya Darevsky (Darevsky, 1957; Darevsky, 1958; Darevsky, 1967). Multiple studies were performed to solve the phylogenetic relationships and population structures of these taxa, and new parthenogenetic species and their progenitor bisexual species were identified.

The parthenogenetic species *D. dahli, D. armeniaca, D. rostombekowi* and *D. unisexualis* were described as parthenogenetic forms in Darevsky (1967). Later, *D. uzzelli* was described in Darevsky and Danielyan (1977). Lastly, *D. bendimahiensis* and *D. sapphirina* were described in the study by Schmidtler et al. (1994).

According to Murphy et al. (2000), genus *Darevskia* is made up of three clades (Figure 1.2): the caucasica, rudis and saxicola clades, and based on mtDNA and allozyme, all parthenogenetic species are formed by hybridization between the caucasica clade and the rudis clade. From the caucasica clade, *D. raddei* and *D. mixta* females are only included in the hybridizations as maternal progenitors whereas *D. valentini* and

D. portschinskii males from the rudis clade are only included as paternal progenitors. *Darevskia* species from the saxicola clade are not involved in the hybridization events. Figure 1.2 indicates the parentage relationships between parthenogenetic species and their parental relatives.

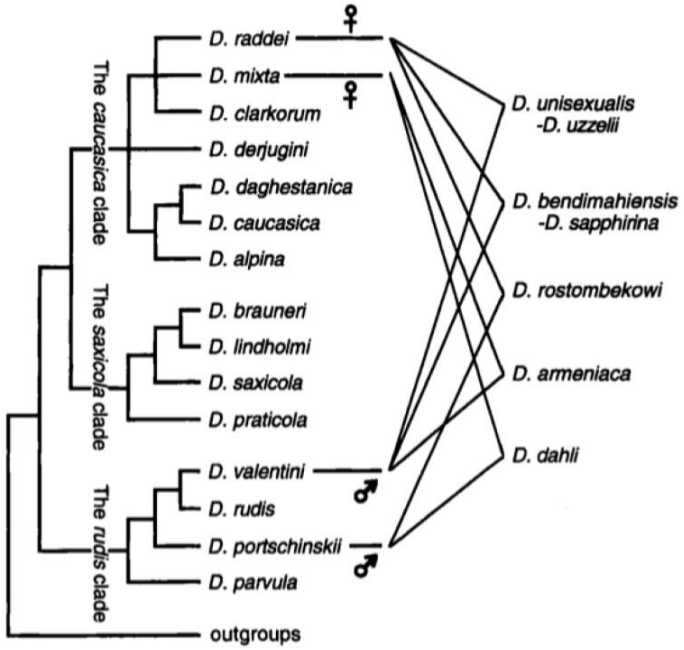

Figure 1.2: Reticulate phylogeny of genus *Darevskia* (Murphy et al., 2000)

One of the maternal species, *D. raddei*, have nominal subspecies *D.raddei raddei* from Turkey, Georgia, Armenia, Azerbaijan, Iran, *D. r. nairensis* from Armenia, Georgia and Turkey (Darevsky, 1967; Bobyn et al., 1996; Murphy et al., 2000), *D. r. vanensis* from eastern Turkey (Eiselt et al., 1993; Grechko et al., 2007; Freitas et al., 2016), *D. r. chaldoranensis* from Iran (Rastegar-Pouyani et al., 2011). Note that the debate on their species status is not yet settled. According to a maximum likelihood phylogenetic tree based on microsatellite markers and mtDNA, among parthenogenetic species, *D. sapphirina* and *D. bendimahiensis* have highest genetic proximity to *D. r. vanensis* than to other putative parental populations, *D. rostombekovi* have highest proximity to *D. r. raddei* from Armenia and Azerbaijan, *D. unisexualis* and *D. uzzelli* have highest proximity to *D. r. raddei* from Kars and central/western Armenia (Tarkhnishvili et al., 2020).

Because parthenogenetic *Darevskia* species appear to have arisen from interspecific

hybridization between bisexual *Darevskia* species, the number of founder hybridization events that took place is an important aspect of their biological history. In a study by Korchagin et al. (2007), the authors found single nucleotide variations (SNVs) in the flanking regions of the microsatellite sequences as well as the microsatellite motifs in the number or length of these repeated sequences. They report that these microsatellite genotypes were shared with the parental species as well, and the divergent genotypes were inherited from the parental species to the hybrid parthenogenetic species. Later, these markers were also utilized to estimate the number of hybridizations that took place. A study by Vergun et al. (2014), based on this method, reported that three founder hybridization events generated one major and two rare clones, and other 8 clones likely arose through post-formation microsatellite mutations from the major clone in the history of *D. dahli*. Major clones are prevalent clones in the species, so they are regarded as coming from hybridization in the birth of the hybrid species while rare clones (i.e. less frequent clones in the hybrid species) are regarded as occurring via either post-formational mutations or secondary hybridizations. Ryskov et al. (2017) reported a single hybridization event and four rare clones in the history of *D. rostombekovi*. Girnyk et al. (2018) carried out a similar study on *D. armeniaca*. They discovered 13 genotypes or presumptive clones of *D. armeniaca*. After comparison of these genotypes between parents and *D. armeniaca*, they reported that three clones could be traced to the parental species, meaning three interspecific hybridization events (one widespread clone, two geographically restricted clones) between parental species *D. valentini* and *D. mixta* in the evolutionary history of *D. armeniaca*. They further described 10 clones that likely originated from post-formational mutations of microsatellites after the species had formed. For *D. unisexualis*, Vergun et al. (2020) estimated a single hybridization event based on a widespread clone using the same method and 11 rare clones explained by mutations that occurred after the hybridization.

The mating of parthenogenetic females with males from their sexual progenitors is called backcross hybridization (Hörandl, 2009). A study by Tarkhnishvili et al. (2020) investigated the hypotheses of a backcross among 7 parthenogenetic *Darevskia* species based on mtDNA and microsatellite data. Using five microsatellite loci, it hypothesized a backcross for *D. dahli* and *D. armeniaca*. Of both parthenogen groups,

the majority of individuals had coincident genotypes at two loci but differed in their most frequent genotypes at the remaining three loci. Based on their observations, the authors propose that rare backcrossing events are the most likely explanation for the observed pattern of genotypes, which involves the incorporation of segments of a different paternal genome into the genome of an existing parthenogenetic form.

When parthenogenetic females mate with males from their progenitors, they can occasionally produce polyploid individuals. In Darevsky and Danielyan's (1968) study, parthenogenetic *D. armeniaca* and *D. unisexualis* females generated triploid hybrids during the study by mating with *D. valentini* males (Figure 1.3) in a location where *D. armeniaca* and *D. unisexualis* were present naturally together but *D. valentini* males were released later. In this case, diploid (2n=38) eggs generated by parthenogenetic females and haploid (n=19) sperms generated by bisexual males produced triploid (3n=57) hybrids. A similar situation was demonstrated by Darevsky et al. (1986), which reported triploid male hybrids of *D. raddei* x *D. rostombekovi*. Moreover, Danielyan et al. (2008) reported, in their monitoring study conducted between 1994 and 2006 on Aragats Mountain (Kuchak population) in central Armenia, a total of 84 *D. valentini* x *D. unisexualis* and 25 *D. valentini* x *D. armeniaca* triploid (3n=57) and tetraploid (4n=76) hybrids. Parthenogenetic *Darevskia* species (2n=38) are all-female (sex chromosomes = WZ) lizards, and the polyploid hybrids in this study were mostly sterile females but there were also fertile female and male polyploid hybrids. Also, there were found intersexual hybrids having female oviducts, male hemipenes and ovotestis gonads, but these were presumably infertile. Hybridizations between these natural polyploid hybrids might occur in hybrid zones where bisexual species and parthenogenetic species live in close proximity or together (Darevsky, 1966; Danielyan et al., 2000).

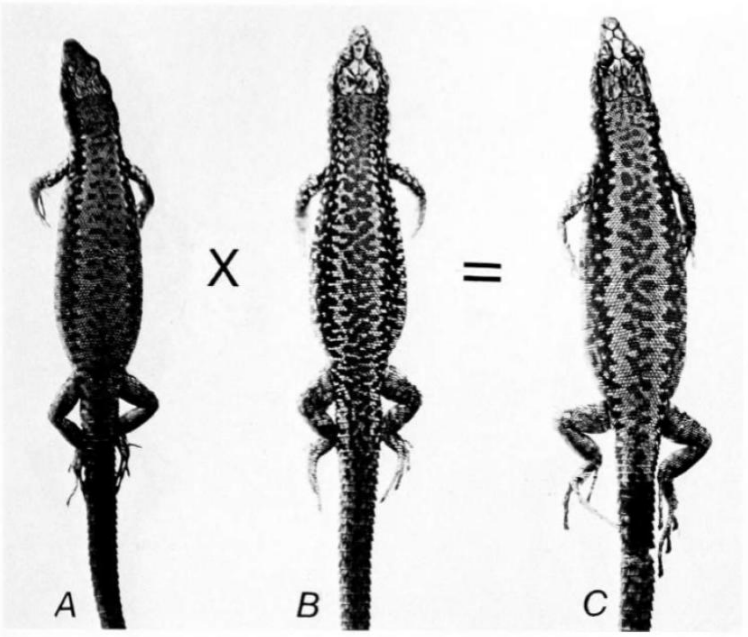

Figure 1.3: Natural hybridization between diploid (2n = 38) parthenogenetic female *D. armeniaca* (A) and diploid (2n = 38) bisexual male *D. valentini* (B). Triploid (3n = 57) hybrid female offspring (C) (Darevsky et al., 1985).

1.3.1 Distribution

The genus *Darevskia* has a wide distribution range. Its range consists of Asia Minor, the whole of the mountain Caucasus, the southern coast of Crimea, northern and north-western Iran to the mountain range Kopet Dagh in Turkmenistan, northwestern, northern and eastern Turkey to Southeast Europe (Darevsky, 1966; Darevsky et al., 1985; Ananjeva et al., 2006; Arnold et al., 2007; Gherghel et al., 2011; Galoyan et al., 2019; Maier et al., 2022).

Among bisexual species inferred to be involved in hybridization events, *D. raddei* and *D. valentini* have a broad distribution in Turkey, Georgia, Armenia, and some parts of Azerbaijan (Freitas et al., 2016; Galoyan et al., 2019; Candan et al., 2021); however, *D. mixta* in Georgia and *D. portschinskii* in Georgia, Armenia, and western Azerbaijan have relatively narrow distributions (Petrosyan, 2020). Among parthenogenetic species, *D. bendimahiensis, D. sapphirina, D. uzzelli* are endemic to Turkey (Schmidtler et al., 1994; Ilgaz, 2019; Tarkhnishvili et al., 2020), *D. rostombekovi*

(spelling following Murphy, 1999) is endemic to Armenia (Ryskov et al., 2017), and *D. dahli* is endemic to Georgia and Armenia (Tuniyev et al., 2020). *D. armeniaca* is distributed in Georgia, Armenia, and Turkey while *D. unisexualis* has a narrower distribution in Armenia and Turkey (Tuniyev et al., 2020). In addition, there are many sympatric locations where bisexuals and unisexuals live together (Darevsky et al., 1985).

Sometimes hybrid species are distributed in different areas than their bisexual parental species; this phenomenon is called "geographical parthenogenesis" (Darevsky et al., 1985; Gaggiotti, 1994; Vrijenhoek and Parker, 2009; Tilquin and Kokko, 2016), and all unisexual hybrids from genus *Darevskia* are an example for geographical parthenogenesis (Darevsky et al., 1985; Tarkhnishvili et al., 2020).

1.3.2 Ecology

Darevskia lizards can be found across a broad habitat range with regard to species from forest and grassland to rocky habitats (Freitas et al., 2016).

Parthenogenetic *Darevskia* species, in terms of ecological choices, have a tendency to live in dryer, colder, or more different climates than their bisexual relatives (Darevsky et al., 1985). According to Darevsky et al. (1985), there are three typical types of parthenogenetic *Darevskia* habitats: (I) dry rocks covered with shrub and grassy vegetation (for *D. rostombekovi*), (II) moderately dry rocks and stony dry riverbeds in the mountain forest zone (for *D. dahli* and *D. armeniaca*), (III) mostly volcanic origin bedrock, stony screes, and large lava fragments in the montane steppe zone (for *D. unisexualis* and sometimes *D. armeniaca*).

According to Murphy et al. 2000, parthenogenetic species can live in optimum to marginal habitats which the bisexuals do not appear to choose. Also, Tarkhnishvili et al.'s study (2010) indicates that *D. dahli* may outcompete its parental species *D. portschinskii* and *D. mixta* and invades a higher proportion of the suitable habitats.

1.4 Objectives of the Thesis

This thesis aims to deeply investigate the evolutionary history of parthenogenetic *Darevskia* species.

In Chapter 3, patrilineal relationships between all parthenogenetic species and their putative bisexual paternal populations are investigated based on chromosome Z sequences.

Chapter 4 focuses on the parental relationship between parthenogenetic *Darevskia* populations in Lake Van basin, *D. sapphirina* and *D. bendimahiensis*, and their putative maternal (*D. raddei*), and putative paternal (*D. valentini*) populations in the same region. Furthermore, I investigate the number of hybridization events between parental species in the evolutionary past of *D. sapphirina* and *D. bendimahiensis*.

CHAPTER 2

GENERAL METHODS

2.1 ddRADseq Method

This study was conducted using double-digest Restriction site-Associated DNA (ddRAD or ddRADseq) marker data. RAD (or RADseq, single-digest RADseq) markers are DNA fragments flanking restriction enzyme (RE) recognition sites in DNA (Miller et al., 2007). After digesting the DNA with one restriction enzyme and adding barcoded adapters, the DNA fragments are sheared and size selection is performed using gel electrophoresis, then the second adapters are ligated and the DNA slivers are amplified (Peterson et al., 2012; Rochette and Catchen, 2017). And accurate size selection is performed in ddRADseq, so the marker sequences only within a specific target size are selected in this method (Peterson et al., 2012). As shown in Figure 2.1, unlike RADseq, ddRADseq uses two restriction enzymes (i.e. double digestion) -both rare and common restriction cut sites- that rule out parts flanked by either very close (a few tens of base pairs) or very remote (a few thousand base pairs) cut sites depending on the specific enzymes used and the characteristics of the genome (see Appendix A for the details of the protocol of ddRAD library).

Compared to whole genome sequencing, because RAD regions are not distributed uniformly, RADseq data of a species includes many gaps across the genome. And they may miss many sites involving SNPs that may be important for local adaptation as shown in Figure 2.2 (Lowry et al., 2016). However, RADseq data is still very useful for evolutionary genetics studies such as phylogenetic analysis, demographic inference, or species delimitation because it can represent genome-wide polymorphism among individuals and populations effectively (Miller et al., 2007; Baird et al.,

2008; Hohenlohe et al., 2010; Ivanov et al., 2021). Also, the ddRADseq method is relatively cheaper (Kirschner et al., 2021). Therefore, it can be preferred as a cost-effective method for evolutionary analysis.

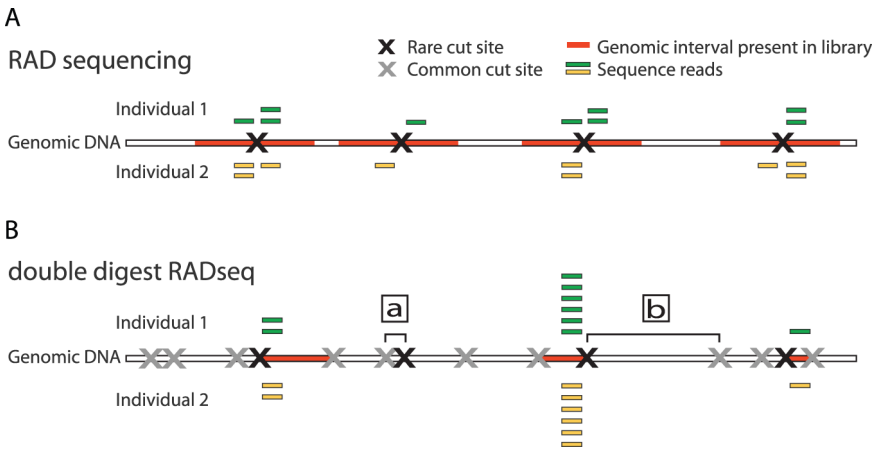

Figure 2.1: RADseq vs. ddRADseq processes. Box "a" (very close) and box "b" (very remote) are restriction cut sites (Peterson et al., 2012).

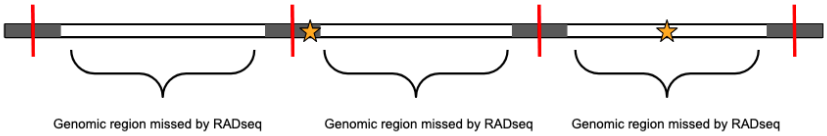

Figure 2.2: Missing loci across the genome in the RADseq method. Red lines mark RAD regions, stars are SNPs (modified from Lowry et al., 2016).

2.2 RAD Data Analysis with Stacks Software in This Study

Stacks (http://creskolab.uoregon.edu/stacks/) is a pipeline constructed for RADseq data analysis (Catchen et al., 2013). From processing the raw data to performing the Stacks analysis, this section includes three steps together with the alignment process:

1. Demultiplexing the multiplexed short reads with the Stacks subprogram "process_radtags".

2. The demultiplexed data is aligned to the reference genome of *D. valentini*

using Bowtie2 v.2.4.1 (Langmead Salzberg, 2012) (https://www.ncbi.nlm.nih.gov/data-hub/genome/GCA_024498535.1/) with the conservative "–end-to-end" setting and default parameter values (i.e. alignment scores calculated for the entire read matching the reference, -6 score penalty for a mismatch, -11 score penalty for a 2-bp gap, -0.6 - 0.6*L total score threshold required to retain the read, where L is the read length). In addition, because it has chromosome level resolution, the data is aligned to the reference genome of *Podarcis muralis* Laurenti, (1768) (Andrade et al., 2019) (https://www.ncbi.nlm.nih.gov/genome/?term=podarcis) for sex chromosome level studies.

3. The ref_map.pl wrapper processes the BAM file of each sample and includes two subprograms of Stacks, "gstacks" and "populations". The "gstacks" algorithm assembles loci according to the alignment positions and calls the SNPs in each sample. The "populations" algorithm calculates population genetics statistics and generates standard output formats such as Phylip, RADpainter, VCF, Fasta, or Plink.

    3.1. The "gtsacks" program uses a sliding window algorithm for looking at RAD data to set a locus at a time. In this algorithm, for each variable site found in the genome, the window is centered over the site with a size defined by 3 sigmas (1 sigma = 150 Kbp in the length of the window by default) using the default window size parameter. This program constructs the loci by combining the single-end reads in the BAM files of each individual moving the window along the chromosome window by window being centered over the next variant site (with overlapping windows). It identifies the polymorphic positions and genotypes by making a gene pool of a locus of all individuals. During this, it uses the "min-mapq" minimum PHRED-scaled mapping quality to consider a read (default: 10), the "max-clipped" maximum soft-clipping level, in a fraction of the read length (default: 0.20), the "max-insert-len" maximum allowed sequencing insert length (default: 1000), the "min_hom_seqs" a minimum number of reads required at a stack to call a homozygous genotype (de-

fault 5), “min_het_seqs” below this minor allele frequency a stack is called a homozygote, above it (but below –max_het_seqs) it is called unknown (default 0.05), and “max_het_seqs” minimum frequency of minor allele to call a heterozygote (default 0.1) (https://catchenlab.life.illinois.edu/stacks/comp/gstacks.php and http://catchenlab.life.illinois.edu/stacks/comp/genotypes.php). In the end, "gstacks" produces two catalog files, one of which comprises the consensus sequence for each assembled locus in the data and one of which comprises genotyping data.

3.2. These two catalog files are used in the "populations" program to compute the population-level statistics and generate the outputs Phylip, GenePop, Fstat, Fasta, Plink, VCF, and RADpainter. The “populations” use them in conjunction with a population map file, which contains sample names and corresponding population names. The “populations” program can apply four essential filters to control the variable sites that are biologically implausible. These are:

   i. maximum observed heterozygosity,
   ii. minimum allele frequency (MAF),
   iii. the minimum fraction of individuals of a single population a locus must be found in to be processed (-r),
   iv. the minimum number of populations that a locus must be found in to be processed (-p).

   The maximum observed heterozygosity and MAF filters help identify false positive SNP calls or erroneous merging of paralogous loci so that more accurate results can be obtained ($0 < MAF < 0.5$, and max_obs_het = 0 as default).

   The "-r" parameter is used as a filter to ensure that, in order for a locus to be processed, a certain percentage of individuals in the given population must have the same locus present in their DNA (default 0).

   The "-p" parameter is a number that determines how many populations must contain the same locus before it can be studied further.

This helps to ensure that any results obtained from analyzing the locus are meaningful and accurate (default 1).

In this study, all parameters above are set to default values for all analyses. We did not apply any filters in order to maximize the efficiency of the data. However, we are aware that this can cause noise in the data. We will discuss this in the Discussion.

Stacks is able to execute *de novo* alignment as well through the "ustacks" program (or the denovo_map.pl wrapper program) for the demultiplexed RADseq data. However, we did not use the *de novo* alignment since we had a very close reference genome aligned with $\sim$95% mapping success; nonetheless, this reference genome was not available at the time we published the chromosome Z article for paternal ancestry (Chapter 3). Also, we have short read sequences so *de novo* alignment does not make a chromosome-level alignment. Therefore, we used the *Podarcis muralis* reference genome as a reference genome of a close species in that study.

2.2.1 Phylogenetic Analysis

Phylogenetic analysis is a method that is often used to classify samples at the species, subspecies, or subpopulation level. The phylogenetic analysis involves using the evolutionary relationships between different organisms to classify them and understand their evolutionary history. Since RADseq data exhibits the variation among even different populations of the same species very well, we employ the phylogenetic method to represent evolutionary relationships among samples in this study.

For the chromosome Z study in Chapter 3, we needed chromosome Z sequences so we aligned our data to the *Podarcis muralis* genome using Bowtie 2 software with the same parameters above. The reference genome of *D. valentini* was not available when we did the chromosome Z study. Also, it does not have a chromosome-level resolution. Therefore, we did not use it for this study.

In the phylogenetic analysis, Bayesian and Maximum Likelihood phylogenies are two widely used approaches. Bayesian phylogenetics is a method for constructing phylogenetic trees that uses Bayesian statistical techniques. In a Bayesian phylogenetic

analysis, the relationships between different organisms are inferred by comparing their genetic data and using complex statistical models. Because it allows researchers to incorporate complex statistical models into their work, it can lead to better accuracy (Holder and Lewis, 2003). Since the Bayesian approach is considered to be more accurate, we built a Bayesian tree based on chromosome Z data using BEAST v.2.6.3 (Bouckaert et al., 2014) on the online platform CIPRES Science Gateway (Miller et al., 2010).

For this, we used BAM files produced from the mapping process. Then to generate a "phylip" file of the nuclear sequence data, we processed these with the Stacks subprogram "populations" using the "phylip" option of the program using the ref_map.pl wrapper module of Stacks. PHYLIP (Felsenstein, 2021), developed by Joseph Felsenstein, is the abbreviation of **PHYL**ogeny **I**nference **P**ackage which builds phylogenetic trees based on different genetic data types such as DNA sequences, protein sequences, distance matrices, gene frequencies, and so forth. It has a typical data file in the format shown in Figure 2.3 which includes aligned sequences of all samples according to their positions in one file. The first number in the header indicates the number of taxa (species, populations, samples etc.) while the second number shows the number of data characters (nucleotides, amino acids etc.) in the same position and first and second columns in the body of the file stand for these respectively (Felsenstein, 2003; Felsenstein, 2021).

```
8       26
OV2.rad13       AANATCNTGACNTCGAGTGNNNNNCC
OV2.rad12       ANCATCNTGACNTCGAGTGCNNTGCC
OV2.rad11       ANCATCCTGANTTCGAGTGNGGTGCC
OV1.rad9        NANATCATGACNTCGAGTGCGNNNCC
OV1.rad11       AACATCATGNCNTCGAGTGCNNTGCC
Ok3.rad2        NCTGNNANATACTGATAGATANCCTG
OK3.rad1        GCTACTNCANNCCCATAGATAACCTG
OD1.rad6        GACATCNTGACTTCGAGTGCGGTGCC
```

Figure 2.3: Input file format example of PHYLIP.

The BEAST software needs an input file "BEAST XML" that is created by the BEAUti program (Bouckaert et al., 2014). Because BEAUti requires a fasta file, we converted the phylip file from Stacks to fasta format using Vim editor in bash, and we produced a BEAST XML file with appropriate evolutionary priors, MCMC number, and the

fasta file using the BEAUti. Here, we chose a strict clock model which assumes that evolution occurs at the same evolutionary rate for every branch of the phylogenetic tree, and a relaxed exponential clock model which assumes that different evolutionary rates act on the branches of the phylogenetic tree. We ran the BEAST both ways to compare the divergence times and topologies. Also, we used the Yule speciation model because it assumes that each species has an equal possibility to bring about a new species and an equal possibility to go extinct at any given time (Steel and McKenzie, 2001); these were the most suitable conditions for our species given that we know little about their speciation rates. So as to build a Bayesian phylogeny, we ran the BEAST v.2.6.3 with the BEAST XML input file and visualized the phylogenetic tree using FigTree software (Rambaut, 2010).

For the phylogenetic analysis based on autosomal sequences in Chapter 4, as our parthenogenetic species are hybrids of two different bisexual species, they have one set of chromosomes from one bisexual species and another set of chromosomes from another bisexual species. We focused part of our analyses on alleles (or haplotypes) called diagnostic alleles that are unique for each parental species, as shown in Figure 2.4. We employed these alleles to build phylogenetic trees of hybrids and either maternal or paternal lineages.

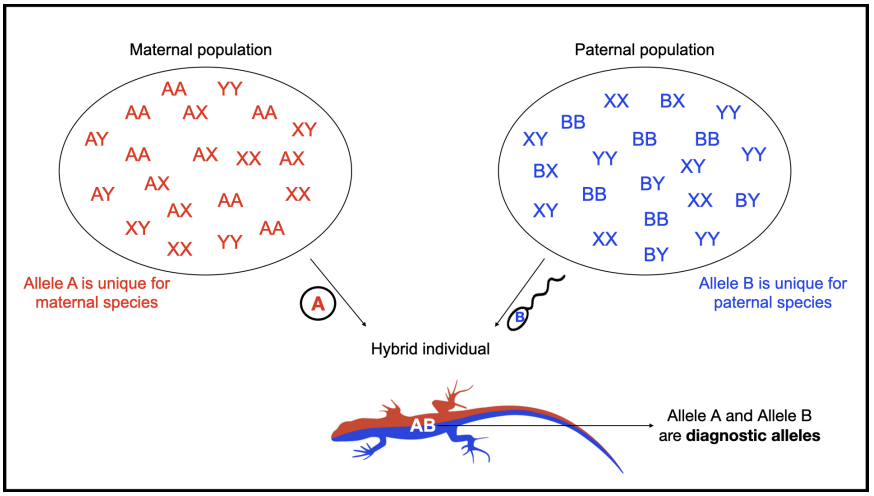

Figure 2.4: Identification of diagnostic alleles. The figure describes four alleles (haplotypes) at a hypothetical locus. A and B are haplotypes unique to only one gene pool of the parental species. X and Y are haplotypes common to both parental species.

To detect the diagnostic alleles, we utilize the custom python script presented in Appendix B-IV. For this, we first produce a fasta file from all aligned reads of all samples that we need using the "populations" program of Stacks with the “–fasta-samples” flag. We create a fasta file for all data of all maternal samples, a separate fasta file for all paternal samples, and a separate fasta file for each hybrid sample after the Stacks process above. We then ran the python script for diagnostic allele detection. During the detection process, the script tracks the steps below:

- Using the ddRAD alleles per individual of the bisexual parental species *D. raddei* and *D. valentini* in fasta format, identify the alleles that are unique to one parent only.
- For each parthenogenetic individual, per heterozygous ddRAD locus, match the alleles found in the hybrid with those in the maternal and paternal pool, respectively. Complete sequence identity is required for the match.
- Export the set of diagnostic maternal and paternal alleles, respectively, per each hybrid individual after filtering out the non-diagnostic sequences.

After the detection process, we obtain two files for each hybrid sample, including maternal and paternal diagnostic alleles separately. These output files have two columns one of which includes the locus number, and one of which includes the allele. Since we need the fasta format of these files for reference alignment, we convert them to fasta format. Later we align them to the *D. valentini* (or *Podarcis muralis*) reference genome to obtain BAM files of these files. In the next step, the Stacks produces a phylip file from these BAM files according to the codes presented in Appendix B-IV. Then we utilized the PhyML 3.3_1 (Guindon et al., 2010; Lemoine et al., 2018) (a phylogeny software based on the Maximum-Likelihood principle) to build a phylogenetic tree. This program uses the Maximum Likelihood (ML) tree reconstruction method. This method searches through all possible trees and calculates their probabilities based on how well they explain the observed data, then selects the tree with the highest probability (maximum likelihood), and it does not use prior/posterior probabilities to calculate the probability of an evolutionary relationship between groups and can work without needing prior knowledge and assumptions differently

from Bayesian tree approach. These properties make the ML tree method relatively faster and because we do not use prior knowledge, we employed the ML phylogeny approach using the PhyML 3.3_1 software. We used this program via the NGPhylogeny web tool (https://ngphylogeny.fr/tools/tool/271/form) (Guindon et al., 2010; Lemoine et al., 2018) with the "optimise parameter" option which adjusts parameters such as branch lengths and substitution rates in order to maximize the likelihood that the tree topology is correct. This program gave an error with the phylip format data file although phylip was one of the available input formats of it. We thus convert the phylip file that we get from the Stacks to fasta format using Vim editor in bash. And we operate the PhyML process using this fasta file. After PhyML concludes the phylogeny calculation, we visualize the tree using the web tool iTOL (Letunic and Bork, 2021) with "rectangular-slanted mode", "ignore the branch lengths", "default leaf sorting", and "display the bootstraps" options using the Newick tree file.

2.2.2 SNP-Based Fst Analysis

Stacks has a subprogram "populations" which can compute AMOVA Fst values of each variable site between a pair of populations. AMOVA (Analysis of Molecular Variance) is a statistical method that is used to partition genetic variation among different groups of individuals or populations. In an AMOVA analysis, the total genetic variation within a sample is partitioned among different levels of the population hierarchy (e.g., among populations within a species, among species within a genus, etc.). The Stacks subprogram "populations" calculates this using the "–fstats" flag. We used it in the ref_map.pl wrapper program and calculated the SNP-based Fst values for *D. bendimahiensis* and *D. sapphirina* samples, separating the individuals either into two nominal species (*D. bendimahiensis* vs. *D. sapphirina*) or into four populations (Pınarlı, Van/Ağrı border, Muradiye, Çaldıran) using different population maps where each sample matches with any population randomly in each iteration.

Stacks generates a file titled populations.fst_summary.tsv which contains summary Fst values between each pair of populations. We collected the Fst value of each SNP from the Stacks output files and merged them in a text file using the shell codes presented in Appendix B-V. We then draw bar plots of this data using the R code (in

Appendix B-V). We also calculated the mean Fst values over these files with the R codes presented in the Appendix.

In addition, we performed a permutation test to test randomness with 500 iterations. In this test, we examine two hypotheses: H0, which proposes that the mean Fst is not random, and H1, which suggests that the mean Fst is random and influenced by other factors and is originally higher. For this test, we created 500 random population maps where each sample matches with any population randomly in each iteration using R the script presented in Appendix B-V. We used these for Fst calculation both between 2 species and among 4 populations, separately. After creating the 500 output directories and finishing the 500 Stack processes using for loop command in the Linux shell presented in Appendix B-V, we collected the Fst values from the server and calculated the permutation test p-value for each pair of populations using the R codes presented in Appendix B-V.

2.2.3 fineRADstructure Analysis

Species have genetic diversity within and among populations. There are many methods to analyze the genetic data of individuals within a population or multiple populations in order to understand patterns of genetic variation and the relationships between different groups of individuals such as STRUCTURE (Falush et al., 2003) and ADMIXTURE (Alexander et al., 2009). In one of these methods, fineStructure, these variations are postulated that they come from different ancestral contributions of some populations to the extant populations (Malinsky et al., 2018). By calculating the coancestry values based on alleles shared among these groups of individuals, we can determine the amount of ancestral contributions from different presumed common ancestors.

For calculating the ancestral contributions, fineRADstructure uses a coancestry matrix of the RADseq alleles. The coancestry matrix is calculated as in Figure 2.5. One individual is taken as a recipient and others are donors in this calculation, the most similar haplotypes to the recipient haplotype, which is called the nearest neighbor haplotype, are counted as a donor. For instance, in Figure 2.5B, the haplotype ACTG is a recipient and the most similar red-colored haplotype is the donor. Likewise, the

haplotype ATTT is a recipient and the green-colored haplotypes are donors. However, if there was no identical haplotype with the recipient haplotype (the haplotype ATAT of individual 2 in Figure 2.5C), the most similar haplotypes (i.e. ATTT, ACAT) were going to be selected as donors. All of them have the closest nucleotide distances (i.e. one nucleotide distance) to the haplotype ATAT. Coancestry values are calculated according to the number of donor haplotypes and donor individuals in the sample population. For example, in Figure 2.5D, a coancestry matrix is shown for a single locus. For individual 1 in the first row, individual 2 gives 0.5 coancestry contribution because there is only one donor haplotype (ACTG) in the population and individual 2 has half of it. In terms of haplotype 2 (ATTT) of individual 1, there are two haplotypes ATTT in the population as most similar haplotypes, and because individual 3 and individual 4 share half of this contribution, they give 1/2/2 = 0.25 ancestral contribution to individual 1. In the coancestry matrix, rows show the coancestry values of the other individual so a total of rows is equal to one and due to that, the fineRADstructure plot is not a symmetrical heatmap, the y-axis only shows the coancestry values of the individuals on the x-axis.

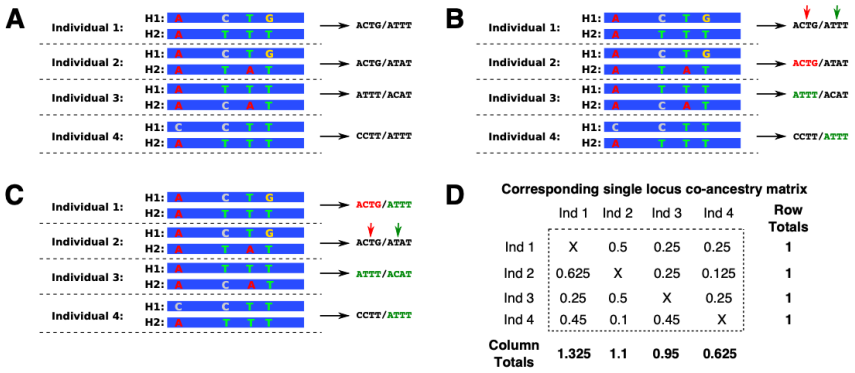

Figure 2.5: Calculation steps of coancestry matrix with four individuals' haplotypes. (A) RADseq haplotypes of four diploid individuals. H1 and H2 are Haplotype 1 and Haplotype 2, respectively. All the variable sites available are included in the haplotypes. (B) Calculation of coancestry values of individual 1 as a recipient. Other individuals are donors in this situation. Red and green arrows show recipient haplotypes. The corresponding colored haplotypes among the potential donors describe the closest matching haplotypes to the recipient. (C) Calculation of coancestry values of individual 2 as a recipient. Other individuals are donors. Red and green arrows show recipient haplotypes. The corresponding colored haplotypes among the potential donors describe the closest matching haplotypes to the recipient. (D) In the coancestry matrix, rows indicate the coancestry values. It is not symmetrical, rows stand for the recipients and all are totally equal to one whereas columns stand for the donors (Malinsky et al., 2016).

To create a fineRADstructure plot, we first used the Stacks subprogram "populations" to produce the fineRADstructure input file "populations.haps.radpainter", with the "radpainter" option which creates the input file of fineRADstructure software. We ran the "populations" in ref_map.pl wrapper program of Stacks with BAM files of samples, and after producing the result file "populations.haps.radpainter", we ran the fineRADstructure program using this file with the shell codes presented in Appendix B-IV and yield the output files of fineRADstructure plot. Then we draw the fineRADstructure heatmap using R scripts shared in the websites presented in Appendix B-IV.

2.3 Analysis of Number of Hybridization Using ddRAD Data for Hybrid Species

We used "Multi-Allelic Loci" (MALO) for estimating the number of founder hybridization events in the history of the hybrid species using RADseq data. Multi-allelic loci represent loci in the gene pool that have more than two haplotypes (Figure 2.6). MALO alleles would not be expected in gene pools of parthenogenetic species. However, some MALO alleles may actually be false positives resulting from hidden paralogs or technical/sequencing errors. We, therefore, study these loci in four different ways in an attempt to minimize the possibility of false positives and to estimate the number of hybridization events in the evolutionary history of the hybrid species.

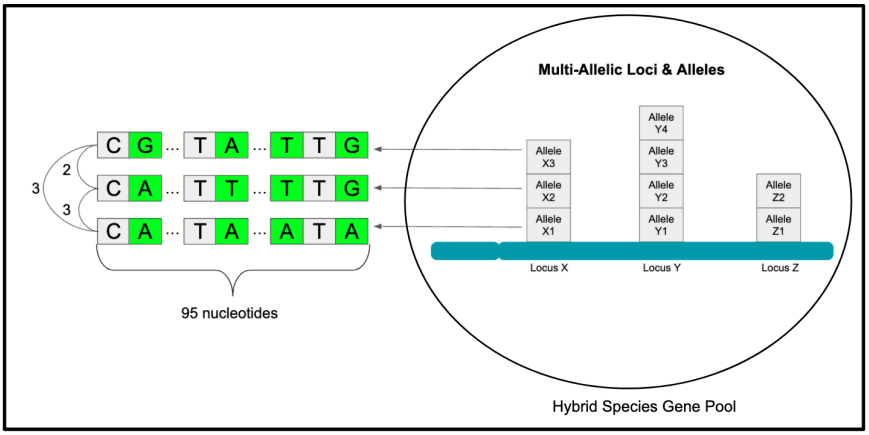

Figure 2.6: Hypothetical examples of multi-allelic loci and their alleles (right), and the number of nucleotide differences between multi-allelic loci alleles (green positions are polymorphic sites) (left).

First, for detecting these loci and their alleles, we prepare maternal, paternal and sample fasta files. For this, we use the first 4 commands (steps 1-4) in the part of "Multi-Allelic Loci Allele Analysis of *D. bendimahiensis* and *D. sapphirina*" (presented in Appendix B-VI). These commands extract the related samples with the "grep" command from the file "populations.samples.fa". Then we run the python script "MALAsHyb.py" for identifying multi-allelic loci alleles in hybrid species and start the python script "MALAsPar.py" for multi-allelic loci alleles in parental species. These scripts simply record if any locus has more than two haplotypes in the gene pool of the species. Because the multi haplotypes might be the result of technical artifacts, all reads including any missing nucleotide (N) were discarded from the

data using the custom python scripts mentioned. After detecting the multi-allelic loci alleles, we calculate the frequencies of these alleles in the related species severally using bash commands (presented in Appendix B-VI). Then we categorize the loci according to how many haplotypes they comprise, using the custom bash code (presented in Appendix B-VI). For instance, if a locus has three haplotypes, it is coded as 3-allelic, if a locus has 4 haplotypes, it is 4-allelic and so forth. We perform the same frequency calculations for the parental multi-allelic loci alleles with the codes presented in Appendix B-VI.

