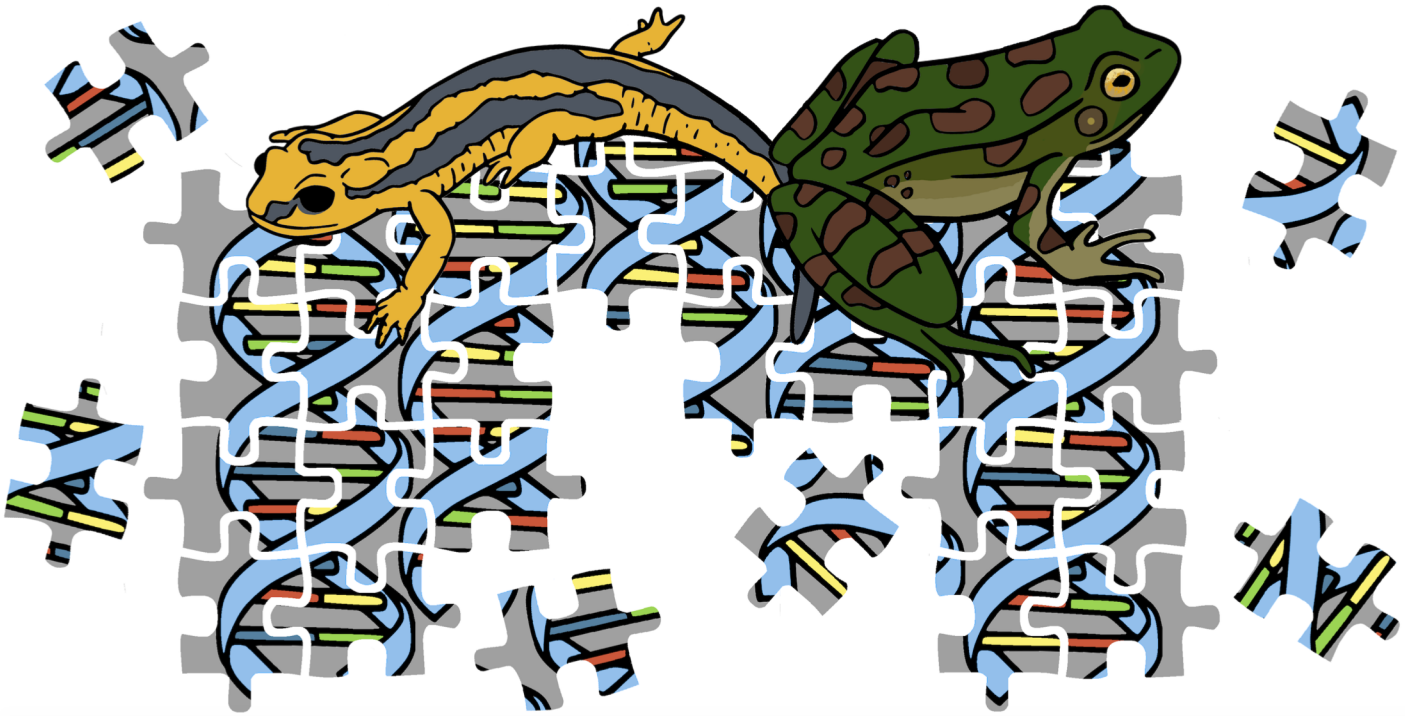


Identifying local adaptation in large amphibian genomes

Kevin Patrick Mulder
Biodiversity, Genetics, and Evolution
Department of Biology
2020

Advisor
Guillermo Velo-Antón, Auxiliary Researcher, CIBIO-InBIO/FCUP

Co-advisor
Anna E. Savage, Assistant Professor, University of Central Florida

Co-advisor
Rayna C. Bell, Assistant Curator of Herpetology, California Academy of Sciences

FOREWORD

According to the General Regulation of Doctoral Programs of the University of Porto (number 2, 4th Article) and the Decree Law 74/2006 (Article 31, 24 of March) revised under the Decree law 230/2009 (14th of September), this thesis includes manuscripts published or in consideration for publication in peer-reviewed scientific journals. These manuscripts are the result of collaborations with several co-authors. The candidate declares that he actively contributed to the ideas and the development of the research work, including the compilation, analysis, results, discussion and writing as in its current publication form. The candidate was supported by the National Foundation for Science and technology (FCT), through a PhD Grant (PD/BD/52604/2014) financed by the European Social Fund and by the National Ministry of Science, Technology and Higher Education, through the Operational Programme Human Capital (POCH), under Portugal 2020. This thesis' research was developed in the context of the Doctoral Programme in Biodiversity, Genetics and Evolution (Faculty of Sciences, Univeristy of Porto). The work was conducted at the Research Centre in Biodiversity and Genetic Resources - InBIO Associate Laboratory (CIBIO-InBIO), the Center for Conservation Genomics at the National Zoological Park and the Division of Amphibians and Reptiles at the National Museum of Natural History, both at the Smithsonian Institution, and the Biology Department of the Univeristy of Central Florida.

FCT Fundação para a Ciência e a Tecnologia
MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E ENSINO SUPERIOR
GOVERNO DA REPÚBLICA PORTUGUESA
PORTUGAL 2020 POCH
UNIÃO EUROPEIA
Fundo Social Europeu

ACKNOWLEDGEMENTS

This PhD journey included multiple advisors, labs, and cities, and this dissertation really is that cliché sum of its parts. Every step along the way could not have been made without the incredible people that were part of it and made it happen. No words can truly show how important you all were on this journey but here a small attempt at a dankjewel, gracias, go raibh maith agat and obrigado.

Guillermo, thank you for accepting me in and being open to this crazy idea of doing genomics on *Salamandra*. A stubborn PhD student that spends too much time on side projects can be a tough hand to be dealt, but thanks to you, we made it happen. With your hard work and ethics (and your focus on sampling gaps!), you are both a great advisor and scientific role model. Our successful salamander trips to Somiedo and Isla de Ons are highlights of this journey and I'm looking forward to the following Salomics steps in this amazing study system that you have built up.

Anna, from the first meeting to talk frogs and MHC many years ago it has been such a pleasure to work with you on all thing's leopard frog. You are so generous with your time, supportive in good times and bad, and always able to see past the maze of results to focus on what is really important. Your enthusiasm when seeing new data makes all those hours in the lab and in the matrix worth it, and you truly inspire me to keep on going! Thank you for advising me from afar, but always making sure you were close.

Rob, thank you for taking a chance on me right out of my masters and for always being so open to new ideas, allowing me to start these crazy projects in your lab, and for your continued support over the years. I really appreciate how you fight for the people in your lab and make sure we are all able to reach our goals. I know you are a secret herpetologist at heart and you are always welcome to drop the birds and join the amphibian and non-avian reptile team!

Rayna, thank you for being both a great advisor and a good friend. You care about the science, and most importantly, you truly care about the people. As a scientist you manage to see things we all missed and connect dots I did not even notice were missing. As a person you are always there for everybody with kind words and good advice, making sure we are all succeeding in both career and life. Thank you for taking me in halfway through and bringing this thesis to a new level, and also keeping me sane and upbeat in the process.

Thanks to the Mulder family, mum, dad and zusie, for not losing faith in me even when I was stressed and frustrated. Thanks for your support and for accepting me doing this thing called a PhD, aka “not a real job”, and for raising me with the values to work hard, enjoy life, and most importantly, allowing me to follow my passions. Mum, it is not (r)evolutionary biology and in your words “kind of boooooooring”, but this thesis is also a little bit your grandchild! Paps, het zijn dan wel geen bomen, maar de appel/kikker is toch een beetje groen? Zusie, altijd lief voor je broertje zelfs als ik het niet verdien, and so proud of you for confronting your foes and holding a salamander! Thanks to the rest of the Mulder and Kerin clan and especially to grandad for inspiring me to try and be a better person.

Thanks to the entire CIBIO team, Walter, Helena, Coelho, Hugo, Rita, Yuri, speciatric speciation, Daniele, James, Marcia, Hosein, Leili, Adri, Mario, Paulo, Sofia, Joana, Barbara, Francesco, Berenice, Lara, Iolanda, João, Ignacio, Nany, Luca and many many more for welcoming me with open arms and always being there for a coffee, a chat, advice, numerous Almada dinners and even more metro, Sr bernardino and car rides to Vairão. Thanks to Teresa for keeping us caffeinated, well fed and laughing. Thanks to the rest of the portuensis; Pedro, Dan, Andra, Daniela, Jessica, Jorge, Howard, Sharon, Stef and Lukas for making me love this city and its inhabitants. Thanks to the growing Salomics family, Adam, Bernardo, Lucía, and especially Marco for translating the abstract and Lorenzo for always being there to help and share a bad joke.

Thanks to the Smithsonian and the two labs that feel like home. Everybody at CCG for starting me on this road of genomics and all the advice, motivation and most importantly friends along the way, special thanks to Madhvi, Courtney, Lilly, Alyssa, Nancy, Carly, Suzette, Natalia, Naoko, Micheal, Loren and Jesus. To NMNH and the Bell lab for adopting me in and making me a better person and scientist. This PhD would not have been half as good without all your love, cheese and critical input; Ivan, Molly, Ed, Kyle, Ryan, Mike, Jeff, Jonathan, KdQ, Addison and Roy. Special thanks to star undergrad Jack Boyette for the stunning cover-image.

Thanks to DC and all its beautiful people that accepted a foreign biologist in their midst. Starting where it all began with the extended ibiza family; Chris, Roberto, Mark, Josh, Samia, Maya, Mar, Don Omar, Justinho, Ali, Rachel, Nathalie, El Centro and Kal! Euclid and all its inhabitants and visitors for being the physical and mental home when finally leaving the lab. A place of rest and dancing, coffee and quesadillas, politics and laughter, dark horse and IPA, bikes and bike-parties, and a group of amazing people; Joe, Maggie, Nika, KC, Claire, Isha, Rachel, Liz, Max, Dana, Fifi.

So many people around the world that I don't see enough but that I know I can always count on and hope to see more of you all now that this thesis-baby has been delivered. Thanks to the amazing MEME family and hoping for many more reunions in the field and at conferences; Moos, Glibuschka, Hectorinho, Ricardovic, Miguel, Fotini. Bergen city en de crew met Wieger en Gijs. Utrecht past and present; Carolientje, Zoe, Iris, Dodo, Bert, Kater, Adam, Paulus en alle allesopeners. The entire Savage lab for being so welcoming to that weird PhD student that comes by every two years, and especially to Vero and Alexa for Frogizona.

Some of you were such a crucial part of this journey that this little booklet would not have happened without you. Roberto, so many adventures, conversations, bike-rides, dancefloors, you even caught a frog! Thanks for being there when I was down and for helping me face my anxieties. Coca, you have seen several high's and low's in this PhD journey but whatever it was you always turned it into a smile and laugh. Ivan, you inspire me and I could not have asked for a better lab brother! Thanks for all the coffee sessions, the hips don't lie dance sessions, and for ever critical figure advice. Rosado, from Lambo-trips to bread-throwing you added positive vibes to the whole journey. Vaz, for finding and creating the interesting in both science and life. Isabel, for providing enough energy and happiness for all of us to get out of our PhD bunkers. Cristela for being so full of kindness for me and everybody around. Stephen for reaching out when I wrongly think I don't need help. Eems for having the best puns and for making the bad times go away. Claudia for knowing what's really important and calling me out in good and in bad. Ana, for being a better friend than I deserve and for being the queen of colours. Duarte for being the encyclopaedia of weird knowledge and jokes I did not know I needed. Omar for bringing the real conversations and always being there to help. Paula for all your positive energy and happiness and making every day a better day. FCUP requires a particular format but for me your names are all on the front page as it is because, and for all of you, that this thesis happened.

Looking back at this difficult and surprising journey and thinking of how lucky I am to have travelled this road with all of you, I can only smile and from deep within say THANK YOU to everybody for being the amazing people that you are.

ABSTRACT

Understanding how species adapt to their local environment is a central question in evolutionary biology. Natural selection acts on the phenotypic level, but it is the corresponding genotype that is passed onto the next generation and can become fixed in populations and species over time. Characterizing the genetic basis of phenotypic traits is thus a crucial step in the study of adaptation that has recently become more attainable due to advances in the field of genomics. Newly developed genomic tools can generate population-level markers across the genome to study variation within and among species, directly linking genetic polymorphisms with the observed phenotypic traits of interest. These tools have, however, not been widely applied to amphibians due to their exceptionally large genomes, and associated challenges with regards to sequencing, assembly and genotyping.

This thesis attempts to improve our understanding of amphibian genomics and the study of local adaptation by applying and optimizing genomic tools in two amphibian systems that show intraspecific variation in adaptive traits. The fire salamander, *Salamandra salamandra*, exhibits two viviparous reproductive modes: larviparity, in which females deliver larvae into nearby waterbodies, and pueriparity, with females delivering fully developed juveniles. This adaptive trait allows salamanders to reproduce in environments lacking water bodies, opening up new habitats with profound ecological and evolutionary implications. The lowland leopard frog (*Rana yavapaiensis*) displays variation in disease susceptibility, with resistance to the fungal disease chytridiomycosis differing widely been individuals and populations and driving adaptation in a disease dominated environment. Both *S. salamandra* and *R. yavapaiensis* have large genomes (~35 Gb and ~6 Gb respectively), show intraspecific variation in adaptive traits and are well studied at different biological levels, but adaptive traits have never been investigated using genomic tools. The two systems are thus prime natural laboratories to assess and optimize the usage of genomic tools on questions of local adaptation in large amphibian genomes.

Following a general introduction on concepts and techniques in the study of adaptation, I summarize the current state of the art on amphibian genomics and report on the challenges posed by large genomes, and discuss some of the potential solutions. By using techniques such as RNA-seq and exome capture with associated bioinformatic analyses, we can focus on the coding regions of the genome and measure functional genetic variation across genomic

space. We can subsequently use these markers to reconstruct the evolutionary history and relationships between populations, characterize gene expression differences between adaptive phenotypes, and identify candidate genes associated with our traits of interest.

To assess the effectiveness of exome capture for ancestral state reconstruction, I reconstructed a phylogeny of the *Salamandra* genus using genome-wide markers and direct observations of births to identify the geographic and phylogenetic extent of larviparity and pueriparity in *Salamandra*. I focused on the clades and regions that display differences in reproductive mode to detect the number and timing of transitions between reproductive modes. I identified five independent transitions from larviparity to pueriparity occurring at different evolutionary timescales ranging from the Pliocene to the Pleistocene. Three of these transitions occurred within *S. salamandra* providing multiple convergent instances of intra-specific variation in reproductive mode that can help control for evolutionary history and identify the genetic basis of reproductive mode shifts.

In the following study we applied RNA-sequencing to two of the three independent transitions within *S. salamandra*, characterizing gene expression in the uterus and oviduct of both larviparous and pueriparous individuals from adjacent localities to detect candidate genes. We identified differentially expressed genes in the uterus that were shared across both transitions and are thus indicative of convergent mechanisms in the evolution of pueriparity. We also identified unique differences in gene expression in both tissues. These single-transition genes may indicate unique genetic components of the convergent phenotypes or they may indicate other environmental or evolutionary differences among pueriparous and larviparous females within a given transition. Many of the identified candidate genes are associated with embryogenesis, cell growth and cell differentiation in other taxa, which is consistent with the described phenotypic differences in embryonic development between larviparity and pueriparity in *S. salamandra*.

We subsequently generated a transcriptome-based exon capture dataset across populations of tolerant, susceptible, extirpated or naïve *Rana yavapaiensis* to assess the effectiveness of exome capture data to detect signals of selection. The pairing of both functional genetic data with phylogeographic sampling enabled us to simultaneously reconstruct the evolutionary history of the populations and detect genetic signatures of local adaptation. I used this strategy to answer questions about population structure, phylogeography, demography and adaptation to disease. I found that lower heterozygosity and allelic richness were associated with increased disease susceptibility and extirpation, but that extirpated and susceptible

populations had higher levels of functionally different private alleles than less threatened populations, highlighting that their loss can reduce future adaptive potential in the species. In addition, I found limited mitochondrial diversity but strong nuclear genetic structure between populations, likely reflecting recent population fragmentation and decline, leading to rapid allele fixation. This backdrop of high genetic drift and population structure likely obscured some F_{ST}-based metrics of adaptation and highlighted some of the challenges of working with natural populations. Nevertheless, by applying multiple independent analyses of signatures of selection, I highlight candidate genes important in local adaptation and disease resistance.

Lastly, I summarize the challenges and opportunities of applying genomic tools to questions of local adaptation in amphibians and discuss important considerations for generating and analysing high-quality data on large genomes. The lack of reference genomes for annotation and quality control was initially challenging, but genome assemblies of related species could be successfully utilized even when phylogenetically distant. Strong population structure and localized, disjunct species ranges also influenced the type of applicable analyses possible, which must be considered during study design. Overall the increase in genomic tools, resources, and bioinformatic pipelines developed these past years has substantially improved the opportunities to study adaptive traits in amphibian with large genomes. The variety of different techniques and analyses in this thesis, applied to two different amphibian systems, shows the promise of studies of adaptation genomics in amphibians.

Keywords: candidate genes, chytridiomycosis, disease, genomics, high-throughput sequencing, intra-specific variation, larviparity, pueriparity, RNA-seq, reproductive mode, sequence capture.

RESUMO

A compreensão de como as espécies se adaptam ao seu ambiente é uma questão central em biologia evolutiva. A seleção natural atua ao nível fenotípico, mas é o genótipo correspondente que é transmitido à geração seguinte e se pode fixar em populações e espécies ao longo do tempo. A caracterização da base genética subjacente a características fenotípicas é, portanto, um passo crucial no estudo da adaptação, tendo-se tornado mais acessível recentemente devido a avanços no campo da genómica. Ferramentas genómicas recém-desenvolvidas conseguem produzir marcadores à escala populacional em todo o genoma para o estudo da variação intra- e interespecífica, permitindo estabelecer conexões diretas entre polimorfismos genéticos e características fenotípicas de interesse. Porém, a aplicação destas ferramentas a anfíbios é ainda limitada, devido aos seus genomas excecionalmente grandes, e aos subsequentes desafios no que diz respeito a sequenciação, montagem e genotipagem.

Esta tese tem como objetivos melhorar o nosso entendimento da genómica de anfíbios e clarificar processos adaptativos através da aplicação e otimização de ferramentas genómicas em dois sistemas de anfíbios que exibem variação intraespecífica em características adaptativas. A salamandra-de-pintas-amarelas (*Salamandra salamandra*) exibe dois modos reprodutivos vivíparos: larviparidade, em que as fêmeas depositam larvas em massas de água, e pueriparidade, em que as fêmeas depositam juvenis pós-metamórficos. Esta característica adaptativa permite a reprodução de salamandras em ambientes onde há escassez de massas de água, possibilitando a ocupação de novos habitats, com profundas implicações ecológicas e evolutivas. A rã *Rana yavapaiensis* apresenta variabilidade na sua suscetibilidade a doenças, existindo amplas diferenças entre indivíduos e populações na resistência à quitridiomicose, levando à adaptação num ambiente dominado pela doença. Tanto a *S. salamandra* como a *R. yavapaiensis* têm genomas grandes (~35 Gb e ~6 Gb, respetivamente), exibem variação intraespecífica em características adaptativas e encontram-se bem estudadas em vários aspetos da sua biologia. No entanto, as suas características adaptativas nunca foram estudadas com recurso a ferramentas genómicas. Ambos os sistemas são, portanto, excelentes laboratórios naturais para testar e otimizar o uso de ferramentas genómicas na resposta a questões de adaptação nos extensos genomas típicos dos anfíbios.

Começando com uma introdução geral aos conceitos e técnicas implicados no estudo de adaptação, sumarizo o atual estado da arte no estudo da genómica de anfíbios e discuto os desafios apresentados por genomas grandes, bem como algumas soluções possíveis. Ao utilizar técnicas como sequenciação de RNA e captura de exoma, associadas a análise bioinformática, conseguimos focar-nos nas regiões codificantes do genoma e quantificar a variabilidade genética funcional no espaço genómico. Subsequentemente, é possível utilizar estes marcadores para reconstruir a história evolutiva e as relações entre populações, caracterizar diferenças em expressão genética entre fenótipos adaptativos e identificar genes candidatos associados às características de interesse.

De modo a determinar a eficácia da captura de exoma na reconstrução de estados ancestrais, reconstruí uma filogenia do género *Salamandra* utilizando marcadores genómicos e observação direta de partos para identificar a extensão geográfica e filogenética da larviparidade e pueriparidade em *Salamandra*. Foquei-me nas clades e regiões que apresentam variabilidade no modo reprodutivo, de modo a quantificar e datar as transições entre estratégias reprodutivas. Identifiquei cinco transições independentes de larviparidade para pueriparidade, que ocorreram a distintas escalas evolutivas entre o Plioceno e o Pleistoceno. Três destas transições ocorreram em *S. salamandra*, representando múltiplos exemplos convergentes de variação intraespecífica no modo reprodutivo, permitindo assim ter um controlo para a história evolutiva e identificar a base genética de transições no modo reprodutivo.

No estudo seguinte aplicámos sequenciação de RNA a duas das três transições independentes em *S. salamandra*, através da caracterização de expressão genética no útero e oviducto de indivíduos larvíparos e pueríparos em localidades adjacentes, de modo a detetar genes candidatos. Identificámos diferenças na expressão genética no útero, comuns a ambas as transições reprodutivas e, como tal, indicativas de mecanismos convergentes na evolução da pueriparidade. Identificámos também, em ambos os órgãos, expressão diferencial entre modos reprodutivos, única para cada transição reprodutiva. Estes genes específicos de cada transição podem indicar componentes genéticas únicas dos fenótipos convergentes ou, alternativamente, diferenças ambientais ou evolutivas entre fêmeas larvíparas e pueríparas em cada transição. Muitos dos genes candidatos identificados estão associados a embriogénese ou a crescimento e diferenciação celular noutros taxa, o que é consistente com as diferenças fenotípicas no desenvolvimento embrionário entre larviparidade e pueriparidade em *S. salamandra*.

Posteriormente, produzimos um conjunto de dados de captura de exoma baseado em transcriptómica para populações tolerantes, suscetíveis, extirpadas ou naive de *Rana yavapaiensis*, de modo a determinar a eficácia de dados de captura de exoma na deteção de sinais de seleção. Através da combinação de dados de genética funcional com amostragem filogeográfica foi possível, simultaneamente, reconstruir a história evolutiva das populações e detetar sinais genéticos de adaptação. Utilizei esta estratégia para responder a questões sobre estrutura populacional, filogeografia, demografia e adaptação a doenças. Descobri que níveis baixos de heterozigotia e riqueza alélica estão associados a maior suscetibilidade a doença e risco de extirpação. No entanto, populações extirpadas e suscetíveis têm níveis mais elevados de alelos privados funcionalmente distintos do que populações menos ameaçadas, indicando que a perda dessas populações pode reduzir o potencial adaptativo nesta espécie. Adicionalmente, descobri diversidade mitocondrial limitada mas simultaneamente uma forte estrutura populacional ao nível do genoma nuclear, que possivelmente resultará de fragmentação e declínio recente das populações, levando a uma rápida fixação alélica. Este contexto de acentuada deriva genética e estrutura populacional pode ter influenciado algumas métricas de adaptação baseadas em F_{ST}, ilustrando assim alguns dos desafios de trabalhar com populações naturais. Não obstante, ao aplicar diversas análises independentes à deteção de sinais de seleção, identifiquei genes candidatos potencialmente relevantes para a adaptação local e resistência a doenças.

Por último, resumo os desafios e oportunidades da aplicação de ferramentas genómicas a questões de adaptação em anfíbios e discuto considerações importantes sobre a produção e análise de dados de alta qualidade em genomas grandes. A falta de genomas de referência para anotação e controlo de qualidade foi um desafio, mas montagens de genomas de espécies aparentadas podem ser utilizados com sucesso, mesmo quando filogeneticamente distantes. Estrutura populacional acentuada e distribuições restritas e disjuntas também influenciam os tipos de análise possíveis, o que deve ser tido em consideração durante o desenho experimental. No geral, o desenvolvimento de ferramentas, recursos e métodos bioinformáticos aplicados à genómica nos últimos anos possibilitou novas oportunidades para o estudo de características adaptativas em anfíbios com grandes genomas. A diversidade de técnicas e análises nesta tese, aplicadas a dois sistemas distintos, apresenta perspetivas promissoras para estudos de adaptação genómica em anfíbios.

Palavras-chave: genes candidatos, quitridiomicose, doença, genómica, sequenciação de larga escala, variabilidade intraespecífica, larviparidade, pueriparidade, sequenciação de RNA, modo reprodutivo, captura de sequência

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1

CHAPTER 2

CHAPTER 3

CHAPTER 4

APPENDIX B

APPENDIX C

APPENDIX D

LIST OF TABLES

CHAPTER 1

CHAPTER 3

CHAPTER 4

APPENDIX B

APPENDIX C

APPENDIX D

LIST OF ABBREVIATIONS

AMP – antimicrobial peptide
ASR – ancestral state reconstruction
Bd – *Batrachochytrium dendrobatidis*
bp – base pair
cDNA – complementary DNA
CDS – coding sequence
COI – cytochrome oxidase subunit 1
CT – computed tomography
CytB – cytochrome b gene
DNA – deoxyribonucleic acid
dsDNA – double stranded DNA
ESS – effective sample sizes
FDR – false discovery rate
F_{ST} – fixation index
GC-content – guanine-cytosine content
gDNA – genomic DNA
Gb – gigabases
GEA – gene-environment associations
GWAS – genome wide association studies
IBD – isolation by distance
IUCN – international union for conservation of nature
LxWxH – length by width by height
MCMC – markov chain monte carlo
MHC – major histocompatibility complex
mRNA – messenger RNA
MSC – multispecies coalescent
mtDNA – mitochondrial
Mya – million years ago
NA – not applicable
ng – nanogram
NCBI – national center for biotechnology information
nuDNA – nuclear DNA

ORF – open reading frame

PCA – principal component analysis

PCR – polymerase chain reaction

qPCR – quantitative polymerase chain reaction

RAD – restriction site associated DNA

RIN – RNA integrity number

RNA – ribonucleic acid

SNP – single nucleotide polymorphism

TLR – toll-like receptors

CHAPTER 1: GENERAL INTRODUCTION

1.1 - Adaptation

The central component of Darwin's theory of evolution by natural selection is that an organism's fitness is dependent on how well it is adapted to its environment (Darwin 1859). Adaptive traits are the phenotypic characters that increase fitness by enhancing survival and subsequent reproductive success (Dobzhansky 1956). Famous examples of adaptive traits are the different shapes of the beaks of Darwin's finches on the Galapagos islands, each continually adapting to the environment of the respective island on which they occur (Lack and David 1983; Lamichhaney et al. 2015). However, not all phenotypic traits are adaptive (Lande 1976; Orr 1998), and understanding the evolution of a trait and how it interacts with the environment is crucial to increasing our understanding of the processes that drive adaptation (Dobzhansky 1956).

Convergent traits that have repeatedly developed in response to similar environmental pressures are likely adaptive (Stern 2013). Comparative phylogenetic methods can identify convergent traits and control for phylogenetic history (Harvey and Purvis 1991; Hansen 2014), and provide the evolutionary framework to test the adaptive value of a phenotypic trait (Larson and Losos 1996). Many studies have focused on evolutionary radiations to identify these adaptive traits (Losos 2009; Salzburger 2009; Kocher 2004), as they allow for multiple comparisons in a limited evolutionary time (Hodges and Derieg 2009). Another approach is to look at within-species variation to determine how individuals and populations are adapted to their local environment (Williams 1966; Savolainen, Lascoux, and Merilä 2013; Kawecki and

Ebert 2004). Studying intra-specific convergence in locally adapted traits can reduce the noise of deep phylogenetic history and help elucidate the causes and consequences of adaptation.

1.1.1 - Local adaptation

The main driver of local adaption is natural selection, with gene flow between localities being an antagonistic force (Savolainen, Pyhäjärvi, and Knürr 2007). The strength of natural selection acting on a population is dependent on both the difference in fitness between phenotypes, and core population principles such as generation time and population size (Lande and Barrowdough 1987; Gandon and Michalakis 2002). Dispersal and subsequent gene flow between environmentally different localities will reduce the effects of local adaptation by homogenizing the gene pool (Lenormand 2002). This interplay between natural selection and gene flow dictate the level of local adaptation and can also affect long-term evolutionary processes such as speciation (Butlin 2010).

Local adaptation can be constrained by a lack of genetic variation, pleiotropic gene effects, and by environmental instability. Natural selection acts on phenotypes arising from standing genetic variation and new mutations in the gene pool, but a lack of genetic variation and in turn phenotypic differences between individuals reduces the variation upon which natural selection works (Lacy 1997). This is problematic for species that have declined and lost genetic variation in the process, reducing their adaptive potential. Pleiotropy reduces the rate of local adaptation when directional selection on one trait can be offset by the negative effects of the same underlying allele on other important traits (Otto 2004). In addition, environmental instability can lead to changes in the strength and direction of natural selection over time, diminishing the occurrence of local adaptation (Melbinger and Vergassola 2015).

Local adaptation has classically been studied by means of reciprocal transplants and common garden experiments (Delph 2018), often involving plants that can easily be manipulated and scored for fitness (Leimu and Fischer 2008). For many species and traits, these experiments are not logistically feasible as transplanted individuals may disperse back, or species may not do well in captivity. Naturally occurring convergence of adaptive traits in replicate populations can serve as a natural laboratory to study local adaptation without the need for experimental manipulation (Barrett, Rogers, and Schluter 2008; Rosenblum et al. 2010). The evolutionary history of the species, as decoded from its genetic data, can be used to uncover signatures of

selection and help elucidate the genotype-phenotype connection of local adaptive traits (Rubin et al. 2012; Evans et al. 2014).

1.1.2 - Adaptation genomics

Advances in genomic research have opened up a new array of tools to detect local adaptation, and connect the specific genes in the genome with observable phenotypes (Storz 2005; Stapley et al. 2010). The intricacies of gene transcription, protein-protein interactions and complex phenotypes complicate this connection, but the growing field of functional and adaptation genomics has made huge strides in uncovering the function of genes and the traits they form (Barrett, Rogers, and Schluter 2008; Elmer and Meyer 2011; Stranger, Stahl, and Raj 2011). Most known genotype-phenotype associations stem from studies on humans and from model organisms selected based on their experimental advantages (Ankeny and Leonelli 2011; Mackay 2014). Famous examples include the LCT gene known to produce the lactase enzyme allowing mammals to break down lactose and consume milk, and sickle-cell anemia caused by a mutation in the haemoglobin gene and the first disease for which both the genetic and molecular defect were characterized (Pauling et al. 1949; Ingram 1957; Serjeant 2010). Artificial selection can also help identify the association between phenotypes and their genotype, as the selective force is known and traits usually evolve faster. For example, artificial selection in pigeons was showcased by Charles Darwin as a proof of principle of evolution by selecting and subsequent breeding of desirable traits (Darwin 1859). This same concept has also been applied to investigate the genetic basis of many traits in domestic animals and livestock with a history of artificial selection such as dogs (Ostrander et al. 2017) and pigs (Rubin et al. 2012).

The function of many genes was initially based on model organisms, and although orthologous genes may have different roles in other species, many gene functions are conserved and gene ontology can help us understand their role in non-model organisms (Ashburner et al. 2000). More recently, the increased amount of genomic resources and techniques available has allowed researchers to directly investigate the function of genes and the genetic basis of adaptive traits in other species and clades (Slate 2005; Elmer and Meyer 2011; Funk, Zamudio, and Crawford 2018). For example, the genetic basis of age of sexual maturity was described based on genome-wide association studies (GWAS) in Atlantic salmon (Barson et al. 2015). Studies in both naturally occurring and experimentally crossed colour morphs has identified numerous genes important in explaining the genetic basis of colour in birds and

reptiles (Lopes et al. 2016; Andrade et al. 2019; Toomey et al. 2017). These studies have shown that intraspecific phenotypic and genetic variation can be used to investigate the functional genetic basis of adaptive traits in non-model clades.

Unravelling the genetic underpinnings of adaptations allow us to ask many questions concerning those adaptations. We can characterize genetic signatures of selection and connect them with environmental conditions. For example, strong selection was found on the lactase gene in humans, which largely corresponded with a concurrent shift to dairy farming in European populations (but see Ségurel and Bon 2017). This can even extend to co-evolutionary dynamics where the signatures of selection are also found on lactose producing genes of co-occurring cattle breeds during the same period (Beja-Pereira et al. 2003). Understanding the genes involved in a phenotype also help in understanding the molecular pathways responsible for this phenotype. Lactose persistence in human adults can be traced back to Single Nucleotide Polymorphisms (SNPs; Enattah et al. 2002; Tishkoff et al. 2007) that have been experimentally shown to enhance lactase production (Troelsen et al. 2003). Knowledge of the important molecular pathways responsible for a phenotype may in turn improve our understanding of the ecological and environmental uses and impacts of adaptive traits. Including adaptation genomics in the study of non-model organisms and natural populations can thus benefit both the fields of genomics and evolutionary ecology.

1.2 - Amphibians as a study system for adaptation

A vertebrate clade that has not yet been widely studied in the field of adaptation genomics is the amphibians. Representing the earliest split among all tetrapods, amphibians have over 300 million years of unique evolution compared to the amniotes. Amphibians are remarkably diverse; as of May 2020, there are over 8,160 described species (AmphibiaWeb 2020) representing about 12% of all vertebrate species. Amphibians are found in a wide variety of habitats and on all continents except Antarctica (but see Mörs, Reguero, and Vasilyan 2020). Due to extensive cryptic species diversity, new genetic data, bioacoustics, and micro CT-scans are fostering species discovery and substantially increasing the number of species known to science (Köhler et al. 2005; Tapley et al. 2018; Vieites et al. 2009; Rakotoarison et al. 2015). Amphibians also display huge phenotypic diversity with a large variety of adaptations in life-history, reproduction, behaviour and many other traits that are not present in other tetrapod orders (Duellman and Trueb 1994). Examples of interesting adaptations in amphibians that merit further exploration include: the biphasic life-history (larvae to metamorph; Schoch 2009), tissue regeneration (McCusker and Gardiner 2011), skin toxins (Clarke 1997), reproductive strategies (Crump 2015; Zamudio et al. 2016) and fungal infection resistance (Ellison et al. 2014).

Amphibians are also the most threatened group of vertebrates, with 41% of all species in classified as in peril by the IUCN (Hoffmann et al. 2010). They, however, receive a disproportionately low level of funding for conservation (Lawler et al. 2006). The reasons for their decline are numerous, including infectious diseases (Scheele et al. 2019), habitat destruction (Cushman 2006) and invasive species (Falaschi et al. 2020). An estimated 156 species have already gone extinct since 1500, and this is likely an undercount (Stuart et al. 2004). Their importance in ecosystem dynamics make this both an ethical and economical concern (Halliday 2008; Whiles et al. 2006). Genetic research can help us identify species and populations in peril (Schoville et al. 2011), in addition to increasing our understanding of their adaptive traits and their function in the ecosystem. It may also help us determine if amphibians are able to adapt to the changing environment and survive the numerous threats they are facing (Pabijan et al. 2020).

Phenotypic diversity and conservation implications make amphibians a prime target for studies on adaptation genomics. They are also relatively easy to sample given their typically large population sizes and characteristics such as site-specific breeding and calling behaviour. With

the exception of the model frog *Xenopus* and model salamander *Ambystoma* they are, however, not common in experimental studies and thus finding the genetic underpinnings of amphibian adaptations would not be possible without applying genomic tools to natural populations.

1.2.1 - Amphibian genetics

Studies on amphibian genetics have long relied on allozymes, traditional Sanger sequencing of one or a few loci, and microsatellites (also known as short tandem repeats). Applying phylogenetic methods and DNA barcoding to amphibians has identified considerable cryptic species diversity (Funk, Caminer, and Ron 2012; Fouquet et al. 2007; Highton 2000; Vences et al. 2005) and improved our understanding of the evolutionary relationships of the different amphibian clades (Hay et al. 1995; Yuan et al. 2016). Microsatellites have been extensively applied to studies of amphibian population genetics (Newman and Squire 2001; Savage, Becker, and Zamudio 2015; Lourenço et al. 2017), paternity analyses (Myers and Zamudio 2004; Steinfartz et al. 2006), and conservation genetics (Jehle and Arntzen 2002; Beebee 2005).

The development of next-generation sequencing methods capable of generating genome-wide data has been slower in amphibians compared to other vertebrate groups (Weisrock et al. 2018). The lack of genomic resources arises from amphibians' large and repetitive genomes, which require more sequencing, expensive long-read technologies, and sophisticated bioinformatic pipelines to assemble. This deficit is hampering large scale phylogenetic comparative methods due to the important position of amphibians at the root of all tetrapods (Vandebergh and Bossuyt 2012). Of the three orders of amphibians (Anura, Caudata and Gymnophiona), only anurans contain high quality chromosome level genomes. As such, high quality reference genome for the other two orders has been called a priority (Koepfli et al. 2015). Given the importance of genomic data in order to investigate patterns of adaptation and the genetic basis of adaptive traits, amphibians could benefit hugely from increased genomic resources (Shaffer et al. 2015; Funk, Zamudio, and Crawford 2018; Pabijan et al. 2020).

1.2.2 - Amphibian genomics

1.2.2.1 - The genomic revolution and amphibian stagnation

Sequencing has produced whole genomes for many species, with numerous more at various stages of assembly and completeness. Based on the 241 vertebrates genomes available in the Ensembl database http://www.ensembl.org as of May 2020 (Ensembl release 100, Yates et al. 2020), the availability of tetrapod genomic resources are concentrated in mammals (104) and to a lesser extent birds (40), and this bias increases when considering the number of unique individuals sequenced (e.g. see McManus et al. 2015; MacDonald et al. 2014). Lagging behind are the non-avian reptiles with 18 annotated genomes, and trailing at the far end are the amphibians with only one Ensembl genome (*Xenopus tropicalis*). The UniProt protein database as of May 2020 is also underrepresented with only 257,183 amphibian sequences compared to 3,494,270 for mammals, a less speciose clade. Additionally, nearly 50% of amphibian protein sequences are based on the genus *Xenopus*, and there is thus a lack of diversity across the other orders and families of this deep clade.

There are many more published and unpublished genomes that do not yet meet Ensembl criteria, but the ratio of genomes by clade is likely similar. These numbers are, however, not reflective of the diversity of species found in these clades (Bonnet, Shine, and Lourdais 2002). The bias in vertebrate genomic resources for reptiles and amphibians is partly explained by the general taxonomic bias in biological research (Bonnet, Shine, and Lourdais 2002; Hecnar 2009). This bias is mostly caused by an increased interest of humans in similar species as well as economically important species (Hecnar 2009). The stagnation in amphibian genomic resources specifically, appears to be mainly due to their extremely large genome sizes that have impeded sequencing studies (see Figure 1.1).