The first approach that we employed was to study some specific 3-allelic loci. We assume mutations are rare events and are uniformly distributed across the genome here. If there is a 3-allelic locus where the alleles have >1 nucleotide distance from each other (distance calculation is exhibited in Figure 2.6), this situation is unlikely to arise via *de novo* mutation or sequencing/technical error happening on other alleles in the gene pool. Therefore, the number of such alleles can help in the assessment of the emergence of these alleles by mutation. We detect these loci using the R codes presented in Appendix B-VI.

2.3.1 Binomial Test For the Number of 3-allelic Loci

In a second approach, we determine the number of 3-allelic loci that have 2 unique alleles in one parental species gene pool and 1 unique allele in another parental species gene pool. Then we classify these loci as maternal or paternal, according to which parent carries 2 alleles. Because if one of two alleles in the parental species appears via mutation from another allele in the parental gene pool or technical error, these alleles are binomially distributed with a 0.5 ratio. Otherwise, if these alleles come to the gene pool via multiple hybridizations from the parental species gene pool, we should most probably see a different ratio than 0.5. We next calculate the ratio between the number of maternal and paternal loci. Finally, we test if this ratio significantly deviates from 0.5.

In this test, our H1 is single hybridization together with *de novo* mutation or sequencing/technical error, and H2 is multiple hybridizations. If we cannot reject a 0.5 ratio, both hypotheses may be correct. If we can reject it, in this case, the hypothesis 2 is

very unlikely. For applying this analysis, we used the R script in the "Binomial Test with 3-MALO" part of Appendix B-VII.

Under the null hypothesis, the outcome pattern of this test may be influenced by a parent's higher allelic diversity, especially when we consider that the reference genome belongs to the paternal species *D. valentini*. To investigate such a potential bias towards higher richness among paternal alleles, we computed the mean number of alleles per locus and allelic richness in hybrid and parental species. For this calculation, we used Stacks's "populations" program with the "–fasta-samples-raw" flag and produced a fasta file for each species, involving all haplotypes detected in each sample for each locus. We then calculated the mean number of alleles per locus, per individual dividing the number of haplotypes by the number of loci for each species using the codes presented in Appendix B-VIII. Since we need a genpop file for the allelic richness calculation, we use Stacks's "populations" program with –genepop flag and calculate the allelic richness for each species with R library "hierfstats" (Goudet, 2005) using the codes presented in Appendix B-VIII.

2.3.2 Correlation Analysis of Frequencies of 3-allelic Loci Alleles

In this part, we hypothesize that if the 3-allelic loci alleles arise by virtue of parallel mutations in both hybrid gene pools and parental gene pools over time, since they occur randomly at different times, their frequencies ought not to be correlated. So according to hypothesis 1, if they are the result of multiple hybridizations, we expect a positive correlation between allele frequencies of hybrid species and parental species. In contrast, under hypothesis 2, if the 3-allelic loci alleles are the result of mutations, we expect the frequencies of the 3-allelic loci alleles do not have a positive correlation with the frequencies of 3-MALO alleles in parental species.

To test these hypotheses, we calculate the allele frequencies of 3-allelic loci alleles in each species and calculate the Spearman's rank correlation using the bash and R codes presented in Appendix B-IX. Also, we perform the same test for the allele frequencies of the 3-allelic loci alleles selected in the binomial test ("1+2" vs. "2+1" alleles) choosing randomly the first allele from the parent that has 2 alleles, and choosing the allele having lowest frequency because they have higher possibility to arise via

hybridization. All calculations are done using the codes presented in Appendix B-IX.

2.3.3 Mean Depth Analysis of 3-MALO vs. 2-MALO Loci Against Paralogous Sequences

We considered the possibility that the multi-allelic loci identified may in fact be hidden paralogs because there may be more than two aligned sequences as a consequence of hidden paralogs. Therefore, we test the mean depth of 3-allelic loci and 2-allelic loci. If the 3-allelic loci alleles may not be by reason of paralogs, we expect the mean depths are not different or the mean depth of 2-allelic loci alleles to be higher than the mean depth of 3-allelic loci alleles. In order to apply this test, we first run the python script "TWOAllelPy_v2.1.py" presented in Appendix B-X, which is a modified version of the script "MALAsHyb.py" used to detect the 2-allelic loci and obtain the 2-allelic loci indices. For the 3-allelic loci indices, we use the result files from the previous MALO process above. After detecting the locus indices, we find the depth of coverages of all 3-allelic and 2-allelic loci with SAMtools (Li et al., 2009). We then filter loci with depths >10 because they are most probably paralogs, then we calculate the mean depths per individual and apply a Mann_Whitney U test to these depths using the R codes presented in Appendix B-X. Also, you can find the R codes to draw the histograms of the depth distributions here.

CHAPTER 3

PRECISE PATERNAL ANCESTRY OF HYBRID UNISEXUAL ZW LIZARDS (GENUS *DAREVSKIA*: LACERTIDAE: SQUAMATA) REVEALED BY Z-LINKED GENOMIC MARKERS

with Alexey Yanchukov, David Tarkhnishvili, Mehmet Kürşat Şahin, Kamil Candan, Marine Murtskhvaladze, Mariam Gabelaia, Giorgi Iankoshvili, Natia Barateli, Çetin Ilgaz, Yusuf Kumlutaş, Ferhat Matur, Faruk Çolak, Marine Arakelyan, Eduard Galoyan

3.1 INTRODUCTION

Obligate parthenogenetic reproduction is widely documented in the order Squamata (lizards and snakes), and in the majority of cases is immediately preceded by hybridization between sexual representatives of phylogenetically distant lineages (Neaves & Baumann, 2011; Dedukh et al., 2020). The majority of extant obligate parthenogenetic hybrid species investigated to date appear to be relatively young, suggesting that parthenogenesis is unlikely to be maintained for very long on the evolutionary timescale (Moreira et al., 2021), despite demonstrated ecological advantages in the short term (Tarkhnishvili et al., 2010).

At the same time, to correctly estimate the time of origin of the parthenogenetic populations, one needs to (1) identify those populations of the sexual species that are most closely related to the true parents of the hybrid parthenogens, and (2) genotype both hybrid and parental populations with sufficient resolution in order to obtain robust age estimates (i.e. based on molecular clock or demographic simulation approaches). The difficulty in fulfilling these requirements leads to wide confidence intervals of

age estimates in most studies on parthenogenetic reptiles (Tarkhnishvili et al., 2010).

The rock lizards of the genus *Darevskia* Arribas, 1999 were the first example of hybrid parthenogenesis discovered in vertebrates (Darevsky, 1958, 1966), and remain one of the most well-studied systems (Murphy et al., 2000; Badaeva et al., 2008; Tarkhnishvili, 2012; Tarkhnishvili et al., 2010, 2020; Freitas et al., 2016; Galoyan et al., 2019; Spangenberg et al., 2020). In *Darevskia*, two maternal (*D. mixta* and *D. raddei)* and two paternal (*D. valentini* and *D. portschinskii*) sexual species have produced seven nominal parthenogenetic species (*D. armeniaca, D. dahli, D. unisexualis, D. rostombekowi, D. uzzelli, D. bendimahiensis* and *D. sapphirina*; Fig. 3.1). Their distribution is centred south of the Lesser Caucasus mountains in Georgia, Armenia and Turkey, as well as the Lake Van basin in eastern Turkey (Fig. 3.2). The hybrid origin of the parthenogenetic *Darevskia* and the identity of their parents at the species level was determined using only a few allozyme and mitochondrial DNA (mtDNA) markers (reviewed in Murphy et al., 2000), and was broadly confirmed in all later genetic studies (Freitas et al., 2016, 2019; Tarkhnishvili et al., 2020). Notably, the parental sexual species themselves were later shown to harbour high levels of genetic diversity and complicated geographical structure (Tarkhnishvili et al., 2020), and at least on the paternal side, they are characterized by widespread interspecific gene flow today as well as in the past (Tarkhnishvili et al., 2013; Freitas et al., 2019). Regarding the origin of the parthenogenetic species, these recent findings imply that the general view of reticulate evolution within *Darevskia* such as summarized two decades ago by Murphy et al. (2000) lacks sufficient detail, in terms of both geographical sampling and genome resolution. While a number of recent studies have compared the genetic distances of various paternal and hybrid parthenogenetic populations using a handful of mitochondrial and nuclear DNA (nDNA) markers, their results lacked genomic resolution (Tarkhnishvili et al., 2020), and only a few species were studied locally (Freitas et al., 2016; Ryskov et al., 2017; Girnyk et al., 2018; Vergun et al., 2020). Estimates of the time of the origin of some parthenogenetic species have been made (Murphy et al., 2000; Freitas et al., 2016; Murtskhvaladze et al., 2020), and in all cases they were based on mtDNA sequences. This effectively measures the time of divergence of the parthenogetic lineage from its nearest maternal parent, but lacks validation based on analysis of the paternal genomes. Considering broad variation of

rates of molecular evolution throughout the tree of life (Ho, 2020), this complicates true estimation of the time of origin of the pathenogenetic species.

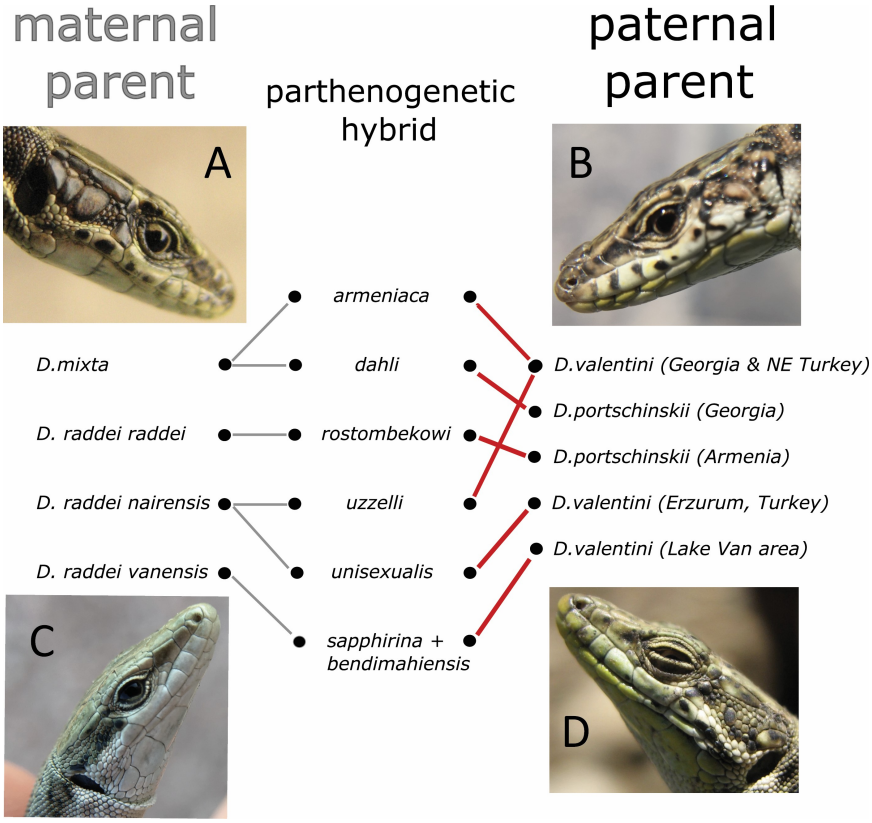

Figure 3.1: Schematic origin of parthenogenetic taxa in the genus *Darevskia*. The list of matrilinear ancestors on the left is taken unmodified from figure 10 in Tarkhnishvili et al. (2020), while details of the patrilinear ancestry on the right were updated using the results of the present study. The insert photographs indicate the sexual ancestral species: maternal *D. mixta* (A) and *D. raddei* (C), and paternal *D. valentini* (B) and *D. portschinskii* (D).

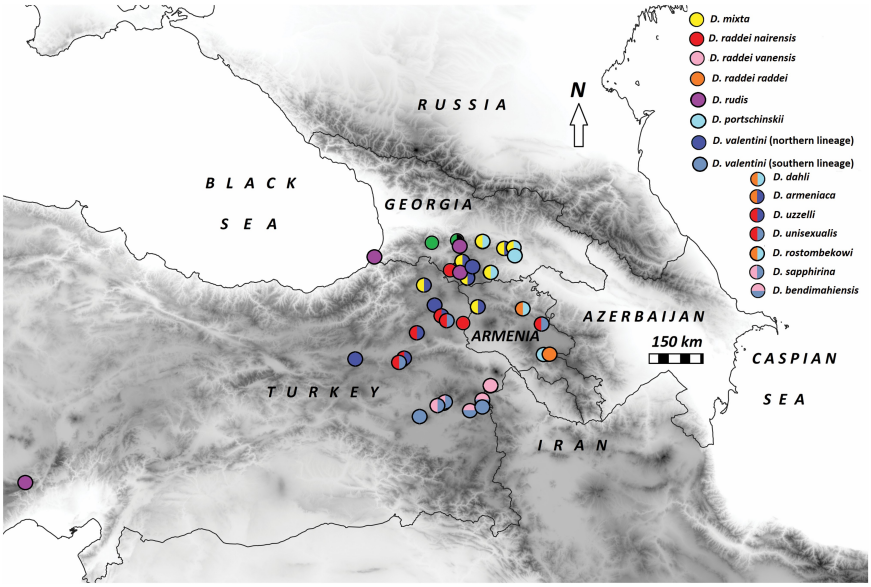

Figure 3.2: Sampling locations of seven parthenogenetic taxa, their presumed ancestral sexual species and their close relatives

For the first time in *Darevskia*, we use high-throughput genotyping [double-digest restriction site associated sequencing (ddRAD-seq)] to obtain genetic data in all parthenogenetic and their putative parental species over the entire area where parthenogenesis occurs. As far as our data allow, we attempt to resolve the exact paternal ancestry of all known parthenogenetic species, and estimate the respective divergence times between the extant parthenogenetic and paternal populations. To this end, we perform phylogenetic analysis using multiple short sequences from one region of the genome, which in parthenogenetic *Darevskia* has been inherited exclusively from the paternal side: the Z chromosome.

Homologous and well-differentiated ZZ/ZW sex chromosomes with shared origin dating back as far as 85 Mya are characteristic for all lacertid lizards (Rovatsos et al., 2016, 2019; Stöck et al., 2021). The ZZ/ZW system possesses several attractive properties compared to XX/XY. The Z chromosome can show a signature of polymorphism similar to the faster evolving X relative to the autosomes (Mank et al., 2007, 2010; but see Axelsson et al., 2004). Assuming an even sex ratio and similar variance in reproductive success between the sexes, the effective population size on Z is expected to be three-quarters that of the autosomes. The Z chromosome spends two-thirds of its time in the males, and therefore should experience a slightly higher

mutation rate than autosomes. Hence, neutral substitutions on the Z should proceed slightly faster than on the autosomes and on the W (Wilson Sayres & Makova, 2011). At the same time, sex chromosomes in general are expected to be less prone to introgression following interspecific hybridization (i.e. due to reduced fitness of the heterogametic sex), which has been demonstrated empirically for X (Macholán et al., 2011; Maroja et al., 2015) as well as for Z (Storchová et al., 2010). In short, the sex chromosomes and the Z chromosome in particular can be expected to more adequately reflect the history of divergence even in cases where incomplete reproductive isolation and secondary hybridization complicate phylogenetic analyses at other parts of the genome.

In *Darevskia*, all-female parthenogenetic populations are diploid and heterogametic (ZW). The presence of female heterogametic chromosomes has been directly confirmed in *Darevskia portschinskii, D. raddei* (Spangenberg et al., 2019; Rovatsos et al., 2019), *D. valentini, D. unisexualis* (Spangenberg et al., 2017, 2020a), *D. armeniaca* (Kupriyanova, 2009) and *D. rostombekowi* (Spangenberg et al., 2020b). Assuming that the hybrid species resulted instantly from the union between the paternal (Z) and maternal (W) gametes, the Z chromosome in parthenogens is then inherited directly and exclusively from the paternal parent. This eliminates the need to phase and correct the assignment of haplotypes to the respective parental genomes: a non-trivial task because many alleles will be shared by both parental populations due to ancestral polymorphism (Tarkhnishvili et al., 2020). In contrast, the Z chromosome is expected to be completely free of maternally derived alleles and can be used in phylogenetic analysis as a single sequence unit – in contrast to the seemingly 'private' alleles on the autosomal sequences, which could still be inherited from the mother. The only differences from the ancestral paternal variant are therefore due to accumulation of novel mutations, which offers unique insight into the evolutionary history of the parthenogenetic populations after hybridization has occurred.

3.2 MATERIALS AND METHODS

Live lizards were collected in Turkey, Georgia and Armenia, and tissue samples (tail tips) were taken with negligible harm to the animals, as described previously

(Tarkhnishvili et al., 2020). The animals were released immediately following the procedure. Sex was identified based on external morphological characters and only adult females were included in the present study to avoid possible bias due to unequal dosage of Z chromosomes between sexes. In total, 99 individuals were analysed using ddRAD-seq and these represented all seven parthenogenetic species as well as their respective sexual parents from multiple localities (Fig. 3.2; Table C.1). In particular, our sampling covered all currently identified, genetically distinct subspecies/-geographical variants of the parental taxa *D. mixta, D. raddei, D. portschinskii* and *D. valentini* found near the parthenogens, as well as *D. rudis rudis* and *D. rudis obscura* from Georgia that are known to exchange genes with both *D. portschinskii* and *D. valentini* (Tarkhnishvili et al., 2013), and a distant *D. rudis bolkardaghica* from the Taurus mountains in southern Turkey, representative of the large distribution range of *D. rudis* (Arribas et al., 2013; Candan et al., 2021). Eleven individuals of *D. raddei*, three of *D. mixta* and one of *D. derjugini*, all representing the maternal clade of parthenogenetic *Darevskia* (Murphy et al., 2000), were used as outgroups in the phylogenetic analysis.

3.2.1 Genomic library preparation and bioinformatic processing

Genomic DNA was extracted from alcohol-preserved tissues as described by Tarkhnishvili et al. (2020), and genomic ddRAD library preparation and sequencing was outsourced to a commercial facility (Floragenex Inc., OR, USA). Briefly, DNA samples were digested with restriction enzymes PstI and MseI (New England Biolabs), and genomic libraries were constructed and multiplexed using a standard protocol (see Appendix A), and sequenced on two lanes of an Illumina HiSeq instrument (100-bp single-end reads). No differences were observed in the FASTA read quality between the two batches of samples. Reads with quality score < 30 were discarded. Raw reads were de-multiplexed and adaptors and indexes removed using Stacks 2.57 software (Rochette et al., 2019), and quality checked in FastQC. We then aligned the reads to the reference genome of *Podarcis muralis* (Andrade et al., 2019), NCBI assembly GCF_004329235.1, using the default mismatch settings in BowTie 2 (Langmead & Salzberg, 2012) and selecting the conservative end- to-end flag to ensure that the entire read matches the reference. Only those reads successfully mapped to the *Po-*

darcis Z chromosome (NCBI assembly CM014761.1) were retained for downstream analyses (Table C.1). The remaining autosomal and unplaced sequences are outside of the scope of the present study but will be used in future analyses. The reference *P. muralis* genome does not include any annotations specific to the W chromosome. After alignment and converting of the individual sequences into BAM format with SAMtools (Li et al., 2009) we performed reference-based assembly in Stacks 2.57 using the software default settings for variant calling. To maximize the number of useful RAD loci per sample, we assumed that each individual comes from its own separate population with size = 1.

3.2.2 Phylogenetic analysis

All Z-linked ddRAD loci were concatenated and used in the phylogenetic analysis. Large genome-wide datasets with a high proportion of missing loci typically perform better relative to small datasets where the number of loci has been sacrificed to increase cross- individual coverage (Wiens & Morrill, 2011; Roure et al., 2013; Talavera et al., 2021). Principal component analysis (PCA) was performed using the adegenet (Jombart 2008) R package, and the co-ancestry matrix was constructed in fineRADstructure (Malinsky et al., 2018). A phylogeny of the Z chromosome was calculated using a Bayesian approach in BEAST v.2.6.3 (Bouckaert et al., 2014) on the CIPRES Science gateway (Miller et al., 2010). Prior to building the trees, the most plausible substitution models were selected using MEGA-X software (Kumar et al., 2018). BEAST was run using the Yule process, and with a random distribution of the offspring number between individuals; Markov chain Monte Carlo iterations = 100 000 000. The strict as well as the relaxed exponential clock models were used to obtain a range of divergence time estimates.

3.2.3 Correlation analysis of the distances

We analysed the association between the geographical and genetic distances for all pairs of parthenogenetic and sexual individuals from the paternal clade. Geographical distances were measured as the Euclidean distance based on the exact geographical

coordinates, between the locations of 60 studied parthenogenetic individuals and each of 21 individual-based locations of *D. portschinskii, D. valentini* and *D. rudis*. Genetic distances were measured as the proportion of nucleotide substitutions within each pair of parthenogenetic and sexual individuals. Since these only represented a subset of all possible individual pairs, the Mantel test could not be applied and Pearson's correlation coefficients, both direct and ln-transformed, were inferred between the geographical and genetic distances.

3.3 RESULTS

After alignment to the *Podarcis* genome with an average mapping success rate of 53%, and the reference-based assembly of the ddRAD loci in Stacks 2.57, a total of 14 588 loci were mapped to the Z chromosome in all 99 female individuals, of which 3659 loci contained 6806 informative single nucleotide polymorphisms (SNPs). Only 56 loci were genotyped in all individuals, but at least 47 (48%) of the studied individuals were genotyped at 12 143 (83%) loci, while at most two individuals were missing at 5302 (36%) loci. Missing data per se do not constitute a major source of bias in phylogenetic analyses, especially when Bayesian methods are employed (Wiens & Morrill, 2011; Roure et al., 2013), and thus our dataset contained sufficient information for further phylogenetic analysis.

The Bayesian tree topology (Fig. 3.5) clearly differentiated two major phylogenetic clades in *Darevskia*, previously designated by Murphy et al. (2000) as 'caucasica' (i.e. including the species *mixta, raddei* and *derjugini*) and 'rudis' (including nominal *rudis, portschinskii* and *valentini*). The sexual species *mixta* and *raddei* from the 'caucasica' clade, although known to be maternal parents of the parthenogenetic taxa, appeared just as distant from them on the Z chromosome as they were from the paternal sexual species of the 'rudis' clade. In contrast, the Z chromosomes of parthenogens clustered closely with their respective paternal species, down to the level of local geographical populations. In particular, *armeniaca* appeared in the same sub-clade with *valentini* samples from Georgia; *dahli* with *portschinskii* collected in Georgia; *rostombekowi* with *portschinskii* south of Lake Sevan in Armenia; *uzzelli* with *valentini* from Ardahan, Turkey; *unisexualis* with *valentini* from Erzu-

rum, Turkey; and two parthenogenetic taxa, *bendimahiensis* and *sapphirina*, with *valentini* populations from Çaldıran (NE of Lake Van, Turkey). The same pairs of paternal sexual and daughter parthenogenetic populations separated clearly on the individual-based PCA and fineRADstructure plots (Fig. 3.3 and Fig. 3.4). In all analyses, *D. sapphirina* and *D. bendimahiensis* appear as a single undifferentiated clade, thus questioning their current status as two separate species (Schmidtler et al., 1994; Tarkhnishvili et al., 2020). The choice of the molecular clock substitution model (strict or relaxed exponential) had no effect on the tree topology.

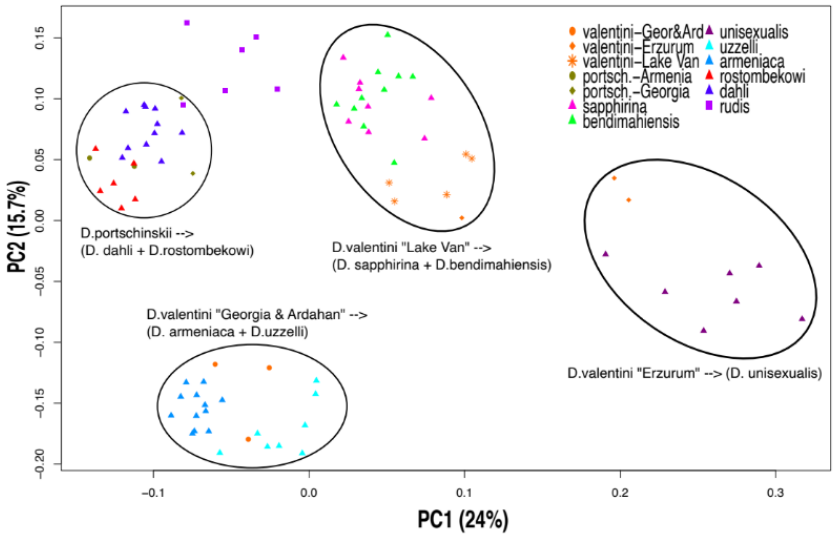

Figure 3.3: Principal Component Analysis on individual Z-linked genotypes. Elipses indicate pairs of parthenogenetic species + their inferred paternal bisexual population.

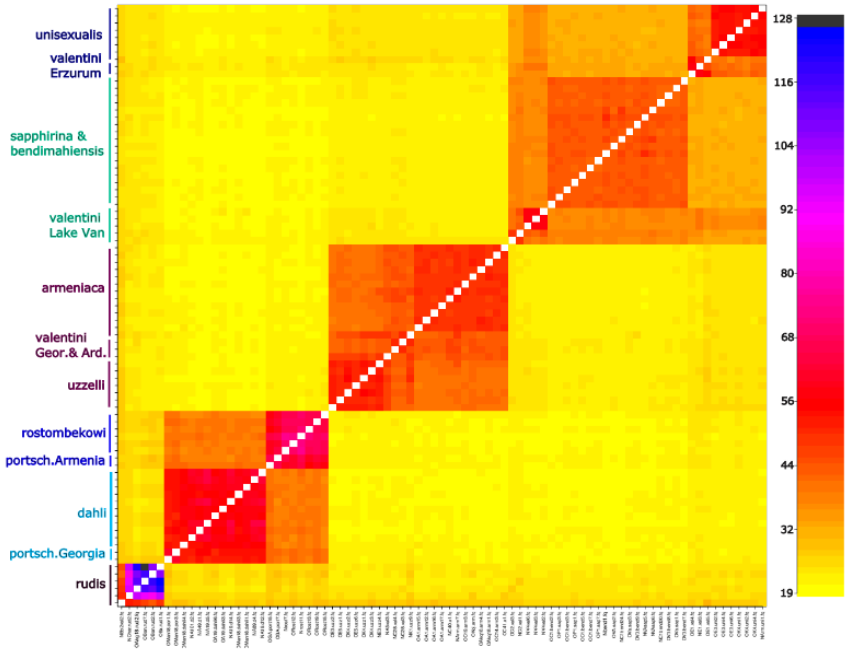

Figure 3.4: fineRADstructure co-ancestry plot constructed from the individual Z-linked genotypes. Species/population names are indicated on the left and the individual samples names are listed along the X-axis. The scale bar on the right represents the absolute co-ancestry values inferred by fineRADstructure.

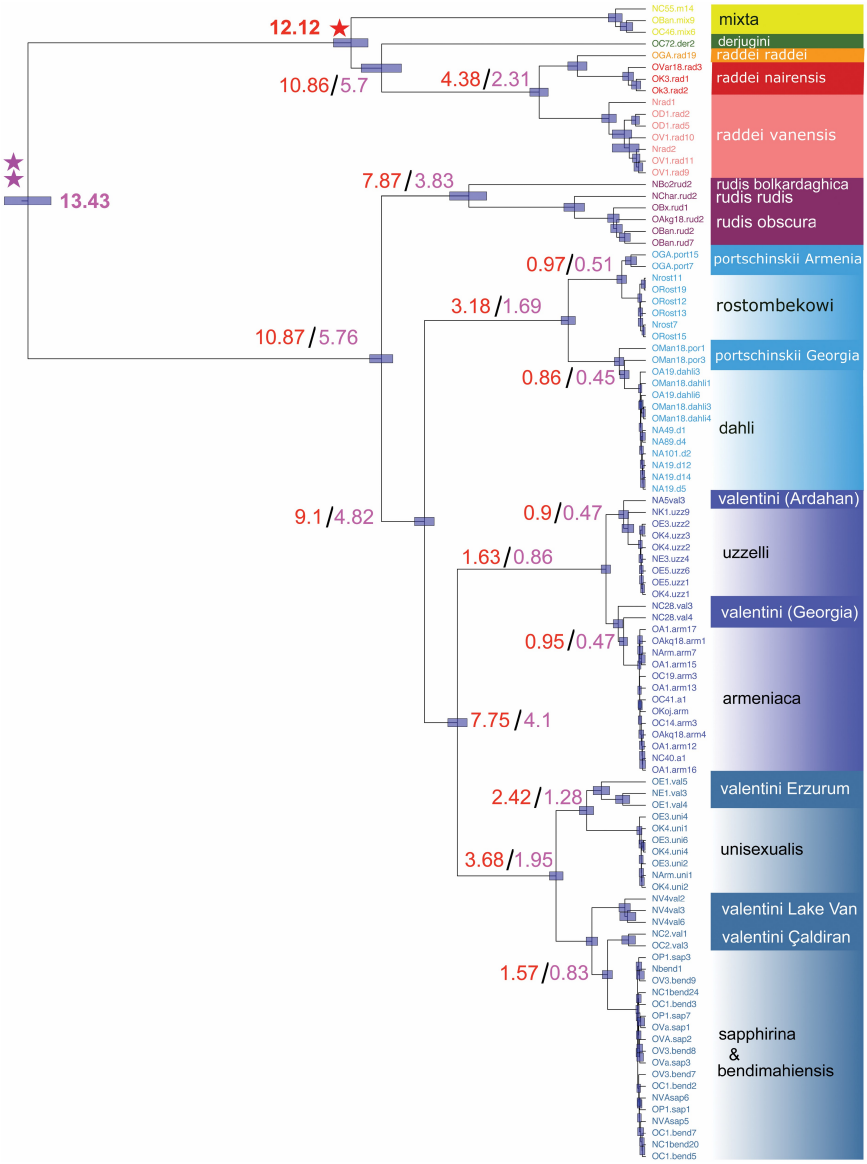

Figure 3.5: Bayesian phylogeny of parthenogenetic taxa and their parental sexual species constructed from Z-linked ddRAD loci. BEAST v.2.6.3 tree, calibrated by (i) the *raddei–mixta* split, indicated by the single red star symbol, at 12.12 Mya (Murtskhvaladze et al., 2020) or (ii) by the *valentini–raddei* split, indicated by the double purple star, at 13.43 Mya (Garcia-Porta et al., 2019). The names of two major clades in the genus *Darevskia* (maternal 'caucasica' and paternal 'rudis') that participated in hybridization and produced parthenogenetic taxa are given according to Murphy et al. (2000). Alternative divergence time estimates (Mya) that followed from the two different calibration points (i–ii) are shown next to the selected nodes in red (i) and purple (ii) font. Bars at the nodes represent 95% highest posterior density intervals. All nodes with ages shown had very high Bayesian support values (>0.99).

We found a strong correspondence of the geographical and genetic distances between the parthenogenetic population and its closest paternal relative. For example, one population of *D. bendimahiensis* was found just 3 km from its closest paternal parent *D. valentini* in Çaldıran (Lake Van area), and the two geographically distant populations of *D. portschinskii* in Georgia and in Armenia clustered locally with their respective daughter species *D. dahli* and *D. rostombekowi* (Fig. 3.2). Overall, the correlation coefficient between the geographical and genetic distance was 0.265 (the total number of distances compared was 1260). After log-transformation of both geographical and genetic distances, the correlation coefficient increased to 0.477, suggesting that genetic distance does not greatly increase for those sexual populations which are very remote from the target parthenogenetic locations (Fig. 3.6).

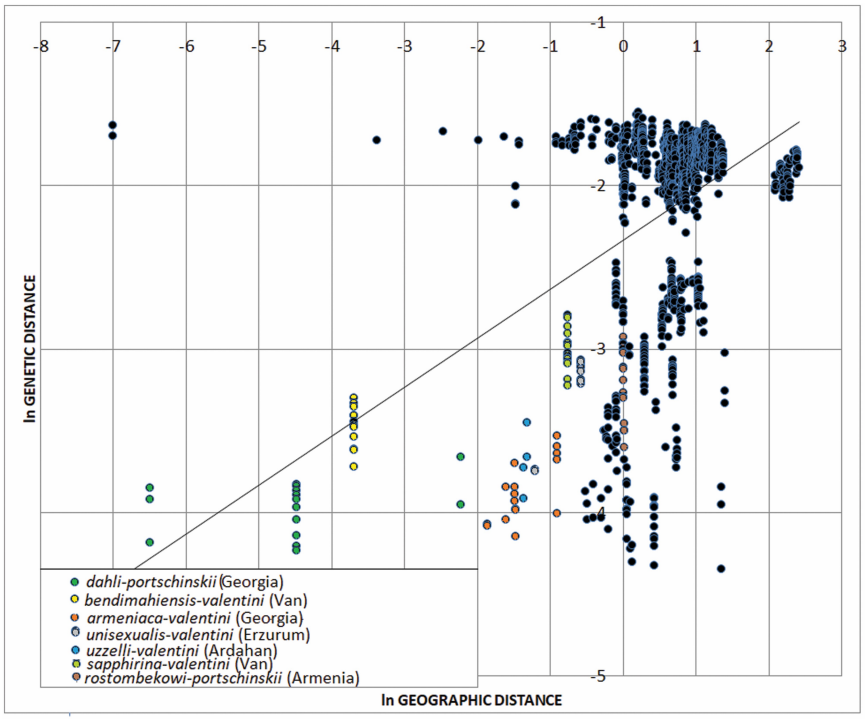

Figure 3.6: Genetic distances between the parthenogenetic and sexual individuals plotted against straight geographical distances between their locations. Numbers are ln-transformed. Distances between the individual parthenogens and their presumed patrilineal ancestral populations are shown with different symbols (see key).

To estimate the node ages on our Z chromosome tree, we calibrated the molecular clock using (1) the previous divergence time estimates between *D. raddei* and *D.*

mixta as reported in three different studies (14.4, 12.12 or 8.76 Mya: respectively Kumar et al., 2017, Murtskhvaladze et al., 2020 and Garcia-Porta et al., 2019), or (2) divergence between *D. valentini* and *D. raddei* (20.6, 18.53 or 13.43 Mya), as reported in the same studies. The resulting divergence time estimates of the parthenogenetic species from their closest extant paternal relative are given in Table 3.1. The timings of four such events (i.e. the divergence of *armeniaca*, *dahli, rostombekowi* and *uzzelli*) grouped within roughly the same period, which we dated to as early as 1.16– 1.84 Mya (median times) under the relaxed exponential clock model calibrated according to the TimeTree online resource (Kumar et al., 2017), or to as recent as 0.45–0.51 Mya under the strict clock model calibrated according to Garcia-Porta et al. (2019). The divergence of *unisexualis* and *bendimahensis–sapphirina* took place within a broader and earlier time frame, i.e. from 2.18–3.43 to 0.83–1.28 Mya, respectively. The 95% probability intervals between these two major groups overlapped when the relaxed exponential clock model was used, but remained separate when the strict clock model was chosen (Table 3.1; Fig. 3.5).

Table 3.1: Estimated divergence times (Mya) of parthenogenetic species from their closest extant paternal population. Two combinations of the calibration point+model used to construct the tree in Figure 3.5 are highlighted in bold.

	Calibrated with *mixta—raddei* split		Calibrated with *valentini—raddei* split	
	T strict clock	Relaxed exponential	T strict clock	Relaxed exponential
Kumar *et al.* (2017)	*14.4±0.55*	*14.4 (6.0–30.1)*	*20.6±0.8*	*20.6 (9.7–41.0)*
dahli–portsch (Georgia)	1.02±0.02	1.45 (0.24–1.7)	0.69±0.01	1.03 (0.21–1.43)
rostom–portsch (Ardahan)	1.16±0.02	1.16 (0.22–1.6)	0.79±0.01	0.82 (0.2–1.5)
armen–val (Georgia)	1.06±0.02	1.60 (0.25–2.0)	0.72±0.01	0.96 (0.2–1.68)

uzzelli–val (Ardahan)	1.05±0.02	1.84 (0.21–2.49)	0.72±0.01	1.30 (0.2–2.46)
unisex–val (Lake Van)	2.88±0.06	3.43 (0.8–4.0)	1.96±0.04	2.42 (0.6–3.2)
bendi–val (Lake Van)	1.87±0.02	2.18 (0.5–2.8)	1.27±0.02	1.53 (0.4–2.5)
sapph–val (Lake Van)	1.87±0.02	2.18 (0.5–2.8)	1.27±0.02	1.53 (0.4–2.5)
Murtskhvaladze *et al.* (2020)	***12.12±0.55***	*12.12 (5.1–25.0)*	*18.53±0.7*	*18.53 (8.0–36.9)*
dahli–portsch (Georgia)	**0.86±0.02**	1.22 (0.23–1.5)	0.62±0.02	0.92 (0.21–1.43)
rostom–portsch (Ardahan)	**0.97±0.02**	0.98 (0.2–1.5)	0.71±0.02	0.74 (0.2–1.5)
armen–val (Georgia)	**0.89±0.02**	1.14 (0.23–1.8)	0.65±0.02	0.86 (0.2–1.67)
uzzelli–val (Ardahan)	**0.89±0.02**	1.55 (0.2–2.47)	0.64±0.02	1.17 (0.19–2.45)
unisex–val (Lake Van)	**2.43±0.06**	2.88 (0.6–3.3)	1.77±0.04	2.17 (0.59–3.1)
bendi–val (Lake Van)	**1.57±0.02**	1.83 (0.41–2.5)	1.14±0.02	1.38 (0.39–2.4)
sapph–val (Lake Van)	**1.57±0.02**	1.83 (0.41–2.5)	1.14±0.02	1.38 (0.39–2.4)
Garsia-Porta *et al.* (2019)	*8.76±0.47*	*8.76 (4.8–21.0)*	***13.43±0.53***	*13.43 (5.6–26.9)*
dahli–portsch (Georgia)	0.62±0.02	0.88 (0.21–1.42)	**0.45±0.01**	0.67 (0.2–1.4)
rostom–portsch (Ardahan)	0.70±0.02	0.71 (0.2–1.5)	**0.51±0.01**	0.53 (0.19–1.62)
armen–val (Georgia)	0.64±0.02	0.72 (0.2–1.66)	**0.47±0.01**	0.62 (0.2–1.65)

uzzelli–val (Ardahan)	0.64±0.02	1.12 (0.19–2.45)	**0.47±0.01**	0.85 (0.18–2.43)
unisex–val (Lake Van)	1.75±0.06	2.08 (0.59–3.1)	**1.28±0.03**	1.58 (0.56–2.8)
bendi – val (Lake Van)	1.14±0.02	1.32 (0.39–2.4)	**0.83±0.01**	1.00 (0.37–2.1)
sapph–val (Lake Van)	1.14±0.02	1.32 (0.39–2.4)	**0.83±0.01**	1.00 (0.37–2.1)

3.4 DISCUSSION

The molecular phylogeny built on the basis of Z chromosome sequences provides insight into paternal ancestry, as well as the place and time of origin of parthenogenesis in *Darevskia* lizards. The first remarkable observation is that the splits of four parthenogenetic species (*armeniaca, dahli, uzzelli* and *rostombekowi*) from their respective paternal ancestor (local populations of *valentini* and *portschinskii*) all occurred within a short period of time. All species in this group are today found in a compact area centred in the Lesser Caucasus (Fig 3.2), and could indeed point to a rapid series of hybridizations with the maternal species *mixta* and *raddei*, probably preceded by simultaneous range expansions due to a climatic event. While our estimates of the split times depend heavily on the choice of clock model and calibration points (Table 3.1), we argue that the strict clock model might be a more reasonable assumption for closely related species (Langley & Fitch, 1974; Tiley et al., 2020). In this case, the middle Pleistocene Climatic Transition, i.e. glacial cycles increasing in amplitude and duration (Clark et al., 2006), could have acted as a trigger for the origin of parthenogenesis around $\sim$1 or $\sim$0.5 Mya in the Lesser Caucasus.

Shared geography (both occur on the south-western side of the Kars–Erzurum Plateau) coupled with the change in climate could also have been responsible for the origin of parthenogenetic *bendimahiensis–sapphirina* and *unisexualis* species. However, their divergence times from the closest extant paternal populations of *valentini* are considerably older compared to those in the Lesser Caucasus and not simultaneous (from 2.18–3.43 to 0.83–1.28 Mya; Table 3.1, Fig. 3.5). If the strict clock model is chosen,

the divergence times overlap roughly with the earliest Pleistocene glacial cycles Gibbard et al. 2010. In fact, our 2.42 Mya estimate of the split of *D. unisexualis* from its paternal ancestor *D. valentini*, under the strict clock model and 12.12 Mya calibration at the *mixta–raddei* node according to Murtskhvaladze et al. (2020), corresponds very closely to its estimated divergence from the maternal ancestor *D. raddei* reported in the same study (2.59 Mya; Murtskhvaladze et al., 2020). In summary, it seems plausible that the earlier origin of *unisexualis* and *bendimahienis–sapphirina* was triggered by a different climatic event(s).

While we have high confidence in the derived topology of our Z chromosome tree, our ability to estimate the true divergence times could be limited by a number of factors. First, the sexual populations examined might not necessarily include the 'true' paternal ancestral lineages, due to them either being extinct or not sampled. This is less of a concern in the nominal subspecies *D. v. valentini* (Georgia and Ardahan in Turkey) and all *portschinskii* samples, which have been studied phylogeographically (Tarkhnishvili et al., 2013; Rato et al., 2021), but is more relevant to *D. valentini* from Erzurum and the Lake Van basin, where more ancient phylogenetic lineages have recently been discovered (Candan et al., 2021). Second, while using the molecular clock, it is hard to avoid relative bias in the estimation of the divergence time between the old and recently diverged lineages. Our calibration is based on different timescales, with the earliest datings of the most basal split between the 'rudis' and 'caucasica' clades (Murphy et al., 2000) around 20 Mya (Kumar et al., 2017; Murtskhvalandze et al., 2020), which corresponds to those from a number of earlier publications (Pyron & Burbrink, 2014; Roquet et al., 2014; Zheng and Wiens, 2016), although we also used the third calibration based on a more recent divergence (13.43 Mya) suggested by Garcia-Porta et al. (2019). If calibrations are placed on the ancient nodes, the ages of young nodes are likely to be overestimated (Tiley et al., 2020). In addition, the largest posterior density intervals for the inferred time tend to show positive skewness for tip branches taxa and negative skewness for basal branches (Beavan et al., 2021). This also explains the unexpectedly long inferred average divergence time of the Z chromosome of some conspecific individuals from the same population, but does not undermine conclusions on the divergence time between the parthenogens and their closest patrilineal ancestor populations. While our results raise the possi-

bility of much older continuous existence of all obligate parthenogenetic forms in *Darevskia* than previously suggested (Moritz et al., 1992; Freitas et al., 2016), achieving more robust conclusions is possible by combining the molecular clock analysis with other approaches. In particular, demographic modelling based on genomic data with large sample sizes in parthenogenetic populations, recently performed on another parthenogenetic lizard species, *Aspidoscellis laredoensis*, produced a maximum age estimate of ∼500 kya (Barley et al., 2022), which coincides with our most recent estimates (0.45–0.51 Mya).

As far as the same sexual species are concerned, the topology of our Z chromosome tree shows two differences compared to the previously suggested phylogenies of *Darevskia*. In particular, *D. derjugini*, which is a sister taxon to *D. mixta* on the full mitogenome tree in Murtskhvaladze et al. (2020), as well as on the transcriptomic data-based tree in Garcia-Porta et al. (2019), is instead found on the same branch with *D. raddei* on our Z chromosome tree (Fig. 3.5). In the second case, the discrepancy includes our results and Murtskhvaladze et al. (2020), on the one side, and Garcia-Porta et al. (2019), on the other, regarding the positions of *D. rudis, D. valentini* and *D. portschinskii* (i.e. the 'rudis' group; Murphy et al., 2000). Our Z chromosome results agree with the mitogenome tree, placing *D. rudis* as a sister taxon to both *D. valentini* and *D. portschinskii*, while the transcriptome data, as well as some earlier mtDNA-based studies, suggest that *D. portschinskii* is an outgroup (Tarkhnishvili et al., 2013; Garcia-Porta et al., 2019; Candan et al., 2021). Further complicating the picture, the node datings within this subclade are drastically different between the mitochondrial and nuclear/genomic DNA- based phylogenies. In particular, while the entire 'rudis' group appears to have diverged from the common ancestor <1 Mya on the mtDNA tree, the most basal of the three species is separated by at least 4–5 Myr (our most recent median time estimates) according to our results (Fig. 3.5; Table 3.1) and to those of Garcia Porta et al. (2019). We suggest that these differences could be explained by hybridization among the three species, as well as by differences in introgression among different parts of the genome. Introgression of mtDNA (Tarkhnishvili et al., 2013) as well as microsatellite genotypes (Freitas et al., 2019) has been demonstrated for all three species of the 'rudis' group. Moreover, the populations which were sampled in different studies could have possessed intro-

gressed vs. non-introgressed haplotypes. In particular, *D. portschinskii* from some Georgian populations, used in the study of Murtskhvaladze et al. (2020), has mitochondrial haplotypes more closely related to *D. rudis obscura* and *D. valentini* from southern Georgia than to the populations from Armenia and south-eastern Georgia. Both mtDNA and autosomal elements may introgress more readily compared to genes located on sex chromosomes (Babik et al., 2005; Macholan et al., 2011; Zieliński et al., 2013; Hassanin, 2015). One could speculate that the introgression of Z chromosomal sequences occurs much more rarely, and populations of *D. portschinskii* and *D. valentini* that are paraphyletic at some mtDNA and nuclear sequences (Tarkhnishvili et al., 2013; Candan et al., 2021, Garcia-Porta et al., 2019) are monophyletic when the Z chromosome is considered.

In conclusion, our approach to constructing the phylogeny using Z-linked genomic markers proved to be very useful in identifying the genetically closest paternal sexual populations of each parthenogenetic species in *Darevskia* rock lizards. We have also pinpointed at least one major period when the hybridization and switch to parthenogenesis occurred almost simultaneously in four different instances, all seemingly confined to the same geographical area in the Lesser Caucasus, while the origin of two other parthenogenetic taxa must have taken place earlier in geological time. In addition, we found that the Z chromosome-based phylogeny of the sexual paternal species themselves also presents different divergence time estimates compared to previous studies based on mtDNA. Further analysis of autosomal genomic markers will provide more resolution into the very complex history of the origin and evolution of parthenogenesis in this group.

CHAPTER 4

THE ORIGIN AND SPECIATION HISTORY OF PARTHENOGENETIC ROCK LIZARDS (*DAREVSKIA SAPPHIRINA* AND *DAREVSKIA BENDIMAHIENSIS* SCHMIDTLER, EISELT & DAREVSKY, 1994) ENDEMIC TO THE BASIN OF LAKE VAN IN TURKEY

with Alexey Yanchukov, Mehmet Somel, David Tarkhnishvili, Mehmet Kürşat Şahin

4.1 INTRODUCTION

The genus of rock lizards *Darevskia* Arribas, 1997 was the first group of terrestrial vertebrates where obligate parthenogenesis had been discovered (Darevsky, 1957). To date, seven unisexual forms have been described (Darevsky 1967; Darevsky & Danielyan, 1977; Murphy et al., 2000). Each of these parthenogenetic forms appears to have resulted from hybridization between two distant phylogenetic branches within the genus, the mitochondrial lineages of which have been estimated to have diverged at least 13.4 Myr ago based on a molecular clock analysis (Yanchukov et al. 2022), with the paternal parent always coming from the clade "rudis" and the maternal parents always coming from the clade "caucasica" (Murphy et al. 2000; Tarkhnishvili et al. 2020, Yanchukov et al. 2022). While the genus *Darevskia* has a broad distribution ranging from the Caucasus to SE Europe (Darevsky et al., 1985; Arnold et al., 2007), all parthenogenetic forms are found in a relatively small area divided between Georgia, Armenia and Eastern Turkey (Tarkhnishvili, 2012).

D. bendimahiensis and *D. sapphirina* are two parthenogenetic forms to have been most recently discovered and given taxonomic species status (Schmidtler et al., 1994). They are the rarest among the seven parthenogenetic forms, with each species only

known from two respective localities in the basin of Lake Van in Eastern Turkey (Schmidtler et al., 1994, Fig. 4.1). The current ranges of the nominal species *D. bendimahiensis* and *D. sapphirina* are allopatric in relation to all other parthenogenetic forms as well as to each other (Fig. 4.1). According to the allozyme, mtDNA, microsatellite, and Z-linked genetic marker data, all parthenogenetic lizard populations in Lake Van basin had originated from hybridization between the local populations of *D. raddei vanensis* (maternal parent) and *D. valentini* (paternal parent, Schmidtler et al., 1994; Murphy et al., 2000, Tarkhnishvili et al. 2020, Yanchukov et al. 2022). In the area north-east of Lake Van, *D. bendimahiensis* and its sexually reproducing parental species can be found in close proximity to each other or even in the same habitat (Tarkhnishvili et al., 2020), but unlike the other parthenogenetic species outside the Lake Van basin (Darevsky & Kulikova, 1964; Arakelyan et al., 2008), no back-cross hybridization with the parental species has ever been reported here.

One of the most disputed questions in the >40 years history of research on parthenogenetic *Darevskia* was the number of so-called "independent clonal lineages" that constitute the all-female unisexual population (Tarkhnishvili et al., 2020). The "clonal lineage" is thereby defined as (i) having originated from a single hybrid (F1) genotype, possibly a single parthenogenetic individual, and (ii) one that derives all its genetic variation to the accumulation of mutations after the onset of parthenogenesis. Since the genetic differences among such clones effectively reflect the difference among the parental individuals in the ancestral sexually reproducing populations, the clones should be identifiable in the parthenogenetic populations as well, if sufficient sequence length is combined with large sample sizes. In previous work, extensive variation in the flanking regions of a few microsatellite loci was taken as evidence of the existence of several clones in parthenogenetic *D. armeniaca, D. dahli* and *D. unisexualis* (Vergun et al., 2014; Vergun et al., 2020; Girnyk et al., 2018), and *D. rostombekowi* (Ryskov et al., 2017). Other authors have pointed out unusually high proportions of microsatellite alleles shared among individuals within as well as between some parthenogenetic species and even suggested that some clonal lineages such as *D. armeniaca* and *D. dahli*, might combine ancestry from more than two bisexual parents (Tarkhnishvili et al. 2020).

Given a high degree of geographic endemism of *D. bendimahiensis* and *D. sapphirina* and their putative parental populations, these lineages constitute a good opportunity to study the evolution of hybridogenenic parthenogenesis in fine detail. In this study, we present a novel genome-wide analysis of several populations of *D. raddei vanensis, D. valentini, D. bendimahiensis* and *D. sapphirina* from the Lake Van area. We aim to gain better insight into the evolution of the hybrid species, asking two main questions:

4.1.1 Are *D. bendimahiensis* and *D. sapphirina* Reciprocally Monophyletic, and Which Maternal and Paternal Populations Contributed To Their Ancestry?

When *D. bendimahiensis* and *D. sapphirina* were first defined by Schmidtler et al., (1994), the authors claimed their species status on the basis of minor morphological differences and their allopatric distribution. However, both parthenogens share 51% to up to 67% of their microsatellite genotypes, preventing the assignment of individuals into any meaningful groups (Tarkhnishvili et al. 2020). Later on, we showed that the two cannot be distinguished based on RAD-seq markers on chromosome Z (Yanchukov et al., 2022). The latter study also revealed that the local population of *D. valentini* from Çaldıran, NE of Lake Van, has the highest Z-chromosome sequence similarity with both parthenogens. Earlier comparison of partial mtDNA sequences revealed that several populations of *D. r. vanensis*, sampled over a large area southeast of Lake Van and extending into Iran, are equally similar to *D. bendimahiensis* and *D. sapphirina* (Freitas et al. 2016). Here we readdress the details of both parents' contribution to the parthenogenetic hybrids, using autosomal ddRAD-seq data.

4.1.2 Was a Single Parental Pair or the Multiple Pairs of Individuals Involved in the Origin of *D. bendimahiensis* and *D. sapphirina*?

The successful hybridization between two highly divergent parental species, which results in the parthenogenetic hybrid lineage is generally regarded as a very rare event (Freitas et al., 2022); likewise, producing a triploid or higher degree polyploid individual resulting from a successful hybridization between a parthenogenetic female

and a sexual male is also rare. For instance, in North-America lizards, *Aspidoscelis sp*, Lutes et al. (2011) reported laboratory syntheses of four parthenogenetic lineages following hybridization between triploid oocytes from a parthenogenetic *A. exsanguis* with haploid sperm from bisexual species *A. inornata*. Yet, this result took > 29 years of continuous breeding attempts, and all the offspring, while viable, appeared to be sterile. In parthenogenetic *Darevskia* species, the apparent presence of multiple clonal lineages would suggest that more than one pair of parental individuals had successfully hybridized. Another interesting observation is that the estimated geological timing of such hybridization events seems to coincide among the different species, possibly due to the physical contact of the parental species following changes in the climate (Freitas et al., 2016; Yanchukov et al., 2022). At present, none of the parental species are sympatric, suggesting that the time windows offering the opportunity for hybridization could have been relatively short. Lastly, it should be mentioned that some parthenogenetic species regularly participate in the back-cross hybridization with their paternal parent *D. valentini*, resulting in sterile triploid or tetraploid offspring (Arakelyan et al., 2008; Danielyan et al., 2008). Theoretically, breaking such sterility barrier and returning to parthenogenetic reproduction in diploid form is very difficult (Hörandl, 2009), and the evidence for gene flow between the parents and the parthenogens has yet to be found in Darevskia (Freitas et al. 2019, 2022). In this study, we test the null hypothesis that a single pair of parental individuals lies at the origin of the entire present-day gene pool of *D. bendimahiensis* and *D. sapphirina.*

4.2 METHODS

4.2.1 Sampling

We collected samples (tail tips of the lizards) from two populations of *D. valentini*, five populations of *D. raddei vanensis*, two populations of *D. bendimahiensis*, and two *D. sapphirina* populations from the locations near Lake Van in Eastern Turkey (Figure 4.1) as shown in Table 4.1 (Table D.1 for coordinates). The entire region has been rigorously studied to ensure that all populations in the region have been identified. DNA was extracted from the tail tips taken with negligible harm to the

animals, as described in Tarkhnishvili et al., 2020. The parental species are always allopatric, while in one location (Muradiye) both maternal *D. raddei vanensis* and the daughter *D. bendimahiensis* occur sympatrically, and in the other location (Çaldıran), the populations of *D. bendimahiensis* and the paternal parent *D. valentini* are c.3 km away from each other. All maps in the study are constituted using the Leaflet R library (Cheng et al., 2022) codes presented in Appendix B-I.

Table 4.1: Samples and populations used in the analysis (n - number of individuals).

Species	**Role in the hybridization**	**Number of populations**	**Location, sample size**
D. raddei	Maternal parent	5	Doğubayazıt (n = 4) Muradiye (n = 2) Saray/Van (n=3) Umut / Yaramış (n=1) Çörekli (n=1)
D. valentini	Paternal parent	2	Çaldıran (n=4) Kızılyusuf/Muş (n=4)
D. bendimahien-sis	Parthenogenetic daughter	2	Çaldıran (n=6) Muradiye (n=4)
D. sapphirina	Parthenogenetic daughter	2	Pınarlı (n=3) Van/Ağrı Border (n=5)

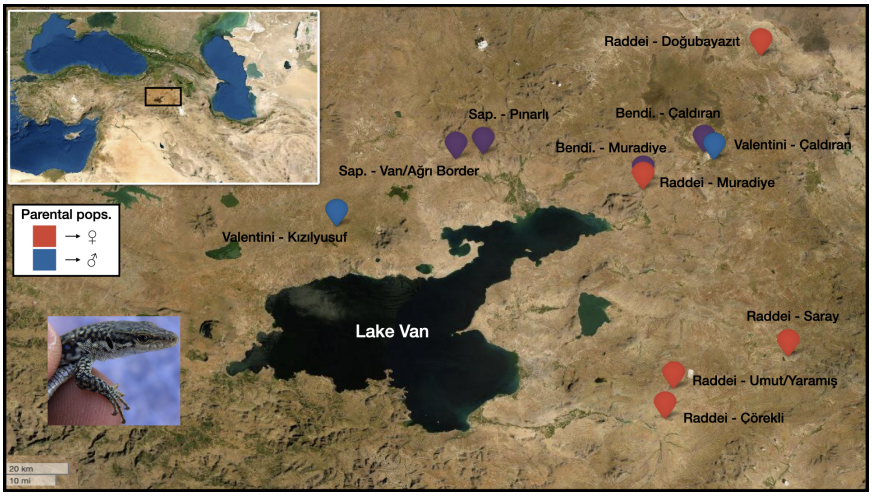

Figure 4.1: Sampling locations of *D. valentini, D. raddei, D. bendimahiensis* and *D. sapphirina* populations around Lake Van (♀: Maternal species, ♂: Paternal species, and the inset photo belongs to a *D. sapphirina* individual).

4.2.2 ddRAD-seq Library Preparation and Sequencing

ddRAD library preparation and sequencing (PstI+MseI restriction enzymes, Single-End reads, 100 bp on Illumina HiSeq 2500) were carried out by Floragenex Inc., Oregon, USA. The details of the library preparation protocol are described in Yanchukov et al. (2022).

4.2.3 ddRAD-seq Data Analysis

As we described in detail in Chapter 2, after demultiplexing the data using the "process_radtags" program in Stacks v. 2.54 (Rochette et al., 2019; Rochette and Catchen, 2017), we got fastq files of samples. Then we performed quality control for the raw reads of each sample using FastQC v0.11.9. We then aligned the reads to the reference genome of *Darevskia valentini* Boettger, (1892) (Ochkalova et al., 2022) (https://www.ncbi.nlm.nih.gov/data-hub/genome/GCA_024498535.1/) and to the reference genome of *Podarcis muralis* Laurenti, (1768) (Andrade et al., 2019) (https://www.ncbi.nlm.nih.gov/genome/?term=podarcis) using Bowtie2 v.2.4.1 (Langmead & Salzberg, 2012) (Figure 4.2) with the conservative "–end-to-end" setting (i.e. alignment scores calculated for the entire read matching the reference with default parameters: -6 score penalty for a mismatch, -11 score penalty for a 2-bp gap, -0.6 - 0.6*L total score threshold required to retain the read, where L is the read length).

Next, we filtered out the reads with an alignment quality score < 30 using Samtools v.1.11 (Li et al., 2009) and calculated the filtering ratio with the codes presented in Appendix B-II. Only the autosomal sequences were kept for downstream analysis, as the comprehensive analysis of the ddRAD loci aligned to chromosome Z was published earlier (Yanchukov et al., 2022). We did, however, compare the number of reads aligned to Z with the number of autosomal reads in order to confirm the ploidy of parthenogenetic individuals. We then drew a ploidy plot (code available in Appendix B-III). The ddRAD loci were assembled at the level of individuals using Stacks v2.54 (i.e. if a locus is present in a single individual, this is enough to catalog a locus).

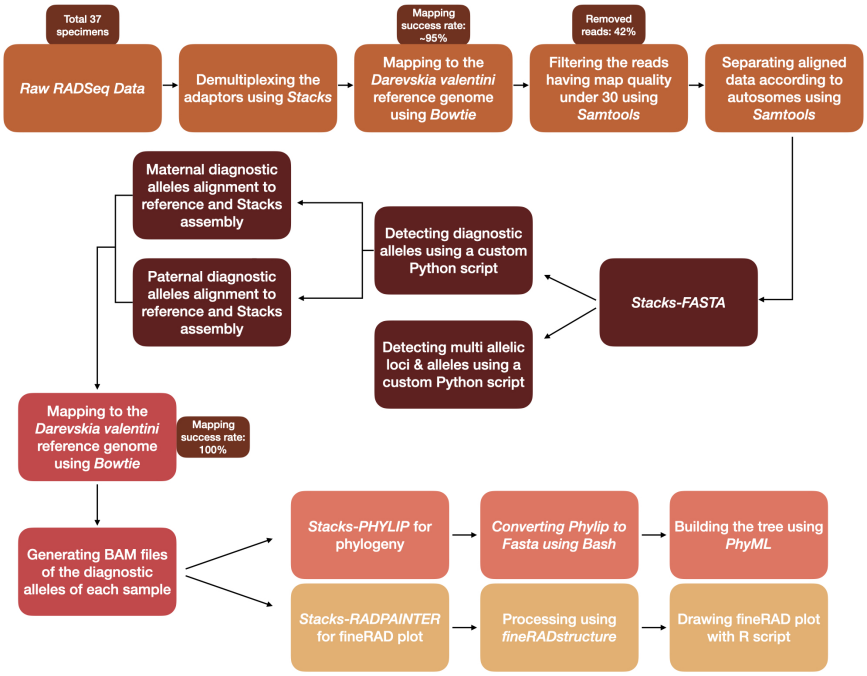

Figure 4.2: A flowchart of RAD-seq data analysis process

4.2.4 Diagnostic Alleles

In each parthenogenetic individual, we focused on the heterozygous loci where one allele is shared exclusively with the maternal parent, and the other allele is shared exclusively with the paternal parent (i.e. diagnostic alleles, Fig. 4.3). A similar approach was used by (Grismer et al. 2014) to phase the nuclear gene haplotypes of the hybrid unisexual *Leiolepis* lizards.

According to the protocol presented in Appendix B-IV, identification and selection of diagnostic alleles were done using a custom Python script (https://github.com/mericerdolu/DiagnoSeq), following the steps below:

(I) Collect the ddRAD alleles per individual of the bisexual parental species (here *D. raddei* and *D. valentini*) in fasta format. This fasta file is generated by running the "populations" program of Stacks 2 with the "–fasta-samples" flag.

(II) For each parthenogenetic individual, investigate whether one allele of each heterozygous ddRAD locus in the hybrid individual is unique for the maternal gene pool and the other allele for the paternal gene pool. Complete sequence

identity is required for the comparison.

(III) For each parthenogenetic individual, investigate whether one allele of each heterozygous ddRAD locus in the hybrid individual matches an allele at the same locus in the maternal gene pool and whether the other allele matches an allele at the same locus in the paternal gene pool. Complete sequence identity is required for the match. The difference from II is that the locus index is included here.

(IV) If both II and III are true, list these alleles as diagnostic alleles from either the mother or father for each hybrid sample. Then generate the set of diagnostic maternal and paternal allele files of each hybrid individual.

We then re-aligned the diagnostic alleles back to the *D. valentini* reference genome (with 100% mapping success using Bowtie 2) and converted them to the BAM format. They were combined with the BAM files of *D. raddei* and *D. valentini*, and two new sets of loci were then re-assembled using the "populations" program of Stacks 2 with the "–phylip" flag. The results were saved in PHYLIP format and the ML phylogenetic trees were built using PhyML v. 3.3_1 (Guindon et al., 2010; Lemoine et al., 2018), and visualized using the web tool iTOL (Letunic and Bork, 2021). The fineRADstructure (Malinsky et al., 2018) was used to construct the co-ancestry matrices and the fineRADstructure plots were visualized using the R script “fineRADstructurePlot.R” (https://github.com/millanek/fineRADstructure/blob/ master/fineRADstructurePlot.R) developed by Milan Malinsky and the R library “FinestructureLibrary.R” (https://github. com/millanek/fineRADstructure/blob/master/FinestructureLibrary.R) developed by Daniel Lawson in R v. 4.0.0 (R Core Team, 2020) and R Studio (RStudio Team, 2020).

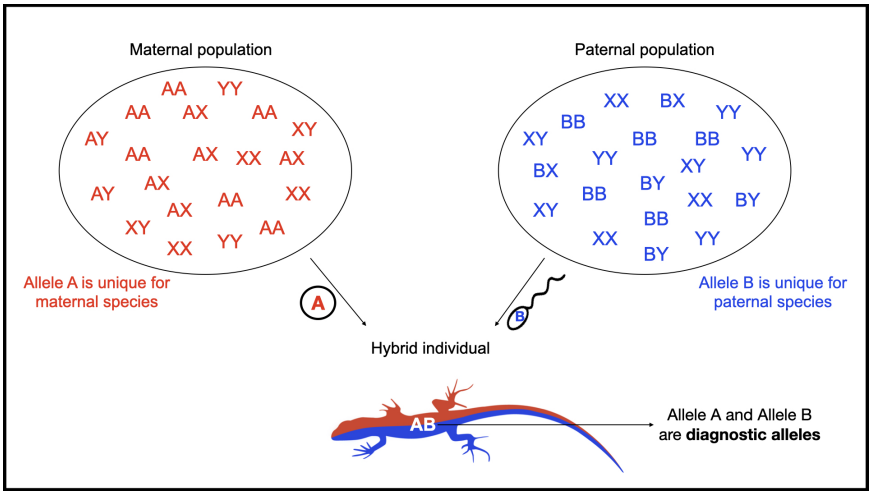

Figure 4.3: Identification of diagnostic alleles (A and B are haplotypes unique for only one parental side, and X and Y are common haplotypes for both parental gene pools at the same locus).

4.2.5 Estimation of Population Structure

We also calculated Fst (per SNP based and as a population average) among the parthenogenetic populations using Stacks 2 and used random permutations of individuals across taxa to test whether the four extant geographic populations and two species of *D. bendimahiensis* and *D. sapphirina* correspond to any meaningful genetic clusters with the codes presented in Appendix B-V.

4.2.6 Analysis of the Number of Hybridization Events in the History of Parthenogenetic Species Using Multi-Allelic Loci

This method is constituted based on the idea that if there has been more than one founder hybridization event in the yore of the hybridogenous parthenogenetic species, there ought to be three or more haplotypes (alleles) at some loci of the gene pool of the hybrid species (see Figure 4.4) because different parental individuals should carry disparate haplotypes at some loci of their genome. Therefore, in this part of the study, we identify the multi-allelic loci of the hybrid gene pool using the custom python script "MALAsHyb.py" according to Appendix B-VI. Whilst detecting different haplotypes at each locus, we consider a complete ddRAD sequence. We have two alternative hy-

potheses (not mutually exclusive) about this phenomenon: (I) these multi-haplotypes are the result of secondary hybridizations, (II) these multi-haplotypes are the result of de novo mutation or sequencing/technical error.

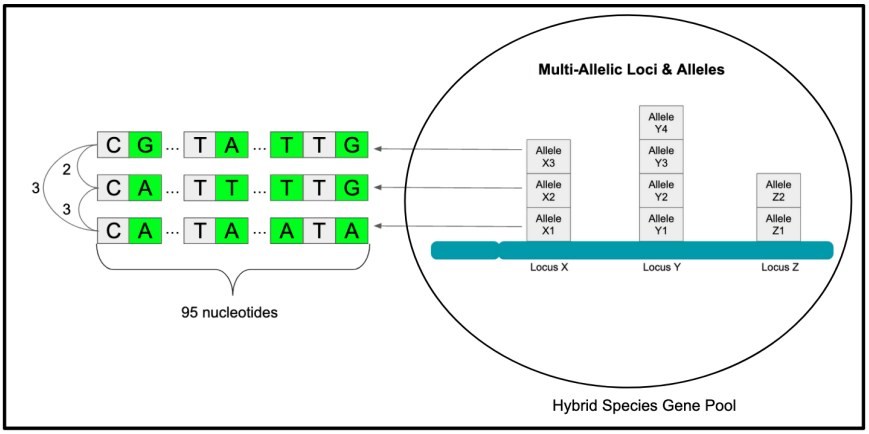

Figure 4.4: Hypothetical examples of multi-allelic loci. Multi-allelic loci and their alleles (right), and the number of nucleotide differences between multi-allelic locus alleles (green positions are polymorphic sites) (left).

To test the hypotheses, we employ three main methods. First, we identify the number of 3-allelic loci that have three alleles differentiating each other >1 nucleotide distances (for the distance calculation, see Figure 4.4) using the R codes presented in Appendix B-VI. Assuming mutations are rare events and are uniformly distributed across the genome, this situation is unlikely to arise via mutations happening on other alleles in the gene pool. Therefore the number of such alleles can help evaluate hypotheses I and II. Secondly, we determine which three allelic loci in the hybrid gene pool have two unique alleles for one parental species, and one unique allele for the other parental species. We then designate them as maternal or paternal locus according to which parental species has two alleles (as shown in Figure 4.5). Later we perform a binomial test to the number of maternal loci vs. the number of paternal loci ratio using the codes presented in Appendix B-VII. If the proportion of maternal vs. paternal loci is not different than 0.5, both hypotheses may be correct but if it is different than 0.5, only hypothesis II is improbable. We applied a binomial test to evaluate these hypotheses. Nevertheless, the motif of the number of maternal and paternal loci may also be caused by allelic diversity in the parental species. We thus calculate the allelic richness and mean number of alleles per locus of parental and

hybrid species using the codes presented in Appendix B-VIII.

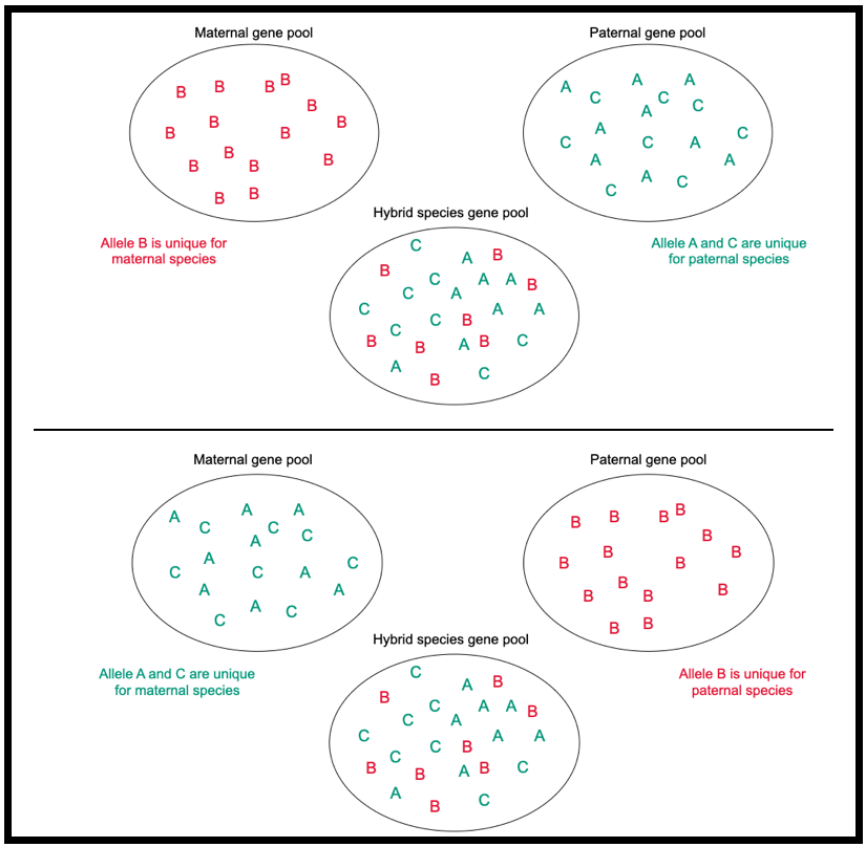

Figure 4.5: Three-allelic loci used in the binomial test. A, B and C are haplotypes at the same locus in the gene pools of the species. The upper panel shows a paternal 3-allelic locus since the paternal species has two haplotypes whereas the lower panel shows a maternal 3-allelic locus.

In the third approach, we calculate Spearman's rank correlations of allele frequencies of 3-allelic loci alleles shared with parental species (Appendix B-IX). In this test, we expect a correlation between the frequencies of the hybrid alleles and parental alleles under hypothesis I, but they would not be expected to show a correlation under hypothesis II. If second alleles appear via parallel mutations in hybrid and parental species, they appear at different times so their frequencies do not change in correlation over time between hybrid and parental species.

Finally, multiple alleles may be because of the paralogs. If the 3-allelic loci alleles are not caused by paralogs, we expect that the mean sequencing depth of the 2-allelic loci vs. 3-allelic loci is not different or the mean depth of 2-allelic loci is higher. So

as to assess this, we compare the mean depths of the loci with the Mann_Whitney U test using the protocol presented in Appendix B-X.

4.3 RESULTS

We produced ddRAD data from 37 individuals from the Lake Van basin. After demultiplexing the data, we obtained the unaligned raw data (the average per-read Phred score = 40, per-sample read: 1,409,381.5 records/sample in the range 884,442-2,379,981) across 11 *D. raddei*, 8 *D. valentini*, 8 *D. sapphirina* and 10 *D. bendimahiensis* individuals collected from around Lake Van.

We first studied whether we have a polyploid individual and the sex of individuals. The ratio of Z-chromosomal reads to autosomal reads in accordance with the alignment to the *Podarcis muralis* genome in all parthenogenetic individuals was close to 0.5 (Figure 4.6), thus confirming that they were diploid females (ZW).

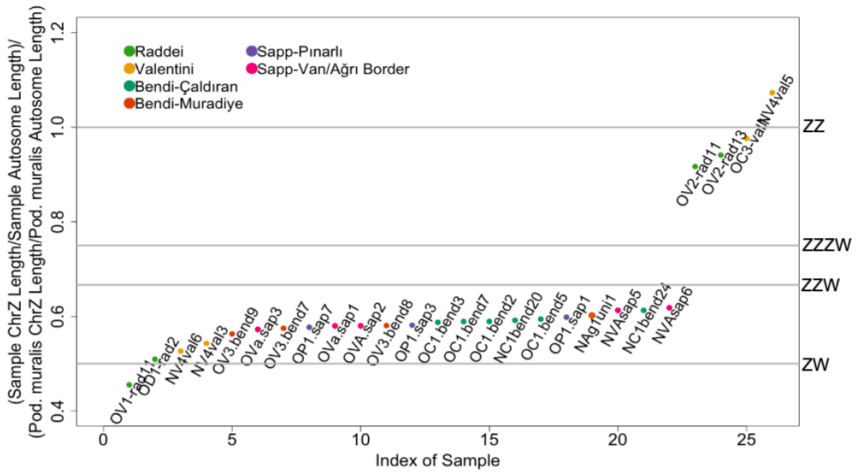

Figure 4.6: Relative coverage of chromosome Z vs. autosomes. The y-axis shows the ratio of the number of mapped nucleotides of chromosome Z/the number of mapped nucleotides of all autosomes of each *D. bendimahiensis* and *D. sapphirina* specimen. Also two females and two males of *D. raddei*, and two females and two males of *D. valentini* specimens were included in the plot in order to indicate what female and male examples of the bisexual species look like in the plot. Gray lines represent the proportion of chromosome Z coverage vs. autosomal coverage under various ploidy levels. We presume that technical effects and paralogs may be shifting the ratios beyond the expected 0.5 and 1.0.

4.3.1 Maternal Ancestry of *D. bendimahiensis* and *D. sapphirina*

A total of 29 individuals were used here: *D. sapphirina, D. bendimahiensis* and the maternal species *D. raddei* in Lake Van basin. We aligned n=34,271 maternal diagnostic alleles (i.e. alleles not found in the paternal gene pool of *D. valentini*; see Methods) to the *D. valentini* reference genome with 100% mapping success rate. Then we filtered reads for having mapping quality <30. Processing this data in Stacks, we obtained 619,135 shared RAD loci among samples and 75,516 SNPs at these loci (see Table D.1 for the number of detected diagnostic alleles for the maternal side).

We built a maximum likelihood phylogenetic tree for *D. bendimahiensis, D. sapphirina* and *D. raddei* populations around Lake Van, using maternal side diagnostic alleles of *D. bendimahiensis* and *D. sapphirina* specimens. We used PhyML v. 3.3_1 to construct a phylogenetic tree (Guindon et al., 2010; Lemoine et al., 2018).

According to the tree (Figure 4.7), *D. raddei* from Doğubayazıt is genetically closest to *D. bendimahiensis* and *D. sapphirina* (with 100% branch support). Thus it is the best candidate for the maternal population of the parthenogens. Interestingly, despite being the geographically closest putative maternal population to the parthenogens, *D. raddei* from Muradiye is genetically the farthermost group. In addition, we saw that *D. bendimahiensis* and *D. sapphirina* samples do not form reciprocally monophyletic clades.

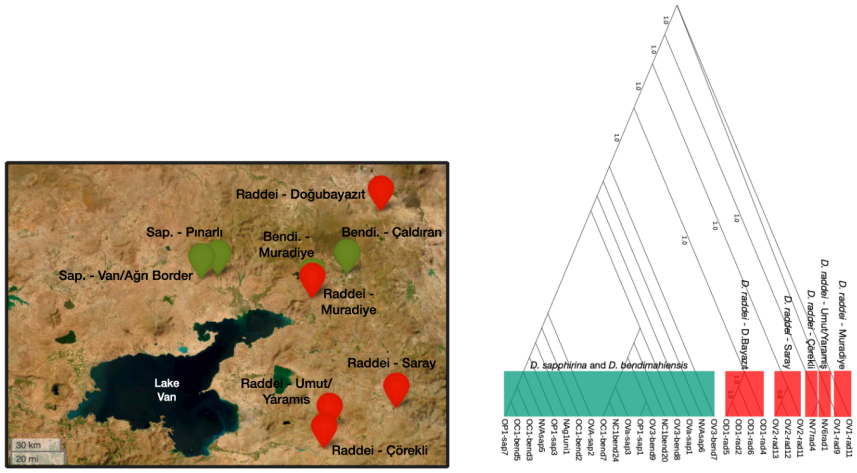

Figure 4.7: A maximum likelihood phylogeny based on diagnostic alleles of *D. bendimahiensis, D. sapphirina* and *D. raddei* populations around Lake Van. The numbers on branches are branch supports (approximate Bayes branch support); support values close to zero are not shown.