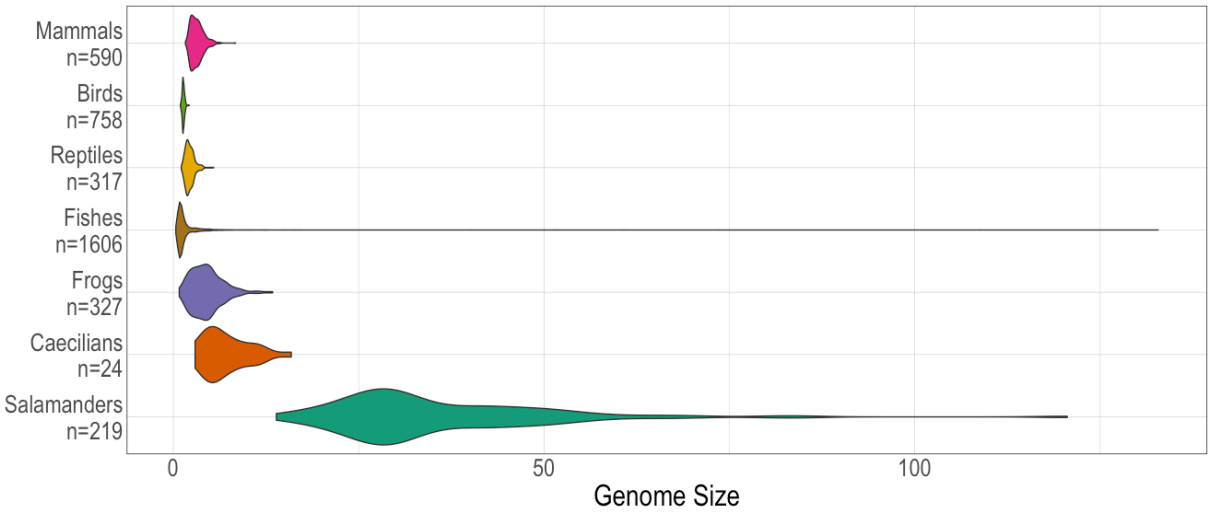
1.2.2.2 - Large genomes of amphibians

Within vertebrates, most amphibian genomes outsize all other clades (Figure 1.1; Sessions 2008; Gregory 2003). On the extreme end, as measured by Fuelgen densitometery, some salamanders in the genus *Necturus* have genomes larger than 100 Gigabases (Gb; roughly 40 times larger than a human genome), whereas some spadefoot toads have genomes similar in size to birds at around 1 Gb (Olmo 1973). The large differences in genome size found in vertebrates does not follow developmental complexity or the number of genes, and this has

been called the C-value paradox (Gregory 2001a). Although there are correlations with factors that appear to influence genome size evolution such as cell size (Gregory 2001b), metabolic rate (Vinogradov 1997) and development speed (Lertzman-Lepofsky, Mooers, and Greenberg 2019; Jockusch 1997; Sessions and Larson 1987), amphibian genome size generally follows a pattern of Brownian motion (Liedtke et al. 2018).

The number of protein coding genes between vertebrate species does not differ nearly as widely as total genome size, and most of the difference thus consists of non-coding sequence (Bird 1995). The large genome sizes seen in amphibians are partly due to longer introns (Smith et al. 2009; Nowoshilow et al. 2018), but can mostly be attributed to genome polyploidization (Mable, Alexandrou, and Taylor 2011), and transposable elements (Sun and Mueller 2014). Both genome duplications and transposable elements do not only increase the amount of sequencing needed, they can also create large problems with genome assembly (Keinath et al. 2015), and paralogs (Mulder et al. 2019), as sequence repeats complicate the assembly process (Treangen and Salzberg 2012).

Figure 1.1. Violin plots of genome size by major vertebrate group, with a focus on amphibians. Genome size data was downloaded from the Animal Genome Size database (Gregory 2020) and from Liedtke et al. (2018). If several scores were available for the same species, the average score was taken. All non-tetrapod vertebrates were combined under the group fish.



The first sequenced amphibian genome, *Xenopus tropicalis*, has a genome size of 1.5 Gb, about half the size of the human genome (Hellsten et al. 2010). This made it feasible to sequence, but also means it is not representative of most amphibian genomes. Some of the

more recently sequenced genomes also sit at the lower end of the amphibian range with *Xenopus laevis* at 3.1 Gb and *Nanorana parkeri* at 2.5 Gb (Sun et al. 2015; Session et al. 2016). From the amphibian genomes that were initially in the G10K pipeline (Koepfli et al. 2015) all but one (*Oophaga pumilio)* were below 5Gb, and notably still did not include any urodeles. The original genome 10K project as set up in 2009 (Haussler et al. 2009) included two urodeles (*Cynops orientalis* and *Andrias davidianus*), but both were removed. These choices with regards to species are a direct consequence of the sequencing and assembly costs of large genomes, forcing scientists to make trade-offs with limited funding.

Since the start of this doctoral project two urodele genomes have been sequenced; *Pleurodeles waltl* (Elewa et al. 2017) and *Ambystoma mexicanum* (Nowoshilow et al. 2018; Smith et al. 2019), both species that are important models for tissue regeneration (McCusker and Gardiner 2011; Matsunami et al. 2019). There have also been at least eight anuran genomes published (Table 1.1): *Xenopus tropicalis* (Hellsten et al. 2010), *Xenopus laevis* (Session et al. 2016), *Nanorana parkeri* (Sun et al. 2015), *Oophaga pumilio* (Rogers et al. 2018), *Rana catesbeiana* (Hammond et al. 2017), *Rhinella marina* (Edwards et al. 2018), *Vibrissaphora ailaonica* (Y. Li et al. 2019) and *Leptobrachium leishanense* (J. Li et al. 2019). Several more amphibian genomes have been sequenced and uploaded to online databases such as NCBI (e.g. the first caecilian *Geotrypetes seraphini*), but the results have not yet been published in peer reviewed journals. As can be noted in Table 1.1, new sequencing and scaffolding technologies such as PacBio and Hi-C greatly increase genome assembly N50 scores, a metric of assembly quality (Ferrarini et al. 2013; Servant et al. 2015). Amphibian whole genome data is increasing both in number and quality, but many studies of adaptation in amphibians without a reference genome must still rely on reduced presentation libraries for genome-scale studies.

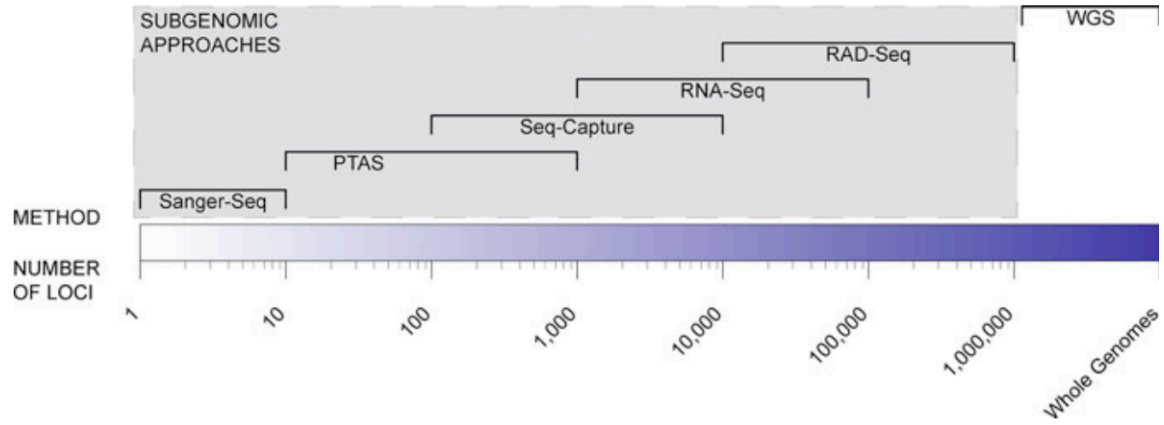
Table 1.1. List of amphibian genomes that have been sequenced and published in peer reviewed journals as of May 2020. Not all methods of calculating genome size and the number of protein coding genes are equal so technical disparities may exist between species.

Species	Published	Genome size (Gb)	Assembly (Gb)	Contig N50 (Kb)	Scaffold N50 (Mb)	Protein coding genes	Method
Xenopus tropicalis	2010	1.5	1.5	71.0	135	27,047	Sanger + BAC
Nanorana parkeri	2015	2.3	2.1	8.1	1.05	21,477	Illumina
Xenopus laevis	2016	3.1	2.7	19.7	136	37,385	Illumina + BAC
Rana catesbeiana	2017	5.8	5.8	5.3	0.05	22,000	Illumina
Pleurodeles waltl	2017	19.4	19.7	1.1	x	19,903	Illumina
Oophaga pumilio	2018	6.8	5.5	0.4	0.07	17,051	Illumina
Rhinella marina	2018	2.4	2.6	0.6	0.19	25,846	Illumina, PacBio
Ambystoma mexicanum	2018	32	32	216.0	3	23,251	PacBio, Bionano
Vibrissaphora ailaonica	2019	3.5	3.5	821	412	26,227	Illumina, PacBio, Hi-C
Leptobrachium leishanense	2019	3.6	3.5	1930	395	23,420	Illumina, PacBio, Hi-C

1.2.3 - Genomic techniques to study adaptation in amphibians

In addition to the promising results in whole genome assembly, new genomic library preparation methods focussing on specific subsets of the whole genome (sub-genomic or reduced representation sequencing) have improved and been optimized for the large genomes of amphibians (Figure 1.2; Weisrock et al. 2018). As sequencing costs have also dropped steadily (Ekblom and Galindo 2011; Steiner et al. 2013), population-level sampling for large amphibian genomes is now more feasible and affordable. These improvements are reducing costs of laboratory work, increasing specificity of reduced representation libraries, and improving the quality of genomic analyses for species without a reference genome. These sub-genomic techniques have different advantages and disadvantages depending on the number of markers needed (Figure 1.2) and the type of research question postulated. I summarize the three intermediate methods; Sequence Capture, RNA-seq and RAD-seq below, including information on how they relate to amphibian genomics and questions on adaptation.

Figure 1.2. Figure from Weisrock et al. (2018), showing different genomic techniques and the approximate number of markers they generate. From low to high number of loci; Sanger sequencing, PTAS (Parallel Tagged Amplicon Sequencing), Sequence Capture, RNA sequencing, RAD sequencing, WGS (Whole Genome Sequencing).



1.2.3.1 - RAD-seq

Restriction site Associated DNA-sequences (RAD) libraries were one of the first library preparation methods to be routinely applied to non-model organisms (Ekblom and Galindo 2011; Baird et al. 2008; Andrews et al. 2016). There are different varieties of RAD-based library preparation (Franchini et al. 2017; Peterson et al. 2012; Wang et al. 2012; Bayona-Vásquez et al. 2019), but they all focus on targeting the same subset of the genome across different individuals by means of restriction enzymes. Originally, there were challenges with applying RAD-seq to species with large genomes because of the limitations of rare cutting enzymes and the presence of repetitive elements that can steal huge chunks of sequencing effort (Wielstra et al. 2014). The development of double-digest RAD-seq (ddRAD; Peterson et al. 2012) has removed some of these limitations and it has now successfully been applied to larger genomes (Gupta et al. 2015; Streicher et al. 2014; Nunziata et al. 2017; Murphy et al. 2018; Hime et al. 2019; Dinis et al. 2019; Burgon et al. 2020). Additional improvements in the size-selection step by means of new machines such as the PippinPrep (automated size selection through gel electrophoresis), have also improved the number of overlapping loci across individuals.

Although RAD libraries can create 1000's of markers for population-level sampling, the loci are spread across the genome and thus by chance most of them will be non-coding and far from the coding sequences that are more likely to be important for the adaptive phenotypes.

As there is still a lack of reference genomes, it is not always possible to map the markers to genomic space. RAD based libraries also require high molecular weight DNA to be successful, and can thus not be applied to museum samples. It is also important to consider and filter for potential paralogs (Mulder et al. 2019) and allele dropout (Gautier et al. 2013). Studies have used RAD based markers to find the functional bases of adaptive traits with success (Burgon et al. 2020), but meta-analyses have found that the median density of RAD-based markers is three orders of magnitudes higher than average linkage disequilibrium, and thus RAD-based studies are likely missing loci that are under selection (Lowry et al. 2017). The majority of amphibian RAD studies are still focussed on population- and phylogenetic questions, but RAD markers are an affordable way to collect genome-wide data for initial studies on the phylogenetic background of adaptive traits and can successfully be applied to large genomes (Weisrock et al. 2018).

1.2.3.2 - RNA-seq

Although genomes vary greatly in their sizes, this is mostly not reflected in an increase in the size of the transcriptome and the number of protein-coding genes (Table 1.1, Hahn and Wray 2002; Taft, Pheasant, and Mattick 2007). The *Xenopus* genome is estimated to have about 20,000 genes which is in the same ballpark as the human genome (Hellsten et al. 2010). Other amphibian transcriptomes that have been published since are all similar in size ranges (e.g. Abdullayev et al. 2013; Habermann et al. 2004; Qiao et al. 2013; Robertson and Cornman 2014; Rodríguez et al. 2017). By targeting only the cDNA of large genomes by means of RNA sequencing (RNA-seq), the most interesting parts of the genome for adaptation, coding regions can be separated from the larger non-coding part of the genome. RNA-seq data also includes relative expression of the different genes and transcripts which can directly be harnessed for questions of differential expression between adaptive traits (Wolf 2013; Pastenes et al. 2017; Savage et al. 2020). The data can also be used to genotype samples, although it is important to control for the biases associated with RNA-seq data (e.g. see Quinn et al. 2013; Lopez-Maestre et al. 2016; Zieliński et al. 2014). The different read depth does make RNA relatively inefficient for marker discovery, although this can be alleviated by enzymatically normalizing the cDNA before sequencing (Christodoulou et al. 2011). The conserved nature of coding sequences also makes them good markers for phylogenetic questions as it facilitates the identification of homologous loci (Rodríguez et al. 2017; Irisarri et al. 2017).

A downside to RNA sequencing is that RNA is less stable and thus harder to collect, store and sequence (Passow et al. 2019), and using RNA-seq from natural populations introduces more biological variance for differential expression analyses (Todd, Black, and Gemmell 2016). Annotation of transcripts can also be challenging due to the lack of reference transcriptomes and genomes (Todd, Black, and Gemmell 2016), although this is improving rapidly for amphibians. Overall, RNA-seq holds great promise for questions on adaptation in non-model organisms and large genomes due to the lack of genomic resources needed, the functional insights provided by differential expression analyses, and the default focus on the coding part of the genome.

1.2.3.3 - Exon-based sequence capture

Once a reference transcriptome or genome has been generated it can be used to develop probes targeting (a subset of) all the exons of the genome (= exome). This is also known as exome capture or exon-based sequence capture and has the twofold advantage that the same subset of loci can be targeted for many individuals, and that it targets the protein coding part of the genome (Hodges et al. 2007). Comparable data allows us to make inferences between individuals, populations and even species (Bi et al. 2012), and by looking at the exons we are immediately targeting potential SNPs directly involved in adaptations. As the technique is not RNA based and does not require high molecular weight DNA, it can be used on a wide range of sample qualities, including on DNA extracted from museum specimens (Ruane and Austin 2017; Cassin-Sackett, Callicrate, and Fleischer 2019; Bi et al. 2013).

Sequence capture enriches the libraries for the selected loci, but the efficiency of enrichment is partly dependent on the initial ratio of target loci to background loci. In large genomes this ratio is inherently lower, and especially the repetitive sequences in the genome can take up substantial parts of the sequencing. Increasing the amount of input DNA and repetitive DNA blocker can improve this efficiency (McCartney-Melstad, Mount, and Shaffer 2016), making it a feasible approach for large amphibian genomes. This makes exome capture an exciting method to look at the genetic underpinning of adaptation in species with large genomes.

1.2.4 - Exploring local adaptation using genomic data

Genomic data and its associated adaptive phenotypic traits can be analysed in a variety of different ways (Hohenlohe, Phillips, and Cresko 2010; Tigano and Friesen 2016). The conceptual framework of these analyses do not differ much when applied to amphibians or large genomes. All have certain benefits and limitations, and often a combination of methods can provide the best evidence of local adaptation and identify candidate genes associated with the trait of interest.

1.2.4.1 - Phylogenetic reconstructions

An important step to investigate any adaptive trait is a better understanding of the evolutionary history of the clade and the trait of interest. Characterizing the phylogenetic relationships between different populations and how phenotypes vary across the tree can identify shared and independent transitions in trait space. Phenotypes that repeatedly evolved are likely adaptive and the timing and location of these transitions can help us understand the evolutionary pressures that might be selecting for it. The options for phylogenetic reconstruction and ancestral state reconstructions are numerous and are extensively discussed in many articles (e.g. Cunningham 1999; Duchêne and Lanfear 2015). Phylogenetic methods have, for example, been applied to squamates to estimate the number and timing of transitions and reversals between reproductive modes, to better understand factors influencing reproductive mode shifts (Griffith et al. 2015; Gomez-Mestre, Pyron, and Wiens 2012).

1.2.4.2 - Differential gene expression analyses

Phenotypic differences are rooted in differences in mRNA expression, and differential expression analyses between phenotypes can identify candidate genes responsible for the observed differences. It is important to have sufficient replicates to identify the signal from biological variation (Schurch et al. 2016), and to control for other potential environmental biases (Todd, Black, and Gemmell 2016; Passow et al. 2019). A recent improvement in gene expression studies has been the use of quasi mapping with programmes such as Kallisto (Pimentel et al. 2017) and Salmon (Zhang et al. 2017), which are both faster and more accurate in estimating expression than the traditional full-mapping approaches. Numerous

methods have been developed to identify differentially expressed genes such as EdgeR (Robinson, McCarthy, and Smyth 2009), DESeq2 (Love, Huber, and Anders 2014) and Sleuth (Pimentel et al. 2017). Studies applying differential expression analyses to adaptive phenotypes have highlighted many candidate genes (McGaugh et al. 2014; Burgon et al. 2020). Although the genetic variation causing these expression differences may be located in other genes or non-coding regions such as gene promotors, some studies have also associated expression differences with specific SNPs mined directly from the RNAseq data (Brown et al. 2018).

1.2.4.3 - F_{ST} outliers

Population structure and differences in gene flow will dictate the average genetic differentiation between populations, often calculated as the fixation index or F_{ST}. If positive selection favours specific alleles for different environments, the genetic differentiation at that locus is expected to be higher than the average differentiation between the populations, whereas balancing selection on loci reduces the differentiation between populations. Using genome-wide SNPs, we can calculate the average F_{ST} and statistically identify any potential outliers (Narum and Hess 2011). This idea has been implemented in programmes such as BayeScan (Foll and Gaggiotti 2008) and OutFLANK (Whitlock and Lotterhos 2015). These software programmes can be run across all populations and loci to identify potential outliers, but can also specifically be applied to certain phenotypic comparisons to associate the outliers with certain adaptive traits. F_{ST} outlier analyses has already been successfully applied to populations of black spruce that also have large genomes (~16Gb; Prunier et al. 2017), illustrating that this method is appropriate for amphibian genomes. Although it is important to control for population structure and demographic history (Narum and Hess 2011), F_{ST} outlier analyses are great tools to identify candidate genes in population level genomic data.

1.2.4.4 - Gene-environment associations

Local adaptation favouring certain alleles in certain environmental conditions can produce patterns that can be identified using gene-environment associations (GEA) methods (Forester et al. 2018). A strong benefit of GEA methods is that they do not require population-level sampling as is generally needed for F_{ST} outliers, and can thus be applied to datasets with more sparse and spread-out sampling. Environmental data can be both locally collected, or

be based on worldwide databases such as Landsat or WorldClim (Fick and Hijmans 2017). They can also be focussed on specific questions of adaptation to environmental conditions, e.g. the evolution of cold tolerance by adaptation to colder environments in anoles (Prates et al. 2018; Campbell-Staton, Edwards, and Losos 2016).

1.2.4.5 - Codon-based signatures of selection

The codon-based translation from mRNA to amino acids means that natural selection acting on the protein does not apply to all base pairs in the same way. Certain genetic mutations do not change the amino acid sequence (synonymous or silent mutations) whereas others do (non-synonymous or missense mutations). Estimating the ratio of non-synonymous to synonymous mutations (called dN/dS between species or pN/pS for intraspecific variation) across a gene can thus tell us if the gene, or parts of the gene, are experiencing positive or balancing selection. This approach is often applied to studies of highly variable markers such as MHC (e.g. Mulder et al. 2017), but can also be applied to genome-wide marker datasets (Oleksyk, Smith, and O'Brien 2010).

1.2.4.6 - Tajima's D

Another statistic that can identify potential signatures of selection is Tajima's D, defined as the difference between the number of pairwise differences and the number of segregating sites (Tajima 1989). A negative Tajima's D occurs when rare alleles are more abundant than expected and this is indicative of either a population expansion or purifying selection, whereas a positive Tajima's D occurs when rare alleles are scarce and this can be the result of population declines or balancing selection. Population declines and expansions should influence Tajima's D similarly across the genome, whereas selection differs between loci. Both the dN/dS ratio and Tajima's D can screen genes for potential signatures of selection and be indicative of adaptation (Biswas and Akey 2006).

1.3 - Amphibian study systems to assess genomic techniques

As new genomic methods are being optimized for organisms with large genomes, it is now possible to look at specific adaptations and functional traits in amphibian systems. Two unique systems that have been extensively studied from a physiological and ecological perspective are the evolution of reproductive mode transitions in the fire salamander (*Salamandra salamandra*), and host-pathogen interactions between the emerging infectious disease chytridiomycosis and the lowland leopard frog (*Rana yavapaiensis*). Although much is known about these systems based on long-term research projects, they have not yet been studied using genomic tools. Both display intra-specific variation in functionally important traits that are likely adaptive, but the genomic basis of these traits are unknown. Given the extensive background knowledge already gathered, these are ideal systems to investigate the genomic basis of adaptation with newly developed genomic tools.

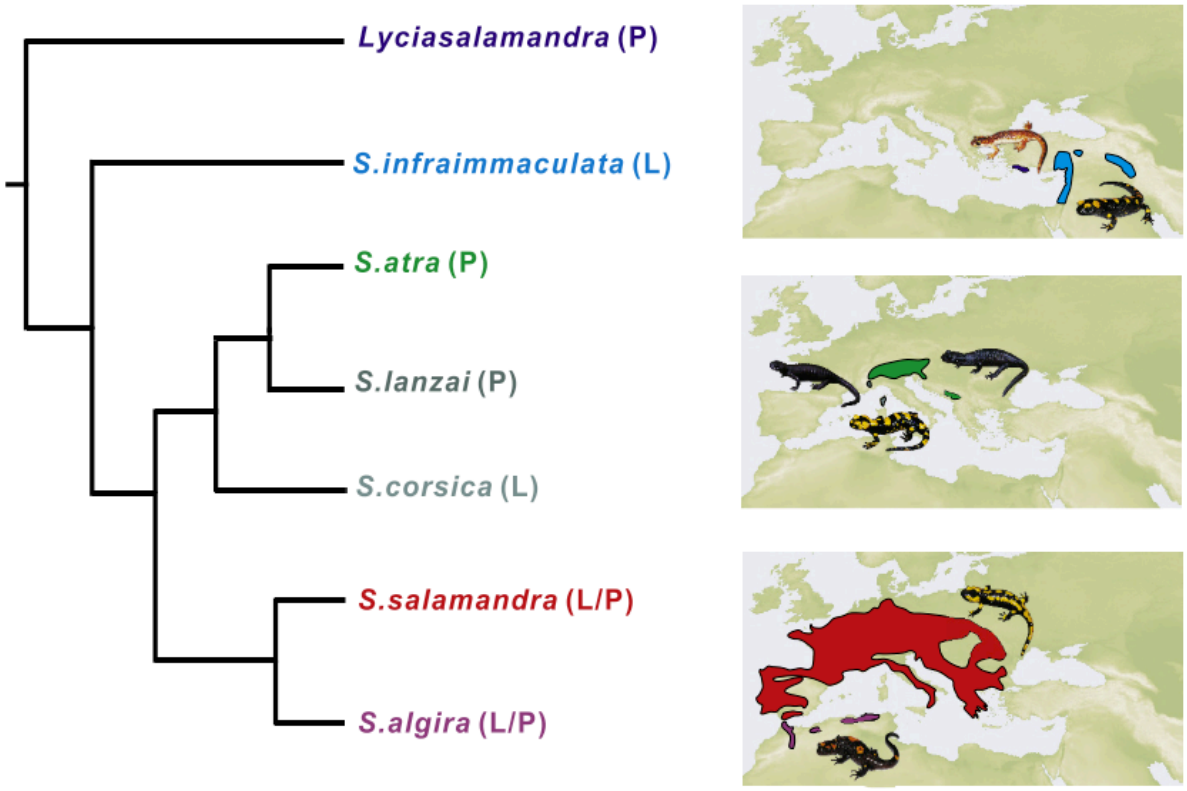
1.3.1 - *Salamandra salamandra* and the evolution of pueriparity

The fire salamanders of the genus *Salamandra* (Linnaeus 1758) are widely distributed across Europe, parts of the Middle East, and northern Africa. They belong to the family Salamandridae and are sister to the genus *Lyciasalamandra*. *Salamandra* contains six described species; the two alpine fire salamanders *S. atra* and *S. lanzai* restricted to alpine regions, *S. corsica* found only on the French island of Corsica, *S. infraimmaculata* with a patched distribution in the Middle East, *S. algira* restricted to northern Africa, and the namesake of the genus *S. salamandra* widely distributed across much of Europe.

Phylogenetic relationships between the different species and subspecies in the genus have been difficult to resolve. Discrepancies between nuclear and mitochondrial phylogenies (Steinfartz, Veith, and Tautz 2000; Vences et al. 2014; Veith et al. 1998), in addition to lack of nuclear data on many subspecies (Beukema et al. 2016), has complicated comparative studies of adaptive traits in the genus. Recent phylogenetic analyses combining multiple lines of evidence stemming from mitochondrial genomes, transcriptomes and ddRAD data reconstructed the species tree of the genus (Figure 1.3; Rodríguez et al. 2017), highlighting the extensive evidence of introgression and/or incomplete lineage sorting complicating the phylogenetic inference. The presence of intraspecific variation in adaptive traits, even within subspecies (Beukema et al. 2016; Velo-Antón et al. 2007) still needs to be resolved before

comparative studies of adaptive traits can be executed based on robust phylogenetic background, especially for the diverse and widespread *S. salamandra*.

Figure 1.3. Figure from Rodriguez et al (2017), depicting the phylogenetic relationships between the different species in the genus Salamandra, and their distributions. Reproductive mode indicated in the brackets (P for Pueriparous, L for larviparous).



Much of the genetic and phenotypic variation within *S. salamandra* is based in the Iberian Peninsula, likely a result of climatic refugia and biogeographic history of the region (García-París et al. 2003; Pereira, Martínez-Solano, and Buckley 2016; Beukema et al. 2016; Abellán and Svenning 2014). Incomplete sampling across subspecies and contact zones, in addition to the lack of large nuclear genetic datasets, has complicated resolving these intra-specific relationships. Lack of genome-wide genetic data is partly due to the large genomes of this genus, estimated to be between 27 and 41 Gb (Gregory 2020), and thus new genomic techniques that can handle large genomes need to be applied to resolve these relationships.

One of the most interesting adaptive traits in need of study is the variation in reproductive mode, with two separate types of viviparous reproduction found in the genus: larviparity and pueriparity (Greven 2003). Larviparous *Salamandra* deliver larvae into nearby waterbodies,

and metamorphosis occurs following the larval stage. Pueriparous females skip this larval stage and deliver fully developed terrestrial juveniles. Pregnancy in both modes takes about 90 days with pueriparous embryos growing faster (Buckley et al. 2007), but the number of offspring is reduced for pueriparous individuals due to presence of unfertilized eggs serving as nutrients and cannibalism of siblings (Buckley et al. 2007; Velo-Antón et al. 2015). This adaptive trait is hypothesized to have evolved due to limited water availability for larval delivery, but comparative data is lacking and genomic tools have not been applied to this system yet. Investigating the multiple independent transitions to pueriparity using genomic tools can help identify the genetic basis of the adaptive transition to pueriparity, and increase our understanding of the evolutionary framework in which it evolved.

1.3.2 - *Rana yavapaiensis* and the evolution of disease resistance

The genus *Rana* in the family of true frogs (Ranidae) has been revised multiple times based on morphological and genetic data, and some of the phylogenetic and taxonomic relationships are still in dispute. Different classifications are used by two major amphibian taxonomic entities. The 6th edition of the Amphibian Species of the World recognises the genus *Lithobates*, encompassing 50 species of new world frogs (Frost 2020). AmphibiaWeb instead classifies *Lithobates* as a subgenus within *Rana*, which itself is recognized as a genus of 105 species, including both old and new world frogs (AmphibiaWeb 2020), based on genetic data from Yuan et al. (2016). There is also cryptic diversity within the clade and likely multiple undescribed species (Yuan et al. 2016; Zaldívar-Riverón, León-Regagnon, and Nieto-Montes De Oca 2004; Hillis, Frost, and Wright 1983; Newman et al. 2012). Because it is based on the most robust phylogenetic analysis to date, we will follow the AmphibiaWeb classification for the remainder of this thesis.

Within *Rana* the majority of species are classified by IUCN as Least Concern (52), with the remaining species being Near Threatened (7), Vulnerable (14), Endangered (8), Critically Endangered (6) Data Deficient or unclassified (17), and *Rana fisheri* is thought to be extinct. Species across the genus are suffering from a variety of threats, including the emerging infectious disease chytridiomycosis (Bradley et al. 2002; Schlaepfer et al. 2007), caused by the invasive global pandemic lineage of the fungus *Batrachochytrium dendrobatidis* (Bd; Scheele et al. 2019). Interestingly, there are differences in susceptibility between species in the genus, varying from the American bullfrog *Rana catesbeiana* which is largely resistant and potentially a vector (Daszak et al. 2004), *Rana chiricahuensis* which is highly susceptible and

critically endangered (Sredl and Jennings 2005; Savage et al. 2018) and the lowland leopard frog, *Rana yavapaiensis* which shows differences in susceptibility across its range (Savage, Sredl, and Zamudio 2011). Differences in susceptibility clearly constitute an adaptive trait, and the variation found within *Rana yavapaiensis* can serve as a natural laboratory to investigate the genetic basis of disease susceptibility.

Rana yavapaiensis is a medium-sized ranid frog that inhabits the southwestern US and northern parts of Mexico. Within the US, populations are locally abundant but some populations have also seen sharp declines (Sredl 1997), for which Bd is at least partly responsible (Bradley et al. 2002). In Arizona, Bd prevalence differs across spatial as well as temporal scales with most outbreaks occurring during the milder temperatures of winter (Schlaepfer et al. 2007; Savage, Sredl, and Zamudio 2011). Bd has been confirmed since at least the early 1990s (Bradley et al. 2002) but there have been anecdotal descriptions of mass die-offs in the 1970s and 80s. This means that most populations are likely in the enzootic phase, and during outbreaks there are significant differences in mortality rates between populations suggesting the evolution of local adaptation (Savage, Sredl, and Zamudio 2011; Savage, Becker, and Zamudio 2015).

Previous research has shown that certain Major Histocompatibility Complex (MHC) genotypes are associated with chytridiomycosis survival (Savage and Zamudio 2011), and that there are differences in immune gene expression between frogs with different disease outcomes (Savage et al. 2020). There is also an extensive research body on immunogenetic variation and the genetic basis of susceptibility to compare to (Hill 2001), including a growing focus on natural populations (Acevedo-Whitehouse and Cunningham 2006). Taken together this system is ideal to investigate the adaptive genetic basis of Bd susceptibility by using genomic tools and genome-wide markers, and identify potential candidate genes important for resistance.

1.4 - Structure and objectives of the thesis

The main objective of this doctoral thesis is to apply and optimize new genomic techniques and analyses to questions of local adaptation in two well studied amphibian study systems that exhibit intra-specific variation in adaptive traits. Both systems have large genomes and have not been previously investigated using genomic tools.

This thesis is divided into five chapters. Chapter 1 is a general introduction to the field of adaptation genomics and how this approach can be applied to amphibian study systems. It briefly discusses the concept of local adaptation and the benefits of intra-specific variation to identify and study adaptive traits. It summarizes new genomic tools and analyses that can be applied to study local adaptation in non-model amphibians with large genomes. Lastly it introduces the two study systems to test these approaches: the fire salamander (*Salamandra salamandra*) and its intra-specific variation in reproductive mode and the lowland leopard frog (*Rana yavapaiensis*) and the variation in disease susceptibility found within the species.

In chapter 2, an exome capture dataset of 1,326 of loci was generated to build a phylogenetic reconstruction of the genus *Salamandra*, with a focus on intra-specific clades within *S. Salamandra* that differ in reproductive mode between larviparity and pueriparity. We mapped reproductive strategy onto the phylogeny to assess the evolutionary history of this adaptive trait. The objectives were to (1) identify the number of independent transitions to pueriparity in the clade, (2) estimate the timing of the transitions, and (3) provide updates on the range of both reproductive modes using DNA barcoding and assessments of live births.

Chapter 3 focused on two independent transitions to pueriparity within *S. salamandra*, the late Pleistocene transition within *S. s. gallaica* on the island of Ons, and the middle Pleistocene transition between *S. s. gallaica* and *S. s. bernardezi*. We collected RNA-seq data from the uterus and oviduct of pregnant females to characterize gene expression differences between both reproductive modes at different temporal scales and in a comparative spatial framework. The objectives were to (1) describe general gene expression patterns of reproductive tissues o*f Salamandra*, (2) characterize gene expression differences between larviparity and pueriparity and distinguish between differential and convergent patterns across the two transitions, and (3) identify candidate genes that may explain the phenotypic adaptive difference between larviparous and pueriparous individuals.

For chapter 4 we applied a transcriptome-based exon capture approach to the *Rana yavapaiensis* system of chytridiomycosis susceptible, resistant and naïve populations. The capture array included both immunogenetic and genome-wide variation of 1,388 loci, and we applied this to 133 individuals across 11 populations. Our objectives were to (1) describe populations genetic structure of *Rana yavapaiensis*, (2) compare several measures of genetic variation between populations to identify the impact of chytridiomycosis, (3) test for general signatures of selection across the genome, and (4) identify genes associated with Bd-resistance that may be related to disease adaptation.

Chapter 5 summarizes the work done in the scope of this thesis and discusses some of the implications on studies of adaptation and applying genomic techniques to large amphibian genomes.

1.5 - References

Abdullayev, Ilgar, Matthew Kirkham, Åsa K. Björklund, András Simon, and Rickard Sandberg. 2013. "A Reference Transcriptome and Inferred Proteome for the Salamander *Notophthalmus viridescens*." *Experimental Cell Research* 319 (8): 1187–97. https://doi.org/10.1016/j.yexcr.2013.02.013.

Abellán, Pedro, and Jens-Christian Svenning. 2014. "Refugia within Refugia - Patterns in Endemism and Genetic Divergence Are Linked to Late Quaternary Climate Stability in the Iberian Peninsula." *Biological Journal of the Linnean Society* 113 (1): 13–28. https://doi.org/10.1111/bij.12309.

Acevedo-Whitehouse, Karina, and Andrew A. Cunningham. 2006. "Is MHC Enough for Understanding Wildlife Immunogenetics?" *Trends in Ecology and Evolution* 21 (8): 433–38. https://doi.org/10.1016/j.tree.2006.05.010.

AmphibiaWeb 2020. University of California, Berkeley, CA, USA. 2020. https://amphibiaweb.org.

Andrade, Pedro, Catarina Pinho, Guillem Pérez i. de Lanuza, Sandra Afonso, Jindrich Brejcha, Carl Johan Rubin, Ola Wallerman, et al. 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): 5633–42. https://doi.org/10.1073/pnas.1820320116.

Andrews, Kimberly R., Jeffrey M. Good, Michael R. Miller, Gordon Luikart, and Paul A. Hohenlohe. 2016. "Harnessing the Power of RADseq for Ecological and Evolutionary Genomics." *Nature Reviews Genetics* 17 (2): 81–92. https://doi.org/10.1038/nrg.2015.28.

Ankeny, Rachel A., and Sabina Leonelli. 2011. "What's so Special about Model Organisms?" *Studies in History and Philosophy of Science Part A* 42 (2): 313–23. https://doi.org/10.1016/j.shpsa.2010.11.039.

Ashburner, Michael, Catherine A. Ball, Judith A. Blake, David Botstein, Heather Butler, J. Michael Cherry, Allan P. Davis, et al. 2000. "Gene Ontology: Tool for the Unification of Biology." *Nature Genetics* 25 (1): 25–29. https://doi.org/10.1038/75556.

Baird, Nathan A., Paul D. Etter, Tressa S. Atwood, Mark C. Currey, Anthony L. Shiver, Zachary A. Lewis, Eric U. Selker, William A. Cresko, and Eric A. Johnson. 2008. "Rapid SNP Discovery and Genetic Mapping Using Sequenced RAD Markers." *PLoS ONE* 3 (10): e3376. https://doi.org/10.1371/journal.pone.0003376.

Barrett, Rowan D.H., Sean M. Rogers, and Dolph Schluter. 2008. "Natural Selection on a

Major Armor Gene in Threespine Stickleback." *Science* 322 (5899): 255–57. https://doi.org/10.1126/science.1159978.

Barson, Nicola J., Tutku Aykanat, Kjetil Hindar, Matthew Baranski, Geir H. Bolstad, Peder Fiske, Céleste Jacq, et al. 2015. "Sex-Dependent Dominance at a Single Locus Maintains Variation in Age at Maturity in Salmon." *Nature* 528 (7582): 405–8. https://doi.org/10.1038/nature16062.

Bayona-Vásquez, Natalia J., Travis C. Glenn, Troy J. Kieran, Todd W. Pierson, Sandra L. Hoffberg, Peter A. Scott, Kerin E. Bentley, et al. 2019. "Adapterama III: Quadruple-Indexed, Double/Triple-Enzyme RADseq Libraries (2RAD/3RAD)." *PeerJ* 2019 (10): 1–25. https://doi.org/10.7717/peerj.7724.

Beebee, Trevor J. C. 2005. "Conservation Genetics of Amphibians." *Heredity* 95 (6): 423–27. https://doi.org/10.1038/sj.hdy.6800736.

Beja-Pereira, Albano, Gordon Luikart, Phillip R. England, Daniel G. Bradley, Oliver C. Jann, Giorgio Bertorelle, Andrew T. Chamberlain, et al. 2003. "Gene-Culture Coevolution between Cattle Milk Protein Genes and Human Lactase Genes." *Nature Genetics* 35 (4): 311–13. https://doi.org/10.1038/ng1263.

Beukema, Wouter, Alfredo G. Nicieza, André Lourenço, and Guillermo Velo-Antón. 2016. "Colour Polymorphism in *Salamandra Salamandra* (Amphibia: Urodela), Revealed by a Lack of Genetic and Environmental Differentiation between Distinct Phenotypes." *Journal of Zoological Systematics and Evolutionary Research* 54 (2): 127–36. https://doi.org/10.1111/jzs.12119.

Bi, Ke, Tyler Linderoth, Dan Vanderpool, Jeffrey M. Good, Rasmus Nielsen, and Craig Moritz. 2013. "Unlocking the Vault: Next-Generation Museum Population Genomics." *Molecular Ecology* 22 (24): 6018–32. https://doi.org/10.1111/mec.12516.

Bi, Ke, Dan Vanderpool, Sonal Singhal, Tyler Linderoth, Craig Moritz, and Jeffrey M Good. 2012. "Transcriptome-Based Exon Capture Enables Highly Cost-Effective Comparative Genomic Data Collection at Moderate Evolutionary Scales." *BMC Genomics* 13 (1): 403. https://doi.org/10.1186/1471-2164-13-403.

Bird, Adrian P. 1995. "Gene Number, Noise Reduction and Biological Complexity." *Trends in Genetics* 11 (3): 94–100. https://doi.org/10.1016/S0168-9525(00)89009-5.

Biswas, Shameek, and Joshua M. Akey. 2006. "Genomic Insights into Positive Selection." *Trends in Genetics* 22 (8): 437–46. https://doi.org/10.1016/j.tig.2006.06.005.

Bonnet, Xavier, Richard Shine, and Olivier Lourdais. 2002. "Taxonomic Chauvinism." *Trends in Ecology and Evolution* 17 (1): 1–3. https://doi.org/10.1016/S0169-5347(01)02381-3.

Bradley, Gregory A., Philip C. Rosen, Michael J. Sredl, Thomas R. Jones, and Joyce E. Longcore. 2002. "Chytridiomycosis in Native Arizona Frogs." *Journal of Wildlife*

Diseases 38 (1): 206–12. https://doi.org/10.7589/0090-3558-38.1.206.

Brown, Anthony P., Lenin Arias-Rodriguez, Muh Ching Yee, Michael Tobler, and Joanna L. Kelley. 2018. "Concordant Changes in Gene Expression and Nucleotides Underlie Independent Adaptation to Hydrogen-Sulfide-Rich Environments." *Genome Biology and Evolution* 10 (11): 2867–81. https://doi.org/10.1093/gbe/evy198.