We also drew a fineRADstructure plot for *D. bendimahiensis*, *D. sapphirina*, and *D. raddei* populations around Lake Van. This plot, in Figure 4.8, again shows (y-axis indicates the recipients) that four *D. raddei* individuals from Doğubayazıt clearly share more coancestry with *D. sapphirina* and *D. bendimahiensis* than other *D. raddei* individuals.

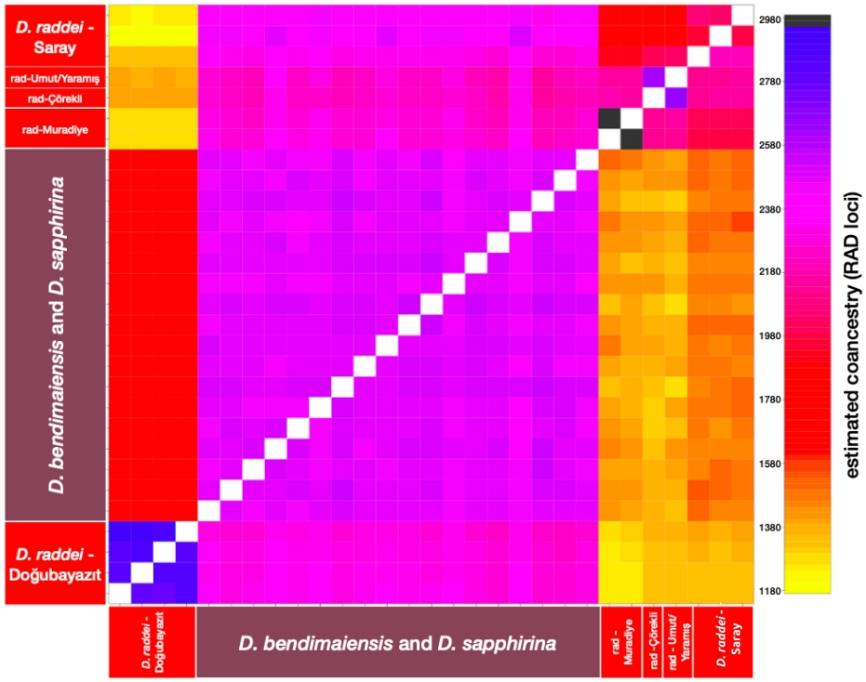

Figure 4.8: A fineRADstructure plot based on diagnostic alleles of *D. bendimahiensis*, *D. sapphirina*, and *D. raddei* populations around Lake Van.

4.3.2 Paternal Ancestry of *D. bendimahiensis* and *D. sapphirina*

A total of 26 individuals was used here: *D. sapphirina, D. bendimahiensis* and paternal species *D. valentini* in Lake Van basin. We aligned the n=34,271 paternal diagnostic alleles (i.e. alleles not found in the paternal gene pool of *D. valentini*; see Methods) to the *D. valentini* reference genome with 100% mapping success rate. Then we applied the filtering process of the reads having mapping quality <30. According to the results of Stacks, we obtained 762,187 RAD shared loci among samples and 126,648 SNPs (see Table D.1 for the number of detected diagnostic alleles for the paternal side).

We built a maximum likelihood phylogenetic tree for *D. bendimahiensis, D. sapphirina* and *D. valentini* populations around Lake Van, using paternal side diagnostic alleles of *D. bendimahiensis* and *D. sapphirina* specimens. For the phylogenetic tree, we used PhyML v. 3.3_1 (Guindon et al., 2010; Lemoine et al., 2018).

According to this tree (Figure 4.9), *D. valentini* from the Çaldıran population is the

genetically closest population to *D. bendimahiensis* and *D. sapphirina*. Thus *D. valentini* from Çaldıran is most likely the paternal population (with 100% branch support value). In addition, among *D. sapphirina* and *D. bendimahiensis* individuals, there is no grouping based on the markers analyzed.

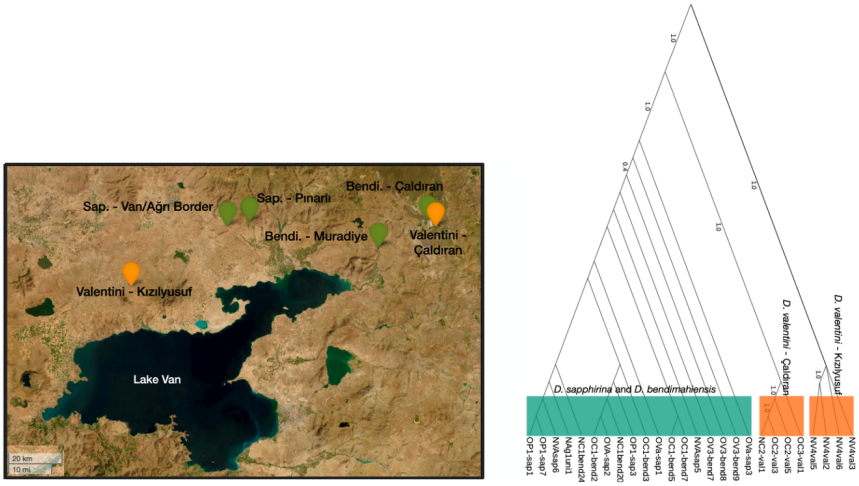

Figure 4.9: A ML phylogeny based on diagnostic alleles of *D. bendimahiensis, D. sapphirina,* and *D. valentini* populations around Lake Van. The numbers on branches are branch supports (approximate Bayes branch support) and support values close to zero are not shown.

We also drew a fineRADstructure plot for the nominal species *D. bendimahiensis, D. sapphirina,* and *D. valentini* populations around Lake Van. This plot (in Figure 4.10) again shows (y-axis indicating the recipients) that *D. valentini* from Çaldıran has a higher genetic similarity with *D. bendimahiensis, D. sapphirina* populations than *D. valentini* in Kızılyusuf. This means that *D. valentini* from Çaldıran is most likely the paternal population of the nominal species *D. bendimahiensis, D. sapphirina*.

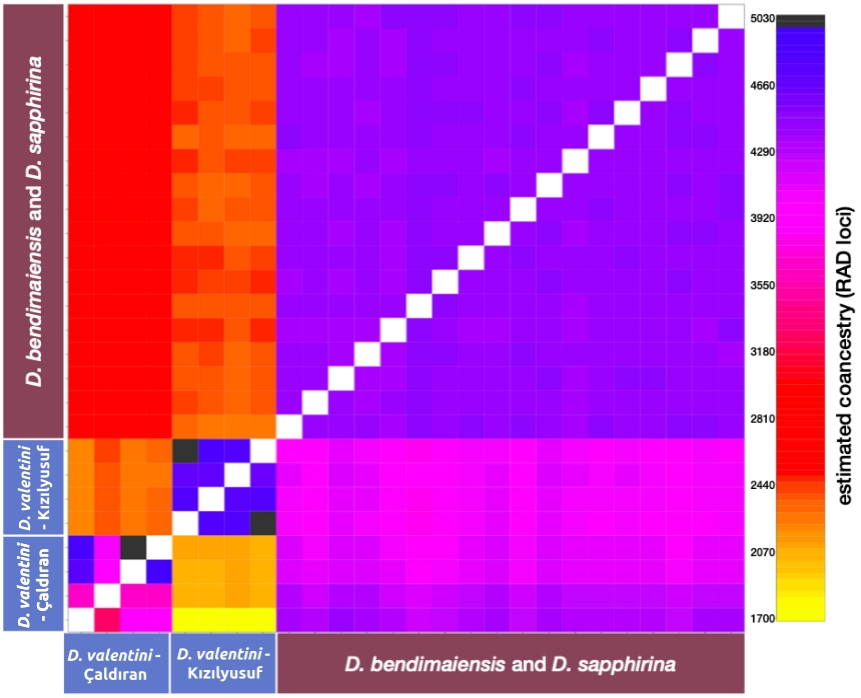

Figure 4.10: A fineRADstructure plot based on diagnostic alleles of *D. bendimahiensis, D. sapphirina*, and *D. valentini* populations around Lake Van.

4.3.3 SNP-Based Fst of *D. bendimahiensis* and *D. sapphirina* Populations

An interesting observation deriving from the phylogenetic analyses presented above was that the *D. sapphirina* and *D. bendimahiensis* groups show no obvious within-group genetic clustering. Both the branching patterns in the phylogenetic trees (Figures 4.7 and 4.9) and the heatmaps of coancestry similarities (Figures 4.8 and 4.10) supported this pattern of lack of reciprocal monophyly.

We, therefore, decided to further test this based on Fst values. For this, we calculated SNP-based Fst between *D. bendimahiensis* (n=10) and *D. sapphirina* (n=8) populations. Across n=396,312 autosomal SNPs identified in n=1,032,862 ddRAD loci, the mean Fst value was 0.03 between *D. bendimahiensis* and *D. sapphirina* species, while the mean Fst values among four *D. bendimahiensis* and *D. sapphirina* populations ranged between 0.04-0.060.05 (see Table 4.2 for pairwise values of 4 populations). The distributions of SNP-based Fst values of *D. bendimahiensis* and *D. sapphirina* species are shown in Figure 4.11. The two *D. bendimahiensis* populations and two

D. sapphirina populations were taken as two separate species in the histogram above and they were taken as four separate populations in the bottom histogram while the Fst calculations were performed.

SNP-based Fst values were mostly zero but there were a small number (0.3%) of high (Fst=1) and some (9.6%) moderate (0.1<=Fst<=0.5) loci. In addition, a permutation test (500 iterations) was performed with individuals assigned randomly into both two species and four groups of the same respective sizes as in the four population samples in *D. bendimahiensis-D. sapphirina*. Our hypotheses for this test were the null hypothesis (H0) that the observed mean Fst is random, i.e., the groups analyzed are not differentiated, and the alternative hypothesis (H1) that the observed mean Fst is not random, i.e., the groups are analyzed are evolutionarily differentiated. For the two putative species, the Fst values were consistently lower than the original Fst estimate (one-sided $p = 0.033$). For the four species test, the observed Fst values were marginally significant when calculated between *D. bendimahiensis* and *D. sapphirina* populations ($0.04 <$ one-sided $p < 0.065$) (Table 4.2).

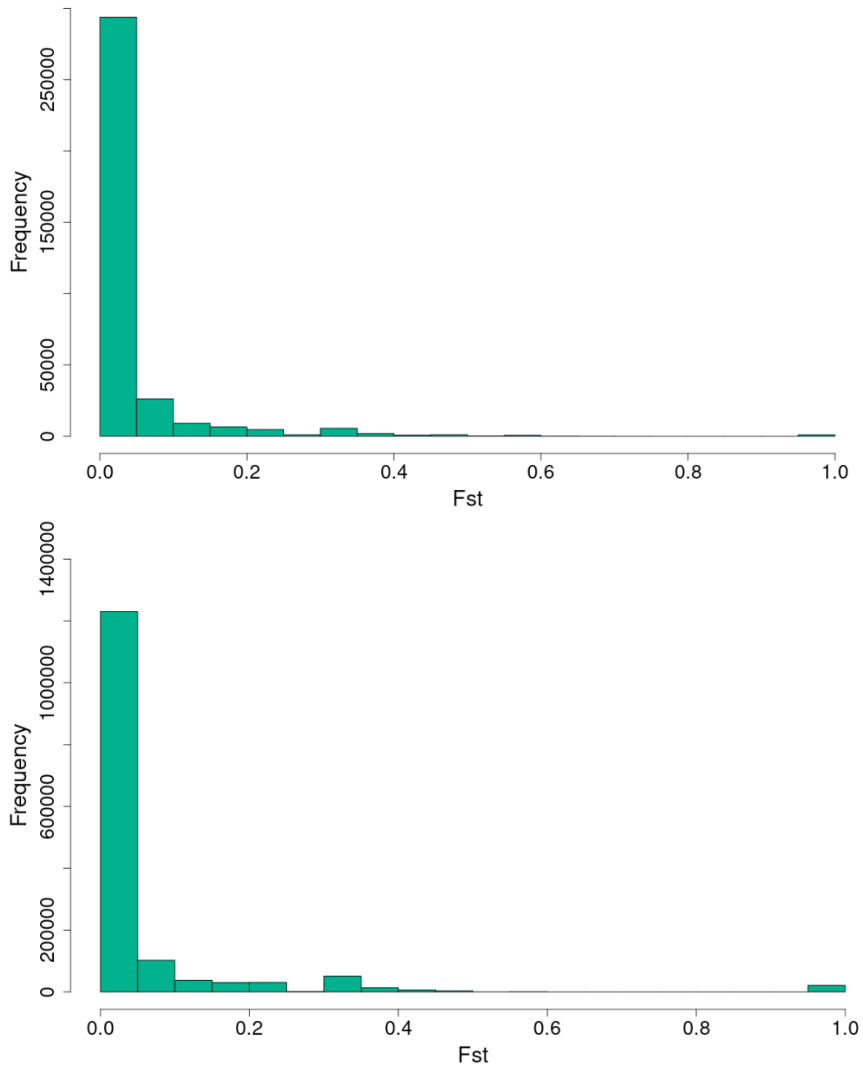

Figure 4.11: Fst distributions based on SNP in *D. bendimahiensis* and *D. sapphirina* as two species (above, 352,988 Fst values and 396,312 SNPs) and 4 populations (bottom, 1,530,146 Fst values (we pooled the Fst values of 6 pairs) and 396,312 SNPs).

Table 4.2: Mean Fst values and randomization (permutation with 500 iterations) test results between *D. bendimahiensis* and *D. sapphirina* populations (no multiple testing correction was conducted, number of SNP: 396,312).

	Bendi. - Çaldıran	Sap. - Pınarlı	Bendi. - Muradiye
Sap. - Van/Ağrı Border	0.05 (p-value=0.044)	0.06 (p-value=0.055)	0.05 (p-value=0.050)
Bendi. - Çaldıran		0.05 (p-value=0.055)	0.04 (p-value=0.050)
Sap. - Pınarlı			0.06 (p-value=0.064)

4.3.4 The Number of Hybridization Events in the Evolution of *D. bendimahiensis* and *D. sapphirina* Populations in Lake Van Basin Based on Multi-Allelic Loci Alleles

One of the most intriguing questions about these nominal species *D. bendimahiensis, D. sapphirina* is the number of hybridizations that have occurred in their history. After alignment of the demultiplexed ddRAD data of *D. bendimahiensis*, *D. sapphirina, D. raddei,* and *D. valentini* specimens to the reference genome of *D. valentini*, we filtered the reads with mapping quality < 30, leaving us with mean 1,409,381.5 (884,442-2,379,981) reads per sample. Then we performed a Stacks assembly. And according to the result of the "populations" program of Stacks, we obtained 1,235,791 ddRAD loci for parthenogenetic (10 *D. bendimahiensis* and 8 *D. sapphirina*), as well as parental (11 *D. raddei* and 8 *D. valentini*) samples.

In order to test the number of hybridizations in the history of the *D. bendimahiensis* and *D. sapphirina*, we decided to use information from multi-allelic loci. Our reasoning was that if there has been more than one founder hybridization event in the history of these species, it is expected that there would be three or more haplotypes (alleles) at certain loci of the hybrid species' gene pool, as different parental individuals typically carry distinct haplotypes at various loci of their genome.

We thus determined the loci having more than two haplotypes in the pool of *D. bendimahiensis* and *D. sapphirina*. All multi-allelic loci (with three or a higher number of alleles) were detected among all RAD-seq loci of 10 *D. bendimahiensis* and 8 *D. sapphirina* specimens. Each specimen has 1,203,742 RAD loci on average (see Appendix B-VI for the calculation method). There are a total of 56,319 alleles across a total of 18,066 multi-allelic loci. Of these, 23% are only present in the maternal (*D. raddei*) species, and 26% of them are only present in the paternal (*D. valentini*) species. Sharing these alleles with the mother and/or father species increases the probability that they have been transmitted via hybridization to the hybrid species, rather than having occurred by de novo mutations. As shown in Figure 4.12, $\sim$90% of these multiallelic loci in *D. bendimahiensis* and *D. sapphirina* populations are 3-allelic. Therefore, we concentrate on 3-allelic loci in this part of the study.

In addition, we counted how many 3-allelic loci have three haplotypes with at least 2 nucleotide distances from each other. We focused on such alleles because again, these loci are more likely to be transmitted from the parental species via hybridization rather than having occurred by de novo mutation on other haplotypes in the gene pool of the hybrid species. Among 16,096 3-allelic loci, there were 310 loci having such three alleles (separated by 2 or more differences).

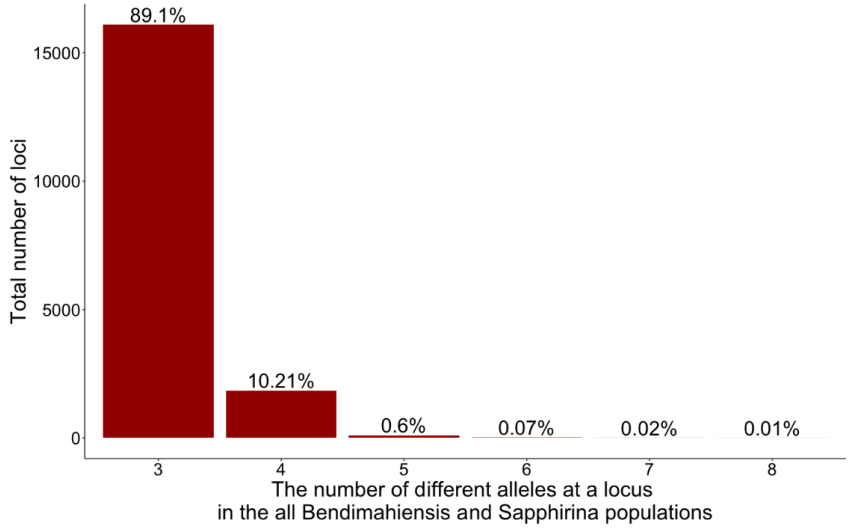

Figure 4.12: Distribution of all multi-allelic locus categories (total 18,066 loci) in *D. bendimahiensis* and *D. sapphirina* populations around Lake Van (the percentages are the proportions of the number of multi-allelic loci in that category relative to the number of all multi-allelic loci).

4.3.5 Analyzing Multiple Hybridization Events by Comparing Maternal and Paternal Contributions with the Binomial Distribution Method

So as to test the multiple hybridization hypothesis, firstly we compiled the list of 3-allelic loci having 2 alleles unique for one parent, and 1 allele unique for another parent (i.e. 2 maternal + 1 paternal or 1 maternal + 2 paternal formats). We named these "maternal" and "paternal" according to which parental species' gene pool contained two alleles. We then performed a binomial test so as to evaluate whether the ratio of the number of maternal loci vs. the number of paternal loci is 0.5. If the 3rd alleles are products of *de novo* mutation, we expect that the ratio would not signifi-

cantly deviate from 0.5. As Table 4.3 shows, we analyzed n=1092 such loci, among which the proportion of maternal alleles was 0.40 (binomial test p-value < 1e-10). This result appears not consistent with the expectation of random mutations creating 3rd alleles, but could instead suggest the introduction of 3rd alleles via secondary hybridization.

Table 4.3: The result of the binomial test on the selected 3-allelic loci (among 18,066 loci).

Number of Loci (maternal)	Number of Loci (paternal)	Prob. of Success	p value
435	657	0.40	1.941e-11

The observed biased pattern here may be the result of the genetic variation present in the parental species. To study this question, we calculated the allelic richness of the parental and hybrid species in this study. In Table 4.4, the allelic richness of the maternal species *D. raddei* and paternal species *D. valentini* are almost the same. The same holds for the mean number of alleles per locus, per individual between parental species. Hence, the higher proportion of 3rd alleles of possible paternal origin is not reflected in the higher diversity of paternal alleles. We also note that, as expected, both allelic richness and mean number of alleles per locus, per individual, are higher in the hybrid species than in parental species. Since the hybrid species has two chromosome sets from two different species, its mean number of alleles per locus, per individual is visibly higher than others.

Table 4.4: Mean number of alleles per locus and allelic richness in hybrid and parental species.

Species	**Mean number of alleles per locus, per individual** (number of alleles/number of loci)	**Allelic richness**
♂ valentini	1.39 (1,469,095/1,057,307)	1.19
♀ raddei	1.38 (1,115,107/808,761)	1.16
bendimahiensis and sapphirina	3.86 (5,196,601/1,345,024)	1.75

4.3.6 Analyzing Multiple Hybridization Events with Correlation of Frequencies of Multi-Allelic Loci Alleles

Of the 3-allelic loci alleles, 23% are only present in the maternal species (*D. raddei*), and 26% are only present in the paternal species (*D. valentini*). These alleles may have been transmitted via hybridization from the parental species or could have arisen via parallel mutations in the parental and hybrid species. In the former case, we would expect the relative frequencies of the 3rd alleles to be correlated between parental and hybrid species (across the loci tested). In the latter case, we would expect no correlation in allele frequencies, which is our null hypothesis (H0).

We found that the frequencies of 3rd alleles were weakly but significantly correlated between the hybrid and both parental species (p-value <0.01) (Table 4.5). The correlation with the paternal species was stronger (r=0.24, across n=657 paternal alleles) than with the maternal species (r=0.11, across n=435 maternal alleles). These correlations would not be expected if the shared 3rd alleles were the product of homoplasy, but appear consistent with a scenario where multiple hybridizations introduced multiple alleles from both paternal and maternal sides.

Table 4.5: Correlation coefficients (Spearman) of the 3-allelic loci alleles in the binomial tests (435 maternal and 657 paternal alleles) (see Table 4.3) between parental and parthenogenetic species.

Species pair	**2 + 1 / 1 + 2** **3-allelic loci minor allele**
valentini - bendi. and sap.	r = 0.24 (p = 4.511e-07)
raddei - bendi. and sap.	r = 0.11 (p = 0.005)

4.3.7 Sequencing Coverage Test of the 3-allelic Loci Alleles Against the Paralogous Sequences

Although the results above appear to support a multiple hybridization scenario, we were also skeptical because of the lack of genetic structure within *D. bendimahiensis* and *D. sapphirina* (Figures 4.7 and 4.9). Indeed, one would expect some degree of clustering between lineages that have derived from different hybridizations. Alternatively, our results on the higher proportion of paternal alleles and correlation between parental and hybrid frequencies could be derived from other sources of genetic variation not considered here, such as paralogs or gene conversion. Indeed, 3rd alleles can be in reality due to paralogous sequences in the data. If true, we expect higher mean depth at the 3-allelic loci than at the 2-allelic loci. Otherwise, we expect equal mean depths. In order to test this hypothesis, we performed a Mann-Whitney U test to the sequencing depths of the 3-allelic loci vs. 2-allelic loci in *D. bendimahiensis* and *D. sapphirina* samples. We used a non-parametric test since the depth values were not normally distributed. The result of the test (p-value = 0.001) implied that the mean depth of 3-allelic loci per individual (mean depth = 2.34, median depth = 2) and the mean depth of 2-allelic loci per individual (mean depth = 2.33, median depth = 2) are significantly different from each other. Hence, we cannot reject that the main reason for the differences is paralogs (see Figure 4.13 for distributions of the depths).

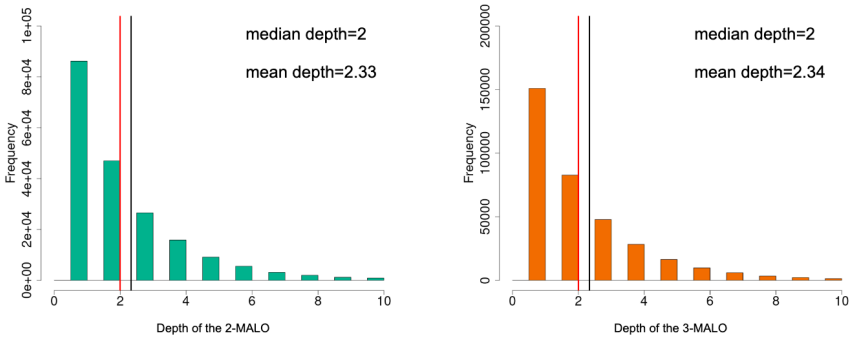

Figure 4.13: Histograms of the depths of 2-allelic loci (197,579 depth values. These are not all such loci, the number of 2-allelic loci is too much, for ease of calculation, 50 loci from different orders are subset from each sample) and 3-allelic loci (349,006 depth values) of *D. bendimahiensis* and *D. sapphirina* samples (vertical black lines are mean depths, red lines are medians).

4.4 DISCUSSION

4.4.1 Maternal and Paternal Ancestry of *D. bendimahiensis* and *D. sapphirina*

In this study, we analyzed a large number of short autosomal ddRAD-seq loci in hybrid parthenogenetic and parental sexually reproducing populations of rock lizards near Lake Van. Among the several geographic populations of both parents included in this study, two with the closest genetic distance to the hybrids were identified unambiguously. Recent studies have demonstrated a strong positive correlation between geographic proximity and the genetic distance in all parent-parthenogen pairs examined over the entire geographic range of parthenogenetic forms in *Darevskia* (Freitas et al. 2016, Tarkhnishvili et al. 2020). This is so in particular on the paternal side, aided by our previous analysis of chromosome Z-linked genetic markers in all seven parthenogenetic forms (Yanchukov et al. 2022). Our current results based on the autosomal markers are in perfect agreement with the latter study. And this confirms that the population of *D. valentini* in Çaldıran shares the highest proportion of genetic ancestry with the actual paternal parent of *D. bendimahiensis* and *D. sapphirina* (Figure 4.9, 4.10). A different pattern, however, was observed on the maternal side. The population of *D. raddei vanensis* from Doğubayazıt is separated by 55.6 km of complex terrain (a 2212 m a.s.l. Tendurek mountain pass) from the nearest known location of

D. bendimahiensis. This population was nevertheless genetically much closer to the latter compared to neighboring populations north-east and east of Lake Van (Figure 4.7, 4.8). This finding might reflect a complex history of distribution range shifts of either the maternal or the parthenogenetic daughter populations. These range shifts were possibly driven by the changes in climate since hybridization (Yanchukov et al. 2022). The timing of this event was independent of the origin of all other parthenogenetic forms and has been estimated as recently as 18-204 kyr (Freitas, 2016) or as early as 1-2 Myr (Yanchukov et al., 2022).

4.4.2 *D. bendimahiensis* and *D. sapphirina* Are Genetically Indistinguishable

The *D. bendimahiensis* and *D. sapphirina* were originally described as two different species by Schmidtler et al., 1994 based on pholidosis characters and the fact that their current distribution is allopatric. Only a single location per species was found at that time, both (Muradyie and Van/Ağrı administrative border) of which have been included in the present study. The pholidosis (i.e. number and shape of scales on certain body parts) in lacertids often exhibits broad intraspecific/intrapopulation variation overlapping with differences among species (Bellati et al., 2011). Furthermore, none of the traits described by the authors is unique to one or the other species. Instead, the variation between them was presented in terms of slightly different proportions of certain character states (Table 1 in Schmidtler et al., 1994). In this paper, we did not aim to re-analyze lizard morphology. However, according to the maternal and paternal phylogeny of two species, our analysis of thousands of autosomal loci failed to detect any geographic structure in the parthenogenetic lizards from the Lake Van area, neither at the level of nominal species *D. bendimahiensis, D. sapphirina* nor among the four locations sampled (Fig. 4.7 and 4.9).

The lack of population genetic structure within *D.sapphirina*/*D. bendimahiensis* based on the autosomes are entirely consistent with the previous analysis based on chromosome Z-linked markers (Yanchukov et al., 2022). However, these ddRAD-seq results are in contrast with the previous findings by Tarkhnishvili et al. (2020), who found consistent differences among the four geographic parthenogenetic populations based on microsatellites and mtDNA (Tarkhnishvili et al., 2020).

Figure 4.14 presents a NJ tree of n=29 *D. sapphirina* and *D. bendimahiensis* individuals (Tarkhnishvili et al., 2020), including some individuals also used in this study. According to that, all samples are coming from one clade but *D. bendimahiensis* appears to be nested within *D. sapphirina*, and each of the four geographic populations forms its own subclade. The clading among four populations of *D. sapphirina* and *D. bendimahiensis* can be the result of differentiation within each population via genetic drift. In addition, a small proportion of the ddRAD loci fixed for alternative alleles in the respective locations (resulting in per-allele Fst = 1, Fig. 4.11) is also consistent with random genetic drift expected in the population composed of the isolated clonal lineages. However, this might be explained by missing alleles in the data due to the low coverage as well. And the mean Fst values (0.03) between both *D. sapphirina* and *D. bendimahiensis* as two nominal species and all pairs of four populations (0.05, 0.06, 0.05, 0.05, 0.04, 0.06, see Table 4.2 for details) are very low. Moreover, we have a significant permutation test result (p = 0.033) for Fst between the two species, and some permutation results for the Fst values among pairs of four populations are all marginally significant. And these may cast doubt on the accuracy of the Fst results.

Fst is calculated based on the observed frequencies of alleles in a population. Low coverage of data may cause some heterozygous loci to appear homozygous. And this may not accurately reflect the true frequencies of alleles in the population which can lead to an incorrect estimate of Fst. And this noise may affect the calculation of the phylogenies as well. Therefore, all of these results with no differentiation in our phylogenies may be caused by the noise of the data due to the low coverage.

Figure 4.14: NJ tree of *D. bendimahiensis* and *D. sapphirina* based on microsatellites (modified from Tarkhnishvili et al., 2020).

In conclusion, the original species description by Schmidtler et al. (1994) does not provide conclusive evidence that the two nominal species *D. bendimahiensis, D. sapphirina* are different morphologically. And all available genetic evidence suggests that *D. bendimahiensis, D. sapphirina* should be viewed as a single taxon according to the 'phylogenetic species' concept (Cracraft, 1983). To correct the existing nomenclature, and follow the order in which the two taxa were described in Schmidtler et al. (1994), we suggest that the former name *D. bendimahiensis* is used only as a synonym of *D. sapphirina*. Nonetheless, when we consider the all results together with the weak correlations of the putative 3rd haplotypes in the 3-allelic loci (see Table 4.5) and the results of the test for the paralogs, we may need to re-analyze the data using coverage filtering.

4.4.3 Multiple Parental Haplotypes Contributed To the Current Genetic Pool of *D. bendimahiensis* and *D. sapphirina*

Because the parthenogenetic *Darveskia* lizards reproduce clonally, a parthenogenetic *Darevskia* population can acquire genetic diversity via mutation or multiple hybrid origins (Murphy et al., 2000; Tarkhnishvili et al., 2020). Thus if there are three haplotypes at 16,096 loci, this may be a sign of two or more hybridizations. Moreover, identifying 310 loci ($\sim$2% of all 3-allelic loci) involving three alleles having >1 nucleotide distance from each other appears to further reduce the possibility of

homoplastic mutations (although the low proportion of this category leaves room for some doubt). Furthermore, finding significant deviations from 0.5 in the proportion of maternal and paternal alleles in the binomial test implies that these third haplotypes might not be the result of random mutation or technical error. Also of 3-allelic loci alleles, $\sim$25% are shared and their uniqueness between parental and hybrids species may indicate these alleles may come from hybridization. Finally, finding positive correlations between 3rd allele frequencies between parental and hybrid species implies that these shared haplotypes may not be due to parallel mutations in hybrid and parental species.

Beyond these observations, which were consistent with multiple hybridization events, we also tested the hypothesis that 3rd loci could represent hidden paralogs. However, the mean and median depths of 2-allelic loci and 3-allelic loci in the data were highly similar, arguing against the major contribution of paralogs to the 3rd allele pool. Furthermore, we cannot see any structure in our phylogenies (see Figures 4.7 and 4.9) as a result of different hybrid origins. And the close ratio to 0.5 in the binomial test result and low correlation values of the correlations of 3-allelic loci alleles used in the binomial test (see 2nd. and 3rd. columns in Table 4.5) do not strongly support the multiple hybridization hypothesis. And this is compatible with the one hybridization result in Tarkhnishvili et al., 2020, though their results do not completely rule out the participation of multiple parental individuals.

In conclusion, our results do not support the null hypothesis of a single hybridization event between two parental individuals from *D. raddei* and *D. valentini*. However, generally, the low coverage of the data might affect the results in a negative manner. It may prevent the healthy evaluation of the analyses such as the test performed for paralogs. Thus in order to claim a more precise result, we need to re-analyze the data using coverage filtering or we may need more analyses with data having high coverage that can make the noise lower.

CHAPTER 5

CONCLUSIONS

Two hypotheses to explain clonal reproduction in vertebrates are the Phylogenetic Constraint Hypothesis (Darevsky, Kupriyanova and Uzzell, 1985) and the Balance Hypothesis (Moritz et al., 1989). In contrast to the former hypothesis, which posits that specific genetic factors in the two parental species play a central role in the creation of a parthenogenetic species, the latter hypothesis focuses on the genetic distance between the two parental species as a key factor in the formation of a parthenogenetic species.

According to the results of the parental population analysis of Chapter 3 (Figure 3.5) and Tarkhnishvili et al. (2020), parthenogenetic species are constrained phylogenetically in terms of their parental bisexual species. Not every two bisexual *Darevskia* species pair can form a parthenogenetic species. Because some specific genetic factors coming from different parental species may be important for creating a parthenogenetic species, this is consistent with the Phylogenetic Constraint Hypothesis. In addition, according to the phylogenies in Figure 3.5 in Chapter 3, and Figures 4.7 and 4.9 in Chapter 4, different subpopulations of parental species may generate different parthenogenetic species. For instance, *D. valentini* from Çaldıran and *D. raddei* from Doğubayazıt are probably parents of *D. bendimahiensis* and *D. sapphirina*, *D. portschinskii* from Armenia (Yanchukov et al., 2022) and *D. raddei raddei* from Armenia (Tarkhnishvili et al., 2020) are most likely progenitors of *D. rostombekovi*. Even though these subpopulations are from the same species, they may form different parthenogenetic species, possibly due to the different genetic factors they have. Therefore, these results may support the Phylogenetic Constraint Hypothesis.

We have two scenarios about the positive correlation between geographic and ge-

netic distances of parental and hybrid species in phylogenetic analysis and fineRADstructure. In scenario (I), the hybridizations occurred ∼1.57 million years ago (95% confidence interval 1.55-1.59), which corresponds to ∼628,000 (620,000-636,000) generations. We note that this estimate is highly dependent on the calibration points used and is subject to broad uncertainty. Nevertheless, if true, it would mean that these populations have not strayed far from each other through this wide time span (Yanchukov et al., 2022). In scenario (II), the hybridizations are comparatively recent events (Freitas et al., 2016). Either scenario may be true. However, our observation that parents and hybrids are frequently in proximity may be seen to support scenario (II).

Another disputed question is the number of hybridization events in the history of *D. bendimahiensis* and *D. sapprhirina*. Although our results support the multiple hybridizations, we could not reach a clear conclusion because the ML phylogeny and fineRADstructure did not produce any visible clusters of parthenogenetic individuals, which would be expected if some individuals were derived from different hybridization events.

This may have two explanations. Either there is only a single hybridization event in the history of *D. bendimahiensis, D. sapphirina*. Or alternatively, there happened multiple hybridizations, but the low depth of the data created too much noise to obscure phylogenetic signals. Repeating these analyses with depth-filtered data may help resolve this issue. In addition, backward/forward simulations and niche modeling may also be performed in the next studies with available data for these processes. However, we did not implement this method in this study.

In conclusion, there are numerous questions about the evolution of parthenogenetic *Darevskia* species. This study contributed to the field by addressing only a few of them. And we need further study of these organisms to better understand the nature of asexual vertebrates and how these species evolved.

REFERENCES

Alexander, D. H., Novembre, J., and Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, **19**, 1655–1664.

Ananjeva, N., L., O., G., K., S., D., A., R., and Barabanov, A. (2006). The Reptiles of Northern Eurasia. Taxonomic Diversity, Distribution, Conservation Status. *Pensoft Series Faunistica No 47*.

Andrade, P., Pinho, C., Pérez i de Lanuza, G., Afonso, S., Brejcha, J., Rubin, C.-J., Wallerman, O., Pereira, P., Sabatino, S. J., Bellati, A., Pellitteri-Rosa, D., Bosakova, Z., Bunikis, I., Carretero, M. A., Feiner, N., Marsik, P., Paupério, F., Salvi, D., Soler, L., While, G. M., Uller, T., Font, E., Andersson, L., and Carneiro, M. (2019). Regulatory changes in pterin and carotenoid genes underlie balanced color polymorphisms in the wall lizard. *Proceedings of the National Academy of Sciences*, **116**(12), 5633–5642. Publisher: Proceedings of the National Academy of Sciences.

Arakelyan, M., Danielyan, F., and Stepanyan, I. (2008). Hybrids of *Darevskia valentini*, *D. armeniaca* and *D. unisexualis* from a sympatric population in Armenia. *Amphibia-Reptilia*, **29**(4), 487–504. Publisher: Brill.

Arakelyan, M., Petrosyan, R., Ilgaz, , Kumlutaş, Y., Durmuş, S. H., Tayhan, Y., and Danielyan, F. (2013). A skeletochronological study of parthenogenetic lizards of genus *Darevskia* from Turkey. *Acta Herpetologica*, **8**(2), 99–104. Number: 2.

Arnold, E., Arribas, O., and Carranza, S. (2019). ARNOLD, E. N.; ARRIBAS, O. & CARRANZA, O. (2007): Systematics of the Palaearctic and Oriental tribe Lacertini (Squamata: Lacertidae: Lacertinae), with description of eight new genera. Zootaxa 1430. 86 pp. [ISSN: 1175-5326 print edition] [ISSN: 1175-5334 online edition].

Arnold, E. N., Arribas, O., and Carranza, S. (2007). Systematics of the palaearctic and oriental lizard tribe lacertini (Squamata: Lacertidae: Lacertinae), with descriptions of eight new genera. *Zootaxa*, **1430**(1), 1–86. Number: 1.

Arribas, O., Ilgaz, , Kumlutaş, Y., Durmuş, S. H., Avci, A., and Üzüm, N. (2013). External morphology and osteology of *Darevskia rudis* (Bedriaga, 1886), with a taxonomic revision of the Pontic and Small-Caucasus populations (Squamata: Lacertidae). *Zootaxa*, **3626**, 401–428.

Avise, J. C. (2004). *The hope, hype reality of genetic engineering: remarkable stories from agriculture, industry, medicine, and the environment*. Oxford University Press, Oxford.

Avise, J. C. (2015). Evolutionary perspectives on clonal reproduction in vertebrate animals. *Proceedings of the National Academy of Sciences*, **112**(29), 8867–8873. Publisher: Proceedings of the National Academy of Sciences.

Axelsson, E., Smith, N. G. C., Sundström, H., Berlin, S., and Ellegren, H. (2004). Male-biased mutation rate and divergence in autosomal, Z-linked and W-linked introns of chicken and turkey. *Molecular Biology and Evolution*, **21**(8), 1538–1547.

Babik, W., Branicki, W., Crnobrnja-Isailović, J., Cogălniceanu, D., Sas, I., Olgun, K., Poyarkov, N. A., Garcia-París, M., and Arntzen, J. W. (2005). Phylogeography of two European newt species–discordance between mtDNA and morphology. *Molecular Ecology*, **14**(8), 2475–2491.

Badaeva, T. N., Malysheva, D. N., Korchagin, V. I., and Ryskov, A. P. (2008). Genetic Variation and de novo Mutations in the Parthenogenetic Caucasian Rock Lizard *Darevskia unisexualis*. *PLOS ONE*, **3**(7), e2730. Publisher: Public Library of Science.

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**(10), e3376.

Barley, A. J., Cordes, J. E., Walker, J. M., and Thomson, R. C. (2022). Genetic diversity and the origins of parthenogenesis in the teiid lizard *Aspidoscelis laredoensis*. *Molecular Ecology*, **31**(1), 266–278.

Beavan, A. J. S., Pisani, D., and Donoghue, P. C. J. (2021). Diversification dynamics of total-, stem-, and crown-groups are compatible with molecular clock estimates of divergence times. *Science Advances*, **7**(24), eabf2257. Publisher: American Association for the Advancement of Science.

Bengtsson, B. (2009). Asex and evolution: a very large-scale overview. In I. Schön, K. Martens, and P. Dijk, editors, *Lost sex*, pages 1–19. Springer.

Booth, W. and Schuett, G. (2016). The emerging phylogenetic pattern of parthenogenesis in snakes. *Biological Journal of the Linnean Society*, **118**, 172–186.

Booth, W., Johnson, D. H., Moore, S., Schal, C., and Vargo, E. L. (2011). Evidence for viable, non-clonal but fatherless Boa constrictors. *Biology Letters*, **7**(2), 253–256.

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 Computational Biology*, **10**(4), e1003537. Publisher: Public Library of Science.

Braconnot, P., Otto-Bliesner, B., Harrison, S., Joussaume, S., Peterchmitt, J.-Y., Abe-Ouchi, A., Crucifix, M., Driesschaert, E., Fichefet, T., Hewitt, C. D., Kageyama, M., Kitoh, A., Laîné, A., Loutre, M.-F., Marti, O., Merkel, U., Ramstein, G., Valdes, P., Weber, S. L., Yu, Y., and Zhao, Y. (2007). Results of PMIP2 coupled simulations of the Mid-Holocene and Last Glacial Maximum – Part 1: experiments and large-scale features. *Climate of the Past*, **3**(2), 261–277. Publisher: Copernicus GmbH.

Candan, K., Kornilios, P., Ayaz, D., Kumlutaş, Y., GüL, S., Caynak, E. Y., and Ilgaz, (2021). Cryptic genetic structure within Valentin's Lizard, *Darevskia valentini*

(Boettger, 1892) (Squamata, Lacertidae), with implications for systematics and origins of parthenogenesis. *Systematics and Biodiversity*, **19**(7), 665–681. Publisher: Taylor & Francis eprint: https://doi.org/10.1080/14772000.2021.1909171.

Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A., and Cresko, W. A. (2013). Stacks: an analysis tool set for population genomics. *Molecular ecology*, **22**(11), 3124–3140.

Chapman, D., Shivji, M., Louis, E., Sommer, J., Fletcher, H., and Prodöhl, P. (2007). Virgin birth in a hammerhead shark. *Biological Letters*, **3**, 425–427.

Cheng, J., Karambelkar, B., and Xie, Y. (2022). *leaflet: Create Interactive Web Maps with the JavaScript 'Leaflet' Library*. R package version 2.1.1.

Ciobanu, D. G., Grechko, V. V., and Darevsky, I. S. (2003). Molecular Evolution of Satellite DNA CLsat in Lizards from the Genus *Darevskia* (Sauria: Lacertidae): Correlation with Species Diversity. *Russian Journal of Genetics*, **39**(11), 1292–1305.

Clark, P. U., Archer, D., Pollard, D., Blum, J. D., Rial, J. A., Brovkin, V., Mix, A. C., Pisias, N. G., and Roy, M. (2006). The middle Pleistocene transition: characteristics, mechanisms, and implications for long-term changes in atmospheric pCO2. *Quaternary Science Reviews*, **25**(23), 3150–3184.

Cracraft, J. (1983). Species Concepts and Speciation Analysis. In R. F. Johnston, editor, *Current Ornithology*, Current Ornithology, pages 159–187. Springer US, New York, NY.

Darevskij, I. S. and Danielyan, F. D. (1977). *Lacerta uzzelli* sp. nov. (Sauria, Lacertidae) a new parthenogenetic species of rock lizard from eastern Turkey. *Trudy zool Inst Leningr*, **74**, 55–59.

Darevsky, I. (1957). Systematics and ecology of rock lizards (*Lacerta saxicola* Eversmann) in Armenia. *Zool Sb AN Armenia SSR*, **10**, 27–57.

Darevsky, I. (1958). Natural parthenogenesis in central subspecies of *Lacerta saxicola* Eversmann. *Doklady Akademii Nauk SSSR*, **122**, 730–732.

Darevsky, I. (1966). Natural parthenogenesis in a polymorphic group of Caucasian Rock Lizards related to *Lacerta saxicola* Eversmann. *Journal of the Ohio Herpetological Society*, **5**, 115–152.

Darevsky, I. (1967). "Skal'nye yashcheritsy Kavkaza", Nauka, Leningrad. (Translated as: Rock lizards of the Caucasus. New Delhi: Indian National Scientific Documentation Centre, 1978).

Darevsky, I., Kupriyanova, L., and Uzzell, T. (1985). Parthenogenesis in reptiles. ed: Gans C, Billet F. volume 15, pages 411–526. Chicago: University of Chicago Press.

Darevsky, I. S. and Danielyan, F. D. (1968). Diploid and triploid progeny arising from natural mating of parthenogenetic *Lacerta armeniaca* and *L. unisexualis* with bisexual *L. saxicola valentini*. *Journal of Herpetology*, **2**(3/4), 65–69.

Dedukh, D., Majtánová, Z., Marta, A., Pšenička, M., Kotusz, J., Klíma, J., Juchno, D., Boron, A., and Janko, K. (2020). Parthenogenesis as a Solution to Hybrid Sterility: The Mechanistic Basis of Meiotic Distortions in Clonal and Sterile Hybrids. *Genetics*, **215**(4), 975–987.

Eiselt, J., Schmidtler, J. F., and Darevsky, I. S. (1993). Untersuchungen an felseidechsen (*Lacerta saxicola*-komplex) in der östlichen türkei. 2. eine neue unterart der *Lacerta raddei* boettger, 1892 (squamata: Sauria: Lacertidae). *Herpetozoa*, **6**(1/2), 65–70.

Falush, D., Stephens, M., and Pritchard, J. K. (2003). Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics*, **164**, 1567–1587.

Feldheim, K. A., Chapman, D. D., Sweet, D., Fitzpatrick, S., Prodöhl, P. A., Shivji, M. S., and Snowden, B. (2010). Shark Virgin Birth Produces Multiple, Viable Offspring. *Journal of Heredity*, **101**(3), 374–377.

Felsenstein, J. (2003). *Inferring phylogenies*. Sinauer Associates.

Felsenstein, J. (2021). Phylip, phylip general information. `https://evolution.genetics.washington.edu/phylip.html`. Accessed: December 3, 2022.

Freitas, S., Rocha, S., Campos, J., Ahmadzadeh, F., Corti, C., Sillero, N., Ilgaz, , Kumlutaş, Y., Arakelyan, M., Harris, D. J., and Carretero, M. A. (2016). Parthenogenesis through the ice ages: A biogeographic analysis of Caucasian rock lizards (genus *Darevskia*). *Molecular Phylogenetics and Evolution*, **102**, 117–127.

Freitas, S. N., Harris, D. J., Sillero, N., Arakelyan, M., Butlin, R. K., and Carretero, M. A. (2019). The role of hybridisation in the origin and evolutionary persistence of vertebrate parthenogens: a case study of *Darevskia* lizards. *Heredity*, **123**(6), 795–808.

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**(2), 107–112.

Fu, J., Murphy, R. W., and Darevsky, I. S. (2000b). Divergence of the Cytochrome b Gene in the *Lacerta raddei* Complex and Its Parthenogenetic Daughter Species: Evidence for Recent Multiple Origins. *Copeia*, **2000**(2), 432–440. Publisher: The American Society of Ichthyologists and Herpetologists.

Fujita, M. K., Singhal, S., Brunes, T. O., and Maldonado, J. A. (2020). Evolutionary Dynamics and Consequences of Parthenogenesis in Vertebrates. *Annual Review of Ecology, Evolution, and Systematics*, **51**(1), 191–214.

Galoyan, E., Bolshakova, A., Abrahamyan, M., Petrosyan, R., and Komarova, V. (2019a). Natural history of Valentin's rock lizard (*Darevskia valentini*) in Armenia. *Zoological Research*, **40**(4), 277–292.

Galoyan, E. A., Tsellarius, E. Y., and Arakelyan, M. S. (2019b). Friend-or-foe? Behavioural evidence suggests interspecific discrimination leading to low probability of hybridization in two coexisting rock lizard species (Lacertidae, *Darevskia*). *Behavioral Ecology and Sociobiology*, **73**(4), 46.

Garcia-Porta, J., Irisarri, I., Kirchner, M., Rodríguez, A., Kirchhof, S., Brown, J. L., MacLeod, A., Turner, A. P., Ahmadzadeh, F., Albaladejo, G., Crnobrnja-Isailovic, J., De la Riva, I., Fawzi, A., Galán, P., Göçmen, B., Harris, D. J., Jiménez-Robles, O., Joger, U., Jovanović Glavaš, O., Karış, M., Koziel, G., Künzel, S., Lyra, M., Miles, D., Nogales, M., Oğuz, M. A., Pafilis, P., Rancilhac, L., Rodríguez, N., Rodríguez Concepción, B., Sanchez, E., Salvi, D., Slimani, T., S'khifa, A., Qashqaei, A. T., Žagar, A., Lemmon, A., Moriarty Lemmon, E., Carretero, M. A., Carranza, S., Philippe, H., Sinervo, B., Müller, J., Vences, M., and Wollenberg Valero, K. C. (2019). Environmental temperatures shape thermal physiology as well as diversification and genome-wide substitution rates in lizards. *Nature Communications*, **10**(1), 4077.

Gibbard, P. L., Head, M. J., Walker, M. J. C., and Stratigraphy, t. S. o. Q. (2010). Formal ratification of the Quaternary System/Period and the Pleistocene Series/Epoch with a base at 2.58 Ma. *Journal of Quaternary Science*, **25**(2), 96–102. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/jqs.1338.

Gibson, A. K., Delph, L. F., and Lively, C. M. (2017). The two-fold cost of sex: experimental evidence from a natural system. *Evolution letters*, **1**(1), 6–15.

Girnyk, A. E., Vergun, A. A., Semyenova, S. K., Guliaev, A. S., Arakelyan, M. S., Danielyan, F. D., Martirosyan, I. A., Murphy, R. W., and Ryskov, A. P. (2018). Multiple interspecific hybridization and microsatellite mutations provide clonal diversity in the parthenogenetic rock lizard *Darevskia armeniaca*. *BMC Genomics*, **19**(1), 979.

Goudet, J. (2005). Hierfstat, a package for R to compute and test hierarchical F-statistics. *Molecular Ecology Notes*, **5**(1), 184–186. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1471-8286.2004.00828.x.

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**(5), 764–775.

Grismer, J. L., Bauer, A. M., Grismer, L. L., Thirakhupt, K., Aowphol, A., Oaks, J. R., Wood, Jr, P. L., Onn, C. K., Thy, N., Cota, M., and Jackman, T. (2014). Multiple origins of parthenogenesis, and a revised species phylogeny for the Southeast Asian butterfly lizards, *Leiolepis*. *Biological Journal of the Linnean Society*, **113**(4), 1080–1093.

Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O. (2010). New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. *Systematic Biology*, **59**(3), 307–321.

Hales, D. F., Wilson, A. C. C., Sloane, M. A., Simon, J.-C., Legallic, J.-F., and Sunnucks, P. (2002). Lack of detectable genetic recombination on the X chromosome

during the parthenogenetic production of female and male aphids. *Genetics Research*, **79**(3), 203–209. Publisher: Cambridge University Press.

Hassanin, A. (2015). The role of Pleistocene glaciations in shaping the evolution of polar and brown bears. Evidence from a critical review of mitochondrial and nuclear genome analyses. *Comptes Rendus Biologies*, **338**(7), 494–501.

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**(15), 1965–1978. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/joc.1276.

Ho, S. Y. W., editor (2020). *The Molecular Evolutionary Clock: Theory and Practice*. Springer International Publishing, Cham.

Hohenlohe, P. A., Bassham, S., Etter, P. D., Stiffler, N., Johnson, E. A., and Cresko, W. A. (2010). Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. *PLoS Genet*, **6**(2), e1000862.

Holder, M. and Lewis, P. (2003). Phylogeny estimation: traditional and Bayesian approaches. *Nat Rev Genet*, **4**, 275–284.

Hörandl, E. (2009). *Geographical Parthenogenesis: Opportunities for Asexuality*, pages 161–186. Springer Netherlands, Dordrecht.

Hörandl, E. (2009). A combinational theory for maintenance of sex. *Heredity*, **103**(6), 445–457.

Ivanov, V., Marusik, Y., Pétillon, J., *et al.* (2021). Relevance of ddRADseq method for species and population delimitation of closely related and widely distributed wolf spiders (Araneae, Lycosidae). *Sci Rep*, **11**, 2177.

Jombart, T. (2008). Adegenet: A R package for the multivariate analysis of genetic markers. *Bioinformatics (Oxford, England)*, **24**(11), 1403–1405.

Kearney, M., Fujita, M., and Ridenour, J. (2009). Lost sex in the reptiles: Constraints and correlations. In I. Schön, K. Martens, and P. Dijk, editors, *Lost Sex*. Springer.

Kirschner, P., Arthofer, W., Pfeifenberger, S., *et al.* (2021). Performance comparison of two reduced-representation based genome-wide marker-discovery strategies in a multi-taxon phylogeographic framework. *Sci Rep*, **11**, 3978.

Kondrashov, A. S. (1988). Deleterious mutations and the evolution of sexual reproduction. *Nature*, **336**(6198), 435–440. Number: 6198 Publisher: Nature Publishing Group.

Korchagin, V. I., Badaeva, T. N., Tokarskaya, O. N., Martirosyan, I. A., Darevsky, I. S., and Ryskov, A. P. (2007). Molecular characterization of allelic variants of (GATA)n microsatellite loci in parthenogenetic lizards *Darevskia unisexualis* (Lacertidae). *Gene*, **392**(1), 126–133.

Kočí, J., Röslein, J., Pačes, J., Kotusz, J., Halačka, K., Koščo, J., Fedorčák, J., Iakovenko, N., and Janko, K. (2020). No evidence for accumulation of deleterious mutations and fitness degradation in clonal fish hybrids: Abandoning sex without regrets. *Molecular Ecology*, **29**(16), 3038–3055. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/mec.15539.

Kumar, S., Stecher, G., Suleski, M., and Hedges, S. B. (2017). TimeTree: A Resource for Timelines, Timetrees, and Divergence Times. *Molecular Biology and Evolution*, **34**(7), 1812–1819.

Kumar, S., Stecher, G., Li, M., Knyaz, C., and Tamura, K. (2018). MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Molecular Biology and Evolution*, **35**(6), 1547–1549.

Kupriyanova, L. (2009). Cytogenetic and genetic trends in the evolution of unisexual lizards. *Cytogenetic and Genome Research*, **127**(2-4), 273–279.

Lampert, K. P. (2008). Facultative parthenogenesis in vertebrates: reproductive error or chance? *Sexual Development: Genetics, Molecular Biology, Evolution, Endocrinology, Embryology, and Pathology of Sex Determination and Differentiation*, **2**(6), 290–301.

Langley, C. H. and Fitch, W. M. (1974). An examination of the constancy of the rate of molecular evolution. *Journal of Molecular Evolution*, **3**(3), 161–177.

Langmead, B. and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, **9**(4), 357–359.

Laskowski, K. L., Doran, C., Bierbach, D., Krause, J., and Wolf, M. (2019). Naturally clonal vertebrates are an untapped resource in ecology and evolution research. *Nature Ecology & Evolution*, **3**(2), 161–169. Number: 2 Publisher: Nature Publishing Group.

Lemoine, F., Domelevo Entfellner, J.-B., Wilkinson, E., Correia, D., Dávila Felipe, M., De Oliveira, T., and Gascuel, O. (2018). Renewing Felsenstein's phylogenetic bootstrap in the era of big data. *Nature*, **556**(7702), 452–456. Number: 7702 Publisher: Nature Publishing Group.

Letunic, I. and Bork, P. (2021). Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Research*, **49**(W1), W293–W296.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics (Oxford, England)*, **25**(16), 2078–2079.

Lowe, C. H. and Wright, J. W. (1966). Evolution of Parthenogenetic Species of Cnemidophorus (Whiptail Lizards) in Western North America. *Journal of the Arizona Academy of Science*, **4**(2), 81–87. Publisher: Arizona-Nevada Academy of Science.

Lowry, D. B., Hoban, S., Kelley, J. L., Lotterhos, K. E., Reed, L. K., Antolin, M. F., and Storfer, A. (2017). Breaking RAD: an evaluation of the utility of restriction site-associated DNA sequencing for genome scans of adaptation. *Mol Ecol Resour*, **17**, 142–152.

Lutes, A. A., Neaves, W. B., Baumann, D. P., Wiegraebe, W., and Baumann, P. (2010). Sister chromosome pairing maintains heterozygosity in parthenogenetic lizards. *Nature*, **464**(7286), 283–286.

Lutes, A. A., Baumann, D. P., Neaves, W. B., and Baumann, P. (2011). Laboratory synthesis of an independently reproducing vertebrate species. *Proceedings of the National Academy of Sciences*, **108**(24), 9910–9915. Publisher: Proceedings of the National Academy of Sciences.

Macculloch, R. D., Murphy, R. W., Kupriyanova, L. A., Darevsky, I. S., and Danielyan, F. D. (1995). Clonal variation in the parthenogenetic rock lizard *Lacerta armeniaca*. *Genome*, **38**(6), 1057–1060.

Macholán, M., Baird, S. J. E., Dufková, P., Munclinger, P., Bímová, B. V., and Piálek, J. (2011). Assessing multilocus introgression patterns: a case study on the mouse X chromosome in central Europe. *Evolution; International Journal of Organic Evolution*, **65**(5), 1428–1446.

Maier, A.-R.-M., Cupșa, D., Ferenți, S., and Cadar, A.-M. (2022). New records of *Darevskia praticola* at the northern limit of its distribution range in Romania. *Herpetozoa*, **35**, 45–50.

Malinsky, M., Trucchi, E., Lawson, D. J., and Falush, D. (2018). RADpainter and fineRADstructure: Population Inference from RADseq Data. *Molecular Biology and Evolution*, **35**(5), 1284–1290.

Mank, J. E., Axelsson, E., and Ellegren, H. (2007). Fast-X on the Z: rapid evolution of sex-linked genes in birds. *Genome Research*, **17**(5), 618–624.

Mank, J. E., Nam, K., and Ellegren, H. (2010). Faster-Z evolution is predominantly due to genetic drift. *Molecular Biology and Evolution*, **27**(3), 661–670.

Maroja, L. S., Larson, E. L., Bogdanowicz, S. M., and Harrison, R. G. (2015). Genes with Restricted Introgression in a Field Cricket (Gryllus firmus/Gryllus pennsylvanicus) Hybrid Zone Are Concentrated on the X Chromosome and a Single Autosome. *G3 (Bethesda, Md.)*, **5**(11), 2219–2227.

Martin, M. and Mendelson, T. (2018). Hybrid sterility increases with genetic distance in snubnose darters (Percidae: Etheostoma). *Environ Biol Fish*, **101**, 215–221.

Martirosyan, I. A., Ryskov, A. P., Petrosyan, V. G., Arakelyan, M. S., Aslanyan, A. V., Danielyan, F. D., Darevsky, I. S., and Tokarskaya, O. N. (2002). Variation of Mini- and Microsatellite DNA Markers in Populations of Parthenogenetic Rock Lizard *Darevskia rostombekovi*. *Russian Journal of Genetics*, **38**(6), 691–698.

Maynard Smith, J. (1971). *The origin and maintenance of sex*, pages 163–175. Aldine-Atherton, Chicago.

Meirmans, S., Meirmans, P., and Kirkendall, L. (2012). The Costs Of Sex: Facing Real-world Complexities. *The Quarterly review of biology*, **87**, 19–40.

Miller, M. A., Pfeiffer, W., and Schwartz, T. (2010). Creating the CIPRES Science Gateway for inference of large phylogenetic trees. In *2010 Gateway Computing Environments Workshop (GCE)*, pages 1–8. ISSN: 2152-1093.

Miller, M. R., Dunham, J. P., Amores, A., Cresko, W. A., and Johnson, E. A. (2007). Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers. *Genome Res*, **17**, 240–248.

Moreira, M. O., Fonseca, C., and Rojas, D. (2021). Parthenogenesis is self-destructive for scaled reptiles. *Biology Letters*, **17**(5), 20210006.

Moritz, C., Donnellan, S., Adams, M., and Baverstock, P. R. (1989). The origin and evolution of parthenogenesis in *Heteronotia binoei* (Gekkonidae): Extensive genotypic diversity among parthenogens. *Evolution; International Journal of Organic Evolution*, **43**(5), 994–1003.

Moritz, C., Uzzell, T., Spolsky, C., Hotz, H., Darevsky, I., Kupriyanova, L., and Danielyan, F. (1992). The material ancestry and approximate age of parthenogenetic species of Caucasian rock lizards (*Lacerta*: Lacertidae). *Genetica*, **87**(1), 53–62.

Muller, H. J. (1932). Some Genetic Aspects of Sex. *The American Naturalist*, **66**(703), 118–138. Publisher: The University of Chicago Press.

Murphy, R., MacCulloch, R., Fu, J., and Darevsky, I. (2000a). Genetic evidence for species status of some Caucasian rock lizards in the *Darevskia saxicola* group. *Amphibia-Reptilia*, **21**(2), 169–176. Publisher: Brill.

Murphy, R. W. (1999). The correct spelling of the Latinized name for Rostombekov's rock lizard. *Amphibia-Reptilia*, **20**(2), 225–226.

Murphy, R. W., Darevsky, I. S., MacCulloch, R. D., Fu, J., Kupriyanova, L. A., Upton, D. E., and Danielyan, F. (1997). Old age, multiple formations or genetic plasticity? Clonal diversity in the uniparental Caucasian rock lizard, *Lacerta dahli*. *Genetica*, **101**(2), 125–130.

Murphy, R. W., Fu, J., Macculloch, R. D., Darevsky, I. S., and Kupriyanova, L. A. (2000b). A fine line between sex and unisexuality: the phylogenetic constraints on parthenogenesis in lacertid lizards. *Zoological Journal of the Linnean Society*, **130**(4), 527–549.

Murtskhvaladze, M., Tarkhnishvili, D., Anderson, C. L., and Kotorashvili, A. (2020). Phylogeny of caucasian rock lizards (*Darevskia*) and other true lizards based on mitogenome analysis: Optimisation of the algorithms and gene selection. *PloS One*, **15**(6), e0233680.

Neaves, W. B. and Baumann, P. (2011). Unisexual reproduction among vertebrates. *Trends in genetics: TIG*, **27**(3), 81–88.

Ochkalova, S., Korchagin, V., Vergun, A., Urin, A., Zilov, D., Ryakhovsky, S., Girnyk, A., Martirosyan, I., Zhernakova, D. V., Arakelyan, M., Danielyan, F., Kliver, S., Brukhin, V., Komissarov, A., and Ryskov, A. (2022). First genome of rock lizard *Darevskia valentini* involved in formation of several parthenogenetic species. *Genes*, **13**(9), 1569.

Parker, E. D. and Selander, R. K. (1976). The organization of genetic diversity in the parthenogenetic lizard *Cnemidophorus tesselatus*. *Genetics*, **84**(4), 791–805.

Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S., and Hoekstra, H. E. (2012). Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLoS One*, **7**(5), e37135.

Petrosyan, V., Osipov, F., Bobrov, V., Dergunova, N., Kropachev, I., Danielyan, F., and Arakelyan, M. (2020a). New records and geographic distribution of the sympatric zones of unisexual and bisexual rock lizards of the genus *Darevskia* in Armenia and adjacent territories. *Biodiversity Data Journal*, page 56030.

Petrosyan, V., Osipov, F., Bobrov, V., Dergunova, N., Omelchenko, A., Varshavskiy, A., Danielyan, F., and Arakelyan, M. (2020b). Species Distribution Models and Niche Partitioning among Unisexual *Darevskia dahli* and Its Parental Bisexual (*D. portschinskii, D. mixta*) Rock Lizards in the Caucasus. *Mathematics*, **8**(8), 1329. Number: 8 Publisher: Multidisciplinary Digital Publishing Institute.

Presgraves, D. and Meiklejohn, C. (2021). Hybrid sterility, genetic conflict and complex speciation: Lessons from the *Drosophila simulans* clade species. *Frontiers in genetics*, **12**, 669045.

Pyron, R. A. and Burbrink, F. T. (2014). Early origin of viviparity and multiple reversions to oviparity in squamate reptiles. *Ecology Letters*, **17**(1), 13–21.

R, C. T. (2020). R: a language and environment for statistical computing.

Rambaut, A. (2010). Figtree v1.3.1.

Rastegar-Pouyani, Nasrullah, Karamiani, Rasoul, Oraei, Hamzeh, Khosrawani-AZAR, and Eskandar (2011). A new subspecies of *Darevskia raddei* Boettger, 1892 Sauria: Lacertidae from the west Azerbaijan province, 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 Herpetol. Res.*, **2**, 216–222.

Rato, C., Stratakis, M., Sousa-Guedes, D., Sillero, N., Corti, C., Freitas, S., Harris, D. J., and Carretero, M. A. (2021). The more you search, the more you find: Cryptic diversity and admixture within the Anatolian rock lizards (Squamata, *Darevskia*). *Zoologica Scripta*, **50**(2), 193–209. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/zsc.12462.

Rice, W. and Friberg, U. (1970). A Graphical Approach to Lineage Selection Between Clonals and Sexuals. In *Lost Sex: The Evolutionary Biology of Parthenogenesis*, pages 75–97. Journal Abbreviation: Lost Sex: The Evolutionary Biology of Parthenogenesis.

Rochette, N. C. and Catchen, J. M. (2017). Deriving genotypes from RAD-seq short-read data using Stacks. *Nature Protocols*, **12**(12), 2640–2659. Number: 12 Publisher: Nature Publishing Group.

Rochette, N. C., Rivera-Colón, A. G., and Catchen, J. M. (2019). Stacks 2: Analytical methods for paired-end sequencing improve RADseq-based population genomics. *Molecular Ecology*, **28**(21), 4737–4754.

Roquet, C., Lavergne, S., and Thuiller, W. (2014). One tree to link them all: a phylogenetic dataset for the European tetrapoda. *PLoS currents*, **6**, ecurrents.tol.5102670fff8aa5c918e78f5592790e48.

Roure, B., Baurain, D., and Philippe, H. (2013). Impact of missing data on phylogenies inferred from empirical phylogenomic data sets. *Molecular Biology and Evolution*, **30**(1), 197–214.

Rovatsos, M., Vukić, J., Altmanová, M., Johnson Pokorná, M., Moravec, J., and Kratochvíl, L. (2016). Conservation of sex chromosomes in lacertid lizards. *Molecular Ecology*, **25**(13), 3120–3126.

Rovatsos, M., Vukić, J., Mrugała, A., Suwala, G., Lymberakis, P., and Kratochvíl, L. (2019). Little evidence for switches to environmental sex determination and turnover of sex chromosomes in lacertid lizards. *Scientific Reports*, **9**(1), 7832.

RStudio, T. (2022). Rstudio.

Ryskov, A. P., Osipov, F. A., Omelchenko, A. V., Semyenova, S. K., Girnyk, A. E., Korchagin, V. I., Vergun, A. A., and Murphy, R. W. (2017). The origin of multiple clones in the parthenogenetic lizard species *Darevskia rostombekowi*. *PloS One*, **12**(9), e0185161.

Sayres, M. A. W. and Makova, K. D. (2011). Genome analyses substantiate male mutation bias in many species. *BioEssays : news and reviews in molecular, cellular and developmental biology*, **33**(12), 938–945.

Schmidtler, J. F., Eiselt, J., and Darevsky, I. S. (1994). *Untersuchungen an Feldeidechsen (Lacerta saxicola-Gruppe) in der östlichen Türkei: 3. Zwei neue parthenogenetische Arten*.

Schut, E., Hemmings, N., and Birkhead, T. (2008). Parthenogenesis in a passerine bird, the zebra finch *Taeniopygia guttata*. *Ibis*, **150**, 197–199.

Simon Y. W. Ho, e. (2020). *The molecular evolutionary clock : theory and practice*. Springer, Cham, Switzerland, 1st ed. 2020. edition.

Spangenberg, V., Arakelyan, M., Galoyan, E., Matveevsky, S., Petrosyan, R., Bogdanov, Y., Danielyan, F., and Kolomiets, O. (2017). Reticulate Evolution of the Rock Lizards: Meiotic Chromosome Dynamics and Spermatogenesis in Diploid and Triploid Males of the Genus *Darevskia*. *Genes*, **8**(6), E149.

Spangenberg, V., Arakelyan, M., Galoyan, E., Pankin, M., Petrosyan, R., Stepanyan, I., Grishaeva, T., Danielyan, F., and Kolomiets, O. (2019). Extraordinary centromeres: differences in the meiotic chromosomes of two rock lizards species *Darevskia portschinskii* and *Darevskia raddei*. *PeerJ*, **7**, e6360.

Spangenberg, V., Arakelyan, M., Cioffi, M. d. B., Liehr, T., Al-Rikabi, A., Martynova, E., Danielyan, F., Stepanyan, I., Galoyan, E., and Kolomiets, O. (2020a). Cytogenetic mechanisms of unisexuality in rock lizards. *Scientific Reports*, **10**(1), 8697.

Spangenberg, V., Kolomiets, O., Stepanyan, I., Galoyan, E., de Bello Cioffi, M., Martynova, E., Martirosyan, I., Grishaeva, T., Danielyan, F., Al-Rikabi, A., Liehr, T., and Arakelyan, M. (2020b). Evolution of the parthenogenetic rock lizard hybrid karyotype: Robertsonian translocation between two maternal chromosomes in *Darevskia rostombekowi*. *Chromosoma*, **129**(3), 275–283.

Steel, M. and McKenzie, A. (2001). Properties of phylogenetic trees generated by yule-type speciation models. *Mathematical biosciences*, **170**(1), 91–112.

Stenberg, P. and Saura, A. (2009a). Cytology of Asexual Animals. In I. Schön, K. Martens, and P. Dijk, editors, *Lost Sex*, pages 63–74. Springer Netherlands, Dordrecht.

Stenberg, P. and Saura, A. (2009b). Cytology of asexual animals. In I. Schön, K. Martens, and P. Dijk, editors, *Lost Sex: The Evolutionary Biology of Parthenogenesis*, pages 63–74. Springer, Dordrecht, The Netherlands.

Storchová, R., Reif, J., and Nachman, M. W. (2010). Female heterogamety and speciation: reduced introgression of the Z chromosome between two species of nightingales. *Evolution; International Journal of Organic Evolution*, **64**(2), 456–471.

Stöck, M., Dedukh, D., Reifová, R., Lamatsch, D. K., Starostová, Z., and Janko, K. (2021). Sex chromosomes in meiotic, hemiclonal, clonal and polyploid hybrid vertebrates: along the 'extended speciation continuum'. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, **376**(1833), 20200103.

Takeuchi, N., Kaneko, K., and Koonin, E. V. (2014). Horizontal gene transfer can rescue prokaryotes from Muller's ratchet: benefit of DNA from dead cells and population subdivision. *G3 (Bethesda, Md.)*, **4**(2), 325–339.

Talavera, G., Lukhtanov, V., Pierce, N. E., and Vila, R. (2022). DNA Barcodes Combined with Multilocus Data of Representative Taxa Can Generate Reliable Higher-Level Phylogenies. *Systematic Biology*, **71**(2), 382–395.

Tarkhnishvili, D. (2012). Evolutionary History, Habitats, Diversification, and Speciation in Caucasian Rock Lizards. In *Advances in Zoology Research, Owen P. Jenkins, Ed*, volume 2, pages 79–120. Nova Science Publishers.

Tarkhnishvili, D., Gavashelishvili, A., Avaliani, A., Murtskhvaladza, M., and Mumladze, L. (2010). Unisexual rock lizard might be outcompeting its bisexual progenitors in the Caucasus. *Biological Journal of the Linnean Society*, **101**(2), 447–460.

Tarkhnishvili, D., Murtskhvaladze, M., and 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**(4), 876–892.

Tarkhnishvili, D., Murtskhvaladze, M., and Anderson, C. L. (2017). Coincidence of genotypes at two loci in two parthenogenetic rock lizards: how backcrosses might trigger adaptive speciation. *Biological Journal of the Linnean Society*, **121**(2), 365–378.

Tarkhnishvili, D., Barateli, N., Murtskhvaladze, M., and Iankoshvili, G. (2020a). Estimating phenotypic heritability of sexual and unisexually reproducing rock lizards (genus *Darevskia*). *Zoologischer Anzeiger*, **285**, 105–113.

Tarkhnishvili, D., Yanchukov, A., Şahin, M. K., Gabelaia, M., Murtskhvaladze, M., Candan, K., Galoyan, E., Arakelyan, M., Iankoshvili, G., Kumlutaş, Y., Ilgaz, , Matur, F., Çolak, F., Erdolu, M., Kurdadze, S., Barateli, N., and Anderson, C. L. (2020b). Genotypic similarities among the parthenogenetic *Darevskia* rock lizards with different hybrid origins. *BMC Evolutionary Biology*, **20**(1), 122.

Tiley, G. P., Poelstra, J. W., Reis, M. d., Yang, Z., and Yoder, A. D. (2020). Molecular Clocks without Rocks: New Solutions for Old Problems. *Trends in Genetics*, **36**(11), 845–856. Publisher: Elsevier.

Tilquin, A. and Kokko, H. (2016). What does the geography of parthenogenesis teach us about sex? *Philosophical Transactions of the Royal Society B: Biological Sciences*, **371**(1706), 20150538. Publisher: Royal Society.

Truong, H. T., Ramos, A. M., Yalcin, F., Ruiter, M. d., Poel, H. J. A. v. d., Huvenaars, K. H. J., Hogers, R. C. J., Enckevort, L. J. G. v., Janssen, A., Orsouw, N. J. v., and Eijk, M. J. T. v. (2012). Sequence-Based Genotyping for Marker Discovery and Co-Dominant Scoring in Germplasm and Populations. *PLOS ONE*, **7**(5), e37565. Publisher: Public Library of Science.

Tuniyev, B. S., Tarkhnishvili, D., Aghasyan, A. L., Bunyatova, S. N., Kamali, K., Mirghazanfari, S. M., Tok, C. V., and Çiçek, K. (2020). *Amphibians and reptiles of the Caucasus*, pages 83–96. WWF, KfW, Tbilisi.

Upton, D. E., Danielyan, F. D., Bobyn, M. L., Darevsky, I. S., Murphy, R. W., Kupriyanova, L. A., and MacCulloch, R. D. (1996). Allozyme variation in populations of *Lacerta raddei* and *Lacerta nairensis* (Sauria: Lacertidae) from Armenia. *Amphibia-Reptilia*, **17**(3), 233–246.

van der Kooi Casper J. and Schwander, T. (2015). Parthenogenesis: Birth of a New Lineage or Reproductive Accident? *Current Biology*, **25**(15), R659–R661.

Vergun, A. A., Martirosyan, I. A., Semyenova, S. K., Omelchenko, A. V., Petrosyan, V. G., Lazebny, O. E., Tokarskaya, O. N., Korchagin, V. I., and Ryskov, A. P. (2014). Clonal Diversity and Clone Formation in the Parthenogenetic Caucasian Rock Lizard *Darevskia dahli*. *PLOS ONE*, **9**(3), e91674. Publisher: Public Library of Science.

Vergun, A. A., Girnyk, A. E., Korchagin, V. I., Semyenova, S. K., Arakelyan, M. S., Danielyan, F. D., Murphy, R. W., and Ryskov, A. P. (2020). Origin, clonal diversity, and evolution of the parthenogenetic lizard *Darevskia unisexualis*. *BMC Genomics*, **21**(1), 351.

Warren, W. C., García-Pérez, R., Xu, S., Lampert, K. P., Chalopin, D., Stöck, M., Loewe, L., Lu, Y., Kuderna, L., Minx, P., Montague, M. J., Tomlinson, C., Hillier, L. W., Murphy, D. N., Wang, J., Wang, Z., Garcia, C. M., Thomas, G. C. W., Volff, J.-N., Farias, F., Aken, B., Walter, R. B., Pruitt, K. D., Marques-Bonet, T., Hahn, M. W., Kneitz, S., Lynch, M., and Schartl, M. (2018). Clonal polymorphism and high heterozygosity in the celibate genome of the *Amazon molly*. *Nature Ecology & Evolution*, **2**(4), 669–679. Number: 4 Publisher: Nature Publishing Group.

Watts, P. C., Buley, K. R., Sanderson, S., Boardman, W., Ciofi, C., and Gibson, R. (2006). Parthenogenesis in Komodo dragons. *Nature*, **444**(7122), 1021–1022.

Wiens, J. J. and Morrill, M. C. (2011). Missing Data in Phylogenetic Analysis: Reconciling Results from Simulations and Empirical Data. *Systematic Biology*, **60**(5), 719–731.

Yanchukov, A., Tarkhnishvili, D., Erdolu, M., Şahin, M. K., Candan, K., Murtskhvaladze, M., Gabelaia, M., Iankoshvili, G., Barateli, N., Ilgaz, , Kumlutaş, Y., Matur, F., Çolak, F., Arakelyan, M., and Galoyan, E. (2022). Precise paternal ancestry of hybrid unisexual ZW lizards (genus *Darevskia*: Lacertidae: Squamata) revealed by Z-linked genomic markers. *Biological Journal of the Linnean Society*, **136**(2), 293–305.

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

Zieliński, P., Nadachowska-Brzyska, K., Wielstra, B., Szkotak, R., Covaciu-Marcov, S. D., Cogălniceanu, D., and Babik, W. (2013). No evidence for nuclear introgression despite complete mtDNA replacement in the Carpathian newt (*Lissotriton montandoni*). *Molecular Ecology*, **22**(7), 1884–1903. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/mec.12225.