Buckley, David, Marina Alcobendas, Mario Garcia-Paris, and Marvalee H. Wake. 2007. "Heterochrony, Cannibalism, and the Evolution of Viviparity in *Salamandra Salamandra*." *Evolution and Development* 9 (1): 105–15. https://doi.org/10.1111/j.1525-142X.2006.00141.x.

Burgon, James D., David R. Vieites, Arne Jacobs, Stefan K. Weidt, Helen M. Gunter, Sebastian Steinfartz, Karl Burgess, Barbara K. Mable, and Kathryn R. Elmer. 2020. "Functional Colour Genes and Signals of Selection in Colour-Polymorphic Salamanders." *Molecular Ecology* 29 (7): 1284–99. https://doi.org/10.1111/mec.15411.

Butlin, Roger K. 2010. "Population Genomics and Speciation." *Genetica* 138 (4): 409–18. https://doi.org/10.1007/s10709-008-9321-3.

Campbell-Staton, Shane C., Scott V. Edwards, and Jonathan B. Losos. 2016. "Climate-Mediated Adaptation after Mainland Colonization of an Ancestrally Subtropical Island Lizard, *Anolis carolinensis*." *Journal of Evolutionary Biology* 29 (11): 2168–80. https://doi.org/10.1111/jeb.12935.

Cassin-Sackett, Loren, Taylor E. Callicrate, and Robert C. Fleischer. 2019. "Parallel Evolution of Gene Classes, but Not Genes: Evidence from Hawai'ian Honeycreeper Populations Exposed to Avian Malaria." *Molecular Ecology* 28 (3): 568–83. https://doi.org/10.1111/mec.14891.

Christodoulou, Danos C., Joshua M. Gorham, Daniel S. Herman, and J. G. Seidman. 2011. "Construction of Normalized RNA-Seq Libraries for next-Generation Sequencing Using the Crab Duplex-Specific Nuclease." *Current Protocols in Molecular Biology* 94: 4–12. https://doi.org/10.1002/0471142727.mb0412s94.

Clarke, B. T. 1997. "The Natural History of Amphibian Skin Secretions, Their Normal Functioning and Potential Medical Applications." *Biological Reviews* 72 (3): 365–79. https://doi.org/10.1111/j.1469-185X.1997.tb00018.x.

Crump, Martha L. 2015. "Anuran Reproductive Modes: Evolving Perspectives." *Journal of Herpetology* 49 (1): 1–16. https://doi.org/10.1670/14-097.

Cunningham, Clifford W. 1999. "Some Limitations of Ancestral Character-State Reconstruction When Testing Evolutionary Hypotheses." *Systematic Biology* 48 (3): 665–74. https://doi.org/10.1080/106351599260238.

Cushman, Samuel A. 2006. "Effects of Habitat Loss and Fragmentation on Amphibians: A

Review and Prospectus." *Biological Conservation* 128 (2): 231–40. https://doi.org/10.1016/j.biocon.2005.09.031.

Darwin, Charles. 1859. *On the Origin of Species by Means of Natural Selection, or, the Preservation of Favoured Races in the Struggle for Life*. London: J Murray.

Daszak, Peter, A. Strieby, Andrew A. Cunningham, Joyce E. Longcore, Corrie C. Brown, and David Porter. 2004. "Experimental Evidence That the Bullfrog (*Rana catesbeiana*) Is a Potential Carrier of Chytridiomycosis, an Emerging Fungal Disease of Amphibians." *Herpetological Journal* 14 (4): 201–7.

Delph, Lynda F. 2018. "The Study of Local Adaptation: A Thriving Field of Research." *Journal of Heredity* 109 (1): 1–2. https://doi.org/10.1093/jhered/esx099.

Dinis, Marco, Khaled Merabet, Fernando Martínez-Freiría, Sebastian Steinfartz, Miguel Vences, James D. Burgon, Kathryn R. Elmer, et al. 2019. "Allopatric Diversification and Evolutionary Melting Pot in a North African Palearctic Relict: The Biogeographic History of *Salamandra algira*." *Molecular Phylogenetics and Evolution* 130: 81–91. https://doi.org/10.1016/j.ympev.2018.10.018.

Dobzhansky, Theodosius. 1956. "What Is an Adaptive Trait?" *The American Naturalist* 90 (855): 337–47.

Duchêne, Sebastian, and Robert Lanfear. 2015. "Phylogenetic Uncertainty Can Bias the Number of Evolutionary Transitions Estimated from Ancestral State Reconstruction Methods." *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution* 324 (6): 517–24. https://doi.org/10.1002/jez.b.22638.

Duellman, William E., and Linda Trueb. 1994. *Biology of Amphibians*. JHU Press.

Edwards, Richard J., Daniel Enosi Tuipulotu, Timothy G. Amos, Denis O'Meally, Mark F. Richardson, Tonia L. Russell, Marcelo Vallinoto, et al. 2018. "Draft Genome Assembly of the Invasive Cane Toad, *Rhinella marina*." *GigaScience* 7 (9): 1–13. https://doi.org/10.1093/gigascience/giy095.

Ekblom, Robert, and Juan Galindo. 2011. "Applications of next Generation Sequencing in Molecular Ecology of Non-Model Organisms." *Heredity* 107 (1): 1–15. https://doi.org/10.1038/hdy.2010.152.

Elewa, Ahmed, Heng Wang, Carlos Talavera-López, Alberto Joven, Gonçalo Brito, Anoop Kumar, L. Shahul Hameed, et al. 2017. "Reading and Editing the *Pleurodeles waltl* Genome Reveals Novel Features of Tetrapod Regeneration." *Nature Communications* 8 (1): 1–9. https://doi.org/10.1038/s41467-017-01964-9.

Ellison, Amy R., Anna E. Savage, Grace V. DiRenzo, Penny Langhammer, Karen R. Lips, and Kelly R. Zamudio. 2014. "Fighting a Losing Battle: Vigorous Immune Response Countered by Pathogen Suppression of Host Defenses in the Chytridiomycosis-

Susceptible Frog *Atelopus zeteki*." *G3: Genes, Genomes, Genetics* 4 (7): 1275–89. https://doi.org/10.1534/g3.114.010744.

Elmer, Kathryn R., and Axel Meyer. 2011. "Adaptation in the Age of Ecological Genomics: Insights from Parallelism and Convergence." *Trends in Ecology and Evolution* 26 (6): 298–306. https://doi.org/10.1016/j.tree.2011.02.008.

Enattah, Nabil Sabri, Timo Sahi, Erkki Savilahti, Joseph D. Terwilliger, Leena Peltonen, and Irma Järvelä. 2002. "Identification of a Variant Associated with Adult-Type Hypolactasia." *Nature Genetics* 30 (2): 233–37. https://doi.org/10.1038/ng826.

Evans, Luke M., Gancho T. Slavov, Eli Rodgers-Melnick, Joel Martin, Priya Ranjan, Wellington Muchero, Amy M. Brunner, et al. 2014. "Population Genomics of *Populus trichocarpa* Identifies Signatures of Selection and Adaptive Trait Associations." *Nature Genetics* 46 (10): 1089–96. https://doi.org/10.1038/ng.3075.

Falaschi, Mattia, Andrea Melotto, Raoul Manenti, and Gentile Francesco Ficetola. 2020. "Invasive Species and Amphibian Conservation." *Herpetologica* 76 (2): in press. https://doi.org/10.1655/Herpetologica-D-19-00063.1.

Ferrarini, Marco, Marco Moretto, Judson A Ward, Nada Šurbanovski, Vladimir Stevanović, Lara Giongo, Roberto Viola, et al. 2013. "An Evaluation of the PacBio RS Platform for Sequencing and de Novo Assembly of a Chloroplast Genome." *BMC Genomics* 14 (1): 670. https://doi.org/10.1186/1471-2164-14-670.

Fick, Stephen E., and Robert J. Hijmans. 2017. "WorldClim 2: New 1-Km Spatial Resolution Climate Surfaces for Global Land Areas." *International Journal of Climatology* 37 (12): 4302–15. https://doi.org/10.1002/joc.5086.

Foll, Matthieu, and Oscar Gaggiotti. 2008. "A Genome-Scan Method to Identify Selected Loci Appropriate for Both Dominant and Codominant Markers: A Bayesian Perspective." *Genetics* 180 (2): 977–93. https://doi.org/10.1534/genetics.108.092221.

Forester, Brenna R., Jesse R. Lasky, Helene H. Wagner, and Dean L. Urban. 2018. "Comparing Methods for Detecting Multilocus Adaptation with Multivariate Genotype–Environment Associations." *Molecular Ecology* 27 (9): 2215–33. https://doi.org/10.1111/mec.14584.

Fouquet, Antoine, Miguel Vences, Marie Dominique Salducci, Axel Meyer, Christian Marty, Michel Blanc, and André Gilles. 2007. "Revealing Cryptic Diversity Using Molecular Phylogenetics and Phylogeography in Frogs of the *Scinax ruber* and *Rhinella margaritifera* Species Groups." *Molecular Phylogenetics and Evolution* 43 (2): 567–82. https://doi.org/10.1016/j.ympev.2006.12.006.

Franchini, Paolo, Daniel Monné Parera, Andreas F. Kautt, and Axel Meyer. 2017. "QuaddRAD: A New High-Multiplexing and PCR Duplicate Removal DdRAD Protocol

Produces Novel Evolutionary Insights in a Nonradiating Cichlid Lineage." *Molecular Ecology* 26 (10): 2783–95. https://doi.org/10.1111/mec.14077.

Frost, Darrel R. 2020. "Amphibian Species of the World: An Online Reference. Version 6.1. American Museum of Natural History." 2020.

Funk, W. Chris, Marcel Caminer, and Santiago R. Ron. 2012. "High Levels of Cryptic Species Diversity Uncovered in Amazonian Frogs." *Proceedings of the Royal Society B: Biological Sciences* 279 (1734): 1806–14. https://doi.org/10.1098/rspb.2011.1653.

Funk, W. Chris, Kelly R. Zamudio, and Andrew J. Crawford. 2018. "Advancing Understanding of Amphibian Evolution, Ecology, Behavior, and Conservation with Massively Parallel Sequencing," 1–44. https://doi.org/10.1007/13836_2018_61.

Gandon, Sylvain, and Y. Michalakis. 2002. "Local Adaptation, Evolutionary Potential and Host-Parasite Coevolution: Interactions between Migration, Mutation, Population Size and Generation Time." *Journal of Evolutionary Biology* 15 (3): 451–62. https://doi.org/10.1046/j.1420-9101.2002.00402.x.

García-París, Mario, Marina Alcobendas, David Buckley, and David B. Wake. 2003. "Dispersal of Viviparity across Contact Zones in Iberian Populations of Fire Salamanders (*Salamandra*) Inferred from Discordance of Genetic and Morphological Traits." *Evolution* 57 (1): 129–43. https://doi.org/10.1111/j.0014-3820.2003.tb00221.x.

Gautier, Mathieu, Karim Gharbi, Timothee Cezard, Julien Foucaud, Carole Kerdelhué, Pierre Pudlo, Jean Marie Cornuet, and Arnaud Estoup. 2013. "The Effect of RAD Allele Dropout on the Estimation of Genetic Variation within and between Populations." *Molecular Ecology* 22 (11): 3165–78. https://doi.org/10.1111/mec.12089.

Gomez-Mestre, Ivan, Robert Alexander Pyron, and John J. Wiens. 2012. "Phylogenetic Analyses Reveal Unexpected Patterns in the Evolution of Reproductive Modes in Frogs." *Evolution* 66 (12): 3687–3700. https://doi.org/10.1111/j.1558-5646.2012.01715.x.

Gregory, T. Ryan. 2001a. "Coincidence, Coevolution, or Causation? DNA Content, Cellsize, and the C-Value Enigma." *Biological Reviews* 76 (1): 65–101. https://doi.org/10.1111/j.1469-185X.2000.tb00059.x.

Gregory, T. Ryan. 2001b. "The Bigger the C-Value, the Larger the Cell: Genome Size and Red Blood Cell Size in Vertebrates." *Blood Cells, Molecules, and Diseases* 27 (5): 830–43. https://doi.org/10.1006/bcmd.2001.0457.

Gregory, T. Ryan. 2003. "Variation across Amphibian Species in the Size of the Nuclear Genome Supports a Pluralistic, Hierarchical Approach to the C-Value Enigma." *Biological Journal of the Linnean Society* 79 (2): 329–39. https://doi.org/10.1046/j.1095-8312.2003.00191.x.

Gregory, T. Ryan. 2020. “Animal Genome Size Database.” 2020.

Greven, H. 2003. “Larviparity and Pueriparity.” In *Reproductive Biology and Phylogeny of Urodela*, 447–75. CRC Press.

Griffith, Oliver W., Daniel G. Blackburn, Matthew C. Brandley, James U. Van Dyke, Camilla M. Whittington, and Michael B. Thompson. 2015. “Ancestral State Reconstructions Require Biological Evidence to Test Evolutionary Hypotheses: A Case Study Examining the Evolution of Reproductive Mode in Squamate Reptiles.” *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution* 324 (6): 493–503. https://doi.org/10.1002/jez.b.22614.

Gupta, Sudhir Kumar, Jongmin Baek, Noelia Carrasquilla-Garcia, and R. Varma Penmetsa. 2015. “Genome-Wide Polymorphism Detection in Peanut Using next-Generation Restriction-Site-Associated DNA (RAD) Sequencing.” *Molecular Breeding* 35 (7). https://doi.org/10.1007/s11032-015-0343-0.

Habermann, Bianca, Anne Gaelle Bebin, Stephan Herklotz, Michael Volkmer, Kay Eckelt, Kerstin Pehlke, Hans Henning Epperlein, Hans Konrad Schackert, Glenis Wiebe, and Elly M. Tanaka. 2004. “An Ambystoma Mexicanum EST Sequencing Project: Analysis of 17,352 Expressed Sequence Tags from Embryonic and Regenerating Blastema CDNA Libraries.” *Genome Biology* 5 (9): R67. https://doi.org/10.1186/gb-2004-5-9-r67.

Hahn, Matthew W., and Gregory A. Wray. 2002. “The G-Value Paradox.” *Evolution and Development* 4 (2): 73–75. https://doi.org/10.1046/j.1525-142X.2002.01069.x.

Halliday, T. R. 2008. “Why Amphibians Are Important.” *International Zoo Yearbook* 42 (1): 7–14. https://doi.org/10.1111/j.1748-1090.2007.00037.x.

Hammond, S. Austin, René L. Warren, Benjamin P. Vandervalk, Erdi Kucuk, Hamza Khan, Ewan A. Gibb, Pawan Pandoh, et al. 2017. “The North American Bullfrog Draft Genome Provides Insight into Hormonal Regulation of Long Noncoding RNA.” *Nature Communications* 8 (1): 1–8. https://doi.org/10.1038/s41467-017-01316-7.

Hansen, Thomas F. 2014. “Use and Misuse of Comparative Methods in the Study of Adaptation.” In *Modern Phylogenetic Comparative Methods and Their Application in Evolutionary Biology*, 351–79. Springer.

Harvey, Paul H, and Andy Purvis. 1991. “Comparing Methods for Explaining Adaptations.” *Nature* 351: 619–24.

Haussler, David, Stephen J. O’Brien, Oliver A. Ryder, F. Keith Barker, Michele Clamp, Andrew J. Crawford, Robert Hanner, et al. 2009. “Genome 10K: A Proposal to Obtain Whole-Genome Sequence for 10000 Vertebrate Species.” *Journal of Heredity* 100 (6): 659–74. https://doi.org/10.1093/jhered/esp086.

Hay, Jennifer M., Ilya Ruvinsky, S Blair Hedges, and Linda R Maxson. 1995. “Phylogenetic

Relationships of Amphibian Families Inferred from DNA Sequences of Mitochondrial 12S and 16S Ribosomal RNA Genes." *Molecular Biology and Evolution*, 928–37. https://doi.org/10.1093/oxfordjournals.molbev.a040270.

Hecnar, Stephen J. 2009. "Human Bias and the Biodiversity Knowledge Base: An Examination of the Published Literature on Vertebrates." *Biodiversity* 10 (1): 18–24. https://doi.org/10.1080/14888386.2009.9712633.

Hellsten, Uffe, Richard M. Harland, Michael J. Gilchrist, David Hendrix, Jerzy Jurka, Vladimir Kapitonov, Ivan Ovcharenko, et al. 2010. "The Genome of the Western Clawed Frog *Xenopus tropicalis*." *Science* 328 (5978): 633–36. https://doi.org/10.1126/science.1183670.

Highton, Richard. 2000. "Detecting Cryptic Species Using Allozyme Data." *The Biology of Plethodontid Salamanders*, 215–41. https://doi.org/10.1007/978-1-4615-4255-1_10.

Hill, Adrian V. S. 2001. "The Genomics and Genetics of Human Infectious Disease Susceptibility." *Annual Review of Genomics and Human Genetics* 2 (33): 373–400. https://doi.org/10.1146/annurev.genom.2.1.373.

Hillis, David M., John S. Frost, and David A. Wright. 1983. "Phylogeny and Biogeography of the *Rana pipiens* Complex: A Biochemical Evaluation." *Systematic Zoology* 32 (2): 132. https://doi.org/10.2307/2413277.

Hime, Paul M., Jeffrey T. Briggler, Joshua S. Reece, and David W. Weisrock. 2019. "Genomic Data Reveal Conserved Female Heterogamety in Giant Salamanders with Gigantic Nuclear Genomes." *G3: Genes, Genomes, Genetics* 9 (10): 3467–76. https://doi.org/10.1534/g3.119.400556.

Hodges, Emily, Zhenyu Xuan, Vivekanand Balija, Melissa Kramer, Michael N. Molla, Steven W. Smith, Christina M. Middle, et al. 2007. "Genome-Wide in Situ Exon Capture for Selective Resequencing." *Nature Genetics* 39 (12): 1522–27. https://doi.org/10.1038/ng.2007.42.

Hodges, Scott A., and Nathan J Derieg. 2009. "Adaptive Radiations: From Field to Genomic Studies." *Proceedings of the National Academy of Sciences* 106: 9947–54. https://doi.org/10.1073./pnas.0901594106.

Hoffmann, Michael, Craig Hilton-Taylor, Ariadne Angulo, Monika Böhm, Thomas M. Brooks, Stuart H.M. Butchart, Kent E. Carpenter, et al. 2010. "The Impact of Conservation on the Status of the World's Vertebrates." *Science* 330 (6010): 1503–9. https://doi.org/10.1126/science.1194442.

Hohenlohe, Paul A., Patrick C. Phillips, and William A. Cresko. 2010. "Using Population Genomics to Detect Selection in Natural Populations: Key Concepts and Methodological Considerations." *International Journal of Plant Sciences* 171 (9): 1059–

71. https://doi.org/10.1086/656306.

Ingram, V M. 1957. “Gene Mutations in Human Haemoglobin: The Chemical Difference between Normal and Sickle Cell Haemoglobin.” *Nature* 180 (4581): 326–28. https://doi.org/10.1038/180326a0.

Irisarri, Iker, Denis Baurain, Henner Brinkmann, Frédéric Delsuc, Jean Yves Sire, Alexander Kupfer, Jörn Petersen, et al. 2017. “Phylotranscriptomic Consolidation of the Jawed Vertebrate Timetree.” *Nature Ecology and Evolution* 1 (9): 1370–78. https://doi.org/10.1038/s41559-017-0240-5.

Jehle, R., and J. W. Arntzen. 2002. “Microsatellite Markers in Amphibian Conservation Genetics.” *Herpetological Journal* 12 (1): 1–9.

Jockusch, E. L. 1997. “An Evolutionary Correlate of Genome Size Change in Plethodontid Salamanders.” *Proceedings of the Royal Society B: Biological Sciences* 264 (1381): 597–604. https://doi.org/10.1098/rspb.1997.0085.

Kawecki, Tadeusz J., and Dieter Ebert. 2004. “Conceptual Issues in Local Adaptation.” *Ecology Letters* 7 (12): 1225–41. https://doi.org/10.1111/j.1461-0248.2004.00684.x.

Keinath, Melissa C., Vladimir A. Timoshevskiy, Nataliya Y. Timoshevskaya, Panagiotis A. Tsonis, S. Randal Voss, and Jeramiah J. Smith. 2015. “Initial Characterization of the Large Genome of the Salamander *Ambystoma mexicanum* Using Shotgun and Laser Capture Chromosome Sequencing.” *Scientific Reports* 5 (1): 16413. https://doi.org/10.1038/srep16413.

Kocher, Thomas D. 2004. “Adaptive Evolution and Explosive Speciation: The Cichlid Fish Model.” *Nature Reviews Genetics* 5 (4): 288–98. https://doi.org/10.1038/nrg1316.

Koepfli, Klaus-peter, Benedict Paten, Genome 10K Community of Scientists, and Stephen J O’Brien. 2015. “The Genome 10K Project: A Way Forward.” *Annual Review of Animal Biosciences* 3: 57–111. https://doi.org/10.1146/annurev-animal-090414-014900.

Köhler, Jörn, David R. Vieites, Ronald M. Bonett, Francisco Hita García, Frank Glaw, Dirk Steinke, and Miguel Vences. 2005. “New Amphibians and Global Conservation: A Boost in Species Discoveries in a Highly Endangered Vertebrate Group.” *BioScience* 55 (8): 693–96. https://doi.org/10.1641/0006-3568(2005)055[0693:NAAGCA]2.0.CO;2.

Lack, David, and Lack David. 1983. *Darwin’s Finches*. CUP Archive.

Lacy, R. C. 1997. “Importance of Genetic Variation to the Viability of Mammalian Populations.” *Journal of Mammalogy* 78 (2): 320–35. https://doi.org/10.2307/1382885.

Lamichhaney, Sangeet, Jonas Berglund, Markus Sällman Almén, Khurram Maqbool, Manfred Grabherr, Alvaro Martinez-Barrio, Marta Promerová, et al. 2015. “Evolution of Darwin’s Finches and Their Beaks Revealed by Genome Sequencing.” *Nature* 518 (7539): 371–75. https://doi.org/10.1038/nature14181.

Lande, Russell. 1976. “Natural Selection and Random Genetic Drift in Phenotypic Evolution.” *Evolution* 30 (2): 314–34. https://doi.org/10.2307/2407703.

Lande, Russell, and G Barrowdough. 1987. “Effective Population Size, Genetic Variation, and Their Use in Population.” In *Viable Populations for Conservation*. Cambridge University Press.

Larson, Allan, and Jonathan B. Losos. 1996. “Phylogenetic Systematics of Adaptation.” In *Adaptation*, 187–220. Academic Press San Diego.

Lawler, Joshua J., Juliann E. Aukema, Jacqualine B. Grant, Benjamin S. Halpern, Peter Kareiva, Cara R. Nelson, Kris Ohleth, et al. 2006. “Conservation Science: A 20-Year Report Card.” *Frontiers in Ecology and the Environment* 4 (9): 473–80. https://doi.org/10.1890/1540-9295(2006)4[473:CSAYRC]2.0.CO;2.

Leimu, Roosa, and Markus Fischer. 2008. “A Meta-Analysis of Local Adaptation in Plants.” *PLoS ONE* 3 (12): 1–8. https://doi.org/10.1371/journal.pone.0004010.

Lenormand, Thomas. 2002. “Gene Flow and the Limits to Natural Selection.” *Trends in Ecology & Evolution* 17 (4): 183–89. https://doi.org/10.1016/S0169-5347(02)02497-7.

Lertzman-Lepofsky, Gavia, Arne Mooers, and Dan A. Greenberg. 2019. “Ecological Constraints Associated with Genome Size across Salamander Lineages.” *Proceedings of the Royal Society B: Biological Sciences* 286 (1911). https://doi.org/10.1098/rspb.2019.1780.

Li, Jun, Haiyan Yu, Wenxia Wang, Chao Fu, Wei Zhang, Fengming Han, and Hua Wu. 2019. “Genomic and Transcriptomic Insights into Molecular Basis of Sexually Dimorphic Nuptial Spines in *Leptobrachium leishanense*.” *Nature Communications* 10 (1): 1–13. https://doi.org/10.1038/s41467-019-13531-5.

Li, Yongxin, Yandong Ren, Dongru Zhang, Hui Jiang, Zhongkai Wang, Xueyan Li, and DIngqi Rao. 2019. “Chromosome-Level Assembly of the Mustache Toad Genome Using Third-Generation DNA Sequencing and Hi-C Analysis.” *GigaScience* 8 (9): 1–10. https://doi.org/10.1093/gigascience/giz114.

Liedtke, H. Christoph, David J. Gower, Mark Wilkinson, and Ivan Gomez-Mestre. 2018. “Macroevolutionary Shift in the Size of Amphibian Genomes and the Role of Life History and Climate.” *Nature Ecology and Evolution* 2 (11): 1792–99. https://doi.org/10.1038/s41559-018-0674-4.

Lopes, Ricardo J., James D. Johnson, Matthew B. Toomey, Mafalda S. Ferreira, Pedro M. Araujo, José Melo-Ferreira, Leif Andersson, Geoffrey E. Hill, Joseph C. Corbo, and Miguel Carneiro. 2016. “Genetic Basis for Red Coloration in Birds.” *Current Biology* 26 (11): 1427–34. https://doi.org/10.1016/j.cub.2016.03.076.

Lopez-Maestre, Hélène, Lilia Brinza, Camille Marchet, Janice Kielbassa, Sylvère Bastien,

Mathilde Boutigny, David Monnin, et al. 2016. “SNP Calling from RNA-Seq Data without a Reference Genome: Identification, Quantification, Differential Analysis and Impact on the Protein Sequence.” *Nucleic Acids Research* 44 (19): 1–13. https://doi.org/10.1093/nar/gkw655.

Losos, Jonathan B. 2009. *Lizards in an Evolutionary Tree: Ecology and Adaptive Radiation of Anoles*. Univ of California Press.

Lourenço, André, David Álvarez, Ian J. Wang, and Guillermo Velo-Antón. 2017. “Trapped within the City: Integrating Demography, Time since Isolation and Population-Specific Traits to Assess the Genetic Effects of Urbanization.” *Molecular Ecology* 26 (6): 1498–1514. https://doi.org/10.1111/mec.14019.

Love, Michael I, Wolfgang Huber, and Simon Anders. 2014. “Moderated Estimation of Fold Change and Dispersion for RNA-Seq Data with DESeq2.” *Genome Biology* 15 (12): 550. https://doi.org/10.1186/s13059-014-0550-8.

Lowry, David B., Sean Hoban, Joanna L. Kelley, Katie E. Lotterhos, Laura K. Reed, Michael F. Antolin, and Andrew Storfer. 2017. “Breaking RAD: An Evaluation of the Utility of Restriction Site-Associated DNA Sequencing for Genome Scans of Adaptation.” *Molecular Ecology Resources* 17 (2): 142–52. https://doi.org/10.1111/1755-0998.12635.

Mable, B. K., M. A. Alexandrou, and M. I. Taylor. 2011. “Genome Duplication in Amphibians and Fish: An Extended Synthesis.” *Journal of Zoology* 284 (3): 151–82. https://doi.org/10.1111/j.1469-7998.2011.00829.x.

MacDonald, Jeffrey R., Robert Ziman, Ryan K C Yuen, Lars Feuk, and Stephen W. Scherer. 2014. “The Database of Genomic Variants: A Curated Collection of Structural Variation in the Human Genome.” *Nucleic Acids Research* 42 (D1): D986–92. https://doi.org/10.1093/nar/gkt958.

Mackay, Trudy F.C. 2014. “Epistasis and Quantitative Traits: Using Model Organisms to Study Gene-Gene Interactions.” *Nature Reviews Genetics* 15 (1): 22–33. https://doi.org/10.1038/nrg3627.

Matsunami, Masatoshi, Miyuki Suzuki, Yoshikazu Haramoto, Akimasa Fukui, Takeshi Inoue, Katsushi Yamaguchi, Ikuo Uchiyama, et al. 2019. “A Comprehensive Reference Transcriptome Resource for the Iberian Ribbed Newt *Pleurodeles waltl*, an Emerging Model for Developmental and Regeneration Biology.” *DNA Research* 26 (3): 217–29. https://doi.org/10.1093/dnares/dsz003.

McCartney-Melstad, Evan, Genevieve G. Mount, and H. Bradley Shaffer. 2016. “Exon Capture Optimization in Amphibians with Large Genomes.” *Molecular Ecology Resources* 16 (5): 1084–94. https://doi.org/10.1111/1755-0998.12538.

McCusker, Catherine, and David M. Gardiner. 2011. “The Axolotl Model for Regeneration and Aging Research: A Mini-Review.” *Gerontology* 57 (6): 565–71. https://doi.org/10.1159/000323761.

McGaugh, Suzanne E., Joshua B. Gross, Bronwen Aken, Maryline Blin, Richard Borowsky, Domitille Chalopin, Hélène Hinaux, et al. 2014. “The Cavefish Genome Reveals Candidate Genes for Eye Loss.” *Nature Communications* 5. https://doi.org/10.1038/ncomms6307.

McManus, Kimberly F., Joanna L. Kelley, Shiya Song, Krishna R. Veeramah, August E. Woerner, Laurie S. Stevison, Oliver A. Ryder, et al. 2015. “Inference of Gorilla Demographic and Selective History from Whole-Genome Sequence Data.” *Molecular Biology and Evolution* 32 (3): 600–612. https://doi.org/10.1093/molbev/msu394.

Melbinger, Anna, and Massimo Vergassola. 2015. “The Impact of Environmental Fluctuations on Evolutionary Fitness Functions.” *Scientific Reports* 5: 1–11. https://doi.org/10.1038/srep15211.

Mörs, Thomas, Marcelo Reguero, and Davit Vasilyan. 2020. “First Fossil Frog from Antarctica: Implications for Eocene High Latitude Climate Conditions and Gondwanan Cosmopolitanism of Australobatrachia.” *Scientific Reports* 10 (1): 1–11. https://doi.org/10.1038/s41598-020-61973-5.

Mulder, Kevin P., Maria Cortazar-Chinarro, D. James Harris, Angelica Crottini, Evan H. Campbell Grant, Robert C. Fleischer, and Anna E. Savage. 2017. “Evolutionary Dynamics of an Expressed MHC Class IIβ Locus in the Ranidae (Anura) Uncovered by Genome Walking and High-Throughput Amplicon Sequencing.” *Developmental and Comparative Immunology* 76: 177–88. https://doi.org/10.1016/j.dci.2017.05.022.

Mulder, Kevin P., Nandadevi Cortes-Rodriguez, Evan H. Campbell Grant, Adrianne Brand, and Robert C. Fleischer. 2019. “North-Facing Slopes and Elevation Shape Asymmetric Genetic Structure in the Range-Restricted Salamander *Plethodon shenandoah*.” *Ecology and Evolution* 9 (9): 5094–5105. https://doi.org/10.1002/ece3.5064.

Murphy, Mason O., Kara S. Jones, Steven J. Price, and David W. Weisrock. 2018. “A Genomic Assessment of Population Structure and Gene Flow in an Aquatic Salamander Identifies the Roles of Spatial Scale, Barriers, and River Architecture.” *Freshwater Biology* 63 (5): 407–19. https://doi.org/10.1111/fwb.13071.

Myers, E. M., and K. R. Zamudio. 2004. “Multiple Paternity in an Aggregate Breeding Amphibian: The Effect of Reproductive Skew on Estimates of Male Reproductive Success.” *Molecular Ecology* 13 (7): 1951–63. https://doi.org/10.1111/j.1365-294X.2004.02208.x.

Narum, Shawn R., and Jon E. Hess. 2011. “Comparison of F_{ST} Outlier Tests for SNP Loci

under Selection." *Molecular Ecology Resources* 11: 184–94. https://doi.org/10.1111/j.1755-0998.2011.02987.x.

Newman, Catherine E., Jeremy A. Feinberg, Leslie J. Rissler, Joanna Burger, and H. Bradley Shaffer. 2012. "A New Species of Leopard Frog (Anura: Ranidae) from the Urban Northeastern US." *Molecular Phylogenetics and Evolution* 63 (2): 445–55. https://doi.org/10.1016/j.ympev.2012.01.021.

Newman, Robert A., and Tina Squire. 2001. "Microsatellite Variation and Fine-Scale Population Structure in the Wood Frog (*Rana sylvatica*)." *Molecular Ecology* 10 (5): 1087–1100. https://doi.org/10.1046/j.1365-294X.2001.01255.x.

Nowoshilow, Sergej, Siegfried Schloissnig, Ji Feng Fei, Andreas Dahl, Andy W.C. Pang, Martin Pippel, Sylke Winkler, et al. 2018. "The Axolotl Genome and the Evolution of Key Tissue Formation Regulators." *Nature* 554 (7690): 50–55. https://doi.org/10.1038/nature25458.

Nunziata, Schyler O., Stacey L. Lance, David E. Scott, Emily Moriarty Lemmon, and David W. Weisrock. 2017. "Genomic Data Detect Corresponding Signatures of Population Size Change on an Ecological Time Scale in Two Salamander Species." *Molecular Ecology* 26 (4): 1060–74. https://doi.org/10.1111/mec.13988.

Oleksyk, Taras K., Michael W. Smith, and Stephen J. O'Brien. 2010. "Genome-Wide Scans for Footprints of Natural Selection." *Philosophical Transactions of the Royal Society B: Biological Sciences* 365 (1537): 185–205. https://doi.org/10.1098/rstb.2009.0219.

Olmo, Ettore. 1973. "Quantitative Variations in the Nuclear DNA and Phylogenesis of the Amphibia." *Caryologia* 25 (April 2015): 43–68. https://doi.org/10.1080/00087114.1973.10796525.

Orr, H. Allen. 1998. "Testing Natural Selection vs. Genetic Drift in Phenotypic Evolution Using Quantitative Trait Locus Data." *Genetics* 149 (4): 2099–2104.

Ostrander, Elaine A., Robert K. Wayne, Adam H. Freedman, and Brian W. Davis. 2017. "Demographic History, Selection and Functional Diversity of the Canine Genome." *Nature Reviews Genetics* 18 (12): 705–20. https://doi.org/10.1038/nrg.2017.67.

Otto, Sarah P. 2004. "Two Steps Forward, One Step Back: The Pleiotropic Effects of Favoured Alleles." *Proceedings of the Royal Society B: Biological Sciences* 271 (1540): 705–14. https://doi.org/10.1098/rspb.2003.2635.

Pabijan, Maciej, Gemma Palomar, Bernardo Antunes, Weronika Antoł, Piotr Zieliński, and Wiesław Babik. 2020. "Evolutionary Principles Guiding Amphibian Conservation." *Evolutionary Applications* 13: 857–78. https://doi.org/10.1111/eva.12940.

Passow, Courtney N., Thomas J.Y. Kono, Bethany A. Stahl, James B. Jaggard, Alex C. Keene, and Suzanne E. McGaugh. 2019. "Nonrandom RNAseq Gene Expression

Associated with RNAlater and Flash Freezing Storage Methods." *Molecular Ecology Resources* 19 (2): 456–64. https://doi.org/10.1111/1755-0998.12965.

Pastenes, Luis, Camilo Valdivieso, Alex Di Genova, Dante Travisany, Andrew Hart, Martín Montecino, Ariel Orellana, et al. 2017. "Global Gene Expression Analysis Provides Insight into Local Adaptation to Geothermal Streams in Tadpoles of the Andean Toad *Rhinella spinulosa*." *Scientific Reports* 7 (1): 1–10. https://doi.org/10.1038/s41598-017-01982-z.

Pauling, L., H. A. Itano, S. J. Singer, and I. C. Wells. 1949. "Sickle Cell Anemia, a Molecular Disease." *Science* 110 (2865): 543–48. https://doi.org/10.1126/science.110.2865.543.

Pereira, Ricardo J., Iñigo Martínez-Solano, and David Buckley. 2016. "Hybridization during Altitudinal Range Shifts: Nuclear Introgression Leads to Extensive Cyto-Nuclear Discordance in the Fire Salamander." *Molecular Ecology* 25 (7): 1551–65. https://doi.org/10.1111/mec.13575.

Peterson, Brant K., Jesse N. Weber, Emily H. Kay, Heidi S. Fisher, and Hopi E. Hoekstra. 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. https://doi.org/10.1371/journal.pone.0037135.

Pimentel, Harold, Nicolas L. Bray, Suzette Puente, Páll Melsted, and Lior Pachter. 2017. "Differential Analysis of RNA-Seq Incorporating Quantification Uncertainty." *Nature Methods* 14 (7): 687–90. https://doi.org/10.1038/nmeth.4324.

Prates, Ivan, Anna Penna, Miguel Trefaut Rodrigues, and Ana Carolina Carnaval. 2018. "Local Adaptation in Mainland Anole Lizards: Integrating Population History and Genome–Environment Associations." *Ecology and Evolution* 8 (23): 11932–44. https://doi.org/10.1002/ece3.4650.

Prunier, Jérôme G., Marc Colyn, Xavier Legendre, Kim F. Nimon, and Marie Christine Flamand. 2015. "Multicollinearity in Spatial Genetics: Separating the Wheat from the Chaff Using Commonality Analyses." *Molecular Ecology* 24 (21): 263–83. https://doi.org/10.1111/mec.13029.

Qiao, Liang, Weizhao Yang, Jinzhong Fu, and Zhaobin Song. 2013. "Transcriptome Profile of the Green Odorous Frog (*Odorrana margaretae*)." *PLoS ONE* 8 (9): e75211. https://doi.org/10.1371/journal.pone.0075211.

Quinn, Emma M., Paul Cormican, Elaine M. Kenny, Matthew Hill, Richard Anney, Michael Gill, Aiden P. Corvin, and Derek W. Morris. 2013. "Development of Strategies for SNP Detection in RNA-Seq Data: Application to Lymphoblastoid Cell Lines and Evaluation Using 1000 Genomes Data." *PLoS ONE* 8 (3). https://doi.org/10.1371/journal.pone.0058815.

Rakotoarison, Andolalao, Angelica Crottini, Johannes Müller, Mark Oliver Rödel, Frank Glaw, and Miguel Vences. 2015. "Revision and Phylogeny of Narrow-Mouthed Treefrogs (*Cophyla*) from Northern Madagascar: Integration of Molecular, Osteological, and Bioacoustic Data Reveals Three New Species." *Zootaxa* 3937 (1): 61–89. https://doi.org/10.11646/zootaxa.3937.1.3.

Robertson, Laura S., and Robert S. Cornman. 2014. "Transcriptome Resources for the Frogs *Lithobates clamitans* and *Pseudacris regilla*, Emphasizing Antimicrobial Peptides and Conserved Loci for Phylogenetics." *Molecular Ecology Resources* 14 (1): 178–83. https://doi.org/10.1111/1755-0998.12164.

Robinson, Mark D., Davis J. McCarthy, and Gordon K. Smyth. 2009. "EdgeR: A Bioconductor Package for Differential Expression Analysis of Digital Gene Expression Data." *Bioinformatics* 26 (1): 139–40. https://doi.org/10.1093/bioinformatics/btp616.

Rodríguez, Ariel, James D. Burgon, Mariana Lyra, Iker Irisarri, Denis Baurain, Leon Blaustein, Bayram Göçmen, et al. 2017. "Inferring the Shallow Phylogeny of True Salamanders (*Salamandra*) by Multiple Phylogenomic Approaches." *Molecular Phylogenetics and Evolution* 115: 16–26. https://doi.org/10.1016/j.ympev.2017.07.009.

Rogers, Rebekah L., Long Zhou, Chong Chu, Roberto Marquez, Ammon Corl, Tyler Linderoth, Layla Freeborn, et al. 2018. "Genomic Takeover by Transposable Elements in the Strawberry Poison Frog." *Molecular Biology and Evolution* 35 (12): 2913–27. https://doi.org/10.1093/molbev/msy185.

Rosenblum, Erica Bree, Holger Römpler, Torsten Schöneberg, and Hopi E. Hoekstra. 2010. "Molecular and Functional Basis of Phenotypic Convergence in White Lizards at White Sands." *Proceedings of the National Academy of Sciences* 107 (5): 2113–17. https://doi.org/10.1073/pnas.0911042107.