APPENDIX A

ddRAD library preparation protocol (perfomed by Floragenex, Inc., OR. USA)

Genomic DNA was digested with the restriction endonucleases PstI and MseI and processed into RAD libraries similar to the method of (Truong et al., 2012). Briefly, 125 ng of genomic DNA was digested for 120 min at 37°C in a 20 µL reaction with 20 units (U) of PstI and 2.75 U of MseI (New England Biolabs [NEB]). After digestion, samples were heat-inactivated for 10 min at 80°C followed by addition of 2.5 µL of 1 µM P1 Adapter and 0.1 µL of 250 µM MseI adapter. PstI P1 adapters each contained a unique multiplex sequence index (barcode), which is read during the first five nucleotides of the Illumina sequence read. P1 and P2 adapters were added to each sample along with 36 units T4 DNA Ligase (high concentration [HC], Enzymatics, Inc), 0.3125 U MseI, 0.25 U PstI in a final reaction volume of 25 µL which was then incubated at 37°C for 180 minutes.

Samples were diluted 1:10 in water and 2.5 µL of this product was used in a PCR amplification with 10x PCR Buffer 1 (Applied Biosystems), 0.3 µL of 50 ng/µL MseI primer, 0.05 µL of PstI primer, and 0.2 U AmpliTaq DNA Polymerase (Applied Biosystems). The PCR program is as follows:

2 min 72°C

13 cycles of:

30 sec 94°C

2 min 67°C, decreasing -0.7°C per cycle

2 min 72°C

37 cycles of:

30 sec 94°C

2 min 58°C

2 min 72°C

Hold 4°C

The amplified product from each sample was pooled and mixed thoroughly. 125 µL of the pooled product was purified with a MinElute Enzymatic Reaction Cleanup Kit (Qiagen), eluted in 15 µL, and run on a 1.5agarose gel. DNA 300 bp to 650 bp was excised manually and purified with the MinElute Gel Extraction Kit (Qiagen).

APPENDIX B

I) MAP OF LAKE VAN SAMPLES:

Run the leaflet commands below in R (Ref: https://rstudio.github.io/leaflet/basemaps.html):

```
library(tidyverse)
library(leaflet)
PopulationsLakeVan <- c("Sapphirina-Van/Ağrı Border", "Sapphirina-Pınarlı", "Bendimahiensis-Muradiye",
"Bendimahiensis-Çaldıran", "Raddei-Muradiye", "Raddei-Doğubayazıt", "Raddei-Saray", "Raddei-Çörekli",
"Raddei-Between Umut & Yaramış", "Valentini-Çaldıran", "Valentini-Kızılyusuf")
LatitudeLakeVan <- c(39.149, 39.164, 39.077,  39.170, 39.057, 39.463, 38.545014, 38.35445588,
38.44788438, 39.147, 38.944)
LongitudeLakeVan <- c(43.096, 43.192, 43.756, 43.969, 43.756, 44.170,  44.265424, 43.83140915,
43.86374338, 44.007, 42.677)
LakeVanMapDF <- data.frame(PopulationsLakeVan, LatitudeLakeVan, LongitudeLakeVan, colors=c("darkpurple",
"darkpurple", "darkpurple", "darkpurple", "red", "red", "red", "red", "red", "darkblue", "darkblue"))
```

Make a map for all Lake Van samples with Leaflet library:

```
leaflet() %>% addTiles() %>%
  addAwesomeMarkers (LakeVanMapDF$LongitudeLakeVan, LakeVanMapDF$LatitudeLakeVan,
  icon= awesomeIcons(icon="ios-close",
  markerColor = LakeVanMapDF$colors), label=LakeVanMapDF$PopulationsLakeVan) %>%
  addProviderTiles(providers$Esri.WorldImagery) %>% addMiniMap(tiles = providers$Esri.WorldImagery,
  position="topleft",
  width=250, height=150, zoomLevelOffset = -5) %>%
  addScaleBar(position = "bottomleft", options = scaleBarOptions( maxWidth = 100,
    metric = TRUE,
    imperial = TRUE,
    updateWhenIdle = TRUE
  ))
```

Map for only Raddei samples:

```
LakeVanMapDFRaddei <- data.frame(PopulationsLakeVan, LatitudeLakeVan, LongitudeLakeVan,
colors=c("darkgreen", "darkgreen", "darkgreen", "darkgreen", "red", "red", "red", "red", "red",
"orange", "orange"))
LakeVanMapDFRaddei <- LakeVanMapDFRaddei[-10:-11,]
leaflet() %>% addTiles() %>%
    addAwesomeMarkers (LakeVanMapDFRaddei$LongitudeLakeVan, LakeVanMapDFRaddei$LatitudeLakeVan,
    icon= awesomeIcons(icon="ios-close", markerColor = LakeVanMapDFRaddei$colors),
    label=LakeVanMapDFRaddei$PopulationsLakeVan)
    %>%
    addProviderTiles(providers$Esri.WorldImagery) %>%
    addScaleBar(position = "bottomleft", options = scaleBarOptions( maxWidth = 100,
                                                                    metric = TRUE,
                                                                    imperial = TRUE,
                                                                    updateWhenIdle = TRUE
    ))
```

Map for only Valentini samples:

```
LakeVanMapDFVal <- data.frame(PopulationsLakeVan, LatitudeLakeVan, LongitudeLakeVan, colors=c("darkgreen", "darkgreen",
"darkgreen", "darkgreen", "red", "red", "red", "red", "orange", "orange"))
LakeVanMapDFVal <- LakeVanMapDFVal[-5:-9,]
leaflet() %>% addTiles() %>%
    addAwesomeMarkers (LakeVanMapDFVal$LongitudeLakeVan, LakeVanMapDFVal$LatitudeLakeVan, icon=
    awesomeIcons(icon="ios-close", markerColor = LakeVanMapDFVal$colors), label=LakeVanMapDFVal$PopulationsLakeVan) %>%
    addProviderTiles(providers$Esri.WorldImagery) %>%
    addScaleBar(position = "bottomleft", options = scaleBarOptions( maxWidth = 100,
                                                                    metric = TRUE,
                                                                    imperial = TRUE,
                                                                    updateWhenIdle = TRUE
    ))
```

II) FILTERING THE <30 MAP QUALITY READS FROM THE BAM FILES AND CALCULATION OF THE FILTERING PERCENTAGE:

1) Filter the reads having <30 map quality from the aligned BAM files using samtools:

```
for i in *bam; do samtools view -b -q 30 $i > /the_path_of_the_directory_of_the_FilteredBAM_files/$i; done
```

2) Convert both the filtered and non-filtered BAM files to the fasta format with the code:

```
for i in *.bam; do /usr/local/sw/samtools-1.11/samtools view -F 4 $i | awk '{OFS="\t"; print ">"$1"\n"$10}' - > $i.fa; done
```

3) Find the total of sequences in fasta files with the code:

```
grep -vc '>' *fa | awk -F':' '{sum+=$2;} END{print sum}'
```

4) Find the percentage using the number of total reads of non-filtered (x1) and filtered (x2) sequences: `(100 * (x1 - x2))/x1`

III) ESTIMATION OF PLOIDY LEVELS OF SPECIMENS USING CHROMO-

SOME Z READ RATIO:

1) Compute the proportion of number of reads of chromosome Z vs. number of reads of autosomes of each sample, and normalize these numbers dividing by number of reads of chromosome Z vs. number of reads of autosomes of *Podarcis muralis*. Then create the data frame below for plotting in R:

```
ChrZRatios_BendiSapp <- data.frame( ratios = c(0.601889061082883, 0.591673842012224, 0.612773807105907,
0.612723057871383, 0.618187131814745, 0.589125633299246, 0.587379470798049, 0.594206946441417,
0.589046471039942, 0.598292292426858, 0.581365459138663, 0.576927022473046, 0.574871277458628,
0.580493413918183, 0.562974258227298, 0.579983806797642, 0.580047770128053, 0.572479331087877,
0.94109030627355, 0.916747441080237, 0.455200974156836, 0.50948988666815,
0.542877521280903, 0.526596918703769, 1.07303554925458, 0.976375124332685),
samples= c("NAg1uni1", "NC1bend20", "NC1bend24", "NVAsap5", "NVAsap6", "OC1.bend2", "OC1.bend3",
"OC1.bend5", "OC1.bend7", "OP1.sap1", "OP1.sap3", "OP1.sap7", "OV3.bend7", "OV3.bend8", "OV3.bend9",
"OVa.sap1", "OVA.sap2", "OVa.sap3", "OV2-rad13","OV2-rad11","OV1-rad11","OD1-rad2",
"NV4val3","NV4val6","NV4val5","OC3-val1"), colors=c("#D95F02", "#1B9E77", "#1B9E77", "#E7298A",
"#E7298A", "#1B9E77", "#1B9E77", "#1B9E77", "#1B9E77", "#7570B3", "#7570B3", "#7570B3", "#D95F02",
"#D95F02", "#D95F02", "#E7298A", "#E7298A", "#E7298A", "#66A61E", "#66A61E", "#66A61E", "#66A61E",
"#E6AB02", "#E6AB02", "#E6AB02", "#E6AB02") )
```

2) Draw the ploidy levels plot in R:

```
par(mar=c(5,7,3,2)) # Adjust the plot margins, Below, Left, Above, Rigth in turn
plot(ChrZRatios_BendiSapp[with(ChrZRatios_BendiSapp, order(ratios)),][1:26,1],
col=ChrZRatios_BendiSapp[with(ChrZRatios_BendiSapp, order(ratios)),][1:26,3], pch=20, cex=1.9, #main="Length
of ChrZ/Length of Autosomes \n Darevskia sapphirina & D. bendimahiensis Samples",
ylab = "(Sample ChrZ Length/Sample Autosome Length)/\n(Pod. muralis ChrZ Length/Pod. muralis Autosome Length)",
xlab = "Index of Sample", xlim=c(0,26), ylim=c(0.4, 1.2), cex.lab=1.9, cex.axis=1.9 )
text(ChrZRatios_BendiSapp[with(ChrZRatios_BendiSapp, order(ratios)),][1:26,1],
ChrZRatios_BendiSapp[with(ChrZRatios_BendiSapp, order
(ratios)),][1:26,2], cex=1.7, pos=1, col="black", srt=60)
legend ("topleft", legend=c("Raddei", "Valentini", "Bendi-Çaldıran", "Bendi-Muradiye",
         "Sapp-Pınarlı", "Sapp-Van/Ağrı Border"),
        col=c( "#66A61E", "#E6AB02", "#1B9E77", "#D95F02", "#7570B3", "#E7298A"), inset=c(0,0), ncol=2,
        pch = 20, cex=1.7, bty="n", x.intersp=0.1, y.intersp = 0.35, text.width = 2.5, horiz = F, pt.cex=4)
        # x.intersp adjust the gap between
        #text and horizontal legend symbols and y.intersp does vertically
abline(h=c(0.5, 0.6666667, 0.75, 1.0 ), col=c("grey"), lty=c(1,1,1,1), lwd=c(3,3,3,3))
```

IV) DIAGNOSTIC ALLELE CODES:

1) assemble all data of samples that you need in one fasta file using Stacks/population –fasta-samples flag with the code:

```
/usr/local/sw/stacks-2.41/bin/ref_map.pl -X populations: --fasta-samples -T 80
--popmap /mnt/NEOGENE1/projects/lizard_2019/ValAligned_Rad_Both_Analysis
/stacks.FASTA.ChrAll_f_ValAli.All/popmap_LakeVanAllSamples.tsv -o /mnt/NEOGENE1/projects/lizard_2019/
ValAligned_Rad_Both_Analysis/stacks.FASTA.ChrAll_f_ValAli.All/ --samples /mnt/NEOGENE1/projects/lizard_2019/
ValAligned_Rad_Both_Analysis/chrAll_f_in1_ValAli/
```

2) grep all paternal individual sequences (valE, valG1, port ... etc.) from the Stacks result file "populations.samples.fa" with the shell code:

```
for i in $(cat val_Samples_Name_List.txt); do grep -A1 $i populations.samples.fa; done >> val.fasta
```

3) grep all maternal (raddei) individuals from the Stacks result file "populations.samples.fa" with the shell code:

```
for i in $(cat rad_Samples_Name_List.txt); do grep -A1 $i populations.samples.fa; done >> rad.fasta
```

4) using the same grep code above, separate each parthenogenetic individual and put it in its own folder. Inside each folder, there is .fa file named after the individual, e.g. OA1.arm17.fa with the shell code:

```
for i in $(cat BendiSapp_Samples_Name_List.txt); do grep -A1 $i populations.samples.fa >> $i.fa; done
```

5) run python script "diagnoSeq.py" (https://github.com/mericerdolu/DiagnoSeq) below for detecting the diagnostic alleles of hybrid individuals with shell command:

```
python3.5 diagnoSeq.py paternalFasta.fa maternalFasta.fa
```

This will take all hybrid fasta files with extension .fa in the directory where the script file is.

```
#DIAGNOSTIC ALLELE DETECTION

# Get the libraries
import pandas as pd
import sys

# Take the parental objects (FASTA files)
with open(sys.argv[1], "r") as p1:
    with open(sys.argv[2], "r") as m1:

        # Load the data
        # Create Pandas object from input fasta files
        p0 = pd.read_csv(p1, sep="\t", header=None)
        m0 = pd.read_csv(m1, sep="\t", header=None)

        # Convert the single-columns fasta format to two-column data frame
        rowIndexp0 = p0.index[:]
        idxp0 = rowIndexp0.values.tolist()

        even_idxp0 = [x for x in idxp0 if x%2 == 0]
        odd_idxp0 = [x for x in idxp0 if x%2 != 0]

        p0DF = pd.DataFrame(columns=['SeqNames', 'Seq'])

        p0DF['SeqNames'] = list(p0.loc[even_idxp0, 0])
        p0DF['Seq'] = list(p0.loc[odd_idxp0, 0])


        rowIndexm0 = m0.index[:]
        idxm0 = rowIndexm0.values.tolist()

        even_idxm0 = [x for x in idxm0 if x%2 == 0]
        odd_idxm0 = [x for x in idxm0 if x%2 != 0]

        m0DF = pd.DataFrame(columns=['SeqNames', 'Seq'])

        m0DF['SeqNames'] = list(m0.loc[even_idxm0, 0])
        m0DF['Seq'] = list(m0.loc[odd_idxm0, 0])

        # Change the seq names as Locus Numbers in 1st column of
        # the data frame of parentals
        for seqName in range(0, len(m0DF)):
            m0DF.at[seqName, 'SeqNames'] = m0DF.iloc[seqName, 0].split('_')[5]

        for seqName in range(0, len(p0DF)):
            p0DF.at[seqName, 'SeqNames'] = p0DF.iloc[seqName, 0].split('_')[5]

        # Export the maternal file
        m0DF.to_csv('m0DF_2col', sep = '\t', index=False, header=None)

        # Export the paternal file
        p0DF.to_csv('p0DF_2col', sep = '\t', index=False, header=None)

## Hybrid code part
# Get hybrid files from the current working directory as "*.fa" extension
import glob
h0_list = glob.glob("*.fa")
for h_file in h0_list:
    h0 = pd.read_csv(h_file, sep="\t", header=None)

    # Convert the single-columns fasta format to two-column data frame
    rowIndexh0 = h0.index[:]
    idxh0 = rowIndexh0.values.tolist()
```

```
    even_idxh0 = [x for x in idxh0 if x%2 == 0]
    odd_idxh0 = [x for x in idxh0 if x%2 != 0]

    h0DF = pd.DataFrame(columns=['SeqNames', 'Seq'])

    h0DF['SeqNames'] = list(h0.loc[even_idxh0, 0])
    h0DF['Seq'] = list(h0.loc[odd_idxh0, 0])

    # Change the SeqNames in the 1st column as a locus number in the hybrid data
    for seqName in range(0, len(h0DF)):
        h0DF.at[seqName, 'SeqNames'] = h0DF.iloc[seqName, 0].split('_')[5]

    # Export the hybrid indv. data frame
    h0DF.to_csv('h0DF_2col', sep = '\t', index=False, header=None)


    # Extract number of locus of the hybrid sample
    locusNumUniqH = h0DF['SeqNames'].unique().tolist()

    # Create paternal and maternal output files for diagnostic alleles
    PatDF = pd.DataFrame(columns=['SeqNames', 'Seq'])
    MatDF = pd.DataFrame(columns=['SeqNames', 'Seq'])


    for line in locusNumUniqH:
        # Get index numbers of locus "line" in hybrid, pataernal and maternal data
        hidx = h0DF[h0DF.SeqNames == line].SeqNames.index.tolist()
        pidx = p0DF[p0DF.SeqNames == line].SeqNames.index.tolist()
        midx = m0DF[m0DF.SeqNames == line].SeqNames.index.tolist()

        if len(pidx) > 0 and len(midx) > 0 and len(hidx) > 0:
            # Check uniqueness of parental segs
            if ~p0DF.loc[pidx, 'Seq'].isin(m0DF.loc[midx, 'Seq']).any().any() and ~m0DF.loc[midx,
            ↪  'Seq'].isin(p0DF.loc[pidx, 'Seq']).any().any():
                # Inspect if hybrid 1st allle is in paternal side
                inspectorPh0 = pd.DataFrame([h0DF.loc[hidx[0],'Seq']]).isin(list(p0DF.loc[pidx,
                ↪  'Seq'])).any().any()
                # Inspect if hybrid 1st allle is in maternal side
                inspectorMh0 = pd.DataFrame([h0DF.loc[hidx[0],'Seq']]).isin(list(m0DF.loc[midx,
                ↪  'Seq'])).any().any()
                # Inspect if hybrid 2nd allle is in paternal side
                inspectorPh1 = pd.DataFrame([h0DF.loc[hidx[1],'Seq']]).isin(list(p0DF.loc[pidx,
                ↪  'Seq'])).any().any()
                # Inspect if hybrid 2nd allle is in maternal side
                inspectorMh1 = pd.DataFrame([h0DF.loc[hidx[1],'Seq']]).isin(list(m0DF.loc[midx,
                ↪  'Seq'])).any().any()
                # Separate the hybrid alleles as diagnostic paternal and maternal
                # and save them in the output files
                if inspectorPh0 and ~inspectorMh0 and ~inspectorPh1 and inspectorMh1:
                    PatDF = PatDF.append(pd.DataFrame(h0DF.loc[hidx[0],:]).T, ignore_index=True)
                    MatDF = MatDF.append(pd.DataFrame(h0DF.loc[hidx[1],:]).T, ignore_index=True)
                elif ~inspectorPh0 and inspectorMh0 and inspectorPh1 and ~inspectorMh1:
                    PatDF = PatDF.append(pd.DataFrame(h0DF.loc[hidx[1],:]).T, ignore_index=True)
                    MatDF = MatDF.append(pd.DataFrame(h0DF.loc[hidx[0],:]).T, ignore_index=True)

    # Export the paternal and maternal diagnostic allele files of the hybrid indv.
    PatDF.to_csv("./{0}_P".format(h_file), sep = '\t', index=False, header=None)
    MatDF.to_csv("./{0}_M".format(h_file), sep = '\t', index=False, header=None)
```

6) After detecting the diagnostic alleles, there will be two output files, one maternal and one paternal, of each hybrid individual. Convert them to fasta format with shell for loop below:

```
for i in *<output_files>; do awk '{ printf ">%s\n%s\n",$1,$2 }' $i > $i.fa; done
```

7) After all of these, align the hybrid fasta files to the *D. valentini* reference (or *Podarcis muralis* reference genome) genome to gain the SAM files with shell command:

```
for i in *<theHybridFastaFiles>; do /usr/local/sw/the_bowtie2_path/bowtie2 -p 3 -f -x
/the_reference_genome_path/ -U /the_input_files_path/$i -S /the_output_files_path/ $i.sam; done
```

8) After alignment, convert all SAM files to BAM and sort them with shell command using Samtools:

```
for i in *.sam; do /usr/local/sw/samtools-1.11/samtools view -bSh $i | /usr/local/sw/samtools-1.11/samtools sort
--threads 10 > $i.bam; done
```

9) Load the paternal and maternal BAM files to the same directory with the hybrid BAM files and make the popmap files for Stacks.

10) For phylogenetic tree, start the Stacks/population process with –phylip flag for generating phylip file with shell command:

```
/usr/local/sw/stacks-2.41/bin/ref_map.pl -X "populations: --phylip" -T 3 --popmap /the_path/
popmap_BendSappVal.tsv -o /the_output_directory_path/ --samples /the_input_BAMfiles_path/paternals_AlignFiles/
```

11) Convert the PHYLIP file to FASTA with the vim code:

```
%s/^/\>/g | %s/^I/\r/g
```

12) Build the phylogeny with this FASTA file using PhyML online tool in the link: https://ngphylogeny.fr/tools/tool/271/form

13) For fineRADstructure analysis, start the Stacks/populations with –radpainter flag process with shell command:

```
/usr/local/sw/stacks-2.53/bin/ref_map.pl -X "populations: --radpainter" -T 3 --popmap
/the_popmap_path/popmap_BendSappVal.tsv -o /the_output_directory_path/stacks.PHYLIP.BendSappValPy/ --samples
/the_input_BAMfiles_path/paternals_AlignFiles/
```

14) Draw the fineRADstructure plot by processing the Stacks/radpainter output files and commands below in terminal:

```
Calculate the co-ancestry matrix with fineRADstructure/RADpainter option:
/usr/local/sw/fineRADstructure-v.0.3.2r109/RADpainter paint populations.haps.radpainter
```

```
Assign individuals to populations with fineRADstructure/finestructure option:
/usr/local/sw/fineRADstructure-v.0.3.2r109/finestructure -x 100000 -y 100000 -z 1000 populations.haps_chunks.out
populations.haps_chunks.mcmc.xml
```

```
Tree building with fineRADstructure/finestructure option:
/usr/local/sw/fineRADstructure-v.0.3.2r109/finestructure -m T -x 1000 populations.haps_chunks.out
populations.haps_chunks.mcmc.xml populations.haps_chunks.mcmcTree.xml
```

```
Finally, plot the results using R script in the links below:
http://cichlid.gurdon.cam.ac.uk/fineRADstructurePlot.R and
http://cichlid.gurdon.cam.ac.uk/FinestructureLibrary.R
If these websites are not accessed, you can use the GitHub page:
https://github.com/millanek/fineRADstructure
```

V) SNP BASED Fst AND BAR PLOT CODES:

1) Run the Stacks/fstats with the code:

```
/usr/local/sw/stacks-2.41/bin/ref_map.pl -X "populations: --fstats --fst_correction" -T 7 --popmap
/mnt/NEOGENE1/projects/lizard_2019/ValAligned_Rad_Both_Analysis/stacks.Fst.ChrAll_f.LakeVan/Species_BendiSapp_Fst/
popmap_LakeVanBendiSappTWOSpecies.tsv -o
/mnt/NEOGENE1/projects/lizard_2019/ValAligned_Rad_Both_Analysis/stacks.Fst.ChrAll_f.LakeVan/Species_BendiSapp_Fst/
--samples /mnt/NEOGENE1/projects/lizard_2019/ValAligned_Rad_Both_Analysis/chrAll_f_in1_ValAli/
```

2) Extract the Fst values from the "populations.fst_Sap-Bendi.tsv" file with the shell code:

```
# This is for 2 species version
awk '{print $8}' populations.fst_Sap-Bendi.tsv | awk '(NR>1)' >> BendiSapp2Species_SNPFst
# This is for 4 populations version
awk '{print $8}' populations.fst_PinarSap-MurBendi.tsv | awk '(NR>1)' >> BendiSapp4Species_SNPFst
awk '{print $8}' populations.fst_CalBendi-MurBendi.tsv | awk '(NR>1)' >> BendiSapp4Species_SNPFst
awk '{print $8}' populations.fst_CalBendi-PinarSap.tsv | awk '(NR>1)' >> BendiSapp4Species_SNPFst
awk '{print $8}' populations.fst_VanAgriSap-MurBendi.tsv | awk '(NR>1)' >> BendiSapp4Species_SNPFst
awk '{print $8}' populations.fst_VanAgriSap-PinarSap.tsv | awk '(NR>1)' >> BendiSapp4Species_SNPFst
awk '{print $8}' populations.fst_VanAgriSap-CalBendi.tsv | awk '(NR>1)' >> BendiSapp4Species_SNPFst
```

3) Draw the distribution of the bar plot of the SNP Fst values (for 2-species and 4 pops of BendiSapp samples) with R script:

```
# THE HISTOGRAM OF THE SNP DISTRIBUTION
# As 2 species
BendiSapp2Pops_SNPFst <- read.table("/mnt/NEOGENE1/projects/lizard_2019/ValAligned_Rad_Both_Analysis
/stacks.Fst.ChrAll_f.LakeVan/Species_BendiSapp_Fst/BendiSapp2Species_SNPFst", sep="\t")
par(mar=c(5,5,3,0)) # In turn Below, Left, Above, Right
hist(BendiSapp2Pops_SNPFst[,1], col="#1B9E77", main=NULL, xlab="Fst", ylim = range(c(0, 300000)),
cex.lab=1.9, cex.main=2, cex.axis=1.7)
#As 4 populations
BendiSapp4Pops_SNPFst <- read.table("/mnt/NEOGENE1/projects/lizard_2019/ValAligned_Rad_Both_Analysis
/stacks.Fst.ChrAll_f.LakeVan/FourPops_BendiSapp_Fst/BendiSapp4Pops_SNPFst", sep="\t")
par(mar=c(5,5,3,0)) # In turn Below, Left, Above, Right
hist(BendiSapp4Pops_SNPFst[,1], col="#1B9E77", main=NULL,
     xlab="Fst", ylim = range(c(0, 1400000)), cex.lab=1.9, cex.main=2, cex.axis=1.7)
```

4) Calculate the mean Fst values for bendi-sapp as 2 species and 4 different pops via formulae in R STUDIO:

```
sum(BendiSapp2Pops_SNPFst)/dim(BendiSapp2Pops_SNPFst)[1]
sum(BendiSapp4Pops_SNPFst)/dim(BendiSapp4Pops_SNPFst)[1]
```

6) Run the Stacks/plink code 500 times for the permutation test with the shell code below:

- Make 500 directories with the shell code in the permutation files directories:

```
for i in permDir; do for ((l =1; l<=500; l++)); do mkdir $i$l; done; done
```

- Make 500 different popmaps for Stacks/fstats processes with R script below for both 2 species version and 4 populations version separately:

```
## Generate popmaps for permutation tests for 4 pops Lake Van BendiSapp population's SNP Fst in Stacks
popmap_LakeVanBendiSappSEPERATED <-
read.delim("/mnt/NEOGENE1/projects/lizard_2019/ValAligned_Rad_Both_Analysis/stacks.Fst.ChrAll_f.LakeVan/
FourPops_BendiSapp_Fst/popmap_LakeVanBendiSappSEPERATED.tsv", header=FALSE, quote="")
popmap_Perm_All <- lapply(1:500, function (x) data.frame(sample(as.character
(popmap_LakeVanBendiSappSEPERATED[,1])), as.character(popmap_LakeVanBendiSappSEPERATED[,2])))
popmap_Perm_All
# Separate the data frames from one compact data frame and write popmaps into the linux directory
for (i in 1:length(popmap_Perm_All)) {assign(paste0("popmap_Perm_", i), as.data.frame(
popmap_Perm_All[[i]], header=NULL))
write.table(assign(paste0("popmap_Perm_", i), as.data.frame(popmap_Perm_All[[i]], header=NULL)),
file=paste0("/mnt/NEOGENE1/projects/lizard_2019/ValAligned_Rad_Both_Analysis/stacks.Fst.ChrAll_f.LakeVan/
FourPops_BendiSapp_Fst/permutationTest_4PopsLakeVan/popmap_Perm_",i,".tsv"),
sep="\t", quote=F, col.names=F, row.names=F)
  }
```

- Run the Stacks/plink for 500 times with shell loop below:

```
for i in {1..500}; do /usr/local/sw/stacks-2.41/bin/ref_map.pl -X "populations: --fstats --fst_correction"
-T 5
--popmap /mnt/NEOGENE1/projects/lizard_2019/ValAligned_Rad_Both_Analysis/stacks.Fst.ChrAll_f.LakeVan/
FourPops_BendiSapp_Fst/permutationTest_4PopsLakeVan/popmap_Perm_$i.tsv -o
/mnt/NEOGENE1/projects/lizard_2019/ValAligned_Rad_Both_Analysis/stacks.Fst.ChrAll_f.LakeVan/
FourPops_BendiSapp_Fst/permutationTest_4PopsLakeVan/permDir$i/ --samples
/mnt/NEOGENE1/projects/lizard_2019/ValAligned_Rad_Both_Analysis/chrAll_f_in1_ValAli/; done
```

7) Perform a permutation test to the SNP Fst values to test randomization with R script below for each population combination separately:

```
# Take the summary Fst values from the server
# This is for 2 species version
for (i in 1:500) {
permFstLakeVan <- data.frame()
assign(paste("permFstLakeVan", i,
sep=""),read.table(paste("/mnt/NEOGENE1/projects/lizard_2019/ValAligned_Rad_Both_Analysis
/stacks.Fst.ChrAll_f.LakeVan/
Species_endiSapp_Fst/permutationTest_2SpeciesLakeVan/permDir", i, "/populations.fst_summary.tsv", sep=""),
sep="\t", quote="",header=T, row.names = 1))
}
# Take SNP Fst values and calculate the p value for 2 species version
permFstAll = sapply(1:500, function(x) get(paste("permFstLakeVan", x, sep=""))[,3][2] )
# Calculate p value
round(mean(as.numeric(as.character(permFstAll))[which(mean(as.numeric(as.character(permFstAll))) <=
as.numeric(as.character(permFstAll)))]), digits=3)
# Find the mean SNP Fst value
round(mean(as.numeric(as.character(permFstAll))), digits=2)
# This is for 4 populations version
for (i in 1:500) {
permFstLakeVan <- data.frame()
assign(paste("permFstLakeVan", i,
sep=""),read.table(paste("/mnt/NEOGENE1/projects/lizard_2019/ValAligned_Rad_Both_Ana
lysis/stacks.Fst.ChrAll_f.LakeVan/FourPops_BendiSapp_Fst/permutationTest_4PopsLakeVa
n/permDir", i, "/populations.fst_summary.tsv", sep=""), sep="\t",
quote="",header=T, row.names = 1))
}
# Take the SNP Fst values between each population pair among 4 pops BendiSapp Fst calculation
permFstAll_VASapXCalBendi <- sapply(1:100, function(x) get(paste("permFstLakeVan", x, sep=""))[,3][2])
permFstAll_VASapXPinarSap <- sapply(1:100, function(x) get(paste("permFstLakeVan", x, sep=""))[,4][2])
permFstAll_VASapXMurBendi <- sapply(1:100, function(x) get(paste("permFstLakeVan", x, sep=""))[,5][2])
permFstAll_CalBendiXPinarSap <- sapply(1:100, function(x) get(paste("permFstLakeVan", x, sep=""))[,4][3])
permFstAll_CalBendiXMurBendi <- sapply(1:100, function(x) get(paste("permFstLakeVan", x, sep=""))[,5][3])
permFstAll_PinarSapXMurBendi <- sapply(1:100, function(x) get(paste("permFstLakeVan", x, sep=""))[,5][4])
#Run this for each population pair above separately
permFstAll = permFstAll_VASapXCalBendi
permFstAll = permFstAll_VASapXPinarSap
permFstAll = permFstAll_VASapXMurBendi
permFstAll = permFstAll_CalBendiXPinarSap
permFstAll = permFstAll_CalBendiXMurBendi
permFstAll = permFstAll_PinarSapXMurBendi
# Find the mean SNP Fst value (Run this for each population pair!)
round(mean(as.numeric(as.character(permFstAll))), digits=2)
# Calculate each p value for each population pair among 4 populations above
round(mean(as.numeric(as.character(permFstAll))[which(mean(as.numeric(as.character(permFstAll))) <=
as.numeric(as.character(permFstAll)))]), digits=3)
```

VI) MULTI-ALLELIC LOCI ANALYSIS OF *D. bendimahiensis* AND *D. sapphirina*:

1) make the valentini fasta "val.ValAli.fasta" file with the code:

```
for i in $(cat val_Samples_Name_List.txt); do grep -A1 $i populations.samples.fa; done >> val.ValAli.fasta (P.S. The
populations.samples.fa is result of Stacks/fasta-samples proccess)
```

2) make the raddei fasta "rad.ValAli.fasta" file with the code:

```
for i in $(cat rad_Samples_Name_List.txt); do grep -A1 $i populations.samples.fa; done >> rad.ValAli.fasta
(P.S. The populations.samples.fa is result of Stacks/fasta-samples proccess)
```

3) extract the BendiSapp sample fasta files with the code:

```
for i in $(cat BendiSapp_Samples_Name_List.txt); do grep -A1 $i populations.samples.fa >> $i.fa; done (P. S. The
populations.samples.fa is result of Stacks/fasta proccess)
```

4) delete the lines including "- -" in the fasta files above with the vim code: :g/–/d because python scripts may not work properly

5) start the multi-allelic loci process for the ValAligned data with "MALAsHyb.py" (https://github.com/mericerdolu/MALAs) python script using the SLURM codes below:

```
#!/bin/bash
#SBATCH -A #your_user_account
#SBATCH -J #name of job
#SBATCH -N 1 #number of nodes to be used
#SBATCH -n 1 # number of tasks (mpi) to be lunched
#SBATCH -c 1 #number of cores per task.. Not mandatory
#SBATCH -p  #name of the job queue
#SBATCH --output=/mnt/NEOGENE1/projects/lizard_2019/ValAligned_Rad_Both_Analysis/triAllelicState
/slurm_diagno_GitHubv.%out
#SBATCH --output=/mnt/NEOGENE1/projects/lizard_2019/ValAligned_Rad_Both_Analysis/triAllelicState
/slurm_diagno_GitHubv.%jerr
python3.5 MALAsHyb.py
```

```
# Multi Allelic Locus Alleles (MALAs) Detection From A Population of A Hybrid Species

# Get libraries
import pandas as pd
import glob

# import the FASTA files of the hybrid indvs. with .fa extension
h0_list = glob.glob("*.fa")
h0 = pd.DataFrame()
for h_file in h0_list:
    h0_pr = pd.read_csv(h_file, sep="\t", header=None)
    h0 = h0.append(h0_pr)

    # Convert the sigle-column fasta data frame to two-column data frame
    rowIndex = h0.index[:]
    idx = rowIndex.values.tolist()

    even_idx = [x for x in idx if x%2 == 0]
    odd_idx = [x for x in idx if x%2 != 0]

    h0DF = pd.DataFrame(columns=['SeqNames', 'Seq'])

    h0DF['SeqNames'] = list(h0.loc[even_idx, 0])
    h0DF['Seq'] = list(h0.loc[odd_idx, 0])

# Change the SeqNames in the 1st column as a locus number in the hybrid data frame
for seqName in range(0, len(h0DF)):
  h0DF.at[seqName, 'SeqNames'] = h0DF.iloc[seqName, 0].split('_')[5]

# Filter the rows including 'N' in Seq column
h0DFNo_N = h0DF[~h0DF.Seq.str.contains('N')]

# Filter rows contain duplicate seqs in the Seq column
h0DFNo_Dup = h0DFNo_N.drop_duplicates(subset=['Seq'])

# Sort the h0DFNo_Dup file as to locus numbers
h0DFNo_DupSort = h0DFNo_Dup.sort_values(by=['SeqNames'], ascending=False)

# Detect tri or more allelic states through the hybrid indvs.
# and saved them in the "h0DF*" output object
for line in h0DFNo_DupSort['SeqNames'].unique():
    lidx = h0DFNo_DupSort.index[h0DFNo_DupSort['SeqNames'] == str(line)].tolist()
    if len(lidx) < 3 :
        h0DFNo_DupSort.drop(lidx, inplace=True)

# Export the result file
h0DFNo_DupSort.to_csv('h0DFMultiAlleles', sep = '\t', index=False, header=None)
```

6) the result file "h0DFMultiAlleles" is produced in the same file with the python script

7) run the python script "MALAsPar.py" (https://github.com/mericerdolu/MALAs) with paternal fasta (val.ValAli.fasta) and maternal fasta (rad.ValAli.fasta) files using

the SLURM code below for parental multi-allelic loci alleles:

```
#!/bin/bash
#SBATCH -A #your_user_account
#SBATCH -J #name of job
#SBATCH -N 1 #number of nodes to be used
#SBATCH -n 1 # number of tasks (mpi) to be lunched
#SBATCH -c 1 #number of cores per task.. Not mandatory
#SBATCH -p #name of the job queue
#SBATCH --output=/mnt/NEOGENE1/projects/lizard_2019/ValAligned_Rad_Both_Analysis/triAllelicState
/slurm_patMultiAllele.%out
#SBATCH --output=/mnt/NEOGENE1/projects/lizard_2019/ValAligned_Rad_Both_Analysis/triAllelicState
/slurm_patMultiAllele.%jerr
python3.5 GitHub.parntMultiAllelePy.py val_ValAli.fasta rad_ValAli.fasta
```

```
# MULTI ALLELLIC LOCUS ALLELES (MALAs) DETECTION IN THE PARENTAL POPULATIONS OF A HYBRID SPECIES

# Get the libraries
import pandas as pd
import sys

# Take the maternal and paternal fasta files including the all data of all individuals
with open(sys.argv[1], "r") as p1:
    with open(sys.argv[2], "r") as m1:

        # Load the data
        # Create Pandas object from input fasta files
        p0 = pd.read_csv(p1, sep="\t", header=None)
        m0 = pd.read_csv(m1, sep="\t", header=None)

        # Convert the single-columns fasta format to two-column data frame
        SeqNamesP = []
        SeqP = []

        for line in range(0, len(p0), 2):
            SeqNamesP.append(p0.iloc[line,0])
            SeqP.append(p0.iloc[line+1,0])

        p0DF = pd.DataFrame(zip(SeqNamesP, SeqP), columns=['SeqNames', 'Seq'])

        SeqNamesM = []
        SeqM = []

        for line in range(0, len(m0), 2):
            SeqNamesM.append(m0.iloc[line,0])
            SeqM.append(m0.iloc[line+1,0])

        m0DF = pd.DataFrame(zip(SeqNamesM, SeqM), columns=['SeqNames', 'Seq'])

        # Change the seq names as Locus Numbers in 1st column of
        # the data frame of parentals
        for seqName in range(0, len(m0DF)):
            m0DF.at[seqName, 'SeqNames'] = m0DF.iloc[seqName, 0].split('_')[5]

        for seqName in range(0, len(p0DF)):
            p0DF.at[seqName, 'SeqNames'] = p0DF.iloc[seqName, 0].split('_')[5]

        # Export the maternal and paternal files with two-column format
        m0DF.to_csv('m0DF_twoCols', sep = '\t', index=False, header=None)
        p0DF.to_csv('p0DF_twoCols', sep = '\t', index=False, header=None)


## Hybrid code part
# Get the hybrid MALAs file from the cureent working directory
import glob
for h_file in glob.glob("h0DFMultiAlleles"):
    h0DF = pd.read_csv(h_file, sep="\t", header=None)
    h0DF.rename(columns={0: 'SeqNames', 1:'Seq'}, inplace=True)


# Detect if each multi-allele of each MALAs locus in the hybrids is present in parental populations
# Create output files PatDF for MALAs in paternal population and MatDF for MALAs in maternal population
# and Mat_PatDF for MALAs in both maternal and paternal populations
PatDF = pd.DataFrame(columns=['SeqNames', 'Seq'])
MatDF = pd.DataFrame(columns=['SeqNames', 'Seq'])
Mat_PatDF = pd.DataFrame(columns=['SeqNames', 'Seq'])


for sq in range(0, len(h0DF)):
    # Inspect if the hybrid allele is in paternal side
    inspectorPh = pd.DataFrame([h0DF.loc[sq,'Seq']]).isin(list(p0DF.loc[:, 'Seq'])).any().any()
    # Inspect if the hybrid alle is in maternal side
    inspectorMh = pd.DataFrame([h0DF.loc[sq,'Seq']]).isin(list(m0DF.loc[:, 'Seq'])).any().any()

    # Save the MALAs as maternal and paternal (if any) in the output files
    if inspectorPh==True:
        PatDF = PatDF.append(pd.DataFrame(h0DF.loc[sq,:]).T, ignore_index=True)

    if inspectorMh==True:
        MatDF = MatDF.append(pd.DataFrame(h0DF.loc[sq,:]).T, ignore_index=True)

    if inspectorMh==True and inspectorPh==True:
        Mat_PatDF = Mat_PatDF.append(pd.DataFrame(h0DF.loc[sq,:]).T, ignore_index=True)

# Export the paternal and maternal MALAs files of the hybrid species
PatDF.to_csv("./P_h0DFMultiAlleles", sep = '\t', index=False, header=None)
MatDF.to_csv("./M_h0DFMultiAlleles", sep = '\t', index=False, header=None)
Mat_PatDF.to_csv("./MP_h0DFMultiAlleles", sep = '\t', index=False, header=None)
```