Ruane, Sara, and Christopher C. Austin. 2017. "Phylogenomics Using Formalin-Fixed and 100+ Year-Old Intractable Natural History Specimens." *Molecular Ecology Resources* 17 (5): 1003–8. https://doi.org/10.1111/1755-0998.12655.

Rubin, Carl Johan, Hendrik Jan Megens, Alvaro Martinez Barrio, Khurram Maqbool, Shumaila Sayyab, Doreen Schwochow, Chao Wang, et al. 2012. "Strong Signatures of Selection in the Domestic Pig Genome." *Proceedings of the National Academy of Sciences* 109 (48): 19529–36. https://doi.org/10.1073/pnas.1217149109.

Salzburger, Walter. 2009. "The Interaction of Sexually and Naturally Selected Traits in the Adaptive Radiations of Cichlid Fishes." *Molecular Ecology* 18 (2): 169–85. https://doi.org/10.1111/j.1365-294X.2008.03981.x.

Savage, Anna E., Carlos G. Becker, and Kelly R. Zamudio. 2015. "Linking Genetic and Environmental Factors in Amphibian Disease Risk." *Evolutionary Applications* 8 (6):

560–72. https://doi.org/10.1111/eva.12264.

Savage, Anna E., Brian Gratwicke, K. Hope, E. Bronikowski, and R.C. Fleischer. 2020. "Immune Activation Is Associated with Susceptibility to the Amphibian Chytrid Fungus." *Molecular Ecology - Accepted*.

Savage, Anna E., Kevin P. Mulder, Taina Torres, and Stuart Wells. 2018. "Lost but Not Forgotten: MHC Genotypes Predict Overwinter Survival despite Depauperate MHC Diversity in a Declining Frog." *Conservation Genetics* 19 (2): 309–22. https://doi.org/10.1007/s10592-017-1001-3.

Savage, Anna E., Michael J. Sredl, and Kelly R. Zamudio. 2011. "Disease Dynamics Vary Spatially and Temporally in a North American Amphibian." *Biological Conservation* 144 (6): 1910–15. https://doi.org/10.1016/j.biocon.2011.03.018.

Savage, Anna E., and K. R. Zamudio. 2011. "MHC Genotypes Associate with Resistance to a Frog-Killing Fungus." *Proceedings of the National Academy of Sciences* 108 (40): 16705–10. https://doi.org/10.1073/pnas.1106893108.

Savolainen, Outi, Martin Lascoux, and Juha Merilä. 2013. "Ecological Genomics of Local Adaptation." *Nature Reviews. Genetics* 14 (11): 807–20. https://doi.org/10.1038/nrg3522.

Savolainen, Outi, Tanja Pyhäjärvi, and Timo Knürr. 2007. "Gene Flow and Local Adaptation in Trees." *Annual Review of Ecology, Evolution, and Systematics* 38 (1): 595–619. https://doi.org/10.1146/annurev.ecolsys.38.091206.095646.

Scheele, Ben C., Frank Pasmans, Lee F. Skerratt, Lee Berger, An Martel, Wouter Beukema, Aldemar A. Acevedo, et al. 2019. "Amphibian Fungal Panzootic Causes Catastrophic and Ongoing Loss of Biodiversity." *Science* 363 (6434): 1459–63. https://doi.org/10.1126/science.aav0379.

Schlaepfer, Martin A., Michael J. Sredl, Phil C. Rosen, and Michael J. Ryan. 2007. "High Prevalence of Batrachochytrium Dendrobatidis in Wild Populations of Lowland Leopard Frogs *Rana yavapaiensis* in Arizona." *EcoHealth* 4 (4): 421–27. https://doi.org/10.1007/s10393-007-0136-y.

Schoch, Rainer R. 2009. "Evolution of Life Cycles in Early Amphibians." *Annual Review of Earth and Planetary Sciences* 37 (1): 135–62. https://doi.org/10.1146/annurev.earth.031208.100113.

Schoville, Sean D., Tate S. Tustall, Vance T. Vredenburg, Adam R. Backlin, Elizabeth Gallegos, Dustin A. Wood, and Robert N. Fisher. 2011. "Conservation Genetics of Evolutionary Lineages of the Endangered Mountain Yellow-Legged Frog, *Rana muscosa* (Amphibia: Ranidae), in Southern California." *Biological Conservation* 144 (7): 2031–40. https://doi.org/10.1016/j.biocon.2011.04.025.

Schurch, Nicholas J., Pietá Schofield, Marek Gierliński, Christian Cole, Alexander Sherstnev, Vijender Singh, Nicola Wrobel, et al. 2016. "How Many Biological Replicates Are Needed in an RNA-Seq Experiment and Which Differential Expression Tool Should You Use?" *Rna* 22 (6): 839–51. https://doi.org/10.1261/rna.053959.115.

Ségurel, Laure, and Céline Bon. 2017. "On the Evolution of Lactase Persistence in Humans." *Annual Review of Genomics and Human Genetics* 18 (1): 297–319. https://doi.org/10.1146/annurev-genom-091416-035340.

Serjeant, Graham R. 2010. "One Hundred Years of Sickle Cell Disease." *British Journal of Haematology* 151 (5): 425–29. https://doi.org/10.1111/j.1365-2141.2010.08419.x.

Servant, Nicolas, Nelle Varoquaux, Bryan R. Lajoie, Eric Viara, Chong Jian Chen, Jean Philippe Vert, Edith Heard, Job Dekker, and Emmanuel Barillot. 2015. "HiC-Pro: An Optimized and Flexible Pipeline for Hi-C Data Processing." *Genome Biology* 16 (1): 1–11. https://doi.org/10.1186/s13059-015-0831-x.

Session, Adam M., Yoshinobu Uno, Taejoon Kwon, Jarrod A. Chapman, Atsushi Toyoda, Shuji Takahashi, Akimasa Fukui, et al. 2016. "Genome Evolution in the Allotetraploid Frog *Xenopus laevis*." *Nature* 538 (7625): 336–43. https://doi.org/10.1038/nature19840.

Sessions, Stanley K. 2008. "Evolutionary Cytogenetics in Salamanders." *Chromosome Research* 16 (1): 183–201. https://doi.org/10.1007/s10577-007-1205-3.

Sessions, Stanley K., and Allan Larson. 1987. "Developmental Correlates of Genome Size in Plethodontid Salamanders and Their Implications for Genome Evolution." *Evolution* 41 (6): 1239. https://doi.org/10.2307/2409090.

Shaffer, H. Bradley, Müge Gidiş, Evan McCartney-Melstad, Kevin M. Neal, Hilton M. Oyamaguchi, Marisa Tellez, and Erin M. Toffelmier. 2015. "Conservation Genetics and Genomics of Amphibians and Reptiles." *Annual Review of Animal Biosciences* 3 (1): 113–38. https://doi.org/10.1146/annurev-animal-022114-110920.

Slate, Jon. 2005. "Quantitative Trait Locus Mapping in Natural Populations: Progress, Caveats and Future Directions." *Molecular Ecology* 14 (2): 363–79. https://doi.org/10.1111/j.1365-294X.2004.02378.x.

Smith, Jeramiah J., Srikrishna Putta, Wei Zhu, Gerald M. Pao, Inder M. Verma, Tony Hunter, Susan V. Bryant, David M. Gardiner, Timothy T. Harkins, and S. Randal Voss. 2009. "Genic Regions of a Large Salamander Genome Contain Long Introns and Novel Genes." *BMC Genomics* 10: 1–11. https://doi.org/10.1186/1471-2164-10-19.

Smith, Jeramiah J., Nataliya Timoshevskaya, Vladimir A. Timoshevskiy, Melissa C. Keinath, Drew Hardy, and S. Randal Voss. 2019. "A Chromosome-Scale Assembly of the Axolotl Genome." *Genome Research* 29 (2): 317–24. https://doi.org/10.1101/gr.241901.118.

Sredl, Michael J. 1997. "Ranid Frog Conservation and Management."

Sredl, Michael J., and R. D. Jennings. 2005. "*Rana chiricahuensis*: Platz and Mecham, 1979, Chiricahua Leopard Frogs." *Amphibian Declines: The Conservation Status of United States Species. University of California Press, Berkeley*, 546–49.

Stapley, Jessica, Julia Reger, Philine G.D. Feulner, Carole Smadja, Juan Galindo, Robert Ekblom, Clair Bennison, Alexander D. Ball, Andrew P. Beckerman, and Jon Slate. 2010. "Adaptation Genomics: The next Generation." *Trends in Ecology and Evolution*. https://doi.org/10.1016/j.tree.2010.09.002.

Steiner, Cynthia C., Andrea S. Putnam, Paquita E.A. Hoeck, and Oliver A. Ryder. 2013. "Conservation Genomics of Threatened Animal Species." *Annual Review of Animal Biosciences* 1 (1): 261–81. https://doi.org/10.1146/annurev-animal-031412-103636.

Steinfartz, Sebastian, K. Stemshorn, D. Kuesters, and D. Tautz. 2006. "Patterns of Multiple Paternity within and between Annual Reproduction Cycles of the Fire Salamander (*Salamandra salamandra*) under Natural Conditions." *Journal of Zoology* 268 (1): 1–8. https://doi.org/10.1111/j.1469-7998.2005.00001.x.

Steinfartz, Sebastian, Michael Veith, and Diethard Tautz. 2000. "Mitochondrial Sequence Analysis of Salamandra Taxa Suggests Old Splits of Major Lineages and Postglacial Recolonizations of Central Europe from Distinct Source Populations of *Salamandra salamandra*." *Molecular Ecology* 9 (4): 397–410. https://doi.org/10.1046/j.1365-294X.2000.00870.x.

Stern, David L. 2013. "The Genetic Causes of Convergent Evolution." *Nature Reviews Genetics* 14 (11): 751–64. https://doi.org/10.1038/nrg3483.

Storz, Jay F. 2005. "Using Genome Scans of DNA Polymorphism to Infer Adaptive Population Divergence." *Molecular Ecology* 14 (3): 671–88. https://doi.org/10.1111/j.1365-294X.2004.02437.x.

Stranger, Barbara E., Eli A. Stahl, and Towfique Raj. 2011. "Progress and Promise of Genome-Wide Association Studies for Human Complex Trait Genetics." *Genetics* 187 (2): 367–83. https://doi.org/10.1534/genetics.110.120907.

Streicher, Jeffrey W., Thomas J. Devitt, Caren S. Goldberg, John H. Malone, Heath Blackmon, and Matthew K. Fujita. 2014. "Diversification and Asymmetrical Gene Flow across Time and Space: Lineage Sorting and Hybridization in Polytypic Barking Frogs." *Molecular Ecology* 23 (13): 3273–91. https://doi.org/10.1111/mec.12814.

Stuart, Simon N., Janice S. Chanson, Neil A. Cox, Bruce E. Young, Ana S.L. Rodrigues, Debra L. Fischman, and Robert W. Waller. 2004. "Status and Trends of Amphibian Declines and Extinctions Worldwide." *Science* 306 (5702): 1783–86. https://doi.org/10.1126/science.1103538.

Sun, Cheng, and Rachel Lockridge Mueller. 2014. "Hellbender Genome Sequences Shed

Light on Genomic Expansion at the Base of Crown Salamanders." *Genome Biology and Evolution* 6 (7): 1818–29. https://doi.org/10.1093/gbe/evu143.

Sun, Yan Bo, Zi Jun Xiong, Xue Yan Xiang, Shi Ping Liu, Wei Wei Zhou, Xiao Long Tu, Li Zhong, et al. 2015. "Whole-Genome Sequence of the Tibetan Frog *Nanorana parkeri* and the Comparative Evolution of Tetrapod Genomes." *Proceedings of the National Academy of Sciences* 112 (11): E1257–62. https://doi.org/10.1073/pnas.1501764112.

Taft, Ryan J., Michael Pheasant, and John S. Mattick. 2007. "The Relationship between Non-Protein-Coding DNA and Eukaryotic Complexity." *BioEssays* 29 (3): 288–99. https://doi.org/10.1002/bies.20544.

Tajima, F. 1989. "Statistical Method for Testing the Neutral Mutation Hypothesis by DNA Polymorphism." *Genetics* 123 (3): 585–95. https://doi.org/PMC1203831.

Tapley, Benjamin, Christopher J. Michaels, Rikki Gumbs, Monika Böhm, Jennifer Luedtke, Paul Pearce-Kelly, and Jodi J.L. Rowley. 2018. "The Disparity between Species Description and Conservation Assessment: A Case Study in Taxa with High Rates of Species Discovery." *Biological Conservation* 220: 209–14. https://doi.org/10.1016/j.biocon.2018.01.022.

Tigano, Anna, and Vicki L. Friesen. 2016. "Genomics of Local Adaptation with Gene Flow." *Molecular Ecology* 25 (10): 2144–64. https://doi.org/10.1111/mec.13606.

Tishkoff, Sarah A., Floyd A. Reed, Alessia Ranciaro, Benjamin F. Voight, Courtney C. Babbitt, Jesse S. Silverman, Kweli Powell, et al. 2007. "Convergent Adaptation of Human Lactase Persistence in Africa and Europe." *Nature Genetics* 39 (1): 31–40. https://doi.org/10.1038/ng1946.

Todd, Erica V., Michael A. Black, and Neil J. Gemmell. 2016. "The Power and Promise of RNA-Seq in Ecology and Evolution." *Molecular Ecology* 25 (6): 1224–41. https://doi.org/10.1111/mec.13526.

Toomey, Matthew B., Ricardo J. Lopes, Pedro M. Araújo, James D. Johnson, Malgorzata A. Gazda, Sandra Afonso, Paulo G. Mota, et al. 2017. "High-Density Lipoprotein Receptor SCARB1 Is Required for Carotenoid Coloration in Birds." *Proceedings of the National Academy of Sciences* 114 (20): 5219–24. https://doi.org/10.1073/pnas.1700751114.

Treangen, Todd J., and Steven L. Salzberg. 2012. "Repetitive DNA and Next-Generation Sequencing: Computational Challenges and Solutions." *Nature Reviews Genetics* 13 (1): 36–46. https://doi.org/10.1038/nrg3117.

Troelsen, Jesper T., Jørgen Olsen, Jette Møller, and Hans Sjöström. 2003. "An Upstream Polymorphism Associated with Lactase Persistence Has Increased Enhancer Activity." *Gastroenterology* 125 (6): 1686–94. https://doi.org/10.1053/j.gastro.2003.09.031.

Vandebergh, Wim, and Franky Bossuyt. 2012. "Radiation and Functional Diversification of

Alpha Keratins during Early Vertebrate Evolution." *Molecular Biology and Evolution* 29 (3): 995–1004. https://doi.org/10.1093/molbev/msr269.

Veith, M, Sebastian Steinfartz, R Zardoya, A Seitz, and A Meyer. 1998. "A Molecular Phylogeny of 'True'Salamanders (Family Salamandridae) and the Evolution of Terrestriality of Reproductive Modes." *Journal of Zoological Systematics and Evolutionary Research* 36 (1–2): 7–16. https://doi.org/10.1111/j.1439-0469.1998.tb00774.x.

Velo-Antón, Guillermo, Mario García-París, P. Galán, and A. Cordero Rivera. 2007. "The Evolution of Viviparity in Holocene Islands: Ecological Adaptation versus Phylogenetic Descent along the Transition from Aquatic to Terrestrial Environments." *Journal of Zoological Systematics and Evolutionary Research* 45 (4): 345–52. https://doi.org/10.1111/j.1439-0469.2007.00420.x.

Velo-Antón, Guillermo, Xavier Santos, Iago Sanmartín-Villar, Adolfo Cordero-Rivera, and David Buckley. 2015. "Intraspecific Variation in Clutch Size and Maternal Investment in Pueriparous and Larviparous *Salamandra salamandra* Females." *Evolutionary Ecology* 29 (1): 185–204. https://doi.org/10.1007/s10682-014-9720-0.

Vences, Miguel, Eugenia Sanchez, J. Susanne Hauswaldt, Daniel Eikelmann, Ariel Rodríguez, Salvador Carranza, David Donaire, et al. 2014. "Nuclear and Mitochondrial Multilocus Phylogeny and Survey of Alkaloid Content in True Salamanders of the Genus *Salamandra* (Salamandridae)." *Molecular Phylogenetics and Evolution* 73 (1): 208–16. https://doi.org/10.1016/j.ympev.2013.12.009.

Vences, Miguel, Meike Thomas, Ronald M. Bonett, and David R. Vieites. 2005. "Deciphering Amphibian Diversity through DNA Barcoding: Chances and Challenges." *Philosophical Transactions of the Royal Society B: Biological Sciences* 360 (1462): 1859–68. https://doi.org/10.1098/rstb.2005.1717.

Vieites, David R., Katharina C. Wollenberg, Franco Andreone, Jörn Köhler, Frank Glaw, and Miguel Vences. 2009. "Vast Underestimation of Madagascar's Biodiversity Evidenced by an Integrative Amphibian Inventory." *Proceedings of the National Academy of Sciences* 106 (20): 8267–72. https://doi.org/10.1073/pnas.0810821106.

Vinogradov, Alexander E. 1997. "Nucleotypic Effect in Homeotherms: Body-Mass Independent Resting Metabolic Rate of Passerine Birds Is Related to Genome Size." *Evolution* 51 (1): 220. https://doi.org/10.2307/2410975.

Wang, Shi, Eli Meyer, John K. Mckay, and Mikhail V. Matz. 2012. "2b-RAD: A Simple and Flexible Method for Genome-Wide Genotyping." *Nature Methods* 9 (8): 808–10. https://doi.org/10.1038/nmeth.2023.

Weisrock, David W., Paul M. Hime, Schyler O. Nunziata, Kara S. Jones, Mason O. Murphy,

Scott Hotaling, and Justin D. Kratovil. 2018. “Surmounting the Large-Genome ‘Problem’ for Genomic Data Generation in Salamanders.” In *Population Genomics: Wildlife*, 1–28. Springer. https://doi.org/10.1007/13836_2018_36.

Whiles, Matt R., Karen R. Lips, Cathy M. Pringle, Susan S. Kilham, Rebecca J. Bixby, Roberto Brenes, Scott Connelly, et al. 2006. “The Effects of Amphibian Population Declines on the Structure and Function of Neotropical Stream Ecosystems.” *Frontiers in Ecology and the Environment* 4 (1): 27–34. https://doi.org/10.1890/1540-9295(2006)004[0027:TEOAPD]2.0.CO;2.

Whitlock, Michael C., and Katie E. Lotterhos. 2015. “Reliable Detection of Loci Responsible for Local Adaptation: Inference of a Null Model through Trimming the Distribution of F_{ST}.” *American Naturalist* 186 (1): S24–36. https://doi.org/10.1086/682949.

Wielstra, B., E. Duijm, P. Lagler, Y. Lammers, W. R. M. Meilink, J. M. Ziermann, and J. W. Arntzen. 2014. “Parallel Tagged Amplicon Sequencing of Transcriptome-Based Genetic Markers for *Triturus* Newts with the Ion Torrent next-Generation Sequencing Platform.” *Molecular Ecology Resources* 14 (5): 1080–89. https://doi.org/10.1111/1755-0998.12242.

Williams, George C. 1966. “Natural Selection, the Costs of Reproduction, and a Refinement of Lack’s Principle.” *The American Naturalist* 100 (916): 687–90. https://doi.org/10.1086/282461.

Wolf, Jochen B.W. 2013. “Principles of Transcriptome Analysis and Gene Expression Quantification: An RNA-Seq Tutorial.” *Molecular Ecology Resources* 13 (4): 559–72. https://doi.org/10.1111/1755-0998.12109.

Yates, Andrew D., Premanand Achuthan, Wasiu Akanni, James Allen, Jamie Allen, Jorge Alvarez-Jarreta, M. Ridwan Amode, et al. 2020. “Ensembl 2020.” *Nucleic Acids Research* 48 (D1): D682–88. https://doi.org/10.1093/nar/gkz966.

Yuan, Zhi Yong, Wei Wei Zhou, Xin Chen, Nikolay A. Poyarkov, Hong Man Chen, Nian Hong Jang-Liaw, Wen Hao Chou, et al. 2016. “Spatiotemporal Diversification of the True Frogs (Genus *Rana*): A Historical Framework for a Widely Studied Group of Model Organisms.” *Systematic Biology* 65 (5): 824–42. https://doi.org/10.1093/sysbio/syw055.

Zaldívar-Riverón, Alejandro, Virginia León-Regagnon, and Adrían Nieto-Montes De Oca. 2004. “Phylogeny of the Mexican Coastal Leopard Frogs of the *Rana berlandieri* Group Based on MtDNA Sequences.” *Molecular Phylogenetics and Evolution* 30 (1): 38–49. https://doi.org/10.1016/S1055-7903(03)00141-6.

Zamudio, Kelly R., Rayna C. Bell, Renato C. Nali, Célio F. B. Haddad, and Cynthia P. A. Prado. 2016. “Polyandry, Predation, and the Evolution of Frog Reproductive Modes.” *The American Naturalist* 188 (S1): S41–61. https://doi.org/10.1086/687547.

Zhang, Chi, Baohong Zhang, Lih Ling Lin, and Shanrong Zhao. 2017. “Evaluation and Comparison of Computational Tools for RNA-Seq Isoform Quantification.” *BMC Genomics* 18 (1): 1–11. https://doi.org/10.1186/s12864-017-4002-1.

Zieliński, Piotr, Katarzyna Dudek, Michał Tadeusz Stuglik, Marcin Liana, and Wiesław Babik. 2014. “Single Sucleotide Polymorphisms Reveal Genetic Structuring of the Carpathian Newt and Provide Evidence of Interspecific Gene Flow in the Nuclear Genome.” *PLoS ONE* 9 (5). https://doi.org/10.1371/journal.pone.0097431.

CHAPTER 2: INDEPENDENT EVOLUTIONARY TRANSITIONS TO PUERIPARITY ACROSS MULTIPLE TIMESCALES IN THE VIVIPAROUS GENUS *SALAMANDRA*

Kevin. P. Mulder[a,b,c,d], Lucía Alarcón-Ríos[e], Alfredo G. Nicieza[e], Robert C. Fleischer[d], Rayna C. Bell[c,f], Guillermo Velo-Antón[a]

[a] *CIBIO/InBIO, Research Center in Biodiversity and Genetic Resources, Rua Padre Armando Quintas 7, Campus Agrário de Vairão, 4485-661 Vairão, Portugal.*
[b] *Departamento de Biologia da Faculdade de Ciências da Universidade do Porto, Rua Campo Alegre, 4169-007, Porto, Portugal.*
[c] *Department of Vertebrate Zoology, National Museum of Natural History, Smithsonian Institution, 1000 Constitution Avenue NW, Washington, DC 20560, USA.*
[d] *Center for Conservation Genomics, Smithsonian Conservation Biology Institute, National Zoological Park, 3001 Connecticut Avenue NW, Washington, DC 20008, USA.*
[e] *Departamento de Biología de Organismos y Sistemas, Universidad de Oviedo UO, Oviedo, Spain*
[f] *Herpetology Department, California Academy of Sciences, 55 Music Concourse Drive, San Francisco CA 94118, USA*

Abstract

The ability to bear live offspring, viviparity, has evolved multiple times across the tree of life and is a remarkable adaptation with profound life-history and ecological implications. Within amphibians the ancestral reproductive mode is oviparity followed by a larval life stage, but viviparity has evolved independently in all three amphibian orders. Two types of viviparous reproduction exist in amphibians; larviparity and pueriparity. Larviparous amphibians deliver larvae into nearby ponds and streams, while pueriparous amphibians deliver fully developed juveniles and thus do not require waterbodies for reproduction. Among amphibians the salamander genus *Salamandra* is remarkable as it exhibits both inter- and intra-specific variation in the occurrence of larviparity and pueriparity. The evolutionary relationships among *Salamandra* lineages, including intra-specific lineages that differ in reproductive mode, are not well resolved which hampers our understanding of how often and when transitions between modes occurred. Furthermore, in species with intra-specific variation, the reproductive mode of a given population can only be confirmed by direct observation of births and thus the prevalence of pueriparous populations is incompletely documented. We used sequence capture to obtain 1,326 loci from 94 individuals from across the geographic range of the genus, focusing on potential reproductive mode transition zones. We also report additional direct observations of pueriparous births for 17 new locations and multiple lineages. We identify at least five independent transitions from the ancestral larviparity to pueriparity among and within species, occurring at different evolutionary timescales ranging from the Pliocene to the late Pleistocene. Four of these transitions occurred within species. We discuss the implications of our findings on the understanding of the evolution of this complex trait, and the potential of using five independent convergent transitions for further studies on the evolutionary pressures and genetic architecture of pueriparity.

Keywords: amphibians, ancestral state reconstruction, reproductive mode, sequence capture, viviparity

2.1 - Introduction

The ability to bear live offspring, viviparity, is a remarkable adaptive trait found across the tree of life, with at least 150 independent transitions to viviparity in vertebrates (Blackburn 2015; Gower et al. 2008; Reynolds, Goodwin, and Freckleton 2002; Helmstetter et al. 2016). Viviparity is associated with better protection of offspring, diversification and the exploitation of new ecological habitats (Helmstetter et al. 2016; Pincheira-Donoso et al. 2013), and the evolution of viviparity can thus have profound effects on the evolutionary trajectory of a given lineage. Within amphibians viviparity has evolved independently in all three amphibian orders, and two types of viviparous reproduction exist; larviparity and pueriparity (sensu Greven, 2003). This variation in viviparous strategies, allow us to understand and test the evolutionary and ecological context of viviparous reproduction using amphibians as a model system.

The ancestral reproductive mode for amphibians is oviparity; following the delivery and external fertilization of eggs, the young hatch as larvae and later undergo metamorphosis to their adult form (Wake 1982). Larviparity, which is documented in four species of salamander (Buckley 2012) and in one frog (Iskandar, Evans, and McGuire 2014), is characterized by internal fertilization, an incubation period, and subsequent delivery of larvae into nearby waterbodies (Greven 2003). It is hypothesized to help reduce egg predation and increase fecundity (Greven 2003). By contrast, pueriparity, in which species skip this larval stage and females deliver fully developed juveniles, is relatively common in caecilians (~34/214 species), and rare in anurans (16/7,164) and urodeles (11/742; Frost 2016; Buckley 2012; Sandberger-Loua, Müller, and Rödel 2017). Hypotheses for the evolution of pueriparity include that it is an evolutionary response to xeric climatic conditions and a corresponding lack of suitable water bodies for larval delivery (Liedtke et al. 2017; García-París et al. 2003; Velo-Antón, Zamudio, and Cordero-Rivera 2012; Beukema et al. 2010), or alternatively, that it is a response to high larval predation and natural larval loss (Greven 2003). Within salamanders, all known cases of larviparity and pueriparity are restricted to two sister genera; *Lyciasalamandra* in which all species are pueriparous, and *Salamandra* which includes six species with multiple representatives of both modes. This provides a unique comparative framework to investigate the genetic architecture and convergent evolution of larviparity and pueriparity.

The genus *Salamandra* is remarkable as it exhibits both inter- and intra-specific variation in the occurrence of larviparity and pueriparity. The two alpine species *S. atra* and *S. lanzai* are strictly pueriparous, the species *S. infraimmaculata* in the Eastern Mediterranean and *S.*

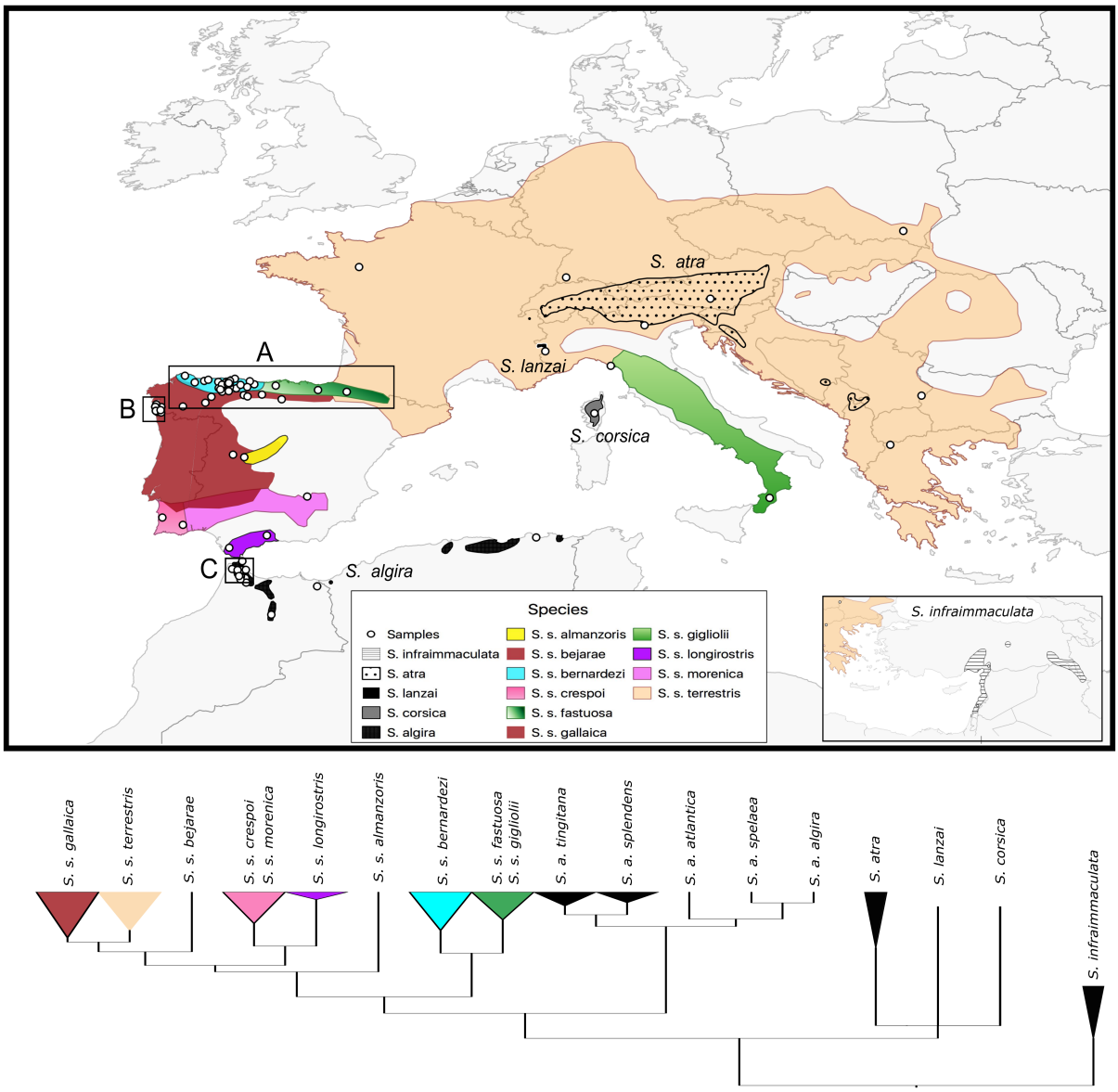
corsica on the Mediterranean island of Corsica are larviparous and the widespread species *S. salamandra* in Europe and *S. algira* in North Africa show intra-specific variation in reproductive mode (Figure 2.1 & Figure 2.2). Although phenotypic plasticity by either epigenetic inheritance or early life-stage imprinting have not been investigated, the delivery of larvae versus juveniles does not appear to be plastic as sexually mature females in controlled lab environments maintain their respective reproductive mode (Velo-Antón et al. 2007, 2015) and heterochronic changes occur in the early stages of embryonic development suggesting that pueriparity is not caused by the retention of larvae until metamorphosis (Buckley et al. 2007). *Salamandra* are thus a compelling system to test comparative evolutionary and ecological hypotheses related to the transition to pueriparity. However, the evolutionary relationships among different lineages, including intra-specific lineages that differ in reproductive mode, are not fully resolved, which hampers our understanding of how often and when transitions between modes occurred. Furthermore, in species with intra-specific variation in reproductive mode, the mode of a given population can only be confirmed by direct observation of births and thus the prevalence of pueriparous populations is incompletely documented.

Several species-level topologies have been proposed for *Salamandra* (Veith et al. 1998; Steinfartz, Veith, and Tautz 2000; Vences et al. 2014), but the most recent and complete genus level dataset points to a single transition to pueriparity in *S. atra*/*S. lanzai* and independent transitions for both *S. salamandra* and *S. algira* (Rodríguez et al. 2017). Within *S. algira* and *S. salamandra* the evolutionary relationships between the different subspecies are not well understood as many are based on morpho-types with limited genetic data supporting the proposed divisions (Figure 2.1). Within *S. salamandra*, pueriparity has been described in three of the 10 main subspecies. *Salamandra s. bernardezi* in northern Spain (Figure 2.1) is considered pueriparous (Buckley et al. 2007; Velo-Antón et al. 2015), but genetic diversity and divergence within the subspecies is high (Beukema et al. 2016; Lourenço et al. 2019), and for large parts of its range there are no direct observations of reproductive mode (Figure 2.2A). *Salamandra s. fastuosa*, found to the east of *S. s. bernardezi* (Figure 2.1 and Figure 2.2A) in northern Spain, includes both larviparous and pueriparous populations (Uotila et al. 2013), and pueriparity in this subspecies has been hypothesized to be due to introgression from pueriparous *S. s. bernardezi* (Figure 2.2A; Garcia-Paris et al. 2003). Additionally, two insular populations of *S. s. gallaica* are pueriparous (Figure 2.2B; Velo-Antón et al. 2007). The island populations are genetically distinct from adjacent mainland populations based on microsatellite markers (Velo-Antón, Zamudio, and Cordero-Rivera 2012; Lourenço, Sequeira, et al. 2018), however, the evolutionary relationships among these populations are unknown. Finally, within *S. algira tingitana* one mitochondrial lineages is documented as

pueriparous, while a second one is both larviparous and pueriparous (Figure 2.2C; Dinis and Velo-Antón 2017), however, nuclear relationships within *S. a. tingitana* are unresolved, and show signs of mitochondrial introgression (Dinis et al. 2019).

We aim to clarify the evolutionary history of *Salamandra* to determine the number and timing of transitions to pueriparity within the genus. We performed sequence capture on 1,326 loci of 94 individuals from across the geographic range of the genus, including all subspecies and pueriparous lineages within *S. salamandra* and *S. algira*, to reconstruct dated phylogenetic relationships. We also assessed the reproductive mode across *S. s. bernardezi* and *S. s. fastuosa* by directly observing delivery of females from 17 localities for which the mode was previously unknown, and summarizing all the pueriparous births previously reported in the literature for these subspecies (Figure 2.2A; Table B2). This study clarifies the number and timing of independent transitions to pueriparity across the genus, increases the known current geographic and phylogenetic extent of larviparity and pueriparity in *Salamandra,* and sets up the basis to further test the evolutionary and ecological context of viviparous reproduction using amphibians as a model system.

Figure 2.1 - Range of all six species and major subspecies of S*alamandra*, both *S. atra* and *S. lanzai* are strictly pueriparous and *S. corsica* and *S. infraimmaculata* are strictly larviparous. Below a schematic of the phylogenetic relationships of species, and *S. salamandra* and *S. algira* subspecies, based on results from this study.



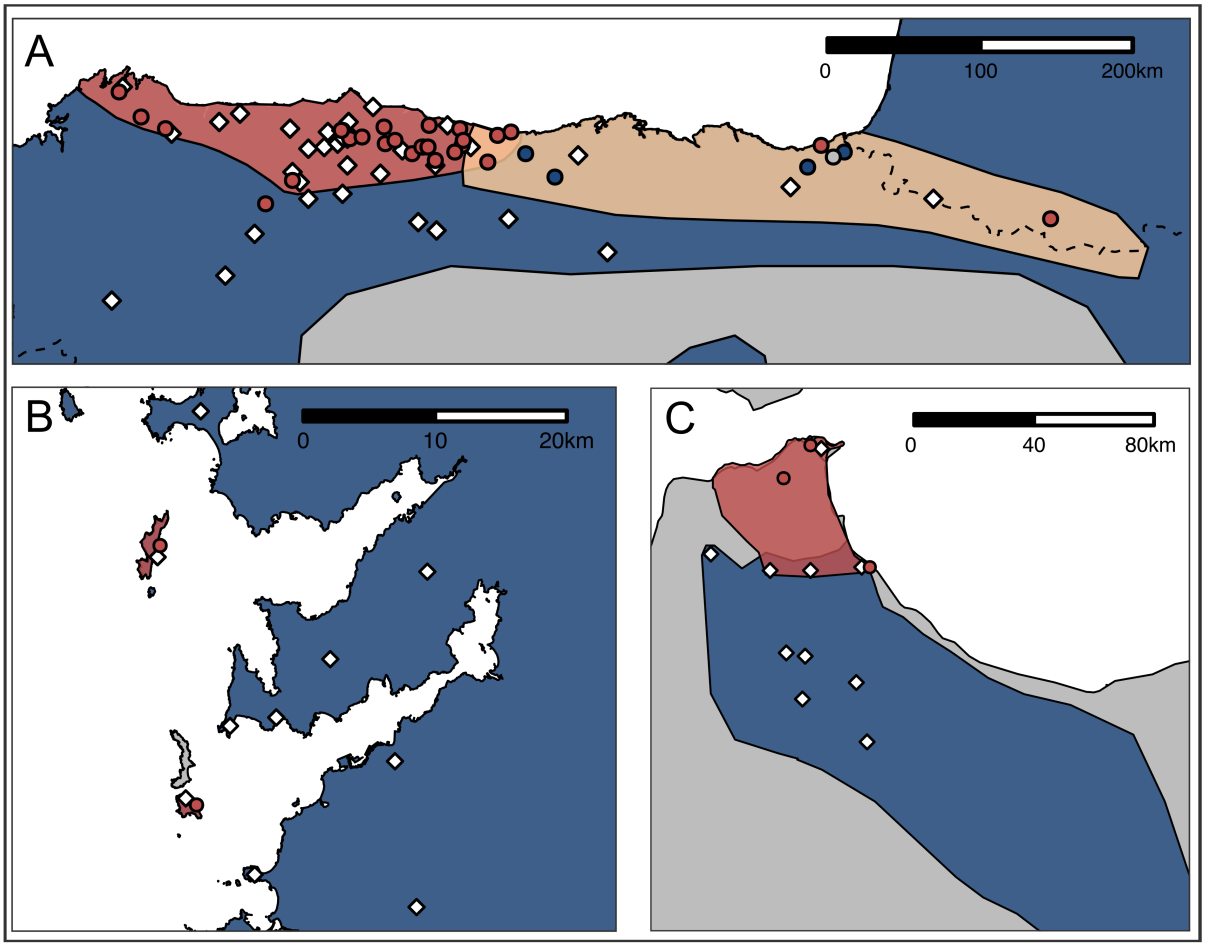
2.2 - Methods

2.2.1 - Field sampling and reproductive mode scoring

We collected toe-clips for genomic analyses from across the range of *Salamandra* (Figure 2.1, Table B1), focusing in and around intraspecific lineages for which the reproductive mode was known based on previous and current work: (a) *S. s. bernardezi*; (b) insular-mainland populations of *S. s. gallaica*, and (c) *S. a. tingitana* (Figure 2.2).

We also collected pregnant salamanders during the reproductive periods between 2015 and 2018 from 20 localities across the ranges of *S. s. bernardezi* (N=35) and *S. s. fastuosa* (N=2), to assess their reproductive modes (see Table B2). We transported females to laboratory facilities at the University of Oviedo and placed them in individual terraria (60x30x40 cm; LxWxH) equipped with a coconut fiber substrate, a container with water, moss, and shelters (bricks or barks). We fed them ad libitum twice a week with crickets and flour worms and collected tail-tips from females for mtDNA barcoding. After parturition, we returned them, together with their offspring, to the place of capture.

Figure 2.2. Sampling localities across three of our focal areas, see Figure 2.1 for location of the insets. Red indicates putative extent of pueriparity, blue larviparity, and in orange is the range of *S. s. fastuosa* for which both reproductive modes have been recorded. Genetic samples indicated by white diamonds, confirmed pueriparous births in with red dots, larviparous births in blue dots and populations with mixed births in grey. *S. salamandra* sampling localities in (a) northern Iberia, (b) Galicia, NW Spain (b), and (c) *S. algira* localities in northern Morocco.



2.2.2 - Sequence array design

We targeted a total of 1,326 loci, including 1,287 loci from transcriptome-based cDNA sequences and 39 nuclear loci available for *Salamandra* on GenBank (see Table B3 for a summary). To select the transcriptome-based loci, we mapped previously collected (see chapter 3) as well as publicly available RNAseq data (NCBI Bioproject PRJNA385088) to the 3,070 orthologous loci identified for *Salamandra* (Rodríguez et al. 2017), to find variable regions with phylogenetic signal. We identified a total of 1,287 unlinked Single Nucleotide Polymorphisms (SNPs) at a variety of phylogenetic levels (between species, between

subspecies and inter and intra-population levels) and sequences of ~175 base pairs (bps) were extracted around these SNPs, targeting a total of 201,026 bps.

For the GenBank based loci we selected sequences between 300-1,540 bps, including longer fragments for more informative loci, for a total of 20,198 bps. To compare the samples to the barcoded individuals with confirmed births (section 2.2.1), we also included a 130 bp fragment for the mitochondrial gene cytochrome B (CytB). Mitochondrial DNA enriches at higher efficiency during sequence capture than nuclear DNA, so at this length we can extract the target sequence plus flanking areas, resulting in an alignment that is comparable in size to previous CytB datasets collected by Sanger sequencing.

Following a quality control pipeline by Arbor Biosciences (Ann Arbor, MI, USA) to filter the probes for GC-content, repetitive elements and hybridization temperature, a total of 5,077 tiled baits of 90bp each were designed across these sequences, tiled at ~5X to increase the sequence capture efficiency. We combined the baits with a separate adaptive locus array for a forthcoming project, for a total of 40,000 baits, and performed joint sequence capture.

2.2.3 - Laboratory methods

2.2.3.1 - Genomic library preparation

We extracted genomic DNA from salamander toe-clips using protein precipitation from 94 tissue samples (Table B1) and eluted the extractions in 100ul of EB buffer. Following quantification with the Qubit 2.0 dsDNA HS Assay kit (Invitrogen, Carlsbad, CA, USA), we sheared 800 ng of DNA to a mean size range of ~ 300 bps with 20 cycles of sonication on the Bioruptor Pico (Diagenode, Liège, Belgium) in 0.2ml tubes at intervals of 30sec on/30 sec off with a short spin after the first 10 cycles.

Sheared DNA was prepared for sequencing following the BEST 2.0 protocol (Carøe et al. 2018) with dual indexed 7bps adapters (Kircher, Sawyer, and Meyer 2012). We added stubby adapters at a 30X excess during ligation, and amplified half of the final solution using 9-10 cycles of indexing PCR. We pooled samples equimolarly in groups of six aiming for a total input of 3000 ng per pool, and ran a subset of pools on the TapeStation 2100 using the High Sensitivity Assay (Agilent technologies, Santa Clara, CA, USA) at several stages of the protocol.

2.2.3.2 - Sequence capture and sequencing

Pooled libraries and baits were hybridized for 36 hours following the MyBaits v3 protocol (https://arborbiosci.com/wp-content/uploads/2017/10/MYbaits-manual-v3.pdf) with twice the recommended amount of cot-1 blocker. Following stringent washes, we re-amplified the pools in two separate PCR reactions for 9-13 cycles. A final pool was prepared for paired end 150 bp sequencing on part of an Illumina NovaSeq S4 run at the UC Davis Genome Center. We pooled samples equimolarly, sourcing DNA from both PCR reactions but with a preference for the reaction with the least number of cycles to reduce the number of PCR duplicates in sequencing.

2.2.3.3 - Barcoding of pregnant females with Cytochrome B

We extracted genomic DNA from tissue samples of collected pregnant females using the EasySpin Genomic DNA Tissue Kit (Citomed, Lisboa, Portugal), following the manufacturer's protocol. We amplified and sequenced a CytB fragment of ca. 1100 bp, following the protocol described in (Beukema et al. 2016) and outsourced DNA sequencing to Genewiz Inc. (Leipzig, Germany).

2.2.4 - Bioinformatic processing

The majority of loci we targeted were based on transcriptome cDNA sequences that do not include introns and thus are not an accurate genomic reference for mapping capture data. To split up the putative exons within a locus we applied the IEB-finder pipeline that identifies intron-exon boundaries by means of mapping scores (Deleury et al. 2019). In short, gDNA reads are mapped against a cDNA reference using a local mapper bwa mem (Li and Durbin 2009), the parts of the read that represent the intron are soft-trimmed. IEB-finder scans a bam file to identify regions that have above average soft-trimming compared to the surrounding region and identifies them as putative exon-intron boundaries (Deleury et al. 2019). We used a representative pool of 12 samples of *S. salamandra* to run the IEB-finder pipeline (parameters -e 0 -c 10 -x 30) and split up our loci into separate exons for all identified boundaries.

We ran the new reference through the SECAPR pipeline (Andermann et al. 2018), using the same 12 samples to identify potential paralogs and duplicate loci. In short, reads were quality filtered and assembled individually using abyss (-k 90; Birol et al. 2009), and the resulting assemblies were compared to the reference by means of reciprocal blast using LASTZ (--min-coverage 80, --min-identity 80; Harris 2007). We manually examined loci that were found to either contain potential paralogs (multiple contigs hitting the same locus), or duplicate loci (one contig hitting two loci). Duplicate loci were often found to be due to short introns/indels that caused IEB-finder to split the locus up, but in which flanking parts of the sequence reads were long enough to bridge this gap and form one assembled contig. We combined these loci into one locus for an updated reference that included the intron. We examined all paralogs to determine if they could unequivocally be split into two loci by shifting the reference sequence. If there was a clear distinction we extracted two separate loci for the updated references but if the paralogs were too similar, we removed the locus.

We ran the updated reference through the SECAPR pipeline again, using all 94 individuals but increasing the LASTZ threshold (--min-coverage 90, --min-identity 90). Following the remapping step, we ran GATK 3.8.1 (McKenna et al. 2010) across all bam files to call high quality SNPs using information from all samples. We used the EMIT-ALL-SITES to also keep non-variable sites. This combined strategy allowed us to include all available evidence to determine SNP quality, and additionally identify loci with a heterozygosity excess that are likely paralogs, while still keeping non-variable sites for phylogenetic analyses. Following strict filtering of low-quality and low coverage SNPs, indels and paralogous loci, alignments were extracted from the vcf file with vcf2phylip (Ortiz 2019) allowing for a maximum of 50% missing data across each site. We concatenated all nuclear loci for a final alignment and analysed the mitochondrial locus separately.

2.2.5 - Phylogenetic reconstruction

2.2.5.1 - Multispecies coalescent analyses of species-level relationships

Given the mixed support for some species-level relationships within the genus (Vences et al. 2014; Rodríguez et al. 2017), we first employed a multispecies coalescent (MSC) approach to infer the species tree for the six *Salamandra* species using one representative sample for every sub-species (23 samples total, Table B1). To verify that the result was not driven by increased sampling in certain species we additionally conducted the analyses with only one

sample per species. Due to the low average length of the loci in our dataset (average 351 bps), we did not pursue gene-tree based methods because poorly-supported gene trees are known to influence subsequent species-tree inference (Salichos and Rokas 2013). Instead, we used a SNP based approach using SNAPP (Bryant et al. 2012), as implemented in Beast 2.6.0 (Bouckaert et al. 2014). SNAPP requires unlinked loci, thus we filtered the SNP dataset to only include one SNP per locus and not allowing for any missing data, resulting in a final dataset of 1,041 SNPs. The SNAPP input file was generated using BEAUti, calculating and sampling the mutation rates U and V from our data, and sampling the coalescence rate with a starting value of 10. Using the estimated age of the genus (Vences et al. 2014), we estimated a starting value for lambda (0.29) using the python script Pyule (available at https://github.com/joaks1/pyule) and applied a uniform distribution. Alpha and Beta for the theta prior were set to default, to explore a wide range of values. We ran SNAPP for 50 million generations, storing the chain every 1,000 trees and assessed convergence using Tracer v1.71 (Rambaut et al. 2018). Removing a burn-in of 10% of the trees we depicted the remaining trees using DensiTree to visualize variation in the posterior distribution of topologies and branch lengths. A maximum clade credibility species tree was subsequently built using TreeAnnotator (Bouckaert et al. 2014) using the same burn-in of 10%.

2.2.5.2 - Phylogenetic reconstruction of intraspecific relationships

We estimated the combined phylogeny and divergence times of our concatenated dataset of the two species with intraspecific variation in reproductive mode, *S. salamandra* (74 samples) and *S. algira* (14 samples), using Bayesian inference in BEAST 2.6.0 (Bouckaert et al. 2014) applying a strict molecular clock. We applied the coalescent constant population tree prior as our dataset was mostly population-based and only included two species. The substitution model for the concatenated alignment was estimated during the BEAST run with bModelTest 1.1.0 (Bouckaert and Drummond 2017). To time-calibrate the phylogeny we applied a prior for the split between *S. algira* and *S. salamandra* at 5.6 mya with a normal distribution and a sigma of 0.13. This corresponds to the estimated time of the Messinian Salinity Crisis which has been hypothesized to be responsible for the divergence of multiple amphibian species pairs across the strait of Gibraltar (Ehl, Vences, and Veith 2019). We ran the MCMC chain twice for 100 million generations sampling every 1,000 generations and monitored convergence using Tracer v1.71 (Rambaut et al. 2018). We combined tree files using LogCombiner v1.8.4, removing the first 10% as burn-in and built a maximum clade credibility tree using TreeAnnotator (Bouckaert et al. 2014).

To confirm our Bayesian analyses, a maximum likelihood tree was inferred using RAxML 8.2.12 (Stamatakis 2014) applying the GTRCAT substitution model on the concatenated nuclear alignment of all samples and starting from ten parsimony informed trees and ten random trees. Bootstrap support was computed on the best scoring tree by means of 100 iterations of rapid bootstrapping (Stamatakis, Hoover, and Rougemont 2008).

2.2.5.3 - Mitochondrial barcoding of pregnant females

To determine the phylogenetic placement of all the pregnant females we used the CytB barcode. Sanger sequence chromatograms for the 37 pregnant females were inspected and aligned using Geneious (Kearse et al. 2012). For the 94 sequence capture samples, we mapped all the filtered reads to the full *S. salamandra* mitochondrial genome (Mulder et al. 2016) using bowtie2 (Langmead and Salzberg 2012) and kept only correctly paired reads. We inspected mapped reads for potential contamination and called consensus sequences requiring a minimum of 12 reads.

All combined sequences were trimmed to 802 bps and aligned using MUSCLE 3.8.425 (Edgar 2004) as implemented in Geneious prime 2019 (Kearse et al. 2012) for a maximum of eight iterations. We performed Bayesian phylogenetic analyses in BEAST version 2.6.1 (Drummond et al. 2012) on the CIPRES Science Gateway (Miller, Pfeiffer, and Schwartz 2010), and selected the optimal nucleotide substitution model (TrN) with JMODELTEST version 2.1.4 (Darriba et al. 2012), under the Bayesian information criterion (BIC). We performed three independent runs using an uncorrelated relaxed clock and a constant population size model as the coalescent tree prior, with a total of 100 million generations per run. We verified parameter convergence by examining the effective sample sizes (ESSs) in TRACER version 1.6 and removed the first 10% as burn-in. We obtained a maximum clade credibility summary tree with Bayesian posterior probabilities (BPP) for each node using TreeAnnotator v 1.8.4, and edited the resulted tree in Figtree version 1.4.3 (http://tree.bio.ed.ac.uk/software/figtree).

2.2.6 - Ancestral state reconstruction of reproductive mode

To estimate the number of transitions between larviparity and pueriparity among lineages of *S. salamandra* and *S. algira*, we performed ancestral state reconstruction (ASR) on our concatenated and dated BEAST phylogeny (Figure B1), coding the two reproductive modes

as discrete characters. Although ASR is usually applied to phylogenies in which all tips are species, the approach can be applied to sub-specific level variation similar to that observed in the *Salamandra* system (Richmond 2006; Joy et al. 2016). We performed stochastic character mapping using SIMMAP (Bollback 2006) to estimate the ancestral reproductive modes across the phylogeny, and to estimate the number of independent transitions to pueriparity. We applied the make.simmap function in phytools 0.6 (Revell 2012), as implemented in R 3.6.3 (R Core Team 2019) on a random subset of 100 trees from the posterior distribution of the BEAST Bayesian inference after removal of the burn-in of 10% (Figure B1). This method simulates character evolution along the phylogeny using an MCMC approach and samples the posterior distribution of transitions to estimate the probabilities of each character state on all nodes. We first used a likelihood-ratio test to compare the two different models of evolution (equal rates, and all-rates-different) between our two discrete characters, using the Akaike Information Criterion (AIC) to choose the best model. With the best-fit model (equal rates) we ran 10,000 MCMC simulations and mapped the posterior probabilities onto the phylogeny to identify the locations of putative reproductive mode transitions.

2.3 - Results

2.3.1 - Reproductive mode assessment

We confirmed pueriparity in all 35 *S. s. bernardezi* females. Clutch sizes were highly variable, with a mean clutch size of 8.9 individuals per female, ranging from one individual to a maximum of 24 juveniles (Table B2). Together with fully metamorphosed juveniles, a few females delivered some gilled individuals, but in a very advanced stage of development. Those individuals had started the colouration change from typical greyish of larvae to black and yellow adult colouration and completed metamorphosis in a short period after parturition. The two *S. s. fastuosa* individuals were confirmed as larviparous as both females delivered fully aquatic larvae, with the typical greyish coloration and morphological traits of *Salamandra* larvae and less variability in the stage of development within clutches compared to the pueriparous births.

2.3.2 - Bioinformatic processing

Illumina NovaSeq S4 sequencing resulted in an average of 8.1 million paired-end reads per individual split between the 1,326 markers included in this study and data for a forthcoming, separate project. After splitting the exons and manual filtering based on the initial SECAPR results on 12 individuals, our reference included 2,363 loci. Removing loci with excess heterozygosity and low coverage reduced this dataset to 2,287 loci (average coverage per individual 36, CI 5.6-78). Our 50% missing data threshold resulted in a total concatenated alignment of 574,577 bps.

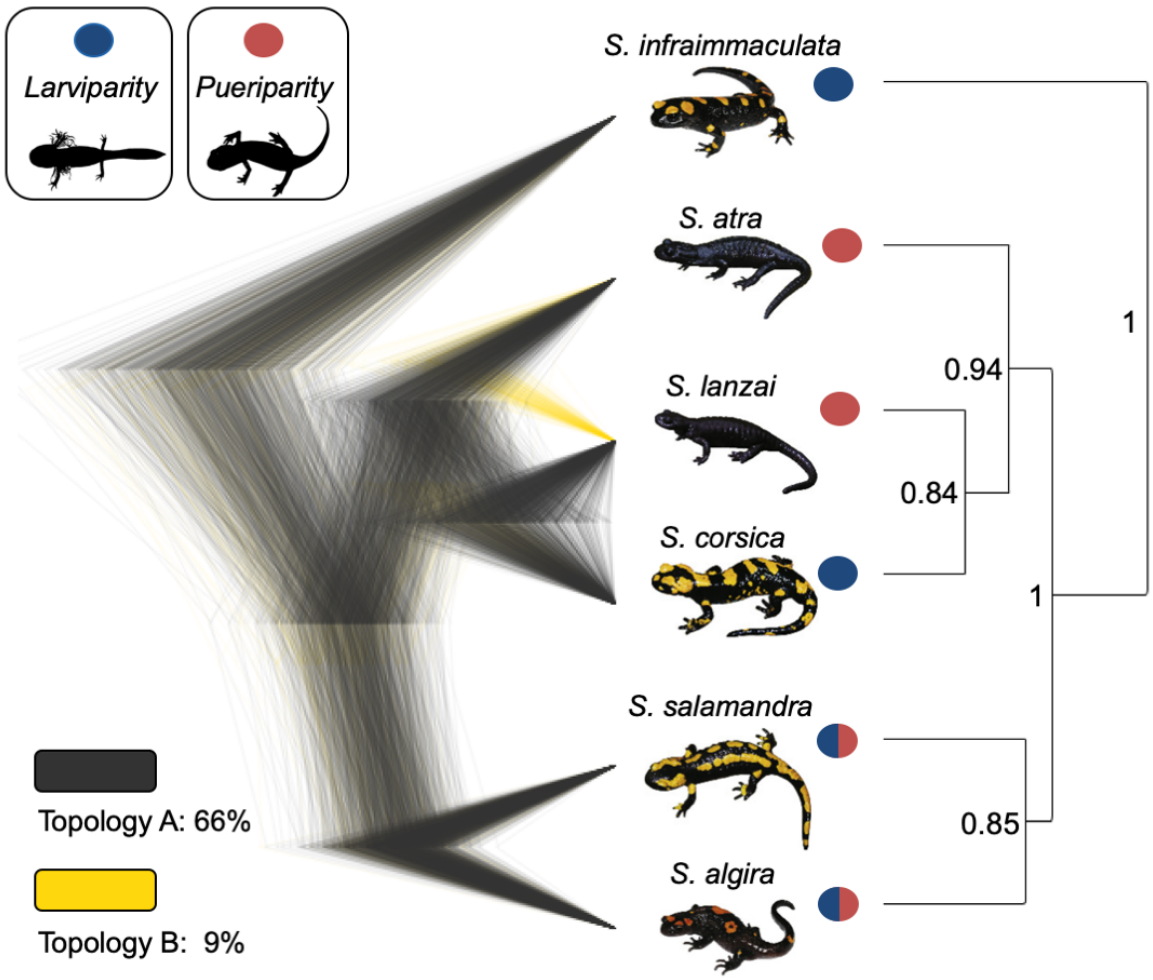
2.3.3 - Phylogenetic reconstruction

2.3.3.1 - Multispecies coalescent tree for species level relationships

The DensiTree plot from the MSC analyses (Figure 2.3) reveals the uncertainty in the relationships among *S. atra, S. corsica* and *S. lanzai*; 66% of all topologies in the posterior distribution of phylogenies place *S. corsica* and *S. lanzai* as sister species, whereas 9% place *S. atra* and *S. lanzai* as sister species. The type locality analysis supported the same top two

topologies with 34% and 31% of all trees respectively (Figure B2). The consensus tree reflects this uncertainty with low support for these nodes. The posterior probability in the *S. salamandra*, *S. algira* node was low in the 23 individual analyses (0.85; Figure 2.3), but high in the SNAPP analysis that only included the type localities (0.98; Figure B2).

Figure 2.3. DensiTree plot of 45000 generated SNAPP trees of the 23 sub-species representing the 6 species in *Salamandra*. The most common topology is in dark grey (66%), and the second most common in yellow (9%). Remaining topologies have been removed for clarity. On the right is the consensus tree as generated by tree-annotator and with posterior probabilities indicated on the nodes.



2.3.3.2 - Dated phylogeny of the concatenated alignment

Bayesian inference with BEAST on the concatenated alignment of 88 samples, resolved the majority of nodes with high support (posterior probability (pp) = 1; Figure 2.3 and Figure B1).

The maximum likelihood tree largely followed the same general topology but node support was lower across some parts of the tree (Figure B3). The nodes at the root of the transitions to pueriparity were equal between both methods and fully supported with a pp = 1 or a bootstrap score of > 99%.

Both species were monophyletic but some recognized subspecies formed paraphyletic groups. For instance, *S. s. gigliolii* sits within the *S. s. fastuosa* clade, which in turn is sister to the *S. s. bernardezi* clade. Likewise, several individuals that were identified as *S. s. bejarae* based on locality and morphology are within the larger *S. s. gallaica* clade, whereas *S. s. bejarae* from the type locality (Candelario) were placed outside of the S*. s. gallaica* clade in a monophyletic lineage (Figure B1).

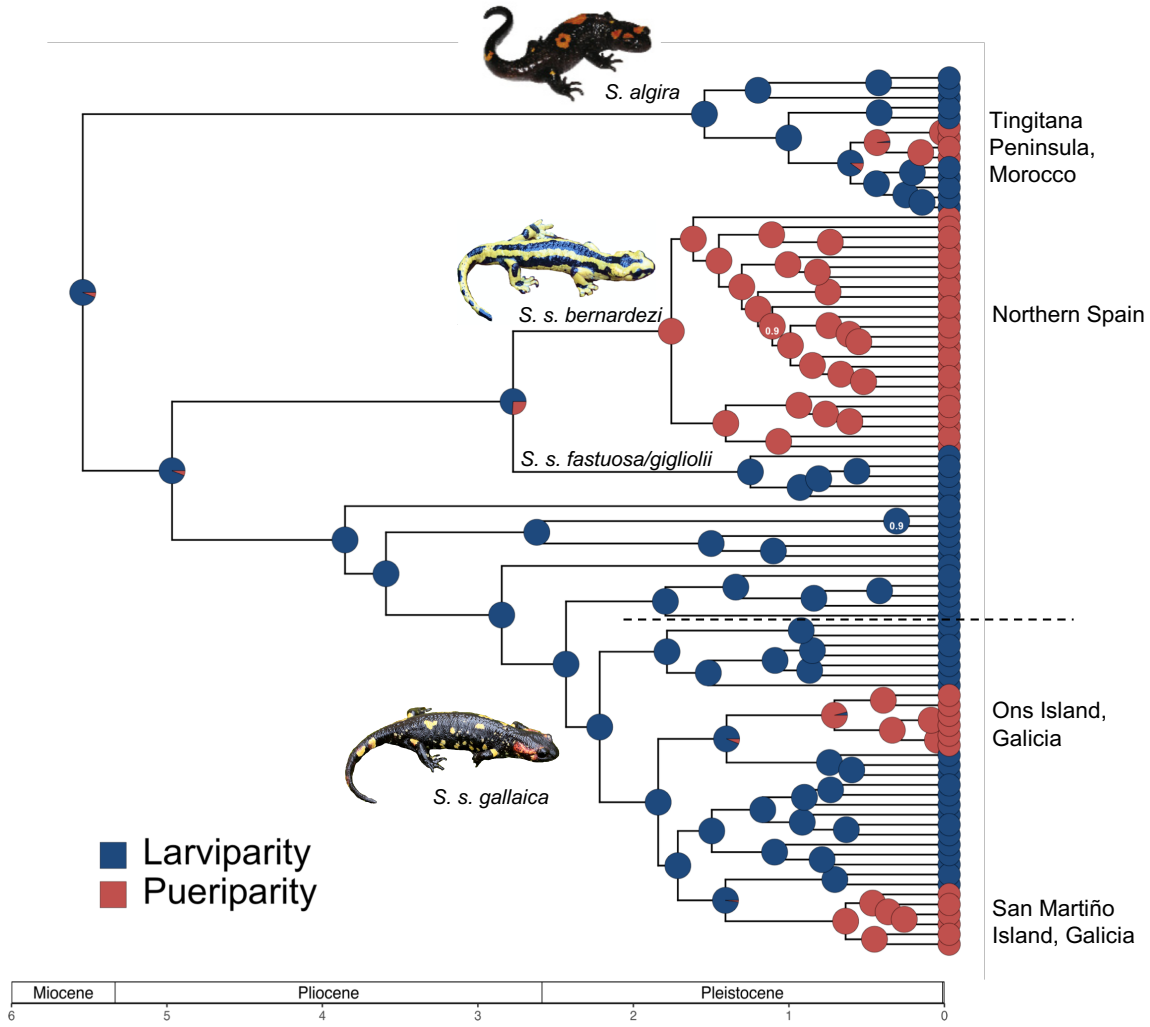
2.3.3.3 - Mitochondrial barcoding of pregnant females

We obtained CytB barcode sequences for 26 pregnant females across the *S. s. bernardezi* (N=25) and *S. s. fastuosa* (N=1) range (see Table B2). They were placed in the *S. s. bernardezi* mtDNA clade and included representatives of all major sublineages within the subspecies, with the exception of two *S. s. bernardezi* samples and the single *S. s. fastuosa* sample, which were identified as *S. s. gallaica* mtDNA. All of them show phenotypic characters typical of *S. s. bernardezi-fastuosa* (striped colouration pattern, round snout shape and small body size; Alarcón-Ríos et al. 2020), supporting the existence of mitochondrial introgression across *S. s. bernardezi* range (Lourenço et al. 2019).

2.3.4 - Ancestral state reconstruction reveals five independent transitions to pueriparity

The equal rates transition rate model was the best fit to our dataset and the Bayesian inference ancestral state reconstruction indicated a total of four independent transitions to pueriparity (Figure 2.3) with no reversals to larviparity. There is additionally at least one transition to pueriparity for *S. atra*/*S. lanzai* based on the MSC analyses. The intraspecific transitions include three independent transitions on the continental islands of Ons and San Martiño and in the northern populations of *S. a tingitana* that likely occurred during the late Pleistocene and one transition in the subspecies *S. s. bernardezi* that occurred in the Early Pleistocene.

Figure 2.4. Bayesian inference based on a concatenated dataset of 574k bps of 88 samples of S. salamandra and S. algira. Node support was 1, unless otherwise stated in white. The tips of the tree are coded by reproductive mode, blue for larviparity and red for the pueriparity. The results of the ancestral state reconstruction are placed on the internal nodes. Sub-species designation and location are included in Figure B1. The extent of the S. s. gallaica clade is indicated by the dashed line.



2.4 - Discussion

Our study provides strong support for a minimum of five independent transitions to pueriparity across multiple timescales in the genus *Salamandra*, indicating that different combinations of climatic and local evolutionary pressures may lead to the development of this complex trait. In addition, we demonstrate that sequence capture using transcriptome-based loci can produce high quality data to solve phylogenetic relationships at both inter- and intra-specific levels, even for the large genomes of urodeles (Gregory 2003; Weisrock et al. 2018). Finally, our increased geographic sampling of documented births and their phylogenetic position confirms that *S. s. bernardezi* is pueriparous across its geographic range, and all phylogenetic sub-lineages, and can be considered fully pueriparous.

2.4.1 - Systematics revision within *Salamandra salamandra*

Many subspecies have been described for *Salamandra salamandra*, but most lack genetic support for these designations. Our phylogenetic analyses suggest that at least two subspecies require systematics revision. *Salamandra s. bejarae* is considered to have a wide range across much of central Iberia (Figure 2.1), but the majority of our samples from that area grouped within the larger *S. s. gallaica* clade. This supports previous results unveiling *S. s. bejarae* as paraphyletic to *S. s. gallaica* across the mountains of the Iberian Central System (Pereira, Martínez-Solano, and Buckley 2016). Given that one *S. s. bejarae* sample (Candelario) very close to the type locality (Béjar) was distinct from *S. s. gallaica* and the other samples of *S. s. bejarae*, it is possible that *S. s. bejarae* is indeed monophyletic, but inhabits a much smaller geographic area than is presently attributed to the subspecies. Likewise, *S. s. gigliolii* occurs entirely within the larger clade of *S. s. fastuosa* from northern Iberia suggesting that its current allopatric distribution in Italy is the result of a past range expansion from an ancestral *S. s. fastuosa* population (Steinfartz, Veith, and Tautz 2000).

2.4.2 - Complex evolutionary history of *S. atra/lanzai/corsica* clade, and uncertainty in the number of transitions to pueriparity

The topology with the highest support as found by multispecies coalescent analyses in SNAPP included *S. lanzai* and *S. corsica* as sister species, and *S. atra* as sister to this clade. However,

a substantial part of the posterior distribution also supported a monophyletic grouping of the two pueriparous species (*S. atra* and *S. lanzai*), corroborating a previous study based on a combined RNAseq and ddRAD dataset of the genus (Rodríguez et al. 2017). This would also be the most parsimonious explanation when considering that both alpine species are geographically close, compared to the insular *S. s. corsica*, and that they are melanic and pueriparous.

The complicated history of *S. corsica*, *S. atra* and *S. lanzai* could be explained by introgression, and at least one instance of introgression is suspected from the mitochondrial tree which shows a different topology than the MSC (Figure B4, and Figure 4 in Rodríguez et al. 2017). This complicated history has been hypothesized to be caused by the Messinian Salinity Crisis (5.6 mya, Pliocene) and the desiccation of the Mediterranean Sea and the simultaneous speciation it induced in this clade of three species, which would entail that the mitochondrial introgression also occurred early in the history of the three species as they have likely been parapatric since then (Vences et al. 2014; Rodríguez et al. 2017). Given the high uncertainty in this node across multiple datasets and analyses we cannot distinguish between, a single transition for both *S. atra* and *S. lanzai*, the possibility of two independent transitions to pueriparity, or that the common ancestor to all three species was pueriparous and that there was a subsequent reversal to larviparity in *S. corsica*.

2.4.3 - A Pleistocene transition in *S. s. bernardezi* and retained larviparity for the sister clade of *S. s. fastuosa* and *S. s. gigliolii*

Salamandra s. bernardezi is considered to be a pueriparous subspecies, but its parity had only been confirmed from a few localities (see Table B2). Our direct observations of births at 17 localities across the subspecies range including several distinct genetic lineages (Figure 2.2A & Figure B4), confirm that pueriparity is likely the prevailing or only parity mode for this subspecies. In our phylogenetic analyses, *S. s. bernardezi* is sister to the combined clade of *S. s. fastuosa* (both larviparity and pueriparity documented) and *S. gigliolii* (only larviparity documented). Although there are confirmations of pueriparity in some *S. s. fastuosa* individuals, it does not appear to be the prevailing mode in this subspecies (see Uotila, Crespo-Diaz, Sanz-Azkue, & Rubio, 2013; presence of larvae across *S. s. fastuosa* range, GVA observations). The original hypothesis by Garcia-Paris et al., (2003) suggests that the appearance of pueriparity, colour pattern and head shape in *S. s. fastuosa* is due to introgression from *S. s. bernardezi* into *S. s. fastuosa* via male biased dispersal which seems

to be the prevailing process to explain mito-nuclear discordances in *S. salamandra* (García-París et al. 2003; Pereira, Martínez-Solano, and Buckley 2016; Bisconti et al. 2018), and the apparent higher philopatric pattern in *S. salamandra* females (Lourenço, Antunes, et al. 2018; Helfer, Broquet, and Fumagalli 2012). This topology and demographic history would correspond with a single transition to pueriparity in *S. s. bernardezi* in the middle Pleistocene (1.78mya, 95% CI:1.66-1.98), retained larviparity in *S. s. fastuosa* and *S. s. gigliolii*, and subsequent introgression of pueriparity. More direct observations of parity mode across the range of *S. s. fastuosa* coupled with nuclear genetic data and demographic modelling are needed to fully explore and test these hypotheses.

2.4.4 - A late Pleistocene transition in *S. algira* followed by mitochondrial introgression

Documented pueriparous births in *S. algira* are highly scarce (Donaire-Barroso and Bogaerts 2001; Donaire-Barroso, Bogaerts, and Herbert 2001; Dinis and Velo-Antón 2017). Most pueriparous populations fall within a single mtDNA sublineage of *S. a. tingitana*, which spreads across the northern Tingitana Peninsula in Morocco (north of the river Oued Martil), where water bodies lack salamander larvae. Pueriparity was also confirmed in one neighbouring population, south of this river, which belongs to a distinct mtDNA sublineage of *S. a. tingitana*, and suggested as another case of mtDNA introgression between sister taxa across a contact zone (Dinis et al. 2019). The produced nuclear phylogeny shows that all the pueriparous populations (Figure 2.2C) form one clade, which suggests a single transition to pueriparity in the late Pleistocene (474 kya, CI: 429-515). All confirmed pueriparous populations are found in areas with little to no surface water, low average precipitation in the coldest quarter, and populations retreating to karstic systems during dry periods and the geographic extent of this single nuclear clade thus overlaps strongly with the predicted distribution of pueriparity based on environmental models (Beukema et al. 2010). This predicted distribution combined with our nuclear and mitochondrial data suggests that pueriparity evolved once above the Quad Martil river and likely expanded south via male biased dispersal to Amsa and Tetouan to colonize this suitable pueriparous habitat (Figure 2.2C).

2.4.5 - Two independent transitions in *S. s. gallaica* on the continental islands of Ons and San Martiño

The insular populations of San Martiño and Ons are both pueriparous (Velo-Antón et al. 2007; Velo-Antón, Zamudio, and Cordero-Rivera 2012; Velo-Antón et al. 2015) but their nuclear phylogenetic relationships to the mainland were unknown. The ASR supports two independent transitions to pueriparity from the ancestral larviparous state in *S. s. gallaica*, showing that San Martiño and Ons populations are not each other's closest relatives and that they independently became separated from the mainland populations. The Ons population is closely related to the larviparous population on the Grove peninsula (a former island reconnected to the mainland with the deposition of river sediments during the XVII-XVIII centuries), which is consistent with bathymetric data that connect those corresponding regions at lower sea levels (Figure 2.2B and Lourenço, Sequeira, et al. 2018). San Martiño is connected to the Monteferro peninsula, which is its geographically closest mainland population but which shows a deeper bathymetric depression compared to the Morrazo peninsula (e.g. Melide and Nerga populations). Interestingly, some females of the larviparous Monteferro population (where larvae are commonly found in water bodies) showed signs of a mixed reproductive mode (laying both fully metamorphosed juveniles and young larvae or larvae at a later developmental stage). Whether this is a retention of a previous pueriparous ancestral state or the result of an ongoing adaptive process to pueriparity is still unknown (Velo-Antón et al. 2015). The estimated divergence dates for the island lineages (Ons: 712kya 95% CI: 656-771 and San Martiño 650 kya, 95% CI: 600-702) is, however, puzzling. Our estimates predate the formation of these islands during the sea level rise after the last glacial maximum during the early Holocene (ca. 8,000 ya), which is assumed as the biogeographic event that disconnect the present insular populations from the mainland counterparts (Velo-Antón et al. 2007). Overestimation of recent node ages is a known bias in divergence dating approaches, especially when using population level genetic data (Ho et al. 2005), skewed dating priors (Phillips 2009; Duchêne, Lanfear, and Ho 2014), and given the uncertainty in the estimated node ages it is possible that the island populations did not become isolated from the adjacent mainland populations until the Holocene. On the other hand, the presence of mixed reproductive individuals in Monteferro opens the hypothesis that insular populations diverged earlier than the formation of the islands, with a posterior extinction of this lineage along the coastal populations. However, this scenario would likely imply divergent mitochondrial groups (as it occurs in *S. s. tingitana*), while there is mtDNA haplotype sharing across populations of this island-mainland system (Velo-Antón et al. 2007; Velo-Antón, Zamudio, and Cordero-Rivera 2012; Lourenço, Sequeira, et al. 2018).

2.5 - Conclusion

Our analyses indicate that the transition to pueriparity has occurred at least five times in the genus *Salamandra.* Transitions to pueriparity arose at different evolutionary time periods ranging from the Pliocene to the late Pleistocene, suggesting that a combination of climatic and local environmental conditions form the evolutionary pressures that lead to this major life-history transition. The number of transitions between reproductive mode is remarkable considering the age of the clade and the number of species. Intra-specific variation in reproductive mode is also rare, and to our knowledge this is the only case in which this occurs in two sister-species. The putative introgression events at multiple phylogenetic levels, as evident from the numerous cases of mito-nuclear discordance, also highlights the potential that reproductive mode shifts can lead to adaptive geographic expansions along suitable habitat. Combining this phylogenetic framework with environmental data can help us understand the evolutionary pressures working on reproductive mode.

Much of the research on viviparity has focused on squamates given the high number of transitions across this large clade (Blackburn 2015). We propose that the genus *Salamandra* and its pueriparous sister clade *Lyciasalamandra* are an excellent parallel case-study to investigate both the transition to viviparity, and the distinction between larviparity and pueriparity. In particular, convergent evolution of pueriparity at multiple timescales is ideal for further research into the adaptive genomic architecture of this complex trait and the evolutionary and ecological context in which it is adaptive.

2.6 - Acknowledgements

We thank Nancy McInerney and Jeff Hunt for facilitating lab work at the Center for Conservation Genomics and the Laboratory of Analytical Biology respectively. Brian Brunelle and Alison Devault at Arbor Biosciences for help and advice on bait design. Lillian Parker was instrumental in developing and optimizing the library prep protocol. Thanks to Duarte Gonçalves and Ivan Prates for their insights on the analyses, Robert Wilson and Addison Wynn for logistical support and André Lourenço, Marco Dinis and Adam Marques for assistance in the lab. Miguel Vences and Sebastian Steinfartz for providing tissue samples from some *Salamandra* lineages and for the pictures of the different species. Portions of the laboratory and computer work were conducted in and with the support of the L.A.B. facilities of the National Museum of Natural History (NMNH). The computations performed for this paper were conducted on the Smithsonian High Performance Cluster (SI/HPC), Smithsonian Institution. https://doi.org/10.25572/SIHPC. Financial support came from the office of the Associate Director of Science at the Smithsonian's National Museum of Natural History, and by National Funds through FCT – Foundation for Science and Technology (EVOVIV: PTDC/BIA-EVF/3036/2012; SALOMICS: PTDC/BIA-EVL/28475/2017). KPM was funded by an FCT predoctoral grants (PD/BD/52604/2014) and GVA by an FCT research contract (IF/01425/2014), all from the Portuguese ''Fundação para a Ciência e a Tecnologia," funded by Programa Operacional Potencial Humano (POPH)—Quadro de Referência Estratégica Nacional (QREN) from the European Social Fund.

2.7 - References

Alarcón-Ríos, Lucía, Alfredo G. Nicieza, Antigoni Kaliontzopoulou, David Buckley, and Guillermo Velo-Antón. 2020. "Evolutionary History and Not Heterochronic Modifications Associated with Viviparity Drive Head Shape Differentiation in a Reproductive Polymorphic Species, *Salamandra salamandra*." *Evolutionary Biology* 47 (1): 43–55. https://doi.org/10.1007/s11692-019-09489-3.

Andermann, Tobias, Ángela Cano, Alexander Zizka, Christine Bacon, and Alexandre Antonelli. 2018. "SECAPR - A Bioinformatics Pipeline for the Rapid and User-Friendly Processing of Targeted Enriched Illumina Sequences, from Raw Reads to Alignments." *PeerJ* 2018 (7): e5175. https://doi.org/10.7717/peerj.5175.

Beukema, Wouter, Alfredo G. Nicieza, André Lourenço, and Guillermo Velo-Antón. 2016. "Colour Polymorphism in *Salamandra salamandra* (Amphibia: Urodela), Revealed by a Lack of Genetic and Environmental Differentiation between Distinct Phenotypes." *Journal of Zoological Systematics and Evolutionary Research* 54 (2): 127–36. https://doi.org/10.1111/jzs.12119.

Beukema, Wouter, Philip De Pous, David Donaire, Daniel Escoriza, Serge Bogaerts, Albertus G. Toxopeus, Cornelis A. J. M. De Bie, Josep Roca, and Salvador Carranza. 2010. "Biogeography and Contemporary Climatic Differentiation among Moroccan *Salamandra algira*." *Biological Journal of the Linnean Society* 101 (3): 626–41. https://doi.org/10.1111/j.1095-8312.2010.01506.x.

Birol, Inanç, Shaun D. Jackman, Cydney B. Nielsen, Jenny Q. Qian, Richard Varhol, Greg Stazyk, Ryan D. Morin, et al. 2009. "De Novo Transcriptome Assembly with ABySS." *Bioinformatics* 25 (21): 2872–77. https://doi.org/10.1093/bioinformatics/btp367.

Bisconti, Roberta, Daniele Porretta, Paola Arduino, Giuseppe Nascetti, and Daniele Canestrelli. 2018. "Hybridization and Extensive Mitochondrial Introgression among Fire Salamanders in Peninsular Italy." *Scientific Reports* 8 (1): 1–10. https://doi.org/10.1038/s41598-018-31535-x.

Blackburn, Daniel. 2015. "Evolution of Vertebrate Viviparity and Specializations for Fetal Nutrition: A Quantitative and Qualitative Analysis." *Journal of Morphology* 276 (8): 961–90. https://doi.org/10.1002/jmor.20272.