8) detect the allele frequencies of each multi allele in bendi-sapp, Raddei and Valentini pops using also SLURM files "MultiFreq1.sh", "MultiFreq2.sh", "MultiFreq3.sh", their results "multiAlleleFreqInHybSH", "multiAlleleFreqInValSH", "multiAlleleFreqInRadSH" are in the same directory of the scripts:

MultiFreq1.sh:

```
#!/bin/bash
#SBATCH -A #your_user_account
#SBATCH -J #name of job
#SBATCH -N 1 #number of nodes to be used
#SBATCH -n 1 # number of tasks (mpi) to be lunched
#SBATCH -c 1 #number of cores per task.. Not mandatory
#SBATCH -p  #name of the job queue
#SBATCH --output=/mnt/NEOGENE1/projects/lizard_2019/ValAligned_Rad_Both_Analysis/
triAllelicState/slurm_multiFreq1.%out
#SBATCH --output=/mnt/NEOGENE1/projects/lizard_2019/ValAligned_Rad_Both_Analysis/
triAllelicState/slurm_multiFreq1.%jerr
for seq in $(awk '{print $2}' h0DFMultiAlleles); do for samp in *.fa; do printf $seq"\t"; printf $samp"\t";
grep -cw $seq $samp; done; done > multiAlleleFreqInHybSHw
```

MultiFreq2.sh:

```
#!/bin/bash
#SBATCH -A #your_user_account
#SBATCH -J #name of job
#SBATCH -N 1 #number of nodes to be used
#SBATCH -n 1 # number of tasks (mpi) to be lunched
#SBATCH -c 1 #number of cores per task.. Not mandatory
#SBATCH -p #name of the job queue
#SBATCH --output=/mnt/NEOGENE1/projects/lizard_2019/ValAligned_Rad_Both_Analysis/triAllelicState
/slurm_multiFreq2.%out
#SBATCH --output=/mnt/NEOGENE1/projects/lizard_2019/ValAligned_Rad_Both_Analysis/triAllelicState
/slurm_multiFreq2.%jerr
for seq in $(awk '{print $2}' h0DFMultiAlleles); do printf $seq"\t"; grep -cw $seq
val_ValAli.NOArdahan.fa; done > multiAlleleFreqInValLakeVanSHw
```

MultiFreq3.sh:

```
#!/bin/bash
#SBATCH -A #your_user_account
#SBATCH -J #name of job
#SBATCH -N 1 #number of nodes to be used
#SBATCH -n 1 # number of tasks (mpi) to be lunched
#SBATCH -c 1 #number of cores per task.. Not mandatory
#SBATCH -p #name of the job queue
#SBATCH --output=/mnt/NEOGENE1/projects/lizard_2019/ValAligned_Rad_Both_Analysis/triAllelicState
/slurm_multiFreq3.%out
#SBATCH --output=/mnt/NEOGENE1/projects/lizard_2019/ValAligned_Rad_Both_Analysis/triAllelicState
/slurm_multiFreq3.%jerr
for seq in $(awk '{print $2}' h0DFMultiAlleles); do printf
$seq"\t"; grep -cw $seq rad_ValAli.NOKars.fasta; done > multiAlleleFreqInRadLakeVanSHw
```

9) detect (3, 4, 5, 6, 7, etc.)-MALO and alleles from the whole MALO file:

- Locus numbers of 3-MALO:

```
for (i in unique(h0Hyb[,1])) { if (sum(h0Hyb[,1] == i) == 3) {triAlLocusNumbers <-
c(triAlLocusNumbers, i) }}  %>%
write.table(as.data.frame(unlist(triAlLocusNumbers)), file="./triAlLocusNumbers", col.names=F,
row.names=F, quote=F) in R
```

- Extract the 3-MALO alleles:

```
for i in $(cat triAlLocusNumbers); do grep -w $i h0DFMultiAlleles | awk '{print $2}'; done >> alleles3to1loc
```

- Draw the plot of the number of MALO of each category with R script:

```
# ggplot version of the histogram of number of MALO
numOfMultialleleBarplot <- c(16096, 1844, 108, 13, 3, 2)
names(numOfMultialleleBarplot) <- c(3,4,5,6,7,8)
library(ggplot2)
library(ggpubr)
library(tidyverse)
library(wesanderson)
ggplt_version_barplot <- data.frame(numOfMultialleleBarplot, malas=c(3,4,5,6,7,8)) %>% group_by(malas)
ggplot(ggplt_version_barplot) +
geom_col(aes( x= factor(malas), y=numOfMultialleleBarplot), fill=wes_palette("Cavalcanti1", 3,
type = c"continuous"))
[3]) + geom_text(aes(y=c(16096, 1844, 108, 13, 3, 2),x = factor(malas), label =
paste0(c(89.1, 10.21, 0.6, 0.07, 0.02, 0.01), "%"), vjust = -0.2), size=9) +
theme_pubr() +
xlab("The number of different alleles at a locus \nin the all Bendimahiensis and Sapphirina populations") +
ylab("Total number of loci") +
theme(plot.title = element_text(color = "black", size = 22, face = "bold", hjust=0.5),
axis.text = element_text(color = "black", size=22),
axis.title = element_text(color = "black", size=27))
```

- Calculate the mean number of loci of samples from the fasta files:

```
for i in *.fa; do awk '!/>/ {count++ } END{print count}' $i; done | awk '{sum+=$1}  END {print sum/NR}'
```

- Filter the 3-MALO alleles freqs with the shell code:

```
for i in $(cat alleles3to1loc); do grep -w $i multiAlleleFreqInHybSHw; done >> TriAlleleFreqInHybSHw
```

- Make the freqs3AlWithSeqs file with R script:

```
freqs3AlWithSeqs <- data.frame()
for (i in seq(1, dim(TriAlleleFreqInHybSHw)[1], 18)) {
    for (fr in sum(TriAlleleFreqInHybSHw[i:sum(i+17), 3])/36) {
      freqs3AlWithSeqs <- rbind(freqs3AlWithSeqs, cbind(TriAlleleFreqInHybSHw[i:sum(i+17),1:3], fr))
    }
}
colnames(freqs3AlWithSeqs) <- c("Sequence", "Sample", "#ofAll", "Freq")
write.table(freqs3AlWithSeqs,
file="/mnt/NEOGENE1/projects/lizard_2019/ValAligned_Rad_Both_Analysis/triAllelicState/freqs3AlWithSeqs",
sep="\t", col.names=F, row.names=F, quote=F)
```

10) Find the number of MALO in parental species with the shell codes below (Run it for each category of MALO file "alleles...to1loc" including the alleles for each parental species and shared alleles with two parental species):

```
for i in $(cat alleles3to1loc); do grep -w $i P_h0DFMultiAlleles; done | wc -l
for i in $(cat alleles3to1loc); do grep -w $i M_h0DFMultiAlleles; done | wc -l
for i in $(cat alleles3to1loc); do grep -w $i MP_h0DFMultiAlleles; done | wc -l
```

11) Calculate the number of 3-MALO with >=2 distances each allele pair with the R

script below:

```
# This is necessary against the rbind error
freqs3AlWithSeqs <- unname(freqs3AlWithSeqs)
triAl_ALXYTriangle <- 0
for (i in seq(1, dim(freqs3AlWithSeqs)[1], 54)) {
if (adist((unique(freqs3AlWithSeqs[i:sum(i,53), 1]))[1,],(unique(freqs3AlWithSeqs[i:sum(i,53),
1]))[2,]) >= 2 &&
adist((unique(freqs3AlWithSeqs[i:sum(i,53), 1]))[1,],(unique(freqs3AlWithSeqs[i:sum(i,53),
1]))[3,]) >= 2 &&
adist((unique(freqs3AlWithSeqs[i:sum(i,53), 1]))[2,],(unique(freqs3AlWithSeqs[i:sum(i,53),
1]))[3,]) >= 2) {
triAl_ALXYTriangle <- triAl_ALXY9Triangle + 1
   }
 }
```

VII) BINOMIAL TEST WITH 3-MALO:

1) Make "val.ValAli.seqs" file with the shell code:

```
grep -v '>' val.ValAli.fasta > val.ValAli.seqs
```

2) Make "rad.ValAli.seqs" file with the shell code:

```
grep -v '>' rad.ValAli.fasta > rad.ValAli.seqs
```

3) Run the code below in R for generating the alleles in the binomial test (2+1 vs. 1+2)

```
# Get the source files
raddei.seqs <- read.delim("/mnt/NEOGENE1/projects/lizard_2019/ValAligned_Rad_Both_Analysis/
triAllelicStrad.ValAli.seqs", sep="\t", quote="", header=F)
val.seqs <- read.delim("/mnt/NEOGENE1/projects/lizard_2019/ValAligned_Rad_Both_Analysis/
triAllelicState.ValAli.seqs", sep="\t", quote="", header=F)
freqs3AlWithSeqs <- read.delim("/mnt/NEOGENE1/projects/lizard_2019/ValAligned_Rad_Both_Analysis/
triAllelicState/freqs3AlWithSeqs", sep="\t", quote="", header=F)
# Separate the tri-MALO alleles as one is paternal, other two are maternal or vice versa, and
calculate the distribution between number of maternals and number of paternals if they have equal ratio or not!
triAl_ALXYBinomsMaternal <- 0
triAl_ALXYBinomsPaternal <- 0
# This is necessary against the rbind error
freqs3AlWithSeqs <- unname(freqs3AlWithSeqs)
for (i in seq(1, dim(freqs3AlWithSeqs)[1], 54)) {
      if ( sum(as.character(unique(freqs3AlWithSeqs[i:sum(i,53), 1])[,1])[1] == val.seqs) > 0 &&
      sum(as.character(unique(freqs3AlWithSeqs[i:sum(i,53), 1])[,1])[1] == raddei.seqs) == 0 &&
      sum(as.character(unique(freqs3AlWithSeqs[i:sum(i,53), 1])[,1])[2] == val.seqs) > 0 &&
      sum(as.character(unique(freqs3AlWithSeqs[i:sum(i,53), 1])[,1])[2] == raddei.seqs) == 0 &&
      sum(as.character(unique(freqs3AlWithSeqs[i:sum(i,53), 1])[,1])[3] == val.seqs) == 0 &&
      sum(as.character(unique(freqs3AlWithSeqs[i:sum(i,53), 1])[,1])[3] == raddei.seqs) > 0) {
      triAl_ALXYBinomsPaternal <- triAl_ALXYBinomsPaternal + 1
} else if (sum(as.character(unique(freqs3AlWithSeqs[i:sum(i,53), 1])[,1])[1] == val.seqs) > 0 &&
      sum(as.character(unique(freqs3AlWithSeqs[i:sum(i,53), 1])[,1])[1] == raddei.seqs) == 0 &&
      sum(as.character(unique(freqs3AlWithSeqs[i:sum(i,53), 1])[,1])[2] == val.seqs) == 0 &&
      sum(as.character(unique(freqs3AlWithSeqs[i:sum(i,53), 1])[,1])[2] == raddei.seqs) > 0 &&
      sum(as.character(unique(freqs3AlWithSeqs[i:sum(i,53), 1])[,1])[3] == val.seqs) > 0 &&
      sum(as.character(unique(freqs3AlWithSeqs[i:sum(i,53), 1])[,1])[3] == raddei.seqs) == 0) {
      triAl_ALXYBinomsPaternal <- triAl_ALXYBinomsPaternal + 1
} else if (sum(as.character(unique(freqs3AlWithSeqs[i:sum(i,53), 1])[,1])[1] == val.seqs) == 0 &&
      sum(as.character(unique(freqs3AlWithSeqs[i:sum(i,53), 1])[,1])[1] == raddei.seqs) > 0 &&
      sum(as.character(unique(freqs3AlWithSeqs[i:sum(i,53), 1])[,1])[2] == val.seqs) > 0 &&
      sum(as.character(unique(freqs3AlWithSeqs[i:sum(i,53), 1])[,1])[2] == raddei.seqs) == 0 &&
      sum(as.character(unique(freqs3AlWithSeqs[i:sum(i,53), 1])[,1])[3] == val.seqs) > 0 &&
      sum(as.character(unique(freqs3AlWithSeqs[i:sum(i,53), 1])[,1])[3] == raddei.seqs) == 0) {
      triAl_ALXYBinomsPaternal <- triAl_ALXYBinomsPaternal + 1
} else if (sum(as.character(unique(freqs3AlWithSeqs[i:sum(i,53), 1])[,1])[1] == val.seqs) == 0 &&
      sum(as.character(unique(freqs3AlWithSeqs[i:sum(i,53), 1])[,1])[1] == raddei.seqs) > 0 &&
      sum(as.character(unique(freqs3AlWithSeqs[i:sum(i,53), 1])[,1])[2] == val.seqs) == 0 &&
      sum(as.character(unique(freqs3AlWithSeqs[i:sum(i,53), 1])[,1])[2] == raddei.seqs) > 0 &&
      sum(as.character(unique(freqs3AlWithSeqs[i:sum(i,53), 1])[,1])[3] == val.seqs) > 0 &&
      sum(as.character(unique(freqs3AlWithSeqs[i:sum(i,53), 1])[,1])[3] == raddei.seqs) == 0) {
      triAl_ALXYBinomsMaternal <- triAl_ALXYBinomsMaternal + 1
} else if (sum(as.character(unique(freqs3AlWithSeqs[i:sum(i,53), 1])[,1])[1] == val.seqs) == 0 &&
      sum(as.character(unique(freqs3AlWithSeqs[i:sum(i,53), 1])[,1])[1] == raddei.seqs) > 0 &&
      sum(as.character(unique(freqs3AlWithSeqs[i:sum(i,53), 1])[,1])[2] == val.seqs) > 0 &&
      sum(as.character(unique(freqs3AlWithSeqs[i:sum(i,53), 1])[,1])[2] == raddei.seqs) == 0 &&
      sum(as.character(unique(freqs3AlWithSeqs[i:sum(i,53), 1])[,1])[3] == val.seqs) == 0 &&
      sum(as.character(unique(freqs3AlWithSeqs[i:sum(i,53), 1])[,1])[3] == raddei.seqs) > 0) {
      triAl_ALXYBinomsMaternal <- triAl_ALXYBinomsMaternal + 1
} else if (sum(as.character(unique(freqs3AlWithSeqs[i:sum(i,53), 1])[,1])[1] == val.seqs) > 0 &&
      sum(as.character(unique(freqs3AlWithSeqs[i:sum(i,53), 1])[,1])[1] == raddei.seqs) == 0 &&
      sum(as.character(unique(freqs3AlWithSeqs[i:sum(i,53), 1])[,1])[2] == val.seqs) == 0 &&
      sum(as.character(unique(freqs3AlWithSeqs[i:sum(i,53), 1])[,1])[2] == raddei.seqs) > 0 &&
      sum(as.character(unique(freqs3AlWithSeqs[i:sum(i,53), 1])[,1])[3] == val.seqs) == 0 &&
      sum(as.character(unique(freqs3AlWithSeqs[i:sum(i,53), 1])[,1])[3] == raddei.seqs) > 0) {
      triAl_ALXYBinomsMaternal <- triAl_ALXYBinomsMaternal + 1
  }
}
```

VIII) ALLELIC RICHNESS AND MEAN NUMBER OF ALLELES PER LOCUS CALCULATION:

1) Run Stacks/fasta-samples-raw process for each species with the code:

```
/usr/local/sw/stacks-2.41/bin/ref_map.pl -X "populations: --fasta-samples-raw" -T 20 --popmap
/mnt/NEOGENE1/projects/lizard_2019/ValAligned_Rad_Both_Analysis/
stacks.FASTA-RAW.ChrAll_f_ValAli.LakeVanRad/popmap_LakeVanR
ad.tsv -o /mnt/NEOGENE1/projects/lizard_2019/ValAligned_Rad_Both_Analysis/
stacks.FASTA-RAW.ChrAll_f_ValAli.LakeVanRad/ --samples /mnt/NEOGENE1/projects/
lizard_2019/ValAligned_Rad_Both_Analysis/chrAll_f_in1_ValAli/
```

2) Calculate the mean number of alleles (number of haplotypes in the gene pool / number of loci in the gene pool):

```
#The first code returns number of haplotypes in the pops (before division sign), and the 2nd code
returns the number of loci in the pops (after division sign), divide them. The fasta file is
from Stacks/fasta-samples-raw above!
grep -v '>' populations.samples-raw.fa | uniq | wc -l / grep '>' populations.samples-raw.fa |
cut -d "_" -f2 | uniq | wc -l
```

3) Run Stacks/genepop process for each species with the code:

```
/usr/local/sw/stacks-2.41/bin/ref_map.pl -X "populations: --genepop" -T 20  --popmap
/mnt/NEOGENE1/projects/lizard_2019/ValAligned_Rad_Both_Analysis/stacks.GENEPOP.ChrAll_f_ValAli.
LakeVanRad/popmap_LakeVanRad
.tsv -o /mnt/NEOGENE1/projects/lizard_2019/ValAligned_Rad_Both_Analysis/stacks.GENEPOP.ChrAll_f_ValAli.LakeVanRad/
--samples /mnt/NEOGENE1/projects/lizard_2019/ValAligned_Rad_Both_Analysis/chrAll_f_in1_ValAli/
```

4) Calculate the allelic richness with R library "hierfstats" using R script below:

```
# Load the GENIND objects result of stacks/genepop process (P.S.: read.genepop FUNC.
# CONVERT THE OBJECT AUTOMATICALLY GENIND OBJETC!), (P.S.: CHANGE (OR COPY THE FILE) NAME OF THE FILE
"populations.snps.genepop" AS "val_ValAligned_LakeVan.snps.gen" WITH ".gen" EXTENSION!)
library(adegenet)
Val_ValAligned_LakeVanGENIND <-
read.genepop("/mnt/NEOGENE1/projects/lizard_2019/ValAligned_Rad_Both_Analysis/
stacks.GENEPOP.ChrAll_f_ValAli.LakeVanVal/val_ValAligned_LakeVan.snps.gen", quiet=TRUE)
Rad_ValAligned_LakeVanGENIND <-
read.genepop("/mnt/NEOGENE1/projects/lizard_2019/ValAligned_Rad_Both_Analysis/
stacks.GENEPOP.ChrAll_f_ValAli.LakeVanRad/Rad_ValAligned_LakeVan.snps.gen", quiet=TRUE)
BendiSapp_ValAligned_LakeVanGENIND <-
read.genepop("/mnt/NEOGENE1/projects/lizard_2019/ValAligned_Rad_Both_Analysis/
stacks.GENEPOP.ChrAll_f_ValAli.LakeVanBendiSapp/BendiSapp_ValAligned_LakeVan.snps.gen", quiet=TRUE)
library(hierfstat)
# Calculate allelic richness
alRich_RadLakeVan <- allelic.richness(Rad_ValAligned_LakeVanGENIND)
alRich_ValLakeVan <- allelic.richness(Val_ValAligned_LakeVanGENIND)
alRich_BendiSappLakeVan <- allelic.richness(BendiSapp_ValAligned_LakeVanGENIND)
# Calculate mean allelic richness for each population/individual(column) and then take mean of all columns
mean(sapply(1:11, function(i) mean(alRich_RadLakeVan$Ar[,i][!(is.na(alRich_RadLakeVan$Ar[,i]))])))
mean(sapply(1:8, function(i) mean(alRich_ValLakeVan$Ar[,i][!(is.na(alRich_ValLakeVan$Ar[,i]))])))
mean(sapply(1:18, function(i) mean(alRich_BendiSappLakeVan$Ar[,i][!(is.na(alRich_BendiSappLakeVan$Ar[,i]))])))
```

IX) CORRELATION ANALYSIS OF FREQUENCIES OF 3-ALLELIC LOCI ALLELES FOR MULTIPLE HYBRIDIZATION ESTIMATION:

1) Import the result files from the SLURM files MultiFreq1.sh: multiAlleleFreqInHybSHw, MultiFreq2.sh: multiAlleleFreqInValLakeVanSHw, MultiFreq3.sh: multi-

AlleleFreqInRadLakeVanSHw (These files include only the number of alleles instead of frequencies, you will calculate the frequencies in R script!)

2) Calculate the frequencies with R script below:

```
# The multi allele frequencies in hybrid, maternal and paternal pops
# And their correlations
freq_Hyb = list()
for (i in seq(1, dim(multiAlleleFreqInHybSHw)[1], 18)) {
    freq_Hyb[i] <- sum(multiAlleleFreqInHybSHw[i:sum(i+17),3])/36 }
freq_Hyb = freq_Hyb[freq_Hyb != "NULL"]
freq_Val = multiAlleleFreqInValLakeVanSHw[,2]/16
freq_Rad = multiAlleleFreqInRadLakeVanSHw[,2]/22
#Make data frames of the frequencies by matching freqs of each allele in available species
val_hybFreqDF = data.frame(unlist(freq_Hyb), unlist(as.list(freq_Val)))
rad_hybFreqDF = data.frame(unlist(freq_Hyb), unlist(as.list(freq_Rad)))
val_radFreqDF = data.frame(unlist(as.list(freq_Val)), unlist(as.list(freq_Rad)))
# NormalDist tests for allele freqs
# IN KS Test: if p < 0.05, we don't believe that our variable follows a normal distribution in our population
ks.test(x=val_hybFreqDF[,1],y='pnorm',alternative='two.sided')
ks.test(x=val_radFreqDF[,2],y='pnorm',alternative='two.sided')
ks.test(x=rad_hybFreqDF[,2],y='pnorm',alternative='two.sided')
# normality test with Q-Q plot
library(ggpubr)
ggqqplot(val_hybFreqDF$unlist.freq_Hyb., ylab= "Freqeuncies of Bendi-Sapp Alleles")
ggqqplot(val_hybFreqDF$unlist.as.list.freq_Val.., ylab= "Freqeuncies of Valentini Alleles")
ggqqplot(rad_hybFreqDF$unlist.as.list.freq_Rad.., ylab="Freqeuncies of Raddei Alleles")
# Perform correlation test (With both Spearman and Pearson)
# Because non-normal dist from KS Test, I used spearman
cor.test(val_hybFreqDF[,1], val_hybFreqDF[,2], method = "spearman")
cor.test(rad_hybFreqDF[,1], rad_hybFreqDF[,2], method = "spearman")
cor.test(val_radFreqDF[,1], val_radFreqDF[,2], method = "spearman")
```

3) Calculation of correlations between allele frequencies of 3-MALO of binomial test:

Calculate and exporting the alleles and allele frequency files "triAl_ALXYBinomsMaternal" and "triAl_ALXYBinomsPaternal" with the code below

```
raddei.seqs <- read.delim("/mnt/NEOGENE1/projects/lizard_2019/ValAligned_Rad_Both_Analysis/
triAllelicState/rad.ValAli.seqs", sep="\t", quote="", header=F)
val.seqs <- read.delim("/mnt/NEOGENE1/projects/lizard_2019/ValAligned_Rad_Both_Analysis/
triAllelicState/val.ValAli.seqs", sep="\t", quote="", header=F)
freqs3AlWithSeqs <- read.delim("/mnt/NEOGENE1/projects/lizard_2019/ValAligned_Rad_Both_Analysis/
triAllelicState/freqs3AlWithSeqs", sep="\t", quote="", header=F)
# Separate the tri-MALO alleles as one is paternal, other two are maternal or vice versa,
and calculate the distribution between number of maternals and number of paternals if they
have equal ratio or not!
triAl_ALXYBinomsMaternal <- data.frame()
triAl_ALXYBinomsPaternal <- data.frame()
# This is necessary against the rbind error
freqs3AlWithSeqs <- unname(freqs3AlWithSeqs)
for (i in seq(1, dim(freqs3AlWithSeqs)[1], 54)) {
if ( sum(as.character(unique(freqs3AlWithSeqs[i:sum(i,53), 1])[,1])[1] == val.seqs) > 0 &&
        sum(as.character(unique(freqs3AlWithSeqs[i:sum(i,53), 1])[,1])[1] == raddei.seqs) == 0 &&
        sum(as.character(unique(freqs3AlWithSeqs[i:sum(i,53), 1])[,1])[2] == val.seqs) > 0 &&
        sum(as.character(unique(freqs3AlWithSeqs[i:sum(i,53), 1])[,1])[2] == raddei.seqs) == 0 &&
        sum(as.character(unique(freqs3AlWithSeqs[i:sum(i,53), 1])[,1])[3] == val.seqs) == 0 &&
        sum(as.character(unique(freqs3AlWithSeqs[i:sum(i,53), 1])[,1])[3] == raddei.seqs) > 0) {
triAl_ALXYBinomsPaternal <- rbind(triAl_ALXYBinomsPaternal, as.matrix(unique(freqs3AlWithSeqs
[i:sum(i, 53),1:4])))
} else if (sum(as.character(unique(freqs3AlWithSeqs[i:sum(i,53), 1])[,1])[1] == val.seqs) > 0 &&
        sum(as.character(unique(freqs3AlWithSeqs[i:sum(i,53), 1])[,1])[1] == raddei.seqs) == 0 &&
        sum(as.character(unique(freqs3AlWithSeqs[i:sum(i,53), 1])[,1])[2] == val.seqs) == 0 &&
        sum(as.character(unique(freqs3AlWithSeqs[i:sum(i,53), 1])[,1])[2] == raddei.seqs) > 0 &&
        sum(as.character(unique(freqs3AlWithSeqs[i:sum(i,53), 1])[,1])[3] == val.seqs) > 0 &&
        sum(as.character(unique(freqs3AlWithSeqs[i:sum(i,53), 1])[,1])[3] == raddei.seqs) == 0) {
        triAl_ALXYBinomsPaternal <- rbind(triAl_ALXYBinomsPaternal, as.matrix(unique(freqs3AlWithSeqs
        [i:sum(i, 53),1:4])))
} else if (sum(as.character(unique(freqs3AlWithSeqs[i:sum(i,53), 1])[,1])[1] == val.seqs) == 0 &&
        sum(as.character(unique(freqs3AlWithSeqs[i:sum(i,53), 1])[,1])[1] == raddei.seqs) > 0 &&
        sum(as.character(unique(freqs3AlWithSeqs[i:sum(i,53), 1])[,1])[2] == val.seqs) > 0 &&
        sum(as.character(unique(freqs3AlWithSeqs[i:sum(i,53), 1])[,1])[2] == raddei.seqs) == 0 &&
        sum(as.character(unique(freqs3AlWithSeqs[i:sum(i,53), 1])[,1])[3] == val.seqs) > 0 &&
        sum(as.character(unique(freqs3AlWithSeqs[i:sum(i,53), 1])[,1])[3] == raddei.seqs) == 0) {
        triAl_ALXYBinomsPaternal <- rbind(triAl_ALXYBinomsPaternal, as.matrix(unique(freqs3AlWithSeqs
        [i:sum(i, 53),1:4])))
} else if (sum(as.character(unique(freqs3AlWithSeqs[i:sum(i,53), 1])[,1])[1] == val.seqs) == 0 &&
        sum(as.character(unique(freqs3AlWithSeqs[i:sum(i,53), 1])[,1])[1] == raddei.seqs) > 0 &&
        sum(as.character(unique(freqs3AlWithSeqs[i:sum(i,53), 1])[,1])[2] == val.seqs) == 0 &&
        sum(as.character(unique(freqs3AlWithSeqs[i:sum(i,53), 1])[,1])[2] == raddei.seqs) > 0 &&
        sum(as.character(unique(freqs3AlWithSeqs[i:sum(i,53), 1])[,1])[3] == val.seqs) > 0 &&
        sum(as.character(unique(freqs3AlWithSeqs[i:sum(i,53), 1])[,1])[3] == raddei.seqs) == 0) {
        triAl_ALXYBinomsMaternal <- rbind(triAl_ALXYBinomsMaternal, as.matrix(unique(freqs3AlWithSeqs
        [i:sum(i, 53),1:4])))
} else if (sum(as.character(unique(freqs3AlWithSeqs[i:sum(i,53), 1])[,1])[1] == val.seqs) == 0 &&
        sum(as.character(unique(freqs3AlWithSeqs[i:sum(i,53), 1])[,1])[1] == raddei.seqs) > 0 &&
        sum(as.character(unique(freqs3AlWithSeqs[i:sum(i,53), 1])[,1])[2] == val.seqs) > 0 &&
        sum(as.character(unique(freqs3AlWithSeqs[i:sum(i,53), 1])[,1])[2] == raddei.seqs) == 0 &&
        sum(as.character(unique(freqs3AlWithSeqs[i:sum(i,53), 1])[,1])[3] == val.seqs) == 0 &&
        sum(as.character(unique(freqs3AlWithSeqs[i:sum(i,53), 1])[,1])[3] == raddei.seqs) > 0) {
        triAl_ALXYBinomsMaternal <- rbind(triAl_ALXYBinomsMaternal, as.matrix(unique(freqs3AlWithSeqs
        [i:sum(i, 53),1:4])))
} else if (sum(as.character(unique(freqs3AlWithSeqs[i:sum(i,53), 1])[,1])[1] == val.seqs) > 0 &&
        sum(as.character(unique(freqs3AlWithSeqs[i:sum(i,53), 1])[,1])[1] == raddei.seqs) == 0 &&
        sum(as.character(unique(freqs3AlWithSeqs[i:sum(i,53), 1])[,1])[2] == val.seqs) == 0 &&
        sum(as.character(unique(freqs3AlWithSeqs[i:sum(i,53), 1])[,1])[2] == raddei.seqs) > 0 &&
        sum(as.character(unique(freqs3AlWithSeqs[i:sum(i,53), 1])[,1])[3] == val.seqs) == 0 &&
        sum(as.character(unique(freqs3AlWithSeqs[i:sum(i,53), 1])[,1])[3] == raddei.seqs) > 0) {
        triAl_ALXYBinomsMaternal <- rbind(triAl_ALXYBinomsMaternal, as.matrix(unique(freqs3AlWithSeqs
        [i:sum(i, 53),1:4])))
  }
}
```

```
# Export the output files
write.table(triAl_ALXYBinomsMaternal,
"/mnt/NEOGENE1/projects/lizard_2019/ValAligned_Rad_Both_Analysis/triAllelicState/triAl_ALXYBinomsMaternal",
quote = F, sep="\t", row.names = F, col.names=F)
write.table(triAl_ALXYBinomsPaternal,
"/mnt/NEOGENE1/projects/lizard_2019/ValAligned_Rad_Both_Analysis/triAllelicState/triAl_ALXYBinomsPaternal",
quote = F, sep="\t", row.names = F, col.names=F)
```

```
# For the parental part
multiAlleleFreqInRad <- read.delim("/mnt/NEOGENE1/projects/lizard_2019/ValAligned_Rad_Both_Analysis
/triAllelicState/multiAlleleFreqInRadLakeVanSHw",
                          sep="\t", quote="", header=F)
multiAlleleFreqInVal <- read.delim("/mnt/NEOGENE1/projects/lizard_2019/ValAligned_Rad_Both_Analysis
/triAllelicState/multiAlleleFreqInValLakeVanSHw",
                               sep="\t", quote="", header=F)
# Find the frequencies of the same alleles in the maternal species
triMALAsAlxyBinomMat <- data.frame()
for (i in 1:dim(triAl_ALXYBinomsMaternal)[1]) {
  triMALAsAlxyBinomMat <- rbind(triMALAsAlxyBinomMat, multiAlleleFreqInRad[which(multiAlleleFreqInRad[,1] ==
  as.list(as.character(triAl_ALXYBinomsMaternal[i,1]))), ])
}
# Find the frequencies of the same alleles in the paternal species
triMALAsAlxyBinomPat <- data.frame()
for (i in 1:dim(triAl_ALXYBinomsPaternal)[1]) {
  triMALAsAlxyBinomPat <- rbind(triMALAsAlxyBinomPat, multiAlleleFreqInVal[which(multiAlleleFreqInVal[,1] ==
  as.list(as.character(triAl_ALXYBinomsPaternal[i,1]))), ])
}
```

```
# Make the matrices for correlation calculation
CorrDFAlxyBinomValAligPat <- data.frame(triAl_ALXYBinomsPaternal[,4], triMALAsAlxyBinomPat[,2]/16)
CorrDFAlxyBinomValAligMat <- data.frame(triAl_ALXYBinomsMaternal[,4], triMALAsAlxyBinomMat[,2]/22)
CorrDFAlxyBinomValAligMatNumeric <- as.matrix(CorrDFAlxyBinomValAligMat)
CorrDFAlxyBinomValAligPatNumeric <- as.matrix(CorrDFAlxyBinomValAligPat)
# Calculate the Spearman correlations
cor.test(as.numeric(CorrDFAlxyBinomValAligMatNumeric[,1]), as.numeric(CorrDFAlxyBinomValAligMatNumeric[,2]),
method="spearman")
cor.test(as.numeric(CorrDFAlxyBinomValAligPatNumeric[,1]), as.numeric(CorrDFAlxyBinomValAligPatNumeric[,2]),
method="spearman")
# Export the output files
write.table(triMALAsAlxyBinomMat, "/mnt/NEOGENE1/projects/lizard_2019/ValAligned_Rad_Both_Analysis
/triAllelicState/triMALAsAlxyBinomMat", quote = F, sep="\t", row.names = F, col.names=F)
write.table(triMALAsAlxyBinomPat, "/mnt/NEOGENE1/projects/lizard_2019/ValAligned_Rad_Both_Analysis
/triAllelicState/triMALAsAlxyBinomPat", quote = F, sep="\t", row.names = F, col.names=F)
```