Bollback, Jonathan P. 2006. "SIMMAP: Stochastic Character Mapping of Discrete Traits on Phylogenies." *BMC Bioinformatics* 7: 1–7. https://doi.org/10.1186/1471-2105-7-88.

Bouckaert, Remco R., and Alexei J. Drummond. 2017. "BModelTest: Bayesian Phylogenetic Site Model Averaging and Model Comparison." *BMC Evolutionary Biology* 17 (1): 1–11.

https://doi.org/10.1186/s12862-017-0890-6.

Bouckaert, Remco R., Joseph Heled, Denise Kühnert, Tim Vaughan, Chieh Hsi Wu, Dong Xie, Marc A. Suchard, Andrew Rambaut, and Alexei J. Drummond. 2014. "BEAST 2: A Software Platform for Bayesian Evolutionary Analysis." *PLoS Computational Biology* 10 (4): 1–6. https://doi.org/10.1371/journal.pcbi.1003537.

Bryant, David, Remco R. Bouckaert, Joseph Felsenstein, Noah A. Rosenberg, and Arindam Roychoudhury. 2012. "Inferring Species Trees Directly from Biallelic Genetic Markers: Bypassing Gene Trees in a Full Coalescent Analysis." *Molecular Biology and Evolution* 29 (8): 1917–32. https://doi.org/10.1093/molbev/mss086.

Buckley, David. 2012. "Evolution of Viviparity in Salamanders (Amphibia, Caudata)." *ELS*. https://doi.org/10.1002/9780470015902.a0022851.

Buckley, David, Marina Alcobendas, Mario Garcia-Paris, and Marvalee H. Wake. 2007. "Heterochrony, Cannibalism, and the Evolution of Viviparity in *Salamandra salamandra*." *Evolution and Development* 9 (1): 105–15. https://doi.org/10.1111/j.1525-142X.2006.00141.x.

Carøe, Christian, Shyam Gopalakrishnan, Lasse Vinner, Sarah S.T. Mak, Mikkel Holger S. Sinding, José A. Samaniego, Nathan Wales, Thomas Sicheritz-Pontén, and M. Thomas P. Gilbert. 2018. "Single-Tube Library Preparation for Degraded DNA." *Methods in Ecology and Evolution* 9 (2): 410–19. https://doi.org/10.1111/2041-210X.12871.

Darriba, Diego, Guillermo L. Taboada, Ramón Doallo, and David Posada. 2012. "JModelTest 2: More Models, New Heuristics and Parallel Computing." *Nature Methods* 9 (8): 772–772. https://doi.org/10.1038/nmeth.2109.

Deleury, Emeline, Thomas Guillemaud, Aurélie Blin, and Eric Lombaert. 2019. "An Evaluation of Pool-Sequencing Transcriptome-Based Exon Capture for Population Genomics in Non-Model Species." *BioRxiv*, 583534. https://doi.org/10.1101/583534.

Dinis, Marco, Khaled Merabet, Fernando Martínez-Freiría, Sebastian Steinfartz, Miguel Vences, James D. Burgon, Kathryn R. Elmer, et al. 2019. "Allopatric Diversification and Evolutionary Melting Pot in a North African Palearctic Relict: The Biogeographic History of *Salamandra algira*." *Molecular Phylogenetics and Evolution* 130: 81–91. https://doi.org/10.1016/j.ympev.2018.10.018.

Dinis, Marco, and Guillermo Velo-Antón. 2017. "How Little Do We Know about the Reproductive Mode in the North African Salamander, *Salamandra algira*? Pueriparity in Divergent Mitochondrial Lineages of *S. a. tingitana*." *Amphibia Reptilia* 38 (4): 540–46. https://doi.org/10.1163/15685381-00003121.

Donaire-Barroso, David, and Sergé Bogaerts. 2001. "Observations on Viviparity of Salamandra Algira in North Morocco." *Herpetologia Candiana*, 147–51.

Donaire-Barroso, David, Sergé Bogaerts, and David Herbert. 2001. “Confirmación de Desarrollo Larvario Completo Intrauterino En *Salamandra algira* (Bedriaga, 1883) Del Noroeste de Marruecos.” *Butlleti de La Societat Catalana d’Herpetologia* 15 (December): 107–10.

Drummond, Alexei J., Marc A. Suchard, Dong Xie, and Andrew Rambaut. 2012. “Bayesian Phylogenetics with BEAUti and the BEAST 1.7.” *Molecular Biology and Evolution* 29 (8): 1969–73. https://doi.org/10.1093/molbev/mss075.

Duchêne, Sebastián, Robert Lanfear, and Simon Y.W. Ho. 2014. “The Impact of Calibration and Clock-Model Choice on Molecular Estimates of Divergence Times.” *Molecular Phylogenetics and Evolution* 78 (1): 277–89. https://doi.org/10.1016/j.ympev.2014.05.032.

Edgar, Robert C. 2004. “MUSCLE: Multiple Sequence Alignment with High Accuracy and High Throughput.” *Nucleic Acids Research* 32 (5): 1792–97. https://doi.org/10.1093/nar/gkh340.

Ehl, Sarah, Miguel Vences, and Michael Veith. 2019. “Reconstructing Evolution at the Community Level: A Case Study on Mediterranean Amphibians.” *Molecular Phylogenetics and Evolution* 134: 211–25. https://doi.org/10.1016/j.ympev.2019.02.013.

Frost, Darrel R. 2020. “Amphibian Species of the World: An Online Reference. Version 6.1. American Museum of Natural History.” 2020.

García-París, Mario, Marina Alcobendas, David Buckley, and David B. Wake. 2003. “Dispersal of Viviparity across Contact Zones in Iberian Populations of Fire Salamanders (*Salamandra*) Inferred from Discordance of Genetic and Morphological Traits.” *Evolution* 57 (1): 129–43. https://doi.org/10.1111/j.0014-3820.2003.tb00221.x.

Gower, David J., Varad Giri, Mahesh S. Dharne, and Yogesh S. Shouche. 2008. “Frequency of Independent Origins of Viviparity among Caecilians (Gymnophiona): Evidence from the First ‘live-Bearing’ Asian Amphibian.” *Journal of Evolutionary Biology* 21 (5): 1220–26. https://doi.org/10.1111/j.1420-9101.2008.01577.x.

Gregory, T. Ryan. 2003. “Variation across Amphibian Species in the Size of the Nuclear Genome Supports a Pluralistic, Hierarchical Approach to the C-Value Enigma.” *Biological Journal of the Linnean Society* 79 (2): 329–39. https://doi.org/10.1046/j.1095-8312.2003.00191.x.

Greven, H. 2003. “Larviparity and Pueriparity.” In *Reproductive Biology and Phylogeny of Urodela*, 447–75. CRC Press.

Harris, Robert S. 2007. “Improved Pairwise Alignment of Genomic DNA.” Pennsylvania State University.

Helfer, V., T. Broquet, and L. Fumagalli. 2012. "Sex-Specific Estimates of Dispersal Show Female Philopatry and Male Dispersal in a Promiscuous Amphibian, the Alpine Salamander (*Salamandra atra*)." *Molecular Ecology* 21 (19): 4706–20. https://doi.org/10.1111/j.1365-294X.2012.05742.x.

Helmstetter, Andrew J., Alexander S.T. Papadopulos, Javier Igea, Tom J.M. Van Dooren, Armand M. Leroi, and Vincent Savolainen. 2016. "Viviparity Stimulates Diversification in an Order of Fish." *Nature Communications* 7: 1–7. https://doi.org/10.1038/ncomms11271.

Ho, Simon Y.W., Matthew J. Phillips, Alan Cooper, and Alexei J. Drummond. 2005. "Time Dependency of Molecular Rate Estimates and Systematic Overestimation of Recent Divergence Times." *Molecular Biology and Evolution* 22 (7): 1561–68. https://doi.org/10.1093/molbev/msi145.

Iskandar, Djoko T., Ben J. Evans, and Jimmy A. McGuire. 2014. "A Novel Reproductive Mode in Frogs: A New Species of Fanged Frog with Internal Fertilization and Birth of Tadpoles." *PLoS ONE* 9 (12): 1–14. https://doi.org/10.1371/journal.pone.0115884.

Joy, Jeffrey B., Richard H. Liang, Rosemary M. McCloskey, T. Nguyen, and Art F.Y. Poon. 2016. "Ancestral Reconstruction." *PLoS Computational Biology* 12 (7): 1–20. https://doi.org/10.1371/journal.pcbi.1004763.

Kearse, Matthew, Richard Moir, Amy Wilson, Steven Stones-Havas, Matthew Cheung, Shane Sturrock, Simon Buxton, et al. 2012. "Geneious Basic: An Integrated and Extendable Desktop Software Platform for the Organization and Analysis of Sequence Data." *Bioinformatics* 28 (12): 1647–49. https://doi.org/10.1093/bioinformatics/bts199.

Kircher, Martin, Susanna Sawyer, and Matthias Meyer. 2012. "Double Indexing Overcomes Inaccuracies in Multiplex Sequencing on the Illumina Platform." *Nucleic Acids Research* 40 (1): 1–8. https://doi.org/10.1093/nar/gkr771.

Langmead, Ben, and Steven L. Salzberg. 2012. "Fast Gapped-Read Alignment with Bowtie 2." *Nature Methods* 9 (4): 357–59. https://doi.org/10.1038/nmeth.1923.

Li, Heng, and Richard Durbin. 2009. "Fast and Accurate Short Read Alignment with Burrows-Wheeler Transform." *Bioinformatics* 25 (14): 1754–60. https://doi.org/10.1093/bioinformatics/btp324.

Liedtke, H. Christoph, Hendrik Müller, Julian Hafner, Johannes Penner, David J. Gower, Tomáš Mazuch, Mark Oliver Rödel, and Simon P. Loader. 2017. "Terrestrial Reproduction as an Adaptation to Steep Terrain in African Toads." *Proceedings of the Royal Society B: Biological Sciences* 284 (1851). https://doi.org/10.1098/rspb.2016.2598.

Lourenço, André, Bernardo Antunes, Ian J. Wang, and Guillermo Velo-Antón. 2018. "Fine-

Scale Genetic Structure in a Salamander with Two Reproductive Modes: Does Reproductive Mode Affect Dispersal?" *Evolutionary Ecology* 32 (6): 699–732. https://doi.org/10.1007/s10682-018-9957-0.

Lourenço, André, João Gonçalves, Filipe Carvalho, Ian J. Wang, and Guillermo Velo-Antón. 2019. "Comparative Landscape Genetics Reveals the Evolution of Viviparity Reduces Genetic Connectivity in Fire Salamanders." *Molecular Ecology* 28 (20): 4573–91. https://doi.org/10.1111/mec.15249.

Lourenço, André, Fernando Sequeira, David Buckley, and Guillermo Velo-Antón. 2018. "Role of Colonization History and Species-Specific Traits on Contemporary Genetic Variation of Two Salamander Species in a Holocene Island-Mainland System." *Journal of Biogeography* 45 (5): 1054–66. https://doi.org/10.1111/jbi.13192.

McKenna, Aaron, Matthew Hanna, Eric Banks, Andrey Sivachenko, Kristian Cibulskis, Andrew Kernytsky, Kiran Garimella, et al. 2010. "The Genome Analysis Toolkit: A MapReduce Framework for Analyzing next-Generation DNA Sequencing Data." *Genome Research* 20 (9): 1297–1303. https://doi.org/10.1101/gr.107524.110.

Miller, Mark A., Wayne Pfeiffer, and Terri Schwartz. 2010. "Creating the CIPRES Science Gateway for Inference of Large Phylogenetic Trees." *2010 Gateway Computing Environments Workshop, GCE 2010*. https://doi.org/10.1109/GCE.2010.5676129.

Mulder, Kevin P., André Lourenço, Miguel Carneiro, and Guillermo Velo-Antón. 2016. "The Complete Mitochondrial Genome of *Salamandra salamandra* (Amphibia: Urodela: Salamandridae)." *Mitochondrial DNA Part B* 1 (1): 880–82. https://doi.org/10.1080/23802359.2016.1253042.

Ortiz, E M. 2019. "Vcf2phylip v2. 0: Convert a VCF Matrix into Several Matrix Formats for Phylogenetic Analysis." *Version V2.*

Pereira, Ricardo J., Iñigo Martínez-Solano, and David Buckley. 2016. "Hybridization during Altitudinal Range Shifts: Nuclear Introgression Leads to Extensive Cyto-Nuclear Discordance in the Fire Salamander." *Molecular Ecology* 25 (7): 1551–65. https://doi.org/10.1111/mec.13575.

Phillips, Matthew J. 2009. "Branch-Length Estimation Bias Misleads Molecular Dating for a Vertebrate Mitochondrial Phylogeny." *Gene* 441 (1–2): 132–40. https://doi.org/10.1016/j.gene.2008.08.017.

Pincheira-Donoso, Daniel, Tom Tregenza, Matthew J. Witt, and Dave J. Hodgson. 2013. "The Evolution of Viviparity Opens Opportunities for Lizard Radiation but Drives It into a Climatic Cul-de-Sac." *Global Ecology and Biogeography* 22 (7): 857–67. https://doi.org/10.1111/geb.12052.

R Core Team. 2019. "R: A Language and Environment for Statistical Computing." Vienna,

Austria.

Rambaut, Andrew, Alexei J. Drummond, Dong Xie, Guy Baele, and Marc A. Suchard. 2018. "Posterior Summarization in Bayesian Phylogenetics Using Tracer 1.7." *Systematic Biology* 67 (5): 901–4. https://doi.org/10.1093/sysbio/syy032.

Revell, Liam J. 2012. "Phytools: An R Package for Phylogenetic Comparative Biology (and Other Things)." *Methods in Ecology and Evolution* 3 (2): 217–23. https://doi.org/10.1111/j.2041-210X.2011.00169.x.

Reynolds, John D., Nicholas B. Goodwin, and Robert P. Freckleton. 2002. "Evolutionary Transitions in Parental Care and Live Bearing in Vertebrates." *Philosophical Transactions of the Royal Society B: Biological Sciences* 357 (1419): 269–81. https://doi.org/10.1098/rstb.2001.0930.

Richmond, Jonathan Q. 2006. "Evolutionary Basis of Parallelism in North American Scincid Lizards." *Evolution and Development* 8 (6): 477–90. https://doi.org/10.1111/j.1525-142X.2006.00121.x.

Rodríguez, Ariel, James D. Burgon, Mariana Lyra, Iker Irisarri, Denis Baurain, Leon Blaustein, Bayram Göçmen, et al. 2017. "Inferring the Shallow Phylogeny of True Salamanders (*Salamandra*) by Multiple Phylogenomic Approaches." *Molecular Phylogenetics and Evolution* 115: 16–26. https://doi.org/10.1016/j.ympev.2017.07.009.

Salichos, Leonidas, and Antonis Rokas. 2013. "Inferring Ancient Divergences Requires Genes with Strong Phylogenetic Signals." *Nature* 497 (7449): 327–31. https://doi.org/10.1038/nature12130.

Sandberger-Loua, Laura, Hendrik Müller, and Mark Oliver Rödel. 2017. "A Review of the Reproductive Biology of the Only Known Matrotrophic Viviparous Anuran, the West African Nimba Toad, *Nimbaphrynoides occidentalis*." *Zoosystematics and Evolution* 93 (1): 105–33. https://doi.org/10.3897/zse.93.10489.

Stamatakis, Alexandros. 2014. "RAxML Version 8: A Tool for Phylogenetic Analysis and Post-Analysis of Large Phylogenies." *Bioinformatics* 30 (9): 1312–13. https://doi.org/10.1093/bioinformatics/btu033.

Stamatakis, Alexandros, Paul Hoover, and Jacques Rougemont. 2008. "A Rapid Bootstrap Algorithm for the RAxML Web Servers." *Systematic Biology* 57 (5): 758–71. https://doi.org/10.1080/10635150802429642.

Steinfartz, Sebastian, Michael Veith, and Diethard Tautz. 2000. "Mitochondrial Sequence Analysis of Salamandra Taxa Suggests Old Splits of Major Lineages and Postglacial Recolonizations of Central Europe from Distinct Source Populations of *Salamandra salamandra*." *Molecular Ecology* 9 (4): 397–410. https://doi.org/10.1046/j.1365-294X.2000.00870.x.

Uotila, E, A Crespo-Diaz, I Sanz-Azkue, and X Rubio. 2013. “Variation in the Reproductive Strategies of *Salamandra salamandra* (Linnaeus, 1758) Populations in the Province of Gipuzkoa (Basque Country).” *Munibe* 61: 91–101.

Veith, M, Sebastian Steinfartz, R Zardoya, A Seitz, and A Meyer. 1998. “A Molecular Phylogeny of ‘True’Salamanders (Family Salamandridae) and the Evolution of Terrestriality of Reproductive Modes.” *Journal of Zoological Systematics and Evolutionary Research* 36 (1–2): 7–16. https://doi.org/10.1111/j.1439-0469.1998.tb00774.x.

Velo-Antón, Guillermo, Mario García-París, P. Galán, and A. Cordero Rivera. 2007. “The Evolution of Viviparity in Holocene Islands: Ecological Adaptation versus Phylogenetic Descent along the Transition from Aquatic to Terrestrial Environments.” *Journal of Zoological Systematics and Evolutionary Research* 45 (4): 345–52. https://doi.org/10.1111/j.1439-0469.2007.00420.x.

Velo-Antón, Guillermo, Xavier Santos, Iago Sanmartín-Villar, Adolfo Cordero-Rivera, and David Buckley. 2015. “Intraspecific Variation in Clutch Size and Maternal Investment in Pueriparous and Larviparous *Salamandra salamandra* Females.” *Evolutionary Ecology* 29 (1): 185–204. https://doi.org/10.1007/s10682-014-9720-0.

Velo-Antón, Guillermo, Kelly R. Zamudio, and Adolfo Cordero-Rivera. 2012. “Genetic Drift and Rapid Evolution of Viviparity in Insular Fire Salamanders (*Salamandra salamandra*).” *Heredity* 108 (4): 410–18. https://doi.org/10.1038/hdy.2011.91.

Vences, Miguel, Eugenia Sanchez, J. Susanne Hauswaldt, Daniel Eikelmann, Ariel Rodríguez, Salvador Carranza, David Donaire, et al. 2014. “Nuclear and Mitochondrial Multilocus Phylogeny and Survey of Alkaloid Content in True Salamanders of the Genus *Salamandra* (Salamandridae).” *Molecular Phylogenetics and Evolution* 73 (1): 208–16. https://doi.org/10.1016/j.ympev.2013.12.009.

Wake, Marvalee H. 1982. “Diversity within a Framework of Constraints. Amphibian Reproductive Modes.” In *Environmental Adaptation and Evolution*, 87–106. Gustav Fischer New York, USA.

Weisrock, David W., Paul M. Hime, Schyler O. Nunziata, Kara S. Jones, Mason O. Murphy, Scott Hotaling, and Justin D. Kratovil. 2018. “Surmounting the Large-Genome ‘Problem’ for Genomic Data Generation in Salamanders.” In *Population Genomics: Wildlife*, 1–28. Springer. https://doi.org/10.1007/13836_2018_36.

CHAPTER 3: RNASEQ ANALYSES ACROSS CONVERGENT REPRODUCTIVE MODES WITHIN *SALAMANDRA SALAMANDRA* HIGHLIGHT CANDIDATE GENES IMPORTANT IN THE EVOLUTIONARY SHIFT FROM LARVIPARITY TO PUERIPARITY

Kevin P. Mulder[a,b,c,d,*], André Lourenço[a,b], Ivan Gomez-Mestre[e], Miguel Carneiro[a], David Buckley[f], Iñigo Martínez-Solano[e,f], Robert C. Fleischer[d], Rayna C. Bell[c,g], Guillermo Velo-Antón[a,*]

[a] *CIBIO/InBIO, Research Center in Biodiversity and Genetic Resources, Rua Padre Armando Quintas 7, Campus Agrário de Vairão, 4485-661 Vairão, Portugal.*
[b] *Departamento de Biologia da Faculdade de Ciências da Universidade do Porto, Rua Campo Alegre, 4169-007, Porto, Portugal.*
[c] *Department of Vertebrate Zoology, National Museum of Natural History, Smithsonian Institution, 1000 Constitution Avenue NW, Washington, DC 20560, USA.*
[d] *Center for Conservation Genomics, Smithsonian Conservation Biology Institute, National Zoological Park, 3001 Connecticut Avenue NW, Washington, DC 20008, USA.*
[e] *Ecology, Evolution, and Development Group, Department of Wetland Ecology, Doñana Biological Station, CSIC, Seville, Spain*
[f] *Departamento de Biodiversidad y Biología Evolutiva, Museo Nacional de Ciencias Naturales MNCN-CSIC, Madrid, Spain*
[g] *Herpetology Department, California Academy of Sciences, 55 Music Concourse Drive, San Francisco, CA 94118, USA*

Abstract

Shifts in reproductive mode occur across the tree of life and represent key adaptations with profound effects on species' life histories and evolutionary trajectories. Understanding the proximate and ultimate causes of these shifts can be challenging due to multiple concurrent neutral and adaptive changes that can arise alongside a shift in reproductive mode. The fire salamander, *Salamandra salamandra* is an amphibian that exhibits intra-specific variation in reproductive mode, allowing us to investigate both larviparity (females give birth to larvae) and pueriparity (females deliver fully-formed offspring) in the same species. Despite yielding fewer offspring, pueriparity is an adaptive innovation that allows individuals to exploit ecological habitats with no available water bodies. *S. salamandra* is larviparous across the majority of its range, but pueriparity evolved twice: during the early Pleistocene within *S. s. bernardezi* populations in the mountains of northern Spain, and again during the late Pleistocene on land-bridge islands inhabited by *S. s. gallaica* in northwestern Spain. To detect candidate genes associated with reproductive mode, we compared gene expression profiles of the uterus and oviduct of pregnant females from adjacent larviparous and pueriparous populations. We identified shared differences in gene expression among pueriparous *S. s. bernardezi* and *S. s. gallaica* relative to their larviparous counterparts. We also identified differences in gene expression between pueriparous and larviparous female`s that were unique to either the mountain or island transition. These single-transition candidate genes may reflect partially unique genetic architectures of the convergent phenotypes or they may indicate other environmental or evolutionary differences among pueriparous and larviparous females within a given transition. Many of the top candidate genes are associated with embryogenesis and cell growth and differentiation, which is consistent with the phenotypic differences in embryonic development between larviparity and pueriparity. This study is an important first step in describing the genetic background of larviparity and pueriparity in a unique comparative system, and provides transcriptome resources and candidate genes that can guide further research into the genomic architecture of this adaptive trait.

Keywords: adaptation genomics, differential expression, independent evolutionary transitions, salamanders, viviparity

3.1 - Introduction

An organism's reproductive mode is an important life history trait, and changes therein often constitute key adaptive innovations with profound effects on species' evolutionary trajectories. Amphibians exhibit tremendous diversity in their modes of reproduction, frequently characterized by complex evolutionary adaptations to new habitats (Zamudio et al. 2016; Crump 2015; Gomez-Mestre, Pyron, and Wiens 2012), but the genomic basis of this diversity in reproductive mode has been largely unexplored (Funk, Zamudio, and Crawford 2018). Despite recent advances in sequencing technology that have helped decipher the genomic architecture of many adaptive traits (Lehner 2013; Singh and Nüsslein-Volhard 2015), phenotypic traits that are not found in model organisms must be studied in natural populations (Ekblom and Galindo 2011; Barson et al. 2015; Steiner et al. 2009).

The ancestral state for amphibians is oviparity with an aquatic larval life stage, but many groups across the three extant amphibian orders (Anura, Caudata and Gymnophiona) have independently evolved viviparity (live-birth). Among viviparous amphibians, there are larviparous species that deliver larvae into waterbodies, whereas others are pueriparous, in which the larval aquatic stage is skipped and females deliver fully developed terrestrial metamorphs (Greven 2003). Pueriparity is a remarkable adaptation for a group that is largely characterized by an aquatic-terrestrial (biphasic) life cycle (Duellman and Trueb 1994), as the evolutionary transition to pueriparity entails semi-independence from water and thus the potential to colonize water-limited habitats. Although it constitutes a clear example of a homoplastic trait that is likely influenced both by genetic constraints (Wake, Wake, and Specht 2011), and environmental factors (Losos 2011), the genetic basis of the shift to pueriparity is unknown.

Pueriparity has independently evolved several times across the three amphibian orders. It occurs in about 15% of caecilians (Gymnophiona; Blackburn, 2015), and is rarely found in frogs (Anura; six out of the ca. 7204 known species; AmphibiaWeb 2020) and salamanders (Caudata; 14 out of the ca. 742 salamander species; AmphibiaWeb 2020). All cases of pueriparity in salamanders occur in the family Salamandridae, specifically in the ten species comprising the genus *Lyciasalamandra* and four in the sister genus *Salamandra* (Buckley 2012). *Salamandra* contains six species, two of which are strictly pueriparous (the Alpine salamanders; *S. lanzai* and *S. atra*), whereas *S. corsica*, *S. infraimmaculata*, *S. algira*, and *S. salamandra* are all referred to as larviparous. However, the latter two species display an

exceptional intraspecific variability in their reproductive mode, with the two strategies, larviparity and pueriparity, co-occurring within the same species (García-París et al. 2003; Velo-Antón et al. 2007; Dinis and Velo-Antón 2017). Pueriparity in *S. salamandra* evolved at least once during the early Pleistocene in the Cantabrian mountains (Chapter 2, Garcia-Paris et al. 2003) and twice independently during the late Pleistocene (Chapter 2, Velo-Antón *et al.*, 2007; Velo-Antón, Zamudio and Cordero-Rivera, 2012; Figure 3.1), potentially due to lack of past surface water for the depositing of larvae. Applying genomic tools to the independent transitions in reproductive mode across *Salamandra* is a powerful approach because these methods can help disentangle the conflicting signals of neutral drift (e.g. Velo-Antón, Zamudio, and Cordero-Rivera 2012) and genetic adaptation, and can uncover the genetic underpinnings of this remarkable shift to pueriparity in a phylogeographic comparative framework (Zamudio, Bell, and Mason 2016).

Within *S. salamandra* the ancestral mode is larviparity, in which females deliver 20 to 80 larvae into nearby water bodies, whereas pueriparity, the derived mode, is characterized by the delivery of 1 to 15 fully terrestrial metamorphs (Velo-Antón et al. 2015). Although less fecund than larviparity, pueriparity provides independence from water bodies (Lourenço et al. 2017; Liedtke et al. 2017), and has important ethological, ecological, physiological and morphological implications (Greven 2003; Buckley et al. 2007; Lourenço et al. 2019). Differences in fecundity between reproductive modes are explained by a series of heterochronic processes arising from the shift to pueriparity, such as the incomplete fertilization of ovulated eggs, accelerated and asynchronous larval development in the reproductive tract, and developing larvae feeding on unfertilized eggs (oophagy) and siblings (adelphophagy; Buckley et al. 2007). Common garden experimental work shows that sexually mature females in controlled lab environments maintain their respective reproductive modes regardless of water availability (Velo-Antón et al. 2015; Buckley et al. 2007) indicating that adaptive genetic changes, rather than the environment, are likely important in explaining this phenotype. This observation is also supported by long-term field work (>20 years) in a local pueriparous population (Ons island) where, despite the presence of water bodies and suitable habitat for larviparous reproduction, females only deliver terrestrial juveniles (Velo-Antón, Zamudio, and Cordero-Rivera 2012; Velo-Antón et al. 2015). Although phenotypic plasticity by either epigenetic inheritance or early life-stage environmental imprinting has not yet been investigated in *Salamandra*, it is clear that reproduction is not environmentally controlled and that gene-expression differences between the reproductive modes are likely.

In vertebrates, studies quantifying morphological and physiological changes associated with pregnancy have focused on the oviduct and uterus (Wourms, Grove, and Lombardi 1988; Biazik et al. 2012; Shine and Guillette 1988; Atkins, Jones, and Guillette 2006; Ramírez-Pinilla et al. 2012). For instance, patterns of gene expression in the uteri of viviparous amniotes are associated with eggshell and placenta formation, gas exchange, nutrient transportation, metabolism and the immune system (Brandley et al. 2012; Whittington et al. 2015; Gao et al. 2019; Foster et al. 2020). However, viviparous reproduction in amphibians differs from that of mammals and reptiles, and the genetic mechanisms underlying morphological and physiological changes are still unknown. In both reproductive modes of *S. salamandra*, the ovulated eggs are coated by a tough egg jelly produced by the different glands along the oviduct, which is key for the subsequent egg fertilization process (Greven 2003). Fertilization occurs in the most caudal portion of the oviduct, the uterus, where the embryos also develop (Greven 2003). In the pueriparous mode not all eggs get fertilized, with up to 50% of the eggs providing additional nutrition for the embryos once they hatch into larvae within the uterus and start feeding (Buckley et al. 2007). The oviduct and uterus are thus the most promising maternal tissue to investigate for differential gene expression that might explain the ontogenetic differences between reproductive modes.

In this study, we apply RNAseq sequencing to a recent and older transition in reproductive mode within *S. salamandra* to identify and quantify the differences in uterine and oviductal gene expression in both a comparative spatial framework (distinct environments) and at different temporal scales (Elmer and Meyer, 2011). We aim to (a) describe general gene expression patterns and identify tissue-specific expression in the reproductive organs of female salamanders, (b) characterize gene expression differences between larviparity and pueriparity and distinguish between both convergent and unique patterns across both transitions, and (c) identify candidate genes associated with the phenotypic differences between larviparous and pueriparous salamanders.

3.2 - Methods

We focused our study on the two regions where *S. salamandra* has independently evolved to pueriparity from the ancestral larviparous state (Chapter 2): (1) the early Pleistocene transition to pueriparity that occurred in *S. s. bernardezi* in the Cantabrian mountains in northern Spain (henceforth called the mountain transition), and (2) the more recent (late Pleistocene) transition within *S. s. gallaica* on two islands (Ons and San Martiño) in SW Galicia, Spain (henceforth called the island transition). To minimize environmental variation that might impact mRNA expression, we sampled geographically adjacent populations of larviparous and pueriparous salamanders. For the island transition, we sampled two mainland larviparous populations of *S. s. gallaica* (Coiro and Monteferro) and the pueriparous island of Ons (16-25 km distance). The pueriparous population of San Martiño island (Chapter 2) could not be sampled due to its low population size (Velo-Antón and Cordero-Rivera 2017). For the mountain transition, we focused on two localities that are in close proximity but separated by a high elevation mountain ridge that impedes gene flow (Velo-Antón, unpublished data): the pueriparous *S. s. bernardezi* population in Somiedo (province of Asturias) and the larviparous *S. s. gallaica* in Orallo (province of Castilla y León; 17 km distance; Figure 3.1).

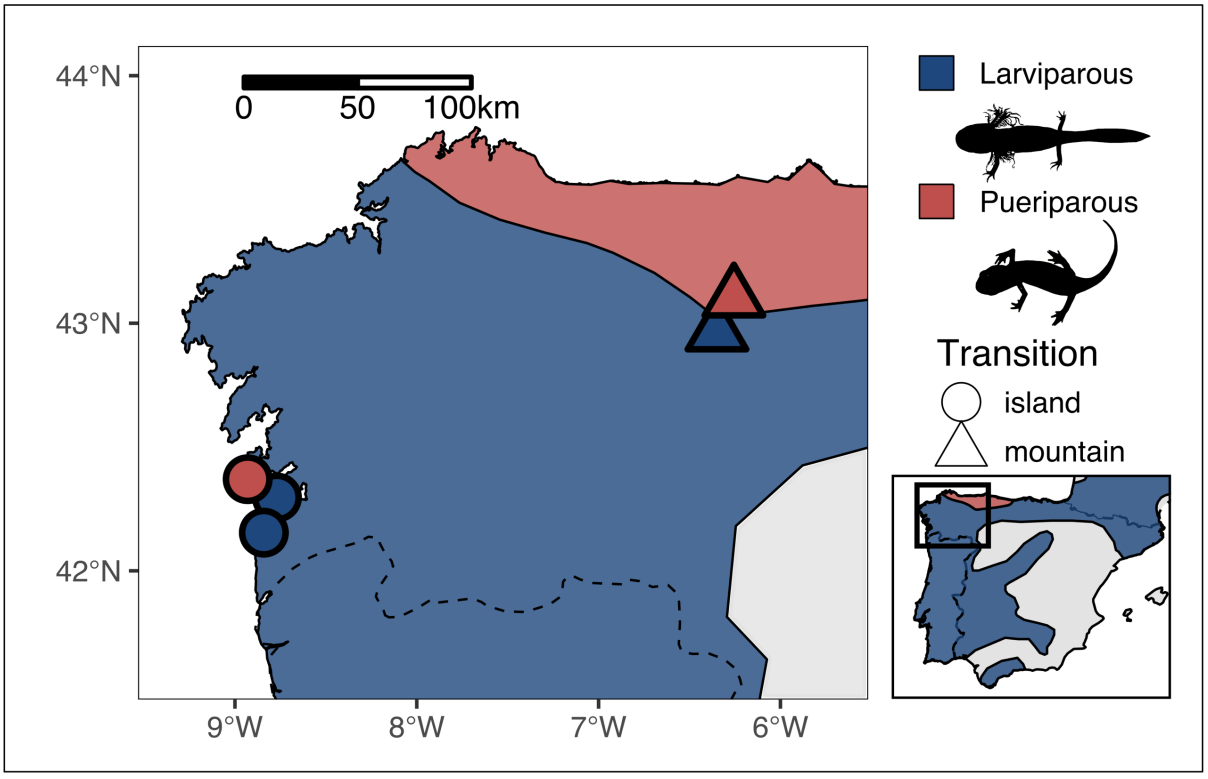
3.2.1 - Field sampling

Sampling for both transitions was completed within a two-week period in October 2016, at which time *Salamandra* females are in the early stages of pregnancy in this region (Table C1; Guillermo Velo-Antón and Buckley 2015). We searched for active pregnant females on two rainy evenings, and sampled the pueriparous and larviparous populations for a given transition zone on the same evening to ensure that we sampled all salamanders at the same stage and activity period. Salamanders were individually housed in a common environment for three days prior to tissue sampling to minimize the environmental variables that could impact gene expression. This also ensured that we were only sampling pregnant salamanders and not individuals with a full crop that appeared pregnant. We sampled a total of seven females for the island transition (three pueriparous females from Ons and four larviparous females from the two mainland localities), and six females for the mountain transition (three salamanders from each sampled locality; Figure 3.1; Table C1). Each female was sacrificed with an overdose of anaesthesia (benzocaine; Ethyl 4-aminobenzoate; Sigma-Aldrich, Darmstadt, Germany), and uterus and oviduct tissues were immediately sampled and stored in liquid

nitrogen, and transferred to -80ºC for long-term storage. We registered the number and stage of the larvae/juveniles/eggs found in each uterus (Table C1). We sampled the right and left sided uterus separately to serve as biological replicates because there were noticeable differences in the number and stage of development of the larvae/juveniles between the two sides of the uterus (Table C1). We only sampled one oviduct per female, always sampling the left-side. All tissue sampling was randomized to reduce possible biases in tissue quality and RNA expression due to order of sampling.

To generate a more complete reference transcriptome we included seven additional tissues from two previously collected, non-pregnant larviparous and pueriparous individuals (Table C1; heart, kidney, lung, liver, muscle, oviduct and uterus). These two individuals were included in the Trinity assembly and annotation to generate a comprehensive reference transcriptome, but were excluded from the differential expression analyses.

Figure 3.1. Study area in northwestern Spain. Larviparous range indicated in blue and pueriparous range in red. The island transition in SW Galicia, Spain included two larviparous mainland populations (Coiro and Monteferro, blue circles) and the pueriparous population on the island of Ons (red circle). Across the mountain transition in the Cantabrian mountains we sampled the larviparous population of S. s. gallaica in Orallo (blue triangle) and the pueriparous population of S. s. bernardezi in Somiedo (red triangle). Symbols and colours are maintained across figures.



3.2.2 - RNA-sequencing

The order of all tissue samples was randomized before starting laboratory work to avoid biasing our results. We extracted total RNA from approximately 25 mg of tissue using the RNeasy kit (Qiagen, Hilden, Germany) and checked RNA integrity on a TapeStation 2200 (Agilent Technologies, Santa Clara, CA, USA). If samples had an RNA Integrity Number (RIN) below 7.5, we re-extracted the tissue. Prior to cDNA synthesis, we enriched for mRNA using the NEBNext® Poly(A) mRNA beads, then double stranded cDNA was generated using the NEBNext first and second strand synthesis kits (all from NEB, Ipswich, Massachusetts, USA). We used an in-house protocol to prepare DNA libraries using double-indexed Nextera-style adapters (Glenn et al. 2019). We quantified libraries using KAPA library quantification kits and pooled samples equimolarly for sequencing.

All uterus and oviduct samples for differential expression analyses were combined in one pool and sequenced across three lanes of a HiSeq4000 using paired-end 100 bp reads at Macrogen (Seoul, South Korea). The two additional reference individuals were sequenced each independently, the larviparous individual on a HiSeq 1500 with paired-end 125 base pair (bp) reads (CIBIO, Portugal) and the pueriparous individual on a HiSeq 4000 with paired-end 150 bp reads (Berkeley, CA, USA).

3.2.3 - Transcriptome assembly and annotation

All bioinformatic processing was performed on the Hydra High Performance Computing Cluster. Demultiplexed reads were filtered and trailing adapters were removed using Trimmomatic 0.33 (Bolger, Lohse, and Usadel 2014) by setting the quality cut-off at Q5, which is considered optimal for transcriptome assembly (MacManes 2014). All tissues were assembled together using Trinity 2.4 (Haas et al. 2013) with default settings. We ran the resulting assembly through BUSCO v2.0 using the vertebrate dataset to assess completeness. Reads were also mapped to the assembly using bowtie 2.2.9 (Langmead and Salzberg 2012) to assess assembly quality.

To remove low-quality transcripts we applied the TransRate pipeline (Smith-Unna et al. 2016), which uses assembly quality and remapping statistics to identify high-quality transcripts. To remove duplicate reads and make a TransRate run computationally feasible, we first applied Trinity's in-silico normalization script to the filtered reads, applying a maximum coverage of

100 (Haas et al. 2013). To annotate transcripts, we ran blastp against the confirmed list of transcripts identified in the *Pleurodeles waltl* genome (Elewa et al. 2017). This list included 123,518 transcripts that were part of 19,903 gene-models. We applied a minimum blast-score of 45 to retain an annotation. Notes on potential gene functions related to embryonic development were gathered from the GeneCards database (Safran et al. 2010), unless indicated otherwise.

3.2.4 - Genetic distance between larviparous and pueriparous population pairs

To estimate genetic distances between the populations in our dataset, we removed the poly-A-tail from the filtered reads using Prinseq-lite 0.20.4 (Schmieder and Edwards 2011) and mapped the reads using bowtie 2.2.9 (Langmead and Salzberg 2012) to a previously identified set of 3,070 transcriptome-derived genes that were considered single locus and phylogenetically informative for the genus *Salamandra* (Rodríguez et al. 2017). We only allowed for concordantly mapped reads and duplicate reads were removed with Picard tools.

Nuclear Single Nucleotide Polymorphisms (SNPs) were called using the Genome Analyses Tool Kit (GATK) using the haplotype caller pipeline (McKenna et al. 2010). SNPs were filtered by minimum depth of 5 reads, minimum SNP quality of 20 and removing sites that showed signs of excess heterozygosity (ExcessHet < 10.0, DP > 5, stand_call_conf > 20.0). We applied additional filtering using vcftools in order to get a strictly filtered dataset of nuclear SNPs (--min-alleles 2 --max-alleles 2 --remove-indels --max-missing 0.6 --mac 2 --minQ 100 --minDP 15 --minGQ 30 --non-ref-ac 5). We ran a PCA on all unlinked nuclear SNPs using custom R scripts to confirm the genetic relationships between the different samples. An additional maximum likelihood tree was constructed on a concatenated alignment of all 3,070 loci, using RAxML 8.2.12 (Stamatakis 2014) applying the GTRCAT substitution model. Bootstrap support was computed on the best scoring tree by means of 100 iterations of rapid bootstrapping (Stamatakis, Hoover, and Rougemont 2008).

3.2.5 - Gene expression

We quantified expression across the transcriptome by quasi-mapping all the uterus and oviduct samples against the filtered reference transcriptome using Salmon v0.8.2 (Patro et al. 2017). Quasi-mapping with Salmon has been shown to be both faster and more accurate in

estimating expression than the traditional full-mapping approaches (Zhang et al. 2017). Transcript counts were imported into R v3.6.3 using tximport 1.14.2 and combined to gene-level counts (Soneson, Love, and Robinson 2016). We used the recommended edgeR offset to normalize for average transcripts length and library size (see Love, Soneson, and Patro 2018).

3.2.5.1 - Expression patterns

To explore general expression patterns across our samples, we transformed the normalized transcript read-counts using the variance stabilizing transformation vst function in the R package DESeq2 v1.26.0 (Love, Huber, and Anders 2014) to make the data homoscedastic. We performed principal component analyses on the homoscedastic data of both oviduct and uterus tissues to explore expression patterns.

3.2.5.2 - Tissue specific expression

Using the size-corrected and library-corrected gene-level counts from tximport, we identified tissue-specific expression using the Tau metric (Yanai et al. 2005). The Tau metric is a measure of how tissue-specific the expression of a given gene is and ranges from 0 (broadly expressed across tissues) to 1 (completely specific to one tissue). This method is considered more robust for identifying tissue-specific genes when comparing tissues with different sample sizes (Kryuchkova-Mostacci and Robinson-Rechavi 2017). As the sample size for uterus and oviduct is larger compared to remaining tissues, there is likely a bias towards finding more genes in these two former tissues. We tried to reduce this bias by using the median value across all samples for a given tissue as this will only include genes found in at least half of the samples, as opposed to the mean, which could be driven by a highly expressed gene in a single individual. We highlighted genes with a high Tau value for uterus- or oviduct-specific expression when comparing all seven tissues that were included in the reference (heart, liver, lung, muscle, kidney, uterus and oviduct). We additionally tested for genes that were specific to reproductive tissues (uterus and oviduct) compared to the other five tissues. Tau was calculated using the tispec 0.99 R package (Condon 2020). We highlighted the top genes, as scored by tispec, showing a high Tau value and high expression, to identify genes that are likely important for uterus and oviduct functioning in *Salamandra*.

3.2.5.3 - Differential expression between larviparity and pueriparity

We applied quasi-likelihood F-tests, for the uterus and oviduct tissues separately, to test for differential expression using the R package edgeR 3.28.1 (Robinson, McCarthy, and Smyth 2009). Trinity transcripts were combined to gene-level counts using the *Pleurodeles* annotations, as gene-level differential expression is considered to be more robust in the absence of a conspecific reference genome (Soneson, Love, and Robinson 2016). We compared expression of all pueriparous vs all larviparous individuals to find genes differentially expressed convergently across both transitions. Additionally, we analysed the two independent transitions separately to find genes differentially expressed across a single transition. We applied the Benjamini & Hochberg false discovery rate at 0.05 (FDR: Benjamini and Hochberg 1995) to correct p-values for all comparisons. Results were visualized by means of a volcano plot to compare edgeR significance values and the log fold change in expression between larviparous and pueriparous samples. The top seven most significant genes per comparison were plotted individually to visualize gene-specific expression differences.

3.3 - Results

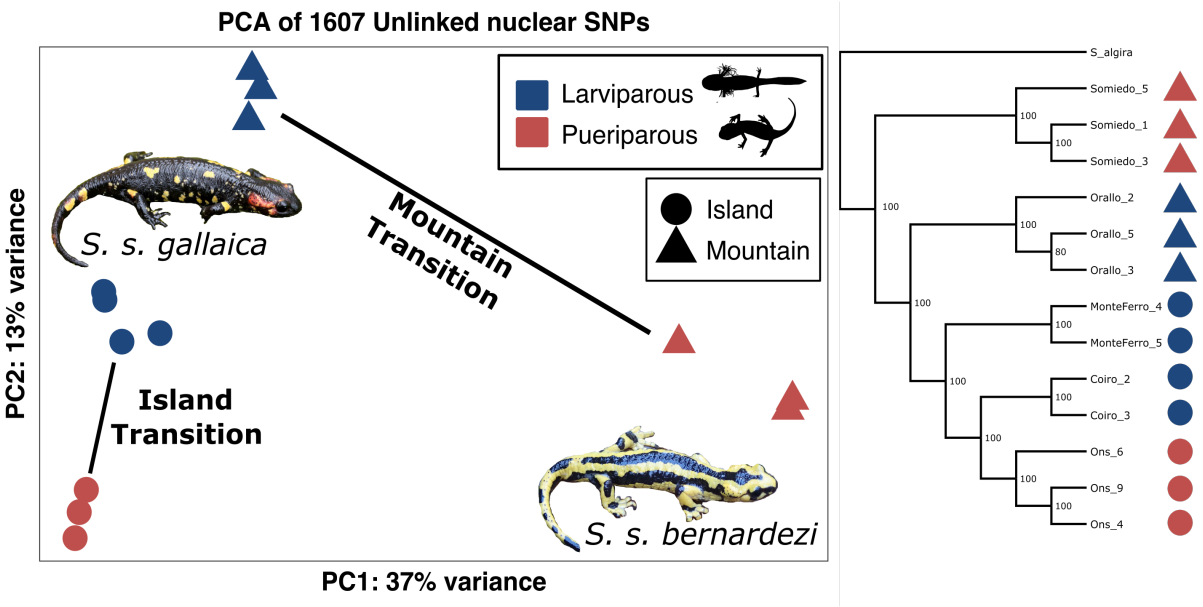
3.3.1 - Reference transcriptome

We analyzed a total of 1.07 billion paired-end Illumina reads, 491 million for the 26 uterus samples, 245 million reads for the 13 oviduct samples and 613 million reads across all 14 tissues for the two reference individuals. The Trinity assembly consisted of 1,402,216 contigs and included complete transcripts for 91.1% of the 2,586 vertebrate BUSCO genes (98% including partial copies). The assembly had a 90% remapping rate with bowtie2 and 73% when only allowing for concordantly mapped reads. TransRate identified 675,041 of transcripts as high-quality and a total of 99,864 transcripts included a *Pleurodeles* annotation.

3.3.2 - Genetic distance between larviparous and pueriparous population pairs

Our SNP calling pipeline resulted in 1,607 high quality unlinked SNPs. Both the phylogenetic tree and the PCA (Figure 3.2) clearly separated the populations into four independent groups. The distance between the pueriparous mountain population (*S. s. bernardezi*) and the three other populations (all *S. s. gallaica*) was the strongest, which is consistent with previous studies (Chapter 2; Burgon et al. in review). Importantly, both the PCA and the maximum likelihood tree indicate that the two pueriparous populations are not closely related. As the ancestral state of the species is larviparity, this suggests that pueriparity evolved twice independently.

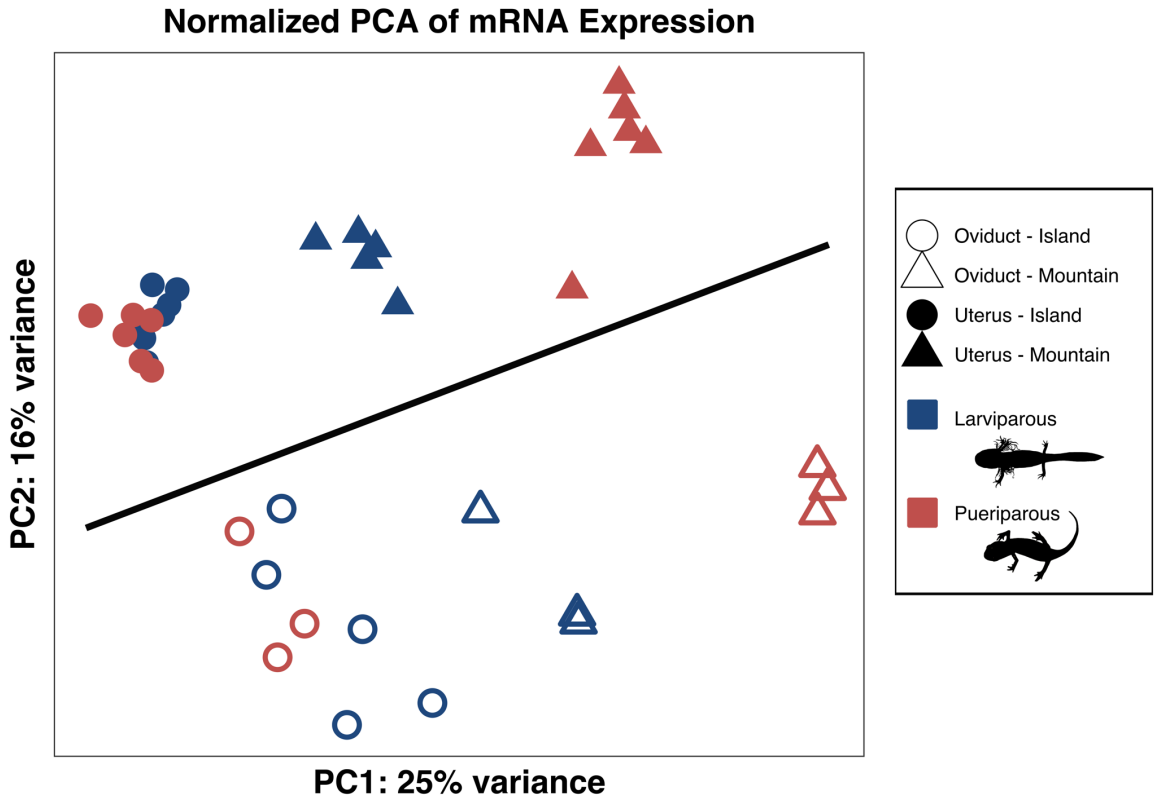
Figure 3.2. Principal component analyses of variation in 1,607 unlinked nuclear SNPs and RAxML phylogenetic reconstruction showing genetic structure in the RNAseq samples used in this study. The deepest split is between the *S. s. bernardezi* population on the bottom-right, and the three populations of *S. s. gallaica* on the left (PC1). The mountain transition is genetically more divergent (PC1) than the island transition (PC2). The maximum likelihood reconstruction shows the same relationships, with the pueriparous *S. s. bernardezi* as the sister clade to all *S. s. gallaica* populations.



3.3.3 - Overall expression patterns

The principal component analysis of mRNA expression shows a split between uterus and oviduct samples (Figure 3.3) but within the same tissue, expression is associated with overall genetic distance as estimated in our nuclear SNP dataset above. The pueriparous *S. s. bernardezi* subspecies had the most divergent expression patterns, the pueriparous and larviparous *S. s. gallaica* samples for the island transition had partially overlapping expression patterns, and the larviparous *S. s. gallaica* samples from the mountain transition showed an intermediate expression pattern.

Figure 3.3. PCA of expression patterns of mRNA transcripts across both oviduct and uterus tissues after variance stabilizing transformation to homoscedastic data using the vst function in the DEseq2 R package. There is a split between the uterus samples in the top and the oviduct samples in the bottom (line added for emphasis).



3.3.4 - Tissue specific expression

Tissue specificity overall was highest for the kidney and lowest for the uterus (421 and 57 specific genes respectively) and intermediate for the oviduct (174; Figure C1). Several of the specific genes in both uterus and oviduct had development and hormonal functions, examples including; DLX6, ADM2, MSX1 and MSX2 (Table 3.1), and the oviduct included several specific carbohydrate sulfotransferase genes (Table 3.1; CHST1, CHST4, CHST6).

Table 3.1. List of highly expressed and tissue specific genes in the uterus and oviduct as identified by the tispec R package. Tau is the tissue-specific score (0 = broadly expressed across tissues, 1 = completely specific to one tissue) and Quant is a relative quantification of expression. Notes added if there was a known function related to embryonic development based on www.genecards.org or a literature search.

Gene	Tau	Quant	Notes on function
			Uterus specific genes
DLX6	1.00	6.1	Forebrain and craniofacial development
UPK1B	0.94	6.5	May play an important role in normal bladder epithelial physiology
TMEM30B	1.00	4.3	-
ADM2	0.93	5.2	Important for the maternal-fetal interface in humans
NIPAL4	0.96	4.7	-
MSX1	0.96	4.7	Acts as a transcriptional repressor. May play a role in limb-pattern formation. Acts in craniofacial development and specifically in odontogenesis.
SCNN1B	0.86	5.8	-
MSX2	1.00	3.8	Acts as a transcriptional regulator in bone development
Pfam:Trypsin	0.91	4.6	-
ADM2B	1.00	3.3	-
			Oviduct specific genes
MUC6	1.00	9.2	Important for DNA replication and can be maternally provided to eggs in Drosophila (Ohno et al. 1998)
YHU2	1.00	8.8	-
BBL021308	1.00	7.4	-
CHST4	1.00	7.4	-
GCNT3	1.00	7.1	Introduce the blood group I antigen during embryonic development
AQP2	1.00	6.7	-
TCTEX1D1	1.00	6.4	-
BBL015215	0.98	6.6	-
CHST6	0.92	6.5	-
CHST1	1.00	5.0	-
			Uterus & oviduct specific genes
CHST9	0.95	8.1	Participates in biosynthesis of glycoprotein hormones lutropin and thyrotropin
B3GAT1	1.00	5.6	Involved in the biosynthesis of L2/HNK-1 carbohydrate epitope on glycoproteins
IVL	0.90	6.9	Part of the insoluble cornified cell envelope (CE) of stratified squamous epithelia
GAL3ST3	1.00	4.8	-
NUDT16	0.88	6.0	-
FAM3D	0.94	4.9	-
CERS2	0.88	5.8	May play a role in the regulation of cell growth
SOX17	0.92	5.1	Plays a key role in the regulation of embryonic development

3.3.5 - Differential expression between larviparity and pueriparity

When comparing uterus expression for all larviparous to all pueriparous individuals, 554 genes were significantly differentially expressed after the FDR correction for multiple testing (Figure 3.4A and 3.4C, Table C2). Of these, 275 were upregulated for pueriparous individuals and 279 were downregulated. For the oviduct there were no significant genes after correcting for multiple testing (Figure 3.4B and 3.4D, Table C3); it is unclear if this is due to lower sample size or biological differences.

For the uterus there were 2,763 genes with significant differential expression when looking at the island transition only (Figure C2, Table C4; 1,292 upregulated and 1,471 downregulated), and 98 genes across the mountain transition (Figure C3, Table C5; 66 upregulated and 43 downregulated). We did not compare the oviduct for the single transitions due to the low number of samples.

Figure 3.4. Volcano plots of (a) uterus transcripts and (b) oviduct transcripts. Across the x-axis is the change in expression (positive means higher expression for pueriparous females). Along the y-axis is the significant value as calculated by edgeR. In yellow are the genes with an FRD value of $p < 0.01$. Indicated by name are the top 7 genes as identified by edgeR. Those top genes have also been depicted in boxplot format for both (c) the uterus tissues and (d) oviduct tissues.

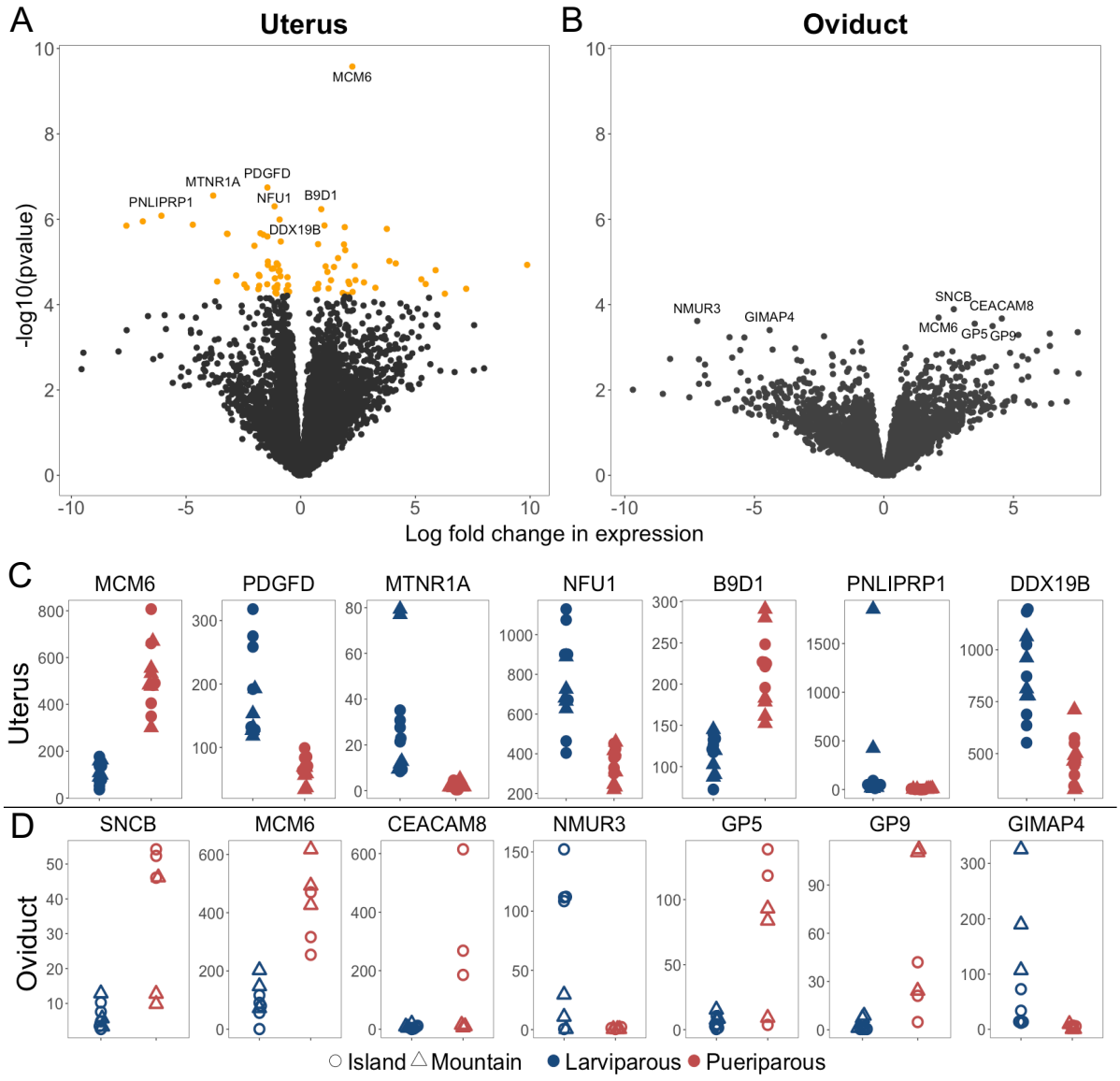


Table 3.2. List of the top differentially expressed genes across our multiple comparisons. LogFC = Log of the Fold Change in expression levels; FDR = False Discovery Rate.

Gene	logFC	FDR	Notes on function
Uterus differentially expressed			
MCM6	2.25	3.5E-06	Important for DNA replication and can be maternally provided to eggs in *Drosophila* (Ohno et al. 1998)
PDGFD	-1.45	1.2E-03	Growth factor that plays an essential role in the regulation of embryonic development, cell proliferation, cell migration, survival and chemotaxis
MTNR1A	-3.82	1.2E-03	Important in the regulation of both circadian rhythms and reproductive cycles
NFU1	-1.14	1.5E-03	-
B9D1	0.89	1.5E-03	Associated with the sonic hedgehog signalling pathway
PNLIPRP1	-6.08	1.7E-03	-
DDX19B	-0.92	1.7E-03	Involved in embryogenesis, spermatogenesis, and cellular growth and division
TMEM56	-6.89	1.7E-03	-
STPG4	-4.71	1.7E-03	Facilitates epigenetic changes in the embryo by means of changing methylation dynamics at several stages of development
MTUS1	1.03	1.7E-03	Developmental regulation of the cardiovascular system (Bundschu and Schuh 2014)
NMUR3	-7.60	1.7E-03	-
TPPP3	1.92	1.7E-03	Required for embryo implantation in mice (Shukla et al. 2018)
Oviduct differentially expressed			
SNCB	2.70	0.59	
MCM6	2.11	0.59	Important for DNA replication and can be maternally provided to eggs in Drosophila (Ohno et al. 1998)
CEACAM8	4.55	0.59	
NMUR3	-7.21	0.59	-
GP5	3.50	0.59	-
GP9	4.19	0.59	-
GIMAP4	-4.41	0.59	-
Uterus - Mountain transition			
SCEL	-8.14	3.3E-03	Correlated with the activation of markers of differentiation in epidermis (Champliaud et al. 2000)
TIGAR	1.62	4.3E-03	Involved in the Warburg effect that can stimulate embryonic growth (Krisher and Prather 2012)
RFWD3	-1.62	5.2E-03	-
BPI	-2.74	5.2E-03	Associated with defence mechanism during bacterial infections of amniotic fluid in humans (Espinoza et al. 2003)
HDAC10	1.74	5.2E-03	-
MCM6	2.02	5.2E-03	Associated with uterine cancer
PCBD2	1.77	7.3E-03	-
Uterus - Island transition			
CCDC175	4.76	8.9E-07	-
PRSS33	-2.37	1.1E-05	-
NMB	7.60	3.1E-05	Can onset labour in pregnant mice (Zhang et al. 2011)
LMNA	2.31	8.5E-05	Important for embryogenesis
PSMC3	-1.39	8.5E-05	Important for embryogenesis (Sakao et al. 2000)
SHISA2	-2.41	9.4E-05	Important in segmental patterning in *Xenopus* embryos (Nagano et al. 2006)
FABP7	-6.39	9.9E-05	Important in neurogenesis for both embryo's and adults

3.4 - Discussion

RNAseq analyses across two independent transitions from larviparity to pueriparity show that there are both convergent and distinct differences in gene expression profile of the uterus across both reproductive modes. Given the independent evolutionary origins and the different time period and environment in which samples were collected, convergent expression patterns are likely associated with reproductive mode. Divergent patterns of gene expression across transitions may also be associated with transitions in reproductive mode; however, we cannot rule out other environmental or genetic factors. Many of the genes that were significantly differentially expressed between reproductive modes are related to embryogenesis and cell growth and differentiation in humans and other model organisms. We highlight and discuss several candidate genes that may be important for embryogenesis in *Salamandra* and underlie the shifts from larviparity to pueriparity in this group.

3.4.1 - Convergent patterns of differential expression across mountain and island transitions to pueriparity

Overall expression patterns in both uterus and oviduct clustered samples by genetic and environmental distance and not by reproductive mode (Figure 3.3); however, several genes in the uterus showed convergent differential expression across both transitions, providing strong evidence that they are related to reproductive mode. This list includes several genes with putative embryonic functions (Table 3.2), though these associations are based on studies performed in model organisms and not much is known about the functions of these genes in amphibians. For instance, MCM6 was upregulated in both the uterus and oviduct of pueriparous individuals across both transitions. This gene is involved in initiating DNA replication and is generally upregulated during the G0 phase of the cell cycle. In *Drosophila*, excess MCM6 mRNA is maternally provided to the eggs (Ohno et al. 1998), and drosophila larvae lacking a functional MCM6 copy do not show developmental problems until metamorphosis when these maternal stores become depleted (Schwed et al. 2002). MCM6 also shows different expression profiles during development between *Ambystoma mexicanum* that retain larval characteristics when sexually mature (paedomorphosis) and *A. tigrinum* that do undergo metamorphosis (Boley 2009). In *Salamandra*, increased maternal MCM6 supply from the uterus might thus promote development and metamorphosis in pueriparous embryos. B9D1 was also consistently upregulated in the uterus of pueriparous individuals. This gene is

required for ciliogenesis, and interestingly it is also associated with the sonic hedgehog signalling pathway, which is important for the developing embryo, including for organogenesis and limb development. The gene TPPP3 is important for embryo implantation in the uterus in mammals and is believed to play a role in signalling between the uterus and the embryo (Shukla et al. 2018). In *Salamandra*, TPPP3 was highly expressed across both reproductive modes, but upregulated in pueriparous females. Although there is no embryo implantation in either mode, TPPP3 in *Salamandra* may be involved in signalling pathways between the embryo and the uterus.

Two genes that were scarcely expressed in pueriparous individuals, but expressed in the uterus tissues of larviparous individuals were STPG4 and MTNR1A, whereas PDGFD and DDX19B were present in pueriparous females but upregulated for the larviparous individuals. Melatonin Receptor 1A (MTNR1A) is important in the regulation of both circadian rhythms and reproductive cycles in mammals (Wang et al. 2017; Migaud, Daveau, and Malpaux 2005); thus, differential expression in *Salamandra* may be related to differences in development speed between larviparous and pueriparous embryos (Buckley et al. 2007). Maternal factor gonad-specific expression gene (STPG4, also called GSE) facilitates epigenetic changes in the embryo by means of changing methylation dynamics at several stages of development (Eckersley-Maslin, Alda-Catalinas, and Reik 2018; Hatanaka et al. 2013). This gene may be involved in the heterochronic development between embryos within a clutch as well as the induction of metamorphosis in pueriparous embryos (Buckley et al. 2007). Platelet derived growth factor D (PDGFD) was upregulated in larviparous females. Platelet derived growth factors in general are important in cell growth and embryonic development, and have been associated with mesoderm patterning of the early embryo in *Xenopus laevis* (Ghil and Chung 1999). Not much is known about the function of DDX19B, but zebrafish CRISPR knockouts of the DDX19 gene results in abnormal apoptosis and cell proliferation causing early death in embryos, highlighting its importance for embryonic development (Shi et al. 2019).

3.4.2 - Potential independent genetic mechanisms in the shift to pueriparity

By analysing both transitions together, we applied a conservative approach to identify genes associated with reproductive mode and not with other evolutionary or environmental factors. However, this approach predisposes us to overlook genes that may be specific for each of the transitions. Given the independent evolutionary origins and disparate timings of these transitions to pueriparity (Chapter 2), the shared pueriparity phenotype may have a distinct (or

partially distinct) underlying genetic architecture (Wittkopp et al. 2003; Steiner, Weber, and Hoekstra 2007). Correspondingly, we found many differentially expressed genes that were unique to the island and mountain transitions (Table 3.2, Figures C2 and C3), several of which have functions associated with embryogenesis. Across the mountain transition, the upregulation of the TIGER gene in the pueriparous *S. s. bernardezi* is interesting, as this gene is associated with embryonic developmental speed in mammals (Krisher and Prather 2012). Thus, in *S. s. bernardezi*, TIGER may be responsible for the embryos faster development compared to their larviparous counterparts (Buckley et al. 2007). Likewise, several genes related to embryogenesis were differentially expressed between the reproductive modes of the island transition (LMNA, PSMC3, SHISA2, FABP7), most interestingly expression of NMB can initiate labour in pregnant mice (Zhang et al. 2011), and we found NMB was upregulated in the pueriparous females on the island of Ons.

These single-transition candidate genes may reflect partially unique genetic architectures of the convergent phenotypes. However, it is far more challenging to associate these candidates with reproductive mode rather than other environmental or evolutionary differences among pueriparous and larviparous females within a given transition. Thus, relative to genes that were differentially expressed in both transitions, we are less confident that the single transition candidates are associated with shifts in reproductive mode in *Salamandra*. Across the mountain transition, for example, *S. s. bernardezi* differ from *S. s. gallaica* in colour patterns and morphology (Velo-Antón and Buckley 2015; Alarcón-Ríos et al. 2020). Likewise, environmental conditions differ between island and continental localities for the two lineages in our island transition. Additionally, there was a difference in timing of the pregnancy between both transitions, both due to different sampling periods, and because of the environmental differences in breeding seasons (Velo-Antón and Buckley 2015). This means that we are likely missing important genes that are either transition-specific, or dependent on timing of pregnancy.

3.4.3 - General gene expression patterns of reproductive tissues in *Salamandra*

General expression patterns were associated with both environmental and evolutionary distance (Figure 3.3). The larviparous and pueriparous *S. s. gallaica* populations of the island transition diverged in the late Pleistocene and are found in the same region, and correspondingly have similar expression patterns. The larviparous *S. s. gallaica* mountain population of Orallo is more closely related to the other two *S. s. gallaica* populations in our

study, but is found in a similar environment as the pueriparous *S. s. bernardezi* and its general expression profile was intermediate between both those groups.

In the reference transcriptomes, the kidney had the highest tissue-specificity in mRNA expression, which is consistent with previous studies in mammals and fishes (Salem et al. 2015; Ramsköld et al. 2009). Tissue specificity was higher in the oviduct than in the uterus (Figure C1), suggesting that in *Salamandra* the oviduct is a more specialized organ. Yet, we did not detect any significantly differentially expressed genes between the oviducts of larviparous and pueriparous females. This null result may reflect the low sample size for this comparison (13 samples), the timing of our sampling (the embryos had already passed through the oviduct to the uterus), or that the oviduct is truly less important in the development of the embryo. The large differences between both the number of larvae/juveniles and the different stages of development found in two uteri of the same individual was surprising and it is not clear if this is related to gene expression or could be related to external factors such as breeding behaviour and/or multiple paternity (Steinfartz et al. 2006; Caspers et al. 2014).

3.4.4 - Caveats and opportunities of studying gene expression in natural populations

Gene expression analyses are often conducted in the laboratory under controlled experimental conditions and on genetically similar individuals. These experimental designs reduce the impact of genetic background, environmental conditions and past experience, which can all greatly impact gene expression (Todd, Black, and Gemmell 2016). Unfortunately, some of the most fascinating phenotypic traits occur in non-model organisms and cannot be studied using these ideal experimental conditions. Examples include traits that are species-specific and those that are not inducible in the laboratory and require field-based research (Anderson et al. 2014; Armengaud et al. 2014; Tagu, Colbourne, and Nègre 2014). Gene expression varies widely between species and environments, and thus the links between a phenotype of interest with changes in gene expression can be challenge to disentangle in cross-species or field based RNAseq experiments. Gene expression in species-specific traits such as viviparity have thus been studied by applying time-series between closely related species that differ in reproductive mode (Griffith et al. 2016; Boswell et al. 2009).

Additionally, gene expression sampled from multiple different environments can also introduce noise into gene expression patterns (Wolf 2013; Wolf et al. 2010). Bringing individuals to a

common garden will harmonize some gene expression patterns, but differences will remain based on past experiences. The intra-specific variation in reproductive mode observed in *S. salamandra* can reduce the effect of phylogenetic background on gene expression and help us study traits that are species-specific. By applying our analyses to two separate transitions and two separate environments within the same species, we reduce the problems associated with both phylogenetic background and environmental specific differences in gene expression, by highlighting differentially expressed genes across both transitions.

3.5 - Conclusion

To our knowledge, this is the first study to characterize gene expression in the reproductive tissues of larviparous and pueriparous organisms and investigate the genetic basis of a remarkable shift in reproductive mode that allows amphibians to colonize water-limited habitats. Shared differences in uterus gene expression across two independent transitions to pueriparity indicate that maternal gene expression is associated with the differences in embryonic development between reproductive modes. We highlight numerous candidate genes that may be important in explaining the key evolutionary transition from larviparity to pueriparity in *Salamandra*. Several of these genes are involved in both embryogenesis and cell growth and differentiation, factors that differ between reproductive modes. Testing for signatures of selection in the coding and regulatory regions of these candidates across the mountain and island transitions to pueriparity in *Salamandra* may provide further confirmation of the importance of these genes. This approach could also be applied to the rest of the *Salamandra* radiation including the second island transition on the island of San Martiño, and to samples from the hybrid zone that can be used for admixture mapping (e.g. in the Basque country, Uotila et al. 2013). The results of this study can form the basis of future genetic screening of *Salamandra* populations by combining genetic and geographic data to better understand the ecological and environmental background in which pueriparity evolved.

3.6 – Acknowledgements

We thank Sara João and Sandra Afonso for facilitating lab work CIBIO. Thanks to Ryan Schott and Anna Savage for their insights on the analyses. We also thank the National Park staff (RPN) that facilitated our trip and lodging on Ons Island. Portions of this study were conducted in and with the support of the L.A.B. facilities of the National Museum of Natural History (NMNH). The computations performed for this paper were conducted on the Smithsonian High Performance Cluster (SI/HPC), Smithsonian Institution. https://doi.org/10.25572/SIHPC. Financial support came from National Funds through FCT – Foundation for Science and Technology (EVOVIV: PTDC/BIA-EVF/3036/2012; SALOMICS: PTDC/BIA-EVL/28475/2017), CIBIO-New-Gen project (ID: 28643; FP7-REGPOT), FEDER / Ministerio de Ciencia, Innovación y Universidades – Agencia Estatal de Investigación, Spain: CGL2017-83131-P). KPM and AL were funded by FCT predoctoral grants (PD/BD/52604/2014 and PD/BD/106060/2015 respectively) and GVA was supported by an FCT research contract (IF/01425/2014), all from the Portuguese ‘‘Fundação para a Ciência e a Tecnologia,” funded by Programa Operacional Potencial Humano (POPH)—Quadro de Referência Estratégica Nacional (QREN) from the European Social Fund. Salamanders were captured, processed, and sacrificed under collection and ethical permits provided by regional governments (Galicia: Ref. 410/2015; Ref. EB016-2018; and Asturias: Ref. 2016/001092; Ref. 2018/0022115).

3.7 - References

Alarcón-Ríos, Lucía, Alfredo G. Nicieza, Antigoni Kaliontzopoulou, David Buckley, and Guillermo Velo-Antón. 2020. "Evolutionary History and Not Heterochronic Modifications Associated with Viviparity Drive Head Shape Differentiation in a Reproductive Polymorphic Species, *Salamandra salamandra*." *Evolutionary Biology* 47 (1): 43–55. https://doi.org/10.1007/s11692-019-09489-3.

AmphibiaWeb 2020. University of California, Berkeley, CA, USA. 2020. https://amphibiaweb.org.

Anderson, Jill T., Maggie R. Wagner, Catherine A. Rushworth, Kasavajhala V.S.K. Prasad, and Thomas Mitchell-Olds. 2014. "The Evolution of Quantitative Traits in Complex Environments." *Heredity* 112 (1): 4–12. https://doi.org/10.1038/hdy.2013.33.

Armengaud, Jean, Judith Trapp, Olivier Pible, Olivier Geffard, Arnaud Chaumot, and Erica M. Hartmann. 2014. "Non-Model Organisms, a Species Endangered by Proteogenomics." *Journal of Proteomics* 105: 5–18. https://doi.org/10.1016/j.jprot.2014.01.007.

Atkins, Natalia, Susan M. Jones, and Louis J. Guillette. 2006. "Timing of Parturition in Two Species of Viviparous Lizard: Influences of β-Adrenergic Stimulation and Temperature upon Uterine Responses to Arginine Vasotocin (AVT)." *Journal of Comparative Physiology B: Biochemical, Systemic, and Environmental Physiology* 176 (8): 783–92. https://doi.org/10.1007/s00360-006-0100-0.

Barson, Nicola J., Tutku Aykanat, Kjetil Hindar, Matthew Baranski, Geir H. Bolstad, Peder Fiske, Céleste Jacq, et al. 2015. "Sex-Dependent Dominance at a Single Locus Maintains Variation in Age at Maturity in Salmon." *Nature* 528 (7582): 405–8. https://doi.org/10.1038/nature16062.

Benjamini, Yoav, and Yosef Hochberg. 1995. "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing." *Journal of the Royal Statistical Society: Series B (Methodological)* 57 (1): 289–300.

Biazik, Joanna M., Scott L. Parker, Christopher R. Murphy, and Michael B. Thompson. 2012. "Uterine Epithelial Morphology and Progesterone Receptors in a Mifepristone-Treated Viviparous Lizard *Pseudemoia entrecasteauxii* (Squamata: Scincidae) During Gestation." *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution* 318 (2): 148–58. https://doi.org/10.1002/jez.b.22003.

Blackburn, Daniel. 2015. "Evolution of Vertebrate Viviparity and Specializations for Fetal Nutrition: A Quantitative and Qualitative Analysis." *Journal of Morphology* 276 (8): 961–

90. https://doi.org/10.1002/jmor.20272.

Boley, MA. 2009. “A Comparative Study of Larval Gene Expression between a Paedomorphic and Metamorphic Species of Ambystomatid Salamander.” University of Kentucky.

Bolger, Anthony M., Marc Lohse, and Bjoern Usadel. 2014. “Trimmomatic: A Flexible Trimmer for Illumina Sequence Data.” *Bioinformatics* 30 (15): 2114–20. https://doi.org/10.1093/bioinformatics/btu170.

Boswell, Mikki G., Melissa C. Wells, Lyndsey M. Kirk, Zhenlin Ju, Ziping Zhang, Rachell E. Booth, and Ronald B. Walter. 2009. “Comparison of Gene Expression Responses to Hypoxia in Viviparous (*Xiphophorus*) and Oviparous (*Oryzias*) Fishes Using a Medaka Microarray.” *Comparative Biochemistry and Physiology - C Toxicology and Pharmacology* 149 (2): 258–65. https://doi.org/10.1016/j.cbpc.2008.11.005.

Brandley, Matthew C., Rebecca L. Young, Dan L. Warren, Michael B. Thompson, and Günter P. Wagner. 2012. “Uterine Gene Expression in the Live-Bearing Lizard, *Chalcides ocellatus*, Reveals Convergence of Squamate Reptile and Mammalian Pregnancy Mechanisms.” *Genome Biology and Evolution* 4 (3): 394–411. https://doi.org/10.1093/gbe/evs013.

Buckley, David. 2012. “Evolution of Viviparity in Salamanders (Amphibia, Caudata).” *ELS*. https://doi.org/10.1002/9780470015902.a0022851.

Buckley, David, Marina Alcobendas, Mario Garcia-Paris, and Marvalee H. Wake. 2007. “Heterochrony, Cannibalism, and the Evolution of Viviparity in *Salamandra salamandra*.” *Evolution and Development* 9 (1): 105–15. https://doi.org/10.1111/j.1525-142X.2006.00141.x.

Bundschu, Karin, and Kai Schuh. 2014. “Cardiovascular ATIP (Angiotensin Receptor Type 2 Interacting Protein) Expression in Mouse Development.” *Developmental Dynamics* 243 (5): 699–711. https://doi.org/10.1002/dvdy.24102.

Caspers, Barbara A., E. Tobias Krause, Ralf Hendrix, Michael Kopp, Oliver Rupp, Katrin Rosentreter, and Sebastian Steinfartz. 2014. “The More the Better - Polyandry and Genetic Similarity Are Positively Linked to Reproductive Success in a Natural Population of Terrestrial Salamanders (*Salamandra salamandra*).” *Molecular Ecology* 23 (1): 239–50. https://doi.org/10.1111/mec.12577.

Champliaud, Marie-France, Howard P. Baden, Manuel Koch, William Jin, Robert E. Burgeson, and Alain Viel. 2000. “Gene Characterization of Sciellin (SCEL) and Protein Localization in Vertebrate Epithelia Displaying Barrier Properties.” *Genomics* 70 (2): 264–68. https://doi.org/https://doi.org/10.1006/geno.2000.6390.

Condon, Kenneth. 2020. “Tispec: Calculates Tissue Specificity from RNA-Seq Data.”

Crump, Martha L. 2015. “Anuran Reproductive Modes: Evolving Perspectives.” *Journal of Herpetology* 49 (1): 1–16. https://doi.org/10.1670/14-097.

Dinis, Marco, and Guillermo Velo-Antón. 2017. “How Little Do We Know about the Reproductive Mode in the North African Salamander, *Salamandra algira*? Pueriparity in Divergent Mitochondrial Lineages of *S. a. Tingitana*.” *Amphibia Reptilia* 38 (4): 540–46. https://doi.org/10.1163/15685381-00003121.

Duellman, William E., and Linda Trueb. 1994. *Biology of Amphibians*. JHU Press.