4) Calculation of correlations between allele frequencies of the minor alleles of 3-MALO alleles included in the binomial test:

```
# Same process above but here we take the minor allele while we take the first allele above.
After taking the input files!
triAl_ALXYMinorBinomsMaternal <- data.frame()
triAl_ALXYMinorBinomsPaternal <- data.frame()
freqs3AlWithSeqs <- unname(freqs3AlWithSeqs)
for (i in seq(1, dim(freqs3AlWithSeqs)[1], 54)) {
  if ( sum(as.character(unique(freqs3AlWithSeqs[i:sum(i,53), 1])[,1])[1] == val.seqs) > 0 &&
       sum(as.character(unique(freqs3AlWithSeqs[i:sum(i,53), 1])[,1])[1] == raddei.seqs) == 0 &&
       sum(as.character(unique(freqs3AlWithSeqs[i:sum(i,53), 1])[,1])[2] == val.seqs) > 0 &&
       sum(as.character(unique(freqs3AlWithSeqs[i:sum(i,53), 1])[,1])[2] == raddei.seqs) == 0 &&
       sum(as.character(unique(freqs3AlWithSeqs[i:sum(i,53), 1])[,1])[3] == val.seqs) == 0 &&
       sum(as.character(unique(freqs3AlWithSeqs[i:sum(i,53), 1])[,1])[3] == raddei.seqs) > 0) {
    if (as.matrix(unique(freqs3AlWithSeqs[i:sum(i,53), 1:4])[1,4]) < as.matrix(unique(freqs3AlWithSeqs
    [i:sum(i,53), 1:4])[19,4])) {triAl_ALXYMinorBinomsPaternal <- rbind(triAl_ALXYMinorBinomsPaternal,
    as.matrix(unique(freqs3AlWithSeqs[i:sum(i,53), 1:4])[1,1:4]))
    } else {triAl_ALXYMinorBinomsPaternal <- rbind(triAl_ALXYMinorBinomsPaternal,
    as.matrix(unique(freqs3AlWithSeqs[i:sum(i,53), 1:4])[19,1:4]))}
  } else if (sum(as.character(unique(freqs3AlWithSeqs[i:sum(i,53), 1])[,1])[1] == val.seqs) > 0 &&
             sum(as.character(unique(freqs3AlWithSeqs[i:sum(i,53), 1])[,1])[1] == raddei.seqs) == 0 &&
             sum(as.character(unique(freqs3AlWithSeqs[i:sum(i,53), 1])[,1])[2] == val.seqs) == 0 &&
             sum(as.character(unique(freqs3AlWithSeqs[i:sum(i,53), 1])[,1])[2] == raddei.seqs) > 0 &&
             sum(as.character(unique(freqs3AlWithSeqs[i:sum(i,53), 1])[,1])[3] == val.seqs) > 0 &&
             sum(as.character(unique(freqs3AlWithSeqs[i:sum(i,53), 1])[,1])[3] == raddei.seqs) == 0) {
    if (as.matrix(unique(freqs3AlWithSeqs[i:sum(i,53), 1:4])[1,4]) < as.matrix(unique(freqs3AlWithSeqs
    [i:sum(i,53), 1:4])[37,4])) {triAl_ALXYMinorBinomsPaternal <- rbind(triAl_ALXYMinorBinomsPaternal,
    as.matrix(unique(freqs3AlWithSeqs[i:sum(i,53), 1:4])[1,1:4]))
    } else {triAl_ALXYMinorBinomsPaternal <- rbind(triAl_ALXYMinorBinomsPaternal,
    as.matrix(unique(freqs3AlWithSeqs[i:sum(i,53), 1:4])[37,1:4]))}
  } else if (sum(as.character(unique(freqs3AlWithSeqs[i:sum(i,53), 1])[,1])[1] == val.seqs) == 0 &&
             sum(as.character(unique(freqs3AlWithSeqs[i:sum(i,53), 1])[,1])[1] == raddei.seqs) > 0 &&
             sum(as.character(unique(freqs3AlWithSeqs[i:sum(i,53), 1])[,1])[2] == val.seqs) > 0 &&
             sum(as.character(unique(freqs3AlWithSeqs[i:sum(i,53), 1])[,1])[2] == raddei.seqs) == 0 &&
             sum(as.character(unique(freqs3AlWithSeqs[i:sum(i,53), 1])[,1])[3] == val.seqs) > 0 &&
             sum(as.character(unique(freqs3AlWithSeqs[i:sum(i,53), 1])[,1])[3] == raddei.seqs) == 0) {
    if (as.matrix(unique(freqs3AlWithSeqs[i:sum(i,53), 1:4])[19,4]) < as.matrix(unique(freqs3AlWithSeqs
    [i:sum(i,53), 1:4])[37,4])) {triAl_ALXYMinorBinomsPaternal <- rbind(triAl_ALXYMinorBinomsPaternal,
    as.matrix(unique(freqs3AlWithSeqs[i:sum(i,53), 1:4])[19,1:4]))
    } else {triAl_ALXYMinorBinomsPaternal <- rbind(triAl_ALXYMinorBinomsPaternal,
    as.matrix(unique(freqs3AlWithSeqs[i:sum(i,53), 1:4])[37,1:4]))}
  } else if (sum(as.character(unique(freqs3AlWithSeqs[i:sum(i,53), 1])[,1])[1] == val.seqs) == 0 &&
             sum(as.character(unique(freqs3AlWithSeqs[i:sum(i,53), 1])[,1])[1] == raddei.seqs) > 0 &&
             sum(as.character(unique(freqs3AlWithSeqs[i:sum(i,53), 1])[,1])[2] == val.seqs) == 0 &&
             sum(as.character(unique(freqs3AlWithSeqs[i:sum(i,53), 1])[,1])[2] == raddei.seqs) > 0 &&
             sum(as.character(unique(freqs3AlWithSeqs[i:sum(i,53), 1])[,1])[3] == val.seqs) > 0 &&
             sum(as.character(unique(freqs3AlWithSeqs[i:sum(i,53), 1])[,1])[3] == raddei.seqs) == 0) {
    if (as.matrix(unique(freqs3AlWithSeqs[i:sum(i,53), 1:4])[1,4]) < as.matrix(unique(freqs3AlWithSeqs
    [i:sum(i,53), 1:4])[19,4])) { triAl_ALXYMinorBinomsMaternal <- rbind(triAl_ALXYMinorBinomsMaternal,
    as.matrix(unique(freqs3AlWithSeqs[i:sum(i,53), 1:4])[1,1:4]))
    }else {triAl_ALXYMinorBinomsMaternal <- rbind(triAl_ALXYMinorBinomsMaternal,
    as.matrix(unique(freqs3AlWithSeqs[i:sum(i,53), 1:4])[19,1:4]))}
  } else if (sum(as.character(unique(freqs3AlWithSeqs[i:sum(i,53), 1])[,1])[1] == val.seqs) == 0 &&
             sum(as.character(unique(freqs3AlWithSeqs[i:sum(i,53), 1])[,1])[1] == raddei.seqs) > 0 &&
             sum(as.character(unique(freqs3AlWithSeqs[i:sum(i,53), 1])[,1])[2] == val.seqs) > 0 &&
             sum(as.character(unique(freqs3AlWithSeqs[i:sum(i,53), 1])[,1])[2] == raddei.seqs) == 0 &&
             sum(as.character(unique(freqs3AlWithSeqs[i:sum(i,53), 1])[,1])[3] == val.seqs) == 0 &&
             sum(as.character(unique(freqs3AlWithSeqs[i:sum(i,53), 1])[,1])[3] == raddei.seqs) > 0) {
    if (as.matrix(unique(freqs3AlWithSeqs[i:sum(i,53), 1:4])[1,4]) < as.matrix(unique(freqs3AlWithSeqs
    [i:sum(i,53), 1:4])[37,4])) {triAl_ALXYMinorBinomsMaternal <- rbind(triAl_ALXYMinorBinomsMaternal,
    as.matrix(unique(freqs3AlWithSeqs[i:sum(i,53), 1:4])[1,1:4]))
    } else {triAl_ALXYMinorBinomsMaternal <- rbind(triAl_ALXYMinorBinomsMaternal,
    as.matrix(unique(freqs3AlWithSeqs[i:sum(i,53), 1:4])[37,1:4]))}
```

```
Continutation of the above!
} else if (sum(as.character(unique(freqs3AlWithSeqs[i:sum(i,53), 1])[,1])[1] == val.seqs) > 0 &&
             sum(as.character(unique(freqs3AlWithSeqs[i:sum(i,53), 1])[,1])[1] == raddei.seqs) == 0 &&
             sum(as.character(unique(freqs3AlWithSeqs[i:sum(i,53), 1])[,1])[2] == val.seqs) == 0 &&
             sum(as.character(unique(freqs3AlWithSeqs[i:sum(i,53), 1])[,1])[2] == raddei.seqs) > 0 &&
             sum(as.character(unique(freqs3AlWithSeqs[i:sum(i,53), 1])[,1])[3] == val.seqs) == 0 &&
             sum(as.character(unique(freqs3AlWithSeqs[i:sum(i,53), 1])[,1])[3] == raddei.seqs) > 0) {
    if (as.matrix(unique(freqs3AlWithSeqs[i:sum(i,53), 1:4])[19,4]) < as.matrix(
    unique(freqs3AlWithSeqs[i:sum(i,53), 1:4])[37,4])) {triAl_ALXYMinorBinomsMaternal <- rbind(
    triAl_ALXYMinorBinomsMaternal, as.matrix(unique(freqs3AlWithSeqs[i:sum(i,53), 1:4])[19,1:4]))
    } else {triAl_ALXYMinorBinomsMaternal <- rbind(triAl_ALXYMinorBinomsMaternal,
    as.matrix(unique(freqs3AlWithSeqs[i:sum(i,53), 1:4])[37,1:4]))}}}
# Export the output files
write.table(triAl_ALXYMinorBinomsMaternal, "/mnt/NEOGENE1/projects/lizard_2019/ValAligned_Rad_Both_Analysis
/triAllelicState/triAl_ALXYMinorBinomsMaternal", quote = F, sep="\t", row.names = F, col.names=F)
write.table(triAl_ALXYMinorBinomsPaternal, "/mnt/NEOGENE1/projects/lizard_2019/ValAligned_Rad_Both_Analysis
/triAllelicState/triAl_ALXYMinorBinomsPaternal", quote = F, sep="\t", row.names = F, col.names=F)
```

```
# For the parental part
multiAlleleFreqInRad <- read.delim("/mnt/NEOGENE1/projects/lizard_2019/ValAligned_Rad_Both_Analysis
/triAllelicState/multiAlleleFreqInRadLakeVanSHw", sep="\t", quote="", header=F)
multiAlleleFreqInVal <- read.delim("/mnt/NEOGENE1/projects/lizard_2019/ValAligned_Rad_Both_Analysis
/triAllelicState/multiAlleleFreqInValLakeVanSHw", sep="\t", quote="", header=F)
# Find the frequencies of the same alleles in the maternal species
triMALAsAlxyBinomMinorMat <- data.frame()
for (i in 1:dim(triAl_ALXYMinorBinomsMaternal)[1]) {
  triMALAsAlxyBinomMinorMat <- rbind(triMALAsAlxyBinomMinorMat, multiAlleleFreqInRad[
  which(multiAlleleFreqInRad[,1] == as.list(as.character(triAl_ALXYMinorBinomsMaternal[i,1]))), ])
}
# Find the frequencies of the same alleles in the paternal species
triMALAsAlxyBinomMinorPat <- data.frame()
for (i in 1:dim(triAl_ALXYMinorBinomsPaternal)[1]) {
  triMALAsAlxyBinomMinorPat <- rbind(triMALAsAlxyBinomMinorPat, multiAlleleFreqInVal[
  which(multiAlleleFreqInVal[,1] == as.list(as.character(triAl_ALXYMinorBinomsPaternal[i,1]))), ])
}
```

```
# Make the matrices for correlation calculation
CorrDFAlxyBinomMinorValAligPat <- data.frame(triAl_ALXYMinorBinomsPaternal[,4],
triMALAsAlxyBinomMinorPat[,2]/16)
CorrDFAlxyBinomMinorValAligMat <- data.frame(triAl_ALXYMinorBinomsMaternal[,4],
triMALAsAlxyBinomMinorMat[,2]/22)
CorrDFAlxyBinomMinorValAligMatNumeric <- as.matrix(CorrDFAlxyBinomMinorValAligMat)
CorrDFAlxyBinomMinorValAligPatNumeric <- as.matrix(CorrDFAlxyBinomMinorValAligPat)
# Calculate the Spearman correlations
cor.test(as.numeric(CorrDFAlxyBinomMinorValAligMatNumeric[,1]),
as.numeric(CorrDFAlxyBinomMinorValAligMatNumeric[,2]), method="spearman")
cor.test(as.numeric(CorrDFAlxyBinomMinorValAligPatNumeric[,1]),
as.numeric(CorrDFAlxyBinomMinorValAligPatNumeric[,2]), method="spearman")
# Export the output files
write.table(triMALAsAlxyBinomMinorMat, "/mnt/NEOGENE1/projects/lizard_2019/
ValAligned_Rad_Both_Analysis/triAllelicState
/triMALAsAlxyBinomMat", quote = F, sep="\t", row.names = F, col.names=F)
write.table(triMALAsAlxyBinomMinorPat, "/mnt/NEOGENE1/projects/lizard_2019/
ValAligned_Rad_Both_Analysis/triAllelicState
/triMALAsAlxyBinomPat", quote = F, sep="\t", row.names = F, col.names=F)
```

X) TEST OF SEQUENCING MEAN DEPTH OF 3-ALLELIC VS. 2-ALLELIC LOCI AGAINST THE PARALOGOUS:

1) Generate subset of each sample fasta file with 100 lines (50 alleles) because the

total number of 2-MALO are too much for the server's memory:

- Different 100 lines from different orders are taken from each sample fasta file!

```
sed -n '1,100p;101q' NC1bend20.fa > NC1bend20Shuff1_ 100.fa
sed -n '101,200p;201q' OVa.sap3.fa > OVa.sap3Shuff101_200.fa
sed -n '201,300p;301q' OVA.sap2.fa > OVA.sap2Shuff201_300.fa
sed -n '301,400p;401q' OVa.sap1.fa > OVa.sap1Shuff301_400.fa
sed -n '401,500p;501q' OV3.bend9.fa > OV3.bend9Shuff401_500.fa
sed -n '501,600p;601q' OV3.bend8.fa > OV3.bend8Shuff501_600.fa
sed -n '601,700p;701q' OV3.bend7.fa > OV3.bend7Shuff601_700.fa
sed -n '701,800p;801q' OP1.sap7.fa > OP1.sap7Shuff701_800.fa
sed -n '801,900p;901q' OP1.sap3.fa > OP1.sap3Shuff801_900.fa
sed -n '901,1000p;1001q' OP1.sap1.fa > OP1.sap1Shuff901_1000.fa
sed -n '1001,1100p;1101q' OC1.bend7.fa > OC1.bend7Shuff1001_1100.fa
sed -n '1101,1200p;1201q' OC1.bend5.fa > OC1.bend5Shuff1101_1200.fa
sed -n '1201,1300p;1301q' OC1.bend3.fa > OC1.bend3Shuff1201_1300.fa
sed -n '1301,1400p;1401q' OC1.bend2.fa > OC1.bend2Shuff1301_1400.fa
sed -n '1401,1500p;1501q' NVAsap6.fa > NVAsap6Shuff1401_1500.fa
sed -n '1501,1600p;1601q' NVAsap5.fa > NVAsap5Shuff1501_1600.fa
sed -n '1601,1700p;1701q' NC1bend24.fa > NC1bend24Shuff1601_1700.fa
sed -n '1701,1800p;1801q' NAg1uni1.fa > NAg1uni1Shuff1701_1800.fa
- Dispose the lines consisting of "--" from the files:
for i in *fa; do grep -v "\-\-" $i > $i.F; done
```

2) To detect 2-MALO and produce file "h0DFTWOAlleles" run the python script "TWOAllelPy_v2.1.py" below using SLURM script below (P.S. This is modified version of the script "MALAsHyb.py" for 2-MALO alleles):

(For 3-MALO, use the previously produced file "triAlLocusNumbers"!)

```
#!/bin/bash
#SBATCH -A merice #your_user_account
#SBATCH -J twoAllele
#SBATCH -N 1 #number of nodes to be used
#SBATCH -n 1 # number of tasks (mpi) to be lunched
#SBATCH -c 1 #number of cores per task.. Not mandatory
#SBATCH -p macaque3 #name of the job queue
#SBATCH --output=/mnt/NEOGENE1/projects/lizard_2019/ValAligned_Rad_Both_Analysis/triAllelicState/
slurm_TwoAll_v2.%out
#SBATCH --output=/mnt/NEOGENE1/projects/lizard_2019/ValAligned_Rad_Both_Analysis/triAllelicState/
slurm_TwoAll_v2.%jerr
python3.5 TWOAllelPy_v2.1.py
```

```python
# TWO Allelic Locus Alleles (MALAs) Detection From Populations of A Hybrid Species

# Get libraries
import pandas as pd
import glob

# import the FASTA files of the hybrid indvs. with .fa extension
h0_list = glob.glob("*.fa")
h0 = pd.DataFrame()
for h_file in h0_list:
    h0_pr = pd.read_csv(h_file, sep="\t", header=None)
    h0 = h0.append(h0_pr)


    # Convert the sigle-column fasta data frame to two-column data frame
    rowIndex = h0.index[:]
    idx = rowIndex.values.tolist()

    even_idx = [x for x in idx if x%2 == 0]
    odd_idx = [x for x in idx if x%2 != 0]

```

```
    h0DF = pd.DataFrame(columns=['SeqNames', 'Seq'])

    h0DF['SeqNames'] = list(h0.loc[even_idx, 0].reset_index(drop=True))
    h0DF['Seq'] = list(h0.loc[odd_idx, 0].reset_index(drop=True))

# Change the SeqNames in the 1st column as a locus number in the hybrid data frame
for seqName in range(0, len(h0DF)):
  h0DF.at[seqName, 'SeqNames'] = h0DF.iloc[seqName, 0].split('_')[5]


# Filter the rows including 'N' in Seq column
h0DFNo_N = h0DF[~h0DF.Seq.str.contains('N')]

# Filter rows contain duplicate seqs in the Seq column
h0DFNo_Dup = h0DFNo_N.drop_duplicates(subset=['Seq'])

# Sort the h0DFNo_Dup file as to locus numbers
h0DFNo_DupSort = h0DFNo_Dup.sort_values(by=['SeqNames'], ascending=False)


# Detect TWO allelic states through the hybrid indvs.
# and saved them in the "h0DF*" output object
for line in h0DFNo_DupSort['SeqNames'].unique():
    lidx = h0DFNo_DupSort.index[h0DFNo_DupSort['SeqNames'] == str(line)].tolist()
    if len(lidx) != 2 :
        h0DFNo_DupSort.drop(lidx, inplace=True)

# Export the result file
h0DFNo_DupSort.to_csv('h0DFTWOAlleles', sep = '\t', index=False, header=None)
```

- Extract the 2-MALO numbers from the h0DFTWOAlleles with the shell code:

```
awk '{print $1}' h0DFTWOAllelesGrp | sort | uniq > twoMALAsLocusNumbers
```

- If they are too much, you can split them to separate files:

```
split -b 1000 twoMALAsLocusNumbers ./twoMALAsLocusNumbersDirectory/
```

3) Begin the depth of coverage extracting the Tri-Allelic loci with the bash code:

```
for i in $(cat /coverages_3Allelic_BendiSapp/BendiSapp_SampleNames.txt); do samtools depth
/mnt/NEOGENE1/projects/lizard_2019/ValAligned_Rad_Both_Analysis/chrAll_f_in1_ValAli/$i.sam.bam |
awk '{print $2, $3}'| grep -w "$(cat /mnt/NEOGENE1/projects/lizard_2019/
ValAligned_Rad_Both_Analysis/triAllelicState/triAlLocusNumbers)"
> $i.TriAllelic.coverage; done
```

4) Begin the depth of coverage extracting the Two-Allelic loci with the bash code:

```
for l in ../twoMALAsLocusNumbersDirectory/*; do for i in $(cat ./BendiSapp_SampleNames.txt); do
samtools depth /mnt/NEOGENE1/projects/lizard_2019/ValAligned_Rad_Both_Analysis/chrAll_f_in1_ValAli
/$i.sam.bam | awk '{print $2, $3}'| grep -w "$(cat /mnt/NEOGENE1/projects/lizard_2019
/ValAligned_Rad_Both_Analysis/triAllelicState/depthAnalysisForParalogsValAli/$l)" >>
$i.TwoAllelic.coverage; done; done
```

5) Make 2-MALO and 3-MALO depth list of loci:

```
for i in *TriAllelic.coverage; do awk '{print $2}' $i >> TriAllelic_BendiSappValAli.coverages; done
for i in *TwoAllelic.coverage; do awk '{print $2}' $i >> TwoAllelic_BendiSappValAli.coverages; done
```

6) Run the R script below to perform a Mann-Whitney U test to mean depths of 2-MALO vs. 3-MALO and to draw histograms of the depth distributions:

```
# BendiSapp Coverages of ValAli data
TwoAllelic_BendiSapp.Coverages <- read.table("/mnt/NEOGENE1/projects/lizard_2019
/ValAligned_Rad_Both_Analysis/triAllelicState/depthAnalysisForParalogsValAli
/TwoAllelic_BendiSappValAli.coverages", quote="", comment.char="")
TriAllelic_BendiSapp.Coverages <- read.table("/mnt/NEOGENE1/projects/lizard_2019
/ValAligned_Rad_Both_Analysis/triAllelicState/depthAnalysisForParalogsValAli
/TriAllelic_BendiSappValAli.coverages", quote="", comment.char="")
DATASET0 <- rbind(data.frame(Class="TwoClass", Depth=TwoAllelic_BendiSapp.Coverages$V1,
stringsAsFactors = T), data.frame(Class="TriClass", Depth=TriAllelic_BendiSapp.Coverages$V1,
stringsAsFactors = T))
library(dplyr)
DATASET1 <- filter(DATASET0, Depth < 11) # filter the dpeths higher than 10 because they are probably paralog
# Check normal distribution
ks.test(DATASET1[which(DATASET1[,1] == "TwoClass"),2], "pnorm")
ks.test(DATASET1[which(DATASET1[,1] == "TriClass"),2], "pnorm")
# Perform Mann-Whitney U Test with wilcox func. FOR COMPAIRING DEPTHS OF 2-MALO and 3-MALO
wilcox.test(Depth ~ Class, data=DATASET1, exact=F)
# Calculate mean depth
mean (DATASET1[which(DATASET1[,1] == "TwoClass"),2])
mean (DATASET1[which(DATASET1[,1] == "TriClass"),2])
# Calculate median depth
median (DATASET1[which(DATASET1[,1] == "TwoClass"),2])
median (DATASET1[which(DATASET1[,1] == "TriClass"),2])
# Histogram of The Depths of 2-MALO
par(mar=c(5,5,3,2)) # In turn Below, Left, Above, Right
hist(DATASET1[which(DATASET1[,1] == "TwoClass"),2],  col="#1B9E77", xlab="Depth of the 2-MALO", main=NULL,
cex.lab=1.9, cex.axis=1.9, ylim=c(0, 100000), xlim = c(0,10))
abline(v=mean(DATASET1[which(DATASET1[,1] == "TwoClass"),2]), col="black", lwd=3)
abline(v=median(DATASET1[which(DATASET1[,1] == "TwoClass"),2]), col="red", lwd=3)
# Histogram of The Depths of 3-MALO
hist(DATASET1[which(DATASET1[,1] == "TriClass"),2],  col="#D95F02", xlab="Depth of the 3-MALO", main=NULL,
cex.lab=1.9, cex.axis=1.9, ylim = c(0,200000), xlim = c(0,10))
abline(v=mean(DATASET1[which(DATASET1[,1] == "TriClass"),2]), col="black", lwd=3)
abline(v=median(DATASET1[which(DATASET1[,1] == "TriClass"),2]), col="red", lwd=3)
```

APPENDIX C

Table C.1: Individual lizard samples used in the study.

sample ID	***Darevskia species / population***	**latitude**	**longi-tude**	**alti-tude, m. a.s.l.**	**No. reads. - all Chr**	**Mapped %**	**No. reads. - ChrZ**	**No. reads. - Auto-somes**	**No. Z-linked ddRAD loci**
NA101.d2	dahli	41.746	44.641	1211	26563776	53.21	492576	26063712	3400
NA19.d12	dahli	41.647	44.695	1207	30063360	55.41	574464	29480160	3577
NA19.d14	dahli	41.647	44.695	1207	27599616	54.12	532992	27059712	3598
NA19.d5	dahli	41.647	44.695	1207	24142368	52.19	463392	23672160	3211
NA49.d1	dahli	41.865	43.974	1055	25275264	52.56	489792	24778176	3549
Nrad1	raddei Van	39.057	43.756	1806	33016128	56.26	570624	32437440	3394
NA5val3	valenini Ardahan	40.792	42.946	2292	24545760	53.08	429792	24110112	3030
NA89.d4	dahli	41.347	44.148	1399	33900576	56.42	648960	33243168	3597
Nbend1	bendimahiensis	39.057	43.756	1806	28803072	54.20	550752	28244832	3594
Nbend2	raddei Van	39.057	43.756	1806	25937664	52.96	448416	25482528	2440
NArm.arm7	armeniaca Armenia	40.77	43.872	1531	26764896	52.95	512736	26246112	3334
NArm.uni1	unisexualis Armenia	40.499	45.275	1918	25943904	52.90	503712	25432512	1531
NBo2rud2	rudis bolkardagica	37.406	34.562	2589	23172480	51.45	431904	22734528	3019

NC1bend20	bendimahiensis	39.17	43.969	2059	25800480	52.14	483360	25311168	3289
NC1bend24	bendimahiensis	39.17	43.969	2059	26066880	52.17	494016	25567104	3132
NC2.val1	valenini Van	39.147	44.007	2049	27685344	54.69	519552	27156768	3520
NC28.val3	valenini Georgia	41.412	43.72	2265	22829088	53.43	438816	22383840	3540
NC28.val4	valenini Georgia	41.412	43.72	2265	30615840	55.72	591456	30014976	3318
NC40.a1	armeniaca	41.265	43.667	1975	23684544	51.30	442752	23235552	2765
NC55.m14	mixta	41.825	42.853	2110	34136736	56.23	661536	33467328	3099
NChar.rud2	rudis rudis	41.551	41.597	150	26679936	53.44	503520	26166912	3387
NE1.val3	valenini Erzurum	39.851	41.278	2300	29305056	55.65	556800	28741248	3212
NE3.uzz4	uzzelli	39.867	42.286	2200	32877888	56.76	624480	32243424	3464
NK1.uzz9	uzzelli	40.327	42.575	2131	19333152	50.15	380640	18946560	2466
Nrost11	rostombekowi	40.745	44.82	1405	28291680	53.27	555360	27727776	3269
Nrost7	rostombekowi	40.745	44.82	1405	24638976	51.93	477024	24154368	3396
NV4val2	valenini Van	38.944	42.677	2213	30225216	55.37	584736	29632992	3413
NV4val3	valenini Van	38.944	42.677	2213	29374560	55.34	558720	28807872	3237
NV4val6	valenini Van	38.944	42.677	2213	27554592	55.40	508704	27039936	3289
NVAsap5	sapphirina	39.149	43.096	1910	26883360	52.48	526944	26349600	3342
NVAsap6	sapphirina	39.149	43.096	1910	25022304	51.98	487104	24529152	3426
OA1.arm12	armeniaca Arda-han	41.116	42.708	1798	27830688	53.66	518400	27305952	2883

OA1.arm13	armeniaca Ardahan	41.116	42.708	1798	7526976	48.99	143136	7382112	2897
OA1.arm15	armeniaca Ardahan	41.116	42.708	1798	25905792	53.55	473856	25425216	3000
OA1.arm16	armeniaca Ardahan	41.116	42.708	1798	23298528	52.87	436704	22854720	2377
OA1.arm17	armeniaca Ardahan	41.116	42.708	1798	23574144	53.10	439584	23128608	2966
OA19.dahli3	dahli	41.647	44.695	1205	31419552	53.70	621888	30790080	2965
OA19.dahli6	dahli	41.647	44.695	1205	22552992	52.40	447072	22099968	3037
OAkq18.arm1	armeniaca	41.364	43.5	1723	29885568	54.56	542112	29335584	3121
OAkq18.arm4	armeniaca	41.364	43.5	1723	28835328	54.14	535872	28292544	3176
OAkg18.rud2	rudis obscura	41.364	43.5	1723	20148384	53.98	351648	19791840	2904
OBan.mix9	mixta	41.873	43.411	830	23419296	52.08	431616	22982784	2957
OBan.rud2	rudis obscura	41.873	43.411	828	25387488	53.04	464832	24915552	2944
OBan.rud7	rudis obscura	41.873	43.411	828	23615232	52.75	467520	23139072	3134
OBx.rud1	rudis obscura	41.873	43.411	828	21572448	51.77	402048	21163392	3232
OC1.bend2	bendimahiensis	39.17	43.969	2059	29565600	54.37	538368	29020416	2945
OC1.bend3	bendimahiensis	39.17	43.969	2059	27282624	54.05	490176	26786688	3291
OC1.bend5	bendimahiensis	39.17	43.969	2059	27722688	53.29	499200	27217056	2805
OC1.bend7	bendimahiensis	39.17	43.969	2059	29701824	53.67	543648	29150016	3099

OC14.arm3	armeniaca	41.373	43.467	1727	25362240	53.23	475584	24881376	2900
OC19.arm3	armeniaca	41.394	43.421	1720	28190880	53.11	516768	27667680	3322
OC2.val3	valenini Van	39.147	44.007	2049	21479712	51.66	405984	21067872	2857
OC41.a1	armeniaca	41.496	43.537	1775	25739616	52.80	489216	25245312	2865
OC46.mix6	mixta	41.788	43.472	1380	25742304	52.61	474144	25261344	2985
OC72.der2	derjugini	41.781	43.241	933	22867104	52.14	424800	22435776	2000
OD1.rad2	raddei Dogubeyazit	39.463	44.17	2200	27350784	52.53	488832	26856096	2980
OD1.rad5	raddei Dogubeyazit	39.463	44.17	2200	28922592	53.08	532512	28383168	3240
OE1.val4	valenini Erzurum	39.851	41.278	2300	24540480	51.11	469056	24064704	2769
OE1.val5	valenini Erzurum	39.851	41.278	2300	21983808	50.05	435552	21543552	2239
OE3.uni2	unisexualis	39.867	42.286	2200	19628640	51.44	372096	19251744	2725
OE3.uni4	unisexualis	39.867	42.286	2200	24073440	51.63	485760	23580384	2935
OE3.uni6	unisexualis	39.867	42.286	2200	21971616	50.97	435360	21529536	2748
OE3.uzz2	uzzelli	39.867	42.286	2200	21242304	51.64	403200	20834496	3052
OE5.uzz1	uzzelli	39.899	42.307	2014	26906208	52.75	533184	26367456	2217
OE5.uzz6	uzzelli	39.899	42.307	2014	31329984	53.28	613728	30707616	3246
OGA.port15	portschinski Armenia	40.069	46.243	1439	24442656	51.91	478560	23957952	2945

OGA.port7	portschinski Armenia	40.069	46.243	1439	25050336	52.52	472896	24569280	2988
OGA.rad19	portschinski Armenia	40.069	46.243	1439	26632320	52.66	479136	26144832	3429
OK3.rad1	raddei	40.506	43.573	1464	21135072	49.57	412608	20717856	3047
Ok3.rad2	raddei	40.506	43.573	1464	23474880	51.54	440448	23030688	3665
OK4.uni1	unisexualis	40.613	43.09	1789	26503680	53.23	483552	26014080	3598
OK4.uni2	unisexualis	40.613	43.09	1789	24904992	50.97	498624	24399072	3209
OK4.uni4	unisexualis	40.613	43.09	1789	24407136	51.09	485664	23914848	2881
OK4.uzz1	uzzelli	40.613	43.09	1789	35055360	52.09	690144	34356384	3363
OK4.uzz2	uzzelli	40.613	43.09	1789	23889888	51.67	472608	23410272	2963
OK4.uzz3	uzzelli	40.613	43.09	1789	22933920	49.73	464832	22463520	2731
OKoj.arm	armeniaca	41.743	44.462	1825	26306784	53.58	493824	25806240	3507
OMan18.dahli1	dahli	41.647	44.684	1186	30085536	54.62	543840	29533920	3554
OMan18.dahli3	dahli	41.647	44.684	1186	30760416	54.02	578592	30171072	3550
OMan18.dahli4	dahli	41.647	44.684	1186	25599744	54.76	469920	25123392	3585
OMan18.por1	portschinski	41.647	44.684	1186	26260704	54.22	489984	25762368	3122
OMan18.por3	portschinski	41.647	44.684	1186	27102144	54.44	489792	26604960	3330
OP1.sap1	sapphirina	39.164	43.192	2154	27178464	52.62	500832	26670528	2837
OP1.sap3	sapphirina	39.164	43.192	2154	29266176	54.74	512352	28747008	3267
OP1.sap7	sapphirina	39.164	43.192	2154	27718368	54.52	490848	27218784	3307

ORost12	rostombekowi	40.751	44.866	1318	24824544	52.19	471744	24346848	3082
ORost13	rostombekowi	40.751	44.866	1318	25147680	52.58	471264	24671136	3375
ORost15	rostombekowi	40.751	44.866	1318	23392224	52.62	436992	22948128	3315
ORost19	rostombekowi	40.751	44.866	1318	26068032	53.11	477216	25584192	3027
OV1.rad10	raddei	39.057	43.756	1806	1430976	45.48	23808	1406784	1836
OV1.rad11	raddei	39.057	43.756	1806	29578080	54.47	495840	29076000	2974
OV1.rad9	raddei	39.057	43.756	1806	27353856	54.79	464640	26883360	3199
OV3.bend7	raddei	39.057	43.756	1806	29113056	54.55	527808	28578528	2978
OV3.bend8	raddei	39.057	43.756	1806	26622432	54.62	472416	26143392	2901
OV3.bend9	raddei	39.057	43.756	1806	23441472	54.64	396192	23039328	3436
OVa.sap1	sapphirina	39.149	43.096	1910	26549088	54.88	472224	26069472	3400
OVA.sap2	sapphirina	39.149	43.096	1910	29748288	53.78	515232	29227104	3186
OVa.sap3	sapphirina	39.149	43.096	1910	27407040	54.90	491424	26909952	3353
OVar18.rad3	raddei	41.377	43.268	1322	24813120	53.03	465504	24339936	3345

APPENDIX D

Table D.1: Lizard samples used in the study.

Species	New Name	Sample	Latitude	Longitude	Population	Assembly : 1 – diagnostic alleles and phylogeny, 2 – multi-allelic loci	# Reads - ♀ Diag. Alleles	# Reads - ♂ Diag. Alleles
sapp./bendi.		NVAsap6	39.149	43.096	Van/Ağrı Border	1 + 2	4040	4040
sapp./bendi.		NVAsap5	39.149	43.096	Van/Ağrı Border	1 + 2	2299	2299
sapp./bendi.		OVa.sap3	39.149	43.096	Van/Ağrı Border	1 + 2	2680	2680
sapp./bendi.		OVA.sap2	39.149	43.096	Van/Ağrı Border	1 + 2	3120	3120
sapp./bendi.		OVa.sap1	39.149	43.096	Van/Ağrı Border	1 + 2	2736	2736
sapp./bendi.		OP1.sap1	39.164	43.192	Pınarlı	1 + 2	2970	2970
sapp./bendi.		NC1bend20	39.170	43.969	Çaldıran	1 + 2	3503	3503

sapp./bendi.		OP1.sap7	39.164	43.192	Pınarlı	1 + 2	2554	2554
sapp./bendi.		OP1.sap3	39.164	43.192	Pınarlı	1 + 2	2993	2993
sapp./bendi.		NC1bend24	39.170	43.969	Çaldıran	1 + 2	3891	3891
sapp./bendi.		OC1.bend2	39.170	43.969	Çaldıran	1 + 2	3356	3356
sapp./bendi.		OC1.bend3	39.170	43.969	Çaldıran	1 + 2	3045	3045
sapp./bendi.		OC1.bend5	39.170	43.969	Çaldıran	1 + 2	3091	3091
sapp./bendi.		OC1.bend7	39.170	43.969	Çaldıran	1 + 2	3249	3249
sapp./bendi.		OV3.bend7	39.057	43.756	Muradiye	1 + 2	3082	3082
sapp./bendi.		OV3.bend8	39.057	43.756	Muradiye	1 + 2	2533	2533
sapp./bendi.		OV3.bend9	39.057	43.756	Muradiye	1 + 2	2212	2212
sapp./bendi.	Nbend1	NAg1uni1	39.057	43.756	Muradiye	1 + 2	4265	4265

raddei	OV2-rad13	38.545014	44.265424	Saray	1 + 2
raddei	OV2-rad11	38.545014	44.265424	Saray	1 + 2
raddei	OV2-rad12	38.545014	44.265424	Saray	1 + 2
raddei	NV6rad1	38.44788438	43.86374338	Between Umut Yaramış	1 + 2
raddei	NV7rad4	38.35445588	43.83140915	Çörekli	1 + 2
raddei	OV1-rad9	39.057	43.756	Muradiye	1 + 2
raddei	OV1-rad11	39.057	43.756	Muradiye	1 + 2
raddei	OD1-rad6	39.463	44.170	Doğubayazıt	1 + 2
raddei	OD1-rad2	39.463	44.170	Doğubayazıt	1 + 2
raddei	OD1-rad4	39.463	44.170	Doğubayazıt	1 + 2

raddei	OD1-rad5	39.463	44.170	Doğubayazıt	1 + 2
valentini	NV4val3	38.944	42.677	Kızılyusuf	1 + 2
valentini	NV4val6	38.944	42.677	Kızılyusuf	1 + 2
valentini	NV4val5	38.944	42.677	Kızılyusuf	1 + 2
valentini	NV4val2	38.944	42.677	Kızılyusuf	1 + 2
valentini	OC3-val1	39.195970	44.021519	Çaldıran	1 + 2
valentini	OC2-val5	39.147	44.007	Çaldıran	1 + 2
valentini	NC2-val1	39.147	44.007	Çaldıran	1 + 2
valentini	OC2-val3	39.147	44.007	Çaldıran	1 + 2

CURRICULUM VITAE

PERSONAL INFORMATION

Surname, Name: Erdolu, Meriç

EDUCATION

Degree	Institution	Year of Graduation
PHD	METU Biology Department	2023
MS	Muğla University Biology Department	2014
BS	Muğla University Biology Department	2011
High School	Adıyaman Rekabet Kurumu Lisesi	2003

FOREIGN LANGUAGE

Advanced English

Latin (Beginner)

WORK AND RESEARCH EXPERIENCE

PhD Project	Resolving Reticulate Speciation In Bisexual And Parthenogenetic Lizards of Genus *Darevskia* In East Anatolia And Caucasus.
Ancient DNA Data Analysis	Population genetics analysis on data from ancient and modern human DNA studies in the laboratory.
Master's Project	Genetic Characterisation of Olive Fruit Fly (*Bactrocera oleae*) Populations Sampled From Aegean Region by Isozyme Markers And Investigation of The Biochemical Basis of Insecticide Resistance In These Populations.

PUBLICATIONS

D. Tarknishvili, A. Yanchukov, MK. Şahin, M. Gabelaia, M. Murtskhvaladza, K. Candan, E. Galoyan, M. Arakelyan, G. Iankoshvili, Y. Kumlutaş, Ç. Ilgaz, F. Matur,

F. Çolak, **M. Erdolu**, S. Kurdadze, N. Barateli, CL. Anderson (2020) "Genotypic similarities among the parthenogenetic Darevskia rock lizards with different hybrid origins" BMC Evol Biol., DOI: 10.1186/s12862-020-01690-9

Yanchukov A., Tarkhnishvili D., **Erdolu M.**, Şahin M.K., Candan K., Murtskhvaladze M., Gabelaia M., Iankoshvili G., Barateli N., Ilgaz, Ç., Kumlutaş, Y., Matur, F., Çolak, F., Arakelyan M., Galoyan E. (2022) "Precise paternal ancestry of hybrid unisexual ZW lizards (genus Darevskia: Lacertidae: Squamata) revealed by Z-linked genomic markers" Biological Journal of the Linnenan Society, DOI: https://doi.org/10.1093/biolinnean/blac023

M. Erdolu, A. Yanchukov, M. Somel, D. Tarkhnishvili, MK. Şahin (2023) "The origin and speciation history of parthenogenetic rock lizards (Darevskia sapphirina and Darevskia benmaniensis Schmidtler, Eiselt Darevsky, 1994) endemic to the basin of Lake Van in Turkey" (In Preparation)

Yanchukov A., Tarkhnishvili D., **Erdolu M.**, Şahin M.K., Ilgaz Ç., Kumlutaş Y., Arakelyan M., Galoyan E. (2023) "Resolving the reticulate evolution, hybridization and gene flow in Armeniaca- Dahli-Uzzelli- Rostombekovi group" (In Preparation)

CONFERENCE/SYMPOSIUM PRESENTATIONS

ORAL PRESENTATIONS:

Tarkhnisvili D., Yanchukov A., Murtskhvaladze M., Kurdadze S., Barateli N., Gabelaia M., Şahin MK., Candan K., Kumlutaş Y., Ilgaz Ç., Çolak F., Matur F., **Erdolu M.**, Arakelyan M., Galoyan E. "Microsatellite Genotypes of Seven Hybridogenetic Unisexual Species of Rock Lizard And Their Putative Bisexual Parents (Genus Darevskia sp.) Suggest Complex Patterns of Hybrid Ancestry" XX Europian Congress of Herpetology, Milan, Italy, 2-19 Sept. 2019, (Abstract Book)

Erdolu M. et al., "Evolutionary History of Asexual Lizard Populations In Lake Van Area", 7th Ecology and Evolutionary Biology Symposium Turkey, Online, Turkey, 16-21 Aug. 2021, (Abstract Book)

Erdolu M. et al., "Multiple parental individuals contributed to common gene pool of the hybridogenic parthenogenetic lizards Darevskia bendimahiensis and D. sap-

phirina", 8th Ecology and Evolutionary Biology Symposium Turkey, Middle East Technical University (METU), Ankara, Turkey, 7-9 Sept. 2022, (Abstract Book)

POSTERS:

Erdolu M. et al., "First Steps To Resolve Reticulate Speciation Process In Bisexual And Parthenogenetic Lizards of Genus Darevskia In East Anatolia And Caucasus", Ecology and Evolutionary Biology Symposium Turkey, İzmir, Turkey, 18-20 Jul. 2018

Erdolu M. et al., "Structure Of Ancestral And Modern Populations In Bisexual And Parthenogenetic Lizards Of Genus Darevskia In East Anatolia And Caucasus", Ecology and Evolutionary Biology Symposium Turkey, Ankara, Turkey, 10-12 Jul. 2019

Somel M., **Erdolu M.**, et al. "Using the male death ratio to estimate COVID-19 burden among excess Istanbul deaths", The 13th International Symposium On Health Informatics And Bioinformatics, 22 - 23 Oct. 2020