Eckersley-Maslin, Melanie A., Celia Alda-Catalinas, and Wolf Reik. 2018. “Dynamics of the Epigenetic Landscape during the Maternal-to-Zygotic Transition.” *Nature Reviews Molecular Cell Biology* 19 (7): 436–50. https://doi.org/10.1038/s41580-018-0008-z.

Ekblom, Robert, and Juan Galindo. 2011. “Applications of next Generation Sequencing in Molecular Ecology of Non-Model Organisms.” *Heredity* 107 (1): 1–15. https://doi.org/10.1038/hdy.2010.152.

Elewa, Ahmed, Heng Wang, Carlos Talavera-López, Alberto Joven, Gonçalo Brito, Anoop Kumar, L. Shahul Hameed, et al. 2017. “Reading and Editing the *Pleurodeles waltl* Genome Reveals Novel Features of Tetrapod Regeneration.” *Nature Communications* 8 (1): 1–9. https://doi.org/10.1038/s41467-017-01964-9.

Elmer, Kathryn R., and Axel Meyer. 2011. “Adaptation in the Age of Ecological Genomics: Insights from Parallelism and Convergence.” *Trends in Ecology and Evolution* 26 (6): 298–306. https://doi.org/10.1016/j.tree.2011.02.008.

Espinoza, Javiera, Tinnakorn Chaiworapongsa, R. Romero, S. Edwin, C. Rathnasabapathy, Ricardo Gomez, Emmanuel Bujold, et al. 2003. “Antimicrobial Peptides in Amniotic Fluid: Defensins, Calprotectin and Bacterial/Permeability-Increasing Protein in Patients with Microbial Invasion of the Amniotic Cavity, Intra-Amniotic Inflammation, Preterm Labor and Premature Rupture of Membranes.” *Journal of Maternal-Fetal and Neonatal Medicine* 13 (1): 2–21. https://doi.org/10.1080/jmf.13.1.2.21.

Foster, Charles S.P., Michael B. Thompson, James U. Van Dyke, Matthew C. Brandley, and Camilla M. Whittington. 2020. “Emergence of an Evolutionary Innovation: Gene Expression Differences Associated with the Transition between Oviparity and Viviparity.” *Molecular Ecology* 29 (7): 1315–27. https://doi.org/10.1111/mec.15409.

Funk, W. Chris, Kelly R. Zamudio, and Andrew J. Crawford. 2018. “Advancing Understanding of Amphibian Evolution, Ecology, Behavior, and Conservation with Massively Parallel Sequencing,” 1–44. https://doi.org/10.1007/13836_2018_61.

Gao, Wei, Yan-Bo Sun, Wei-Wei Zhou, Zi-Jun Xiong, Luonan Chen, Hong Li, Ting-Ting Fu, et al. 2019. “Genomic and Transcriptomic Investigations of the Evolutionary Transition from Oviparity to Viviparity.” *Proceedings of the National Academy of Sciences*,

201816086. https://doi.org/10.1073/pnas.1816086116.

García-París, Mario, Marina Alcobendas, David Buckley, and David B. Wake. 2003. "Dispersal of Viviparity across Contact Zones in Iberian Populations of Fire Salamanders (*Salamandra*) Inferred from Discordance of Genetic and Morphological Traits." *Evolution* 57 (1): 129–43. https://doi.org/10.1111/j.0014-3820.2003.tb00221.x.

Ghil, Jung Sun, and Hae Moon Chung. 1999. "Evidence That Platelet Derived Growth Factor (PDGF) Action Is Required for Mesoderm Patterning in Early Amphibian (*Xenopus laevis*) Embryogenesis." *International Journal of Developmental Biology* 43 (4): 329–34. https://doi.org/10.1387/ijdb.10470649.

Glenn, Travis C., Roger A. Nilsen, Troy J. Kieran, Jon G. Sanders, Natalia J. Bayona-Vásquez, John W. Finger, Todd W. Pierson, et al. 2019. "Adapterama I: Universal Stubs and Primers for 384 Unique Dual-Indexed or 147,456 Combinatorially-Indexed Illumina Libraries (ITru & INext)." *PeerJ* 2019 (10): 1–31. https://doi.org/10.7717/peerj.7755.

Gomez-Mestre, Ivan, Robert Alexander Pyron, and John J. Wiens. 2012. "Phylogenetic Analyses Reveal Unexpected Patterns in the Evolution of Reproductive Modes in Frogs." *Evolution* 66 (12): 3687–3700. https://doi.org/10.1111/j.1558-5646.2012.01715.x.

Greven, H. 2003. "Larviparity and Pueriparity." In *Reproductive Biology and Phylogeny of Urodela*, 447–75. CRC Press.

Griffith, Oliver W., Matthew C. Brandley, Katherine Belov, and Michael B. Thompson. 2016. "Reptile Pregnancy Is Underpinned by Complex Changes in Uterine Gene Expression: A Comparative Analysis of the Uterine Transcriptome in Viviparous and Oviparous Lizards." *Genome Biology and Evolution* 8 (10): 3226–39. https://doi.org/10.1093/gbe/evw229.

Haas, Brian J., Alexie Papanicolaou, Moran Yassour, Manfred Grabherr, Philip D. Blood, Joshua Bowden, Matthew Brian Couger, et al. 2013. "De Novo Transcript Sequence Reconstruction from RNA-Seq Using the Trinity Platform for Reference Generation and Analysis." *Nature Protocols* 8 (8): 1494–1512. https://doi.org/10.1038/nprot.2013.084.

Hatanaka, Yuki, Natsumi Shimizu, Satoshi Nishikawa, Mikiko Tokoro, Seung Wook Shin, Takuji Nishihara, Tomoko Amano, et al. 2013. "GSE Is a Maternal Factor Involved in Active DNA Demethylation in Zygotes." *PLoS ONE* 8 (4). https://doi.org/10.1371/journal.pone.0060205.

Jacobs, Arne, Madeleine Carruthers, Andrey Yurchenko, Natalia V. Gordeeva, Sergey S. Alekseyev, Oliver Hooker, Jong S. Leong, et al. 2020. "Parallelism in Eco-Morphology and Gene Expression despite Variable Evolutionary and Genomic Backgrounds in a

Holarctic Fish." *PLoS Genetics* 16 (4): e1008658. https://doi.org/10.1371/journal.pgen.1008658.

Krisher, Rebecca L., and Randall S. Prather. 2012. "A Role for the Warburg Effect in Preimplantation Embryo Development: Metabolic Modification to Support Rapid Cell Proliferation." *Molecular Reproduction and Development* 79 (5): 311–20. https://doi.org/10.1002/mrd.22037.

Kryuchkova-Mostacci, Nadezda, and Marc Robinson-Rechavi. 2017. "A Benchmark of Gene Expression Tissue-Specificity Metrics." *Briefings in Bioinformatics* 18 (2): 205–14. https://doi.org/10.1093/bib/bbw008.

Langmead, Ben, and Steven L. Salzberg. 2012. "Fast Gapped-Read Alignment with Bowtie 2." *Nature Methods* 9 (4): 357–59. https://doi.org/10.1038/nmeth.1923.

Lehner, Ben. 2013. "Genotype to Phenotype: Lessons from Model Organisms for Human Genetics." *Nature Reviews Genetics* 14 (3): 168–78. https://doi.org/10.1038/nrg3404.

Liedtke, H. Christoph, Hendrik Müller, Julian Hafner, Johannes Penner, David J. Gower, Tomáš Mazuch, Mark Oliver Rödel, and Simon P. Loader. 2017. "Terrestrial Reproduction as an Adaptation to Steep Terrain in African Toads." *Proceedings of the Royal Society B: Biological Sciences* 284 (1851). https://doi.org/10.1098/rspb.2016.2598.

Losos, Jonathan B. 2011. "Convergence, Adaptation, and Constraint." *Evolution* 65 (7): 1827–40. https://doi.org/10.1111/j.1558-5646.2011.01289.x.

Lourenço, André, David Álvarez, Ian J. Wang, and Guillermo Velo-Antón. 2017. "Trapped within the City: Integrating Demography, Time since Isolation and Population-Specific Traits to Assess the Genetic Effects of Urbanization." *Molecular Ecology* 26 (6): 1498–1514. https://doi.org/10.1111/mec.14019.

Lourenço, André, João Gonçalves, Filipe Carvalho, Ian J. Wang, and Guillermo Velo-Antón. 2019. "Comparative Landscape Genetics Reveals the Evolution of Viviparity Reduces Genetic Connectivity in Fire Salamanders." *Molecular Ecology* 28 (20): 4573–91. https://doi.org/10.1111/mec.15249.

Love, Michael I., Charlotte Soneson, and Rob Patro. 2018. "Swimming Downstream: Statistical Analysis of Differential Transcript Usage Following Salmon Quantification." *F1000Research* 7 (0): 952. https://doi.org/10.12688/f1000research.15398.1.

Love, Michael I, Wolfgang Huber, and Simon Anders. 2014. "Moderated Estimation of Fold Change and Dispersion for RNA-Seq Data with DESeq2." *Genome Biology* 15 (12): 550. https://doi.org/10.1186/s13059-014-0550-8.

MacManes, Matthew D. 2014. "On the Optimal Trimming of High-Throughput MRNA Sequence Data." *Frontiers in Genetics* 5 (13): 1–7.

https://doi.org/10.3389/fgene.2014.00013.

McKenna, Aaron, Matthew Hanna, Eric Banks, Andrey Sivachenko, Kristian Cibulskis, Andrew Kernytsky, Kiran Garimella, et al. 2010. "The Genome Analysis Toolkit: A MapReduce Framework for Analyzing next-Generation DNA Sequencing Data." *Genome Research* 20 (9): 1297–1303. https://doi.org/10.1101/gr.107524.110.

Migaud, Martine, Agnès Daveau, and Benoît Malpaux. 2005. "MTNR1A Melatonin Receptors in the Ovine Premammillary Hypothalamus: Day-Night Variation in the Expression of the Transcripts." *Biology of Reproduction* 72 (2): 393–98. https://doi.org/10.1095/biolreprod.104.030064.

Nagano, Takashi, Shoko Takehara, Maiko Takahashi, Shinichi Aizawa, and Akihito Yamamoto. 2006. "Shisa2 Promotes the Maturation of Somitic Precursors and Transition to the Segmental Fate in *Xenopus* Embryos." *Development* 133 (23): 4643–54. https://doi.org/10.1242/dev.02657.

Ohno, Katsuhito, Fumiko Hirose, Yoshihiro H. Inoue, Haruhiko Takisawa, Satoru Mimura, Yoshitami Hashimoto, Tohru Kiyono, Yasuyoshi Nishida, and Akio Matsukage. 1998. "CDNA Cloning and Expression during Development of Drosophila Melanogaster MCM3, MCM6 and MCM7." *Gene* 217 (1–2): 177–86. https://doi.org/10.1016/S0378-1119(98)00358-8.

Patro, Rob, Geet Duggal, Michael I. Love, Rafael A. Irizarry, and Carl Kingsford. 2017. "Salmon Provides Fast and Bias-Aware Quantification of Transcript Expression." *Nature Methods* 14 (4): 417–19. https://doi.org/10.1038/nmeth.4197.

Ramírez-Pinilla, Martha Patricia, Scott L. Parker, Christopher R. Murphy, and Michael B. Thompson. 2012. "Uterine and Chorioallantoic Angiogenesis and Changes in the Uterine Epithelium during Gestation in the Viviparous Lizard, *Niveoscincus conventryi* (Squamata: Scincidae)." *Journal of Morphology* 273 (1): 8–23. https://doi.org/10.1002/jmor.11002.

Ramsköld, Daniel, Eric T. Wang, Christopher B. Burge, and Rickard Sandberg. 2009. "An Abundance of Ubiquitously Expressed Genes Revealed by Tissue Transcriptome Sequence Data." *PLoS Computational Biology* 5 (12): 1–11. https://doi.org/10.1371/journal.pcbi.1000598.

Robinson, Mark D., Davis J. McCarthy, and Gordon K. Smyth. 2009. "EdgeR: A Bioconductor Package for Differential Expression Analysis of Digital Gene Expression Data." *Bioinformatics* 26 (1): 139–40. https://doi.org/10.1093/bioinformatics/btp616.

Rodríguez, Ariel, James D. Burgon, Mariana Lyra, Iker Irisarri, Denis Baurain, Leon Blaustein, Bayram Göçmen, et al. 2017. "Inferring the Shallow Phylogeny of True Salamanders (*Salamandra*) by Multiple Phylogenomic Approaches." *Molecular*

Phylogenetics and Evolution 115: 16–26. https://doi.org/10.1016/j.ympev.2017.07.009.

Safran, Marilyn, Irina Dalah, Justin Alexander, Naomi Rosen, Tsippi Iny Stein, Michael Shmoish, Noam Nativ, et al. 2010. "GeneCards Version 3: The Human Gene Integrator." *Database* 2010: 1–16. https://doi.org/10.1093/database/baq020.

Sakao, Yoshimitsu, Taro Kawai, Osamu Takeuchi, Neal G. Copeland, Debra J. Gilbert, Nancy A. Jenkins, Kiyoshi Takeda, and Shizuo Akira. 2000. "Mouse Proteasomal ATPases Psmc3 and Psmc4: Genomic Organization and Gene Targeting." *Genomics* 67 (1): 1–7. https://doi.org/10.1006/geno.2000.6231.

Salem, Mohamed, Bam Paneru, Rafet Al-Tobasei, Fatima Abdouni, Gary H. Thorgaard, Caird E. Rexroad, and Jianbo Yao. 2015. "Transcriptome Assembly, Gene Annotation and Tissue Gene Expression Atlas of the Rainbow Trout." *PLoS ONE* 10 (3): 1–27. https://doi.org/10.1371/journal.pone.0121778.

Schmieder, Robert, and Robert Edwards. 2011. "Quality Control and Preprocessing of Metagenomic Datasets." *Bioinformatics* 27 (6): 863–64. https://doi.org/10.1093/bioinformatics/btr026.

Schwed, Gina, Noah May, Yana Pechersky, and Brian R. Calvi. 2002. "Drosophila Minichromosome Maintenance 6 Is Required for Chorion Gene Amplification and Genomic Replication." *Molecular Biology of the Cell* 13 (2): 607–20. https://doi.org/10.1091/mbc.01-08-0400.

Shi, Can, Linzhu Bao, Yao Zu, Jianfeng Ren, Weiming Li, and Qinghua Zhang. 2019. "Zebrafish Ddx19 Deficiency Causes Serious Apoptosis and Cell Proliferation." *Research Square*. https://doi.org/10.21203/rs.2.10830/v1.

Shine, Richard, and Louis J. Guillette. 1988. "The Evolution of Viviparity in Reptiles: A Physiological Model and Its Ecological Consequences." *Journal of Theoretical Biology* 132 (1): 43–50. https://doi.org/10.1016/S0022-5193(88)80189-9.

Shukla, Vinay, Pooja Popli, Jyoti Bala Kaushal, Kanchan Gupta, and Anila Dwivedi. 2018. "Uterine TPPP3 Plays Important Role in Embryo Implantation via Modulation of β-Catenin." *Biology of Reproduction* 99 (5): 982–99. https://doi.org/10.1093/biolre/ioy136.

Singh, Ajeet Pratap, and Christiane Nüsslein-Volhard. 2015. "Zebrafish Stripes as a Model for Vertebrate Colour Pattern Formation." *Current Biology* 25 (2): R81–92. https://doi.org/10.1016/j.cub.2014.11.013.

Smith-Unna, Richard, Chris Boursnell, Rob Patro, Julian M. Hibberd, and Steven Kelly. 2016. "TransRate: Reference-Free Quality Assessment of de Novo Transcriptome Assemblies." *Genome Research* 26 (8): 1134–44. https://doi.org/10.1101/gr.196469.115.

Soneson, Charlotte, Michael I. Love, and Mark D. Robinson. 2016. "Differential Analyses for

RNA-Seq: Transcript-Level Estimates Improve Gene-Level Inferences." *F1000Research* 4 (3): 1521. https://doi.org/10.12688/f1000research.7563.2.

Stamatakis, Alexandros. 2014. "RAxML Version 8: A Tool for Phylogenetic Analysis and Post-Analysis of Large Phylogenies." *Bioinformatics* 30 (9): 1312–13. https://doi.org/10.1093/bioinformatics/btu033.

Stamatakis, Alexandros, Paul Hoover, and Jacques Rougemont. 2008. "A Rapid Bootstrap Algorithm for the RAxML Web Servers." *Systematic Biology* 57 (5): 758–71. https://doi.org/10.1080/10635150802429642.

Steiner, Cynthia C., Holger Römpler, Linda M. Boettger, Torsten Schöneberg, and Hopi E. Hoekstra. 2009. "The Genetic Basis of Phenotypic Convergence in Beach Mice: Similar Pigment Patterns but Different Genes." *Molecular Biology and Evolution* 26 (1): 35–45. https://doi.org/10.1093/molbev/msn218.

Steiner, Cynthia C., Jesse N. Weber, and Hopi E. Hoekstra. 2007. "Adaptive Variation in Beach Mice Produced by Two Interacting Pigmentation Genes." *PLoS Biology* 5 (9): 1880–89. https://doi.org/10.1371/journal.pbio.0050219.

Steinfartz, Sebastian, K. Stemshorn, D. Kuesters, and D. Tautz. 2006. "Patterns of Multiple Paternity within and between Annual Reproduction Cycles of the Fire Salamander (*Salamandra salamandra*) under Natural Conditions." *Journal of Zoology* 268 (1): 1–8. https://doi.org/10.1111/j.1469-7998.2005.00001.x.

Tagu, Denis, John K. Colbourne, and Nicolas Nègre. 2014. "Genomic Data Integration for Ecological and Evolutionary Traits in Non-Model Organisms." *BMC Genomics* 15 (1). https://doi.org/10.1186/1471-2164-15-490.

Todd, Erica V., Michael A. Black, and Neil J. Gemmell. 2016. "The Power and Promise of RNA-Seq in Ecology and Evolution." *Molecular Ecology* 25 (6): 1224–41. https://doi.org/10.1111/mec.13526.

Uotila, E, A Crespo-Diaz, I Sanz-Azkue, and X Rubio. 2013. "Variation in the Reproductive Strategies of *Salamandra salamandra* (Linnaeus, 1758) Populations in the Province of Gipuzkoa (Basque Country)." *Munibe* 61: 91–101.

Velo-Antón, Guillermo, and David Buckley. 2015. "Salamandra Común - *Salamandra salamandra* (Linnaeus, 1758)." *Enciclopedia Virtual de Los Vertebrados Españoles. Museo Nacional de Ciencias Naturales, Madrid*.

Velo-Antón, Guillermo, and Adolfo Cordero-Rivera. 2017. "Ethological and Phenotypic Divergence in Insular Fire Salamanders: Diurnal Activity Mediated by Predation?" *Acta Ethologica*. https://doi.org/10.1007/s10211-017-0267-2.

Velo-Antón, Guillermo, Mario García-París, P. Galán, and A. Cordero Rivera. 2007. "The Evolution of Viviparity in Holocene Islands: Ecological Adaptation versus Phylogenetic

Descent along the Transition from Aquatic to Terrestrial Environments." *Journal of Zoological Systematics and Evolutionary Research* 45 (4): 345–52. https://doi.org/10.1111/j.1439-0469.2007.00420.x.

Velo-Antón, Guillermo, Xavier Santos, Iago Sanmartín-Villar, Adolfo Cordero-Rivera, and David Buckley. 2015. "Intraspecific Variation in Clutch Size and Maternal Investment in Pueriparous and Larviparous *Salamandra salamandra* Females." *Evolutionary Ecology* 29 (1): 185–204. https://doi.org/10.1007/s10682-014-9720-0.

Velo-Antón, Guillermo, Kelly R. Zamudio, and Adolfo Cordero-Rivera. 2012. "Genetic Drift and Rapid Evolution of Viviparity in Insular Fire Salamanders (*Salamandra salamandra*)." *Heredity* 108 (4): 410–18. https://doi.org/10.1038/hdy.2011.91.

Wake, David B., Marvalee H. Wake, and Chelsea D. Specht. 2011. "Homoplasy: From Detecting Pattern and Mechanism of Evolution." *Science* 331 (6020): 1032–35. https://doi.org/10.1126/science.1188545.

Wang, Shu Juan, Wen Ju Liu, Li Ke Wang, Xun Sheng Pang, and Li Guo Yang. 2017. "The Role of Melatonin Receptor MTNR1A in the Action of Melatonin on Bovine Granulosa Cells." *Molecular Reproduction and Development* 84 (11): 1140–54. https://doi.org/10.1002/mrd.22877.

Whittington, Camilla M., Oliver W. Griffith, Weihong Qi, Michael B. Thompson, and Anthony B. Wilson. 2015. "Seahorse Brood Pouch Transcriptome Reveals Common Genes Associated with Vertebrate Pregnancy." *Molecular Biology and Evolution* 32 (12): 3114–31. https://doi.org/10.1093/molbev/msv177.

Wittkopp, Patricia J., Barry L. Williams, Jayne E. Selegue, and Sean B. Carroll. 2003. "Drosophila Pigmentation Evolution: Divergent Genotypes Underlying Convergent Phenotypes." *Proceedings of the National Academy of Sciences* 100 (4): 1808–13. https://doi.org/10.1073/pnas.0336368100.

Wolf, Jochen B.W. 2013. "Principles of Transcriptome Analysis and Gene Expression Quantification: An RNA-Seq Tutorial." *Molecular Ecology Resources* 13 (4): 559–72. https://doi.org/10.1111/1755-0998.12109.

Wolf, Jochen B.W., Till Bayer, Bernhard Haubold, Markus Schilhabel, Philip Rosenstiel, and Diethard Tautz. 2010. "Nucleotide Divergence vs. Gene Expression Differentiation: Comparative Transcriptome Sequencing in Natural Isolates from the Carrion Crow and Its Hybrid Zone with the Hooded Crow." *Molecular Ecology* 19: 162–75. https://doi.org/10.1111/j.1365-294X.2009.04471.x.

Wourms, John P, Bryon D Grove, and Julian Lombardi. 1988. "The Maternal-Embryonic Relationship in Viviparous Fishes." In *Fish Physiology*, 11:1–134. Elsevier.

Yanai, Itai, Hila Benjamin, Michael Shmoish, Vered Chalifa-Caspi, Maxim Shklar, Ron Ophir,

Arren Bar-Even, et al. 2005. “Genome-Wide Midrange Transcription Profiles Reveal Expression Level Relationships in Human Tissue Specification.” *Bioinformatics* 21 (5): 650–59. https://doi.org/10.1093/bioinformatics/bti042.

Zamudio, Kelly R., Rayna C. Bell, and Nicholas A. Mason. 2016. “Phenotypes in Phylogeography: Species’ Traits, Environmental Variation, and Vertebrate Diversification.” *Proceedings of the National Academy of Sciences* 113 (29): 8041–48. https://doi.org/10.1073/pnas.1602237113.

Zamudio, Kelly R., Rayna C. Bell, Renato C. Nali, Célio F. B. Haddad, and Cynthia P. A. Prado. 2016. “Polyandry, Predation, and the Evolution of Frog Reproductive Modes.” *The American Naturalist* 188 (S1): S41–61. https://doi.org/10.1086/687547.

Zhang, Chi, Baohong Zhang, Lih Ling Lin, and Shanrong Zhao. 2017. “Evaluation and Comparison of Computational Tools for RNA-Seq Isoform Quantification.” *BMC Genomics* 18 (1): 1–11. https://doi.org/10.1186/s12864-017-4002-1.

Zhang, Wei-She, Qing-Sheng Xie, Xin-Hua Wu, and Qing-Hua Liang. 2011. “Neuromedin B and Its Receptor Induce Labor Onset and Are Associated with the RELA (NFKB P65)/IL6 Pathway in Pregnant Mice.” *Biology of Reproduction* 84 (1): 113–17. https://doi.org/10.1095/biolreprod.110.085746.

CHAPTER 4: QUANTIFYING SELECTION AND GENETIC DRIFT IN A DESERT-ADAPTED AMPHIBIAN DECLINING FROM CHYTRIDIOMYCOSIS

Kevin. P. Mulder[a,b,c,d], Robert C. Fleischer[d], Anna E. Savage[e]

[a] *CIBIO/InBIO, Research Center in Biodiversity and Genetic Resources, Rua Padre Armando Quintas 7, Campus Agrário de Vairão, 4485-661 Vairão, Portugal.*
[b] *Departamento de Biologia da Faculdade de Ciências da Universidade do Porto, Rua Campo Alegre, 4169-007, Porto, Portugal.*
[c] *Center for Conservation Genomics, Smithsonian Conservation Biology Institute, National Zoological Park, 3001 Connecticut Avenue NW, Washington, DC 20008, USA.*
[d] *Department of Vertebrate Zoology, National Museum of Natural History, Smithsonian Institution, 1000 Constitution Avenue NW, Washington, DC 20560, USA.*
[e] *Department of Biology, University of Central Florida, Orlando, USA.*

Abstract

Local adaptation of populations to parallel evolutionary pressures can provide a natural comparative framework to investigate the genomic underpinnings of phenotypic traits. However, understanding the phylogenetic and demographic history of the system is crucial to provide the context in which natural selection is acting. Pairing phylogeographic sampling with functional genetic data is therefore ideal for answering questions about selection in natural populations because this approach enables simultaneous analysis of the evolutionary history of the populations and the genetic signatures of selection. We used this strategy to answer questions about population structure, phylogeography, demography and adaptation across the extant range of the lowland leopard frog (*Rana yavapaiensis*). We generated a genome-wide exon capture dataset, weighting our exon capture towards loci involved in immune function and skin integrity to address questions about adaptation to the fungal disease chytridiomycosis. We compared genome-wide polymorphisms across 11 populations in Arizona, USA, that represent most of the extant species range and that varied from chytridiomycosis susceptible to tolerant, including a chytridiomycosis naïve control population. Our results reveal that lower heterozygosity and allelic richness were associated with increased disease susceptibility and extirpation. However, both extirpated and susceptible populations had higher levels of functionally different private alleles than less threatened populations, potentially reducing future adaptive potential. In addition, we found strong nuclear genetic structure between populations paired with limited mitochondrial diversity, which may reflect recent population fragmentation and decline leading to rapid allele fixation. This backdrop of high genetic drift and population structure likely obscured some F_{ST}-based metrics of adaptation. However, by applying multiple independent analyses of signatures of selection we highlight potential candidate genes important in local adaptation and disease resistance. Our approach shows how functional genetic data combined with phylogeographic datasets can successfully be applied to both questions on demographic history and adaptation on non-model organisms.

Keywords: F_{ST}-outliers, functional genetic variation, host-pathogen, immune-genes, *Rana*

4.1 - Introduction

Understanding the heritable basis of adaptation is a central aim of evolutionary genomics. Similar selective pressures acting on multiple populations can induce parallel events of local adaptation, and this can occur via homologous or analogous genetic mechanisms. Comparing adaptive genomic signatures among genetically distinct natural populations in a comparative framework can offer a natural laboratory to identify the genomic underpinnings of adaptive phenotypic traits. However, accurate detection of genomic regions with signatures of selection linked to adaptive traits (Beaumont 2005; Alves et al. 2019) requires accounting for the phylogenetic and demographic history of the system and the effects of population structure and genetic drift on the genome (Lacy 1987; Hudson et al. 2016). Combining phenotypic information on traits of interest with functional genetic data, all within a phylogeographic context, is what fully enables us to understand the evolutionary history of a system and simultaneously analyse patterns of genetic drift and signatures of selection (Zamudio, Bell, and Mason 2016; Cassin-Sackett, Callicrate, and Fleischer 2019).

The introduction of infectious pathogens can strongly impact population dynamics and genetic drift (Smith, Sax, and Lafferty 2006), as well as imposing strong selective pressures, potentially leading to adaptation (Alves et al. 2019). Anthropogenic changes have increased the occurrence of emerging infectious disease outbreaks in wildlife (Daszak, Cunningham, and Hyatt 2001), and especially in amphibians this has caused global declines (Daszak et al. 1999). The emerging infectious disease chytridiomycosis, caused by the invasive global pandemic lineage of the fungus *Batrachochytrium dendrobatidis* (Bd) has impacted amphibian populations worldwide (Scheele et al. 2019), although some species are showing signs of recovery (Voyles et al. 2018). Bd growth and chytridiomycosis infection intensity are influenced by climatic factors, and differences in disease prevalence and intensity can be partly explained by environmental factors (Kriger, Pereoglou, and Hero 2007; Brem and Lips 2008). However, host genetics are also an important component of differential susceptibilities between species (Gahl, Longcore, and Houlahan 2012; Eskew et al. 2018), populations (Savage, Becker, and Zamudio 2015; Tobler and Schmidt 2010), and individuals (Savage and Zamudio 2011). Most studies have focused on neutral genetic variation (Albert et al. 2015), but the influence of adaptive host genetics has also been investigated in laboratory settings (Savage and Zamudio 2011; Savage et al. 2020), or using single markers like the major histocompatibility complex (MHC) in natural populations (Savage et al. 2018, 2019). However, population level genomic datasets of putatively Bd-adaptive genes are still rare due to the lack

of genomic resources for amphibians, and the difficulty of assessing disease resistance in the field.

The lowland leopard frog (*Rana yavapaiensis*) is a locally abundant amphibian, historically found in mesic habitats across the lowland desert regions of southwestern North America. The species has declined across its range (Clarkson and Rorabaug 1989), and these frogs are now limited to small streams in more pristine habitats with a yearlong water supply. Recent population declines can at least in part be explained by the arrival of Bd (Savage, Sredl, and Zamudio 2011; Bradley et al. 2002). *Rana yavapaiensis* is currently known to occur in Arizona, USA, and in northern parts of Sonora, Mexico (Brennan and Holycross 2006; Rorabaugh and Lemos-Espinal 2016). There are historical records in the neighbouring states of California, Nevada, New Mexico and Utah, but they are now likely extirpated from these regions (Brennan and Holycross 2006; Painter et al. 2017; Pauly et al. 2020). Within Arizona, populations continue to decline and become extirpated, particularly in the southeastern part of the state (Sredl 1997). Die-offs from Bd occur during the cooler months of winter (November – February), but there are differences in disease susceptibility among individuals and among populations (Savage and Zamudio 2016; Savage, Sredl, and Zamudio 2011). This system of multiple populations with differing histories and Bd susceptibilities provides an excellent natural laboratory to study the effects of disease and extirpations on neutral genetic variation and phylogeography and the potential signatures of adaptation and resistance to Bd.

Previous work on *R. yavapaiensis* documented high winter Bd infection in all populations (Savage, Sredl, and Zamudio 2011), and different levels of Bd-associated winter mortality across populations (Savage, Sredl, and Zamudio 2011; Savage and Zamudio 2016) that corresponds to differences in susceptibility assessed from controlled laboratory infection trials (Savage and Zamudio 2011). The Muleshoe Ranch population (MR; Figure 4.1) can be considered a Bd control population because these frogs live in geothermal hot springs where the water temperature remains sufficiently high in winter to inhibit Bd growth (Forrest and Schlaepfer 2011). Thus, MR is the sole population that is potentially representative of the demographic and genetic characteristics of an *R. yavapaiensis* population prior to the introduction of Bd. Microsatellite data has shown that all sampled localities are independently evolving populations (Savage, Becker, and Zamudio 2015), but the phylogenetic relationships between the populations are unknown. Variation in MHC class II alleles does not follow this neutral population structure, and may reflect local adaptation (Savage and Zamudio 2011, 2016).

Here, we used transcriptome-based exon capture to characterize immunogenetic and genome-wide genetic variation in 1,388 loci across 11 populations and 133 individuals of *Rana yavapaiensis*. Samples stemmed from a 10-year timespan and include two populations that were extirpated during our sampling period, enabling us to study the effects of the current distribution and historical Bd susceptibility on genetic variation and putative Bd-resistance genes. Our aims were to (a) describe population genetic structure and phylogeography of the species across space, (b) compare population genetic variation among locations with contrasting Bd histories, (c) test for signatures of selection across all populations, and (d) identify SNPs that are associated with the intensity of Bd infections and may thus be related to disease adaptation.

4.2 - Methods

Our study focusses on 11 populations across Arizona (Figure 4.1) that have shown differential susceptibility to Bd-infections in the last 14 years (see Table 4.1). Although the Mexico distribution of *R. yavapaiensis* is poorly characterized, our sampled populations represent the current US distribution of the species with the exception of one known population which occurs further north (Surprise Canyon; Oláh-Hemmings *et al.* 2010). Winter Bd die-offs are more severe in the southeastern populations, but some of the northwestern populations also show high disease susceptibility (Bradley et al. 2002; Savage, Sredl, and Zamudio 2011), and populations from both sides of the distribution show variation in disease susceptibility in the lab (Savage and Zamudio 2011). Thus, to evaluate these populations in the context of disease susceptibility, we categorized each population into one of three groups based on our previous field and lab studies: (1) tolerant, where Bd is present but there is no evidence of winter die-offs, (2) variable, where Bd causes mortality in some individuals but not others, both in the field and in controlled lab experiments, and (3) susceptible, where annual winter die-offs occur and for some of these populations, complete susceptibility has been confirmed in the lab (Table 4.1). For three populations we do not have sufficient data to assess Bd susceptibility due to observing few or no individuals during winter sampling efforts (Aliso Spring, Two Mile Tank & Dix Creek).

Figure 4.1. (a) Map of the 11 *Rana yavapaiensis* sampling localities across Arizona, USA, with chytridiomycosis disease susceptibility indicated by shape. Elevation is shaded in greyscale and waterways are indicated in white. Two populations (TV and AS) have since been extirpated. Inset shows extent of sampling area in southwestern United States and northwestern Mexico. (b) PCA of 2,248 unlinked SNPs as calculated by snpgdsPCA. Samples coloured by locality.

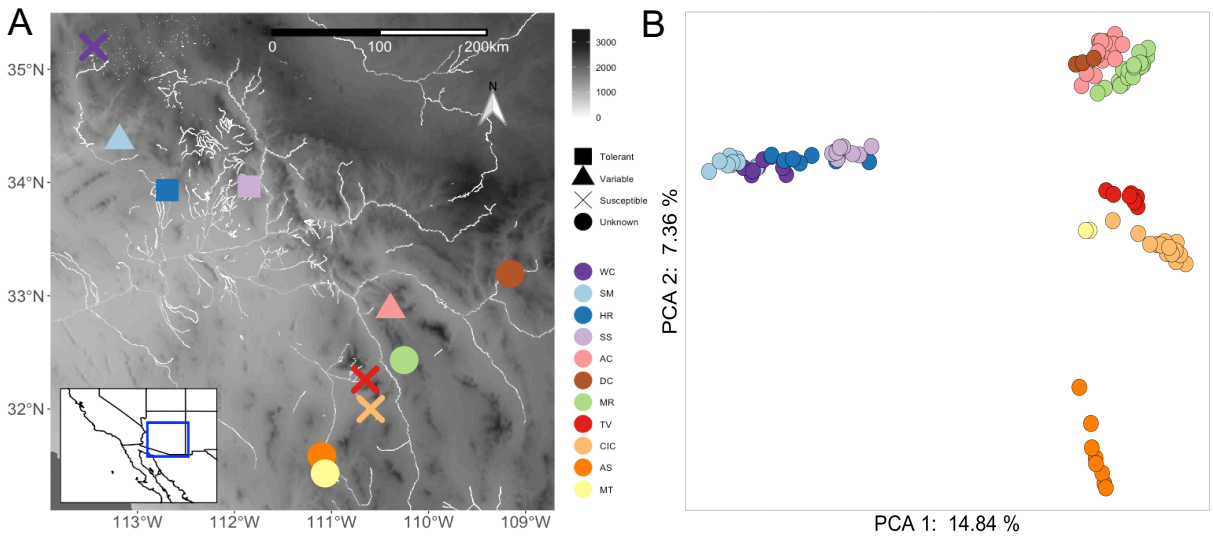


Table 4.1. *Rana yavapaiensis* susceptibility to Bd across Arizona populations based on previous work [1,2,3].

Pop. ID	Population	n	Experimental susceptibility[1]	Winter field mortality[2]	Winter field mortality[3]	Susceptibility designation	Notes
AC	Aravaipa Canyon	23	41% survival	5%	8%	Variable	
AS	Aliso Spring	10	NA	NA	NA	Unknown	Presumed extirpated
CIC	Cienega Creek	15	No survival	40%	20%	Susceptible	
HR	Hassayampa River	11	NA	0%	0%	Tolerant	
MR	Muleshoe Ranch	28	No survival	0% *	0% *	Control	Geothermal spring keeps water warm year-round
MT	Two Mile Tank	2	NA	NA	NA	Unknown	
DC	Dix Creek	3	NA	NA	NA	Unknown	
SM	Santa Maria River	11	27% survival	0%	0%	Variable	
SS	Seven Springs	11	NA	0%	0%	Tolerant	
TV	Tanque Verde Canyon	10	NA	60%	60%	Susceptible	Presumed extirpated
WC	Willow Creek	9	No survival	10%	10%	Susceptible	

* some mortality found outside of the hot springs, in surrounding waterbodies

[1] Savage and Zamudio 2011

[2] Savage, Sredl, and Zamudio 2011

[3] Savage and Zamudio 2016

4.2.1 - Field sampling

We conducted visual encounter surveys at nine perennial stream localities in Arizona, USA from June 2006 through January 2011, sampling twice per year during summer and winter months (detailed in Savage et al. 2011; Savage and Zamudio 2016). We re-surveyed each of these localities in December and January of 2015-2016, December and January of 2016-2017, and in July 2017. Both sampling periods are beyond the initial arrival of Bd in the system which has been present in *Rana yavapaiensis* since at least the early 1990s (Bradley et al. 2002). The extent and location of flowing water varied across sampling events, thus we always covered the same total transect of potential flow to standardize sampling effort. We also continued to survey locations through 2017 even when no frogs were observed for several consecutive sampling events. During each survey, we hand-captured as many metamorphosed *R. yavapaiensis* individuals as possible, including juvenile and adult frogs. We collected a toe clip from the second digit of the left hand of each frog and immediately

placed it in 100% ethanol for DNA preservation. Individuals were then released at the site of capture. We also obtained *R. yavapaiensis* toe clips from two other localities sampled by Arizona Game and Fish Department staff in 2011 (M. Sredl, pers. comm.) for a total of 140 individual *R. yavapaiensis* from 11 localities, and included one *R. chiricahuensis* sample (M. Sredl, pers. comm.) as an outgroup species for phylogenetic analysis.

4.2.2 - Sequence array design

We used a previously generated reference transcriptome (Savage et al. 2020) to design a species-specific sequence array. To identify Single Nucleotide Polymorphisms (SNPs) that were variable between or within populations of *R. yavapaiensis*, we mapped RNAseq sequences from skin and spleen tissues from 12 field-sampled individuals across the species range (Christodoulides et al. 2020), back to the reference to identify SNPs and used BaitsTools v1.3.0 (Campana 2018) to extract the sequences. We used a multi-pronged approach to subset the genes to select both genome-wide markers in addition to putative Bd-associated genes in our array. The full array of 1,388 sequences had a mean length of 386 bps (range: 92-1,716 bps).

4.2.2.1 - Putative Bd-resistance markers

We identified 807 loci that were putative candidate genes for disease resistance and were included in the array. The majority of sequences (746) were derived from the *Rana yavapaiensis* reference transcriptome (Savage et al. 2020). We focused on transcripts from both spleen and skin tissue that were found to be differentially expressed between *R. yavapaiensis* individuals that differ in Bd susceptibility (Savage et al. 2020). In that study, multiple egg-masses from the Muleshoe Ranch (MR) population were lab-raised and juvenile frogs were inoculated with Bd to identify genes differentially expressed at early and late infection stages between frogs showing different levels of susceptibility. This included 514 transcripts that were differentially expressed in both tissues, as well as 95 transcripts that were differentially expressed in one tissue only, but have a putative immune function. Additional transcripts were added based on a literature review of studies on Bd-host dynamics for a total of 746 transcripts. We also included 13 toll-like receptors (TLRs), 37 antimicrobial peptide sequences (AMPs), and both class I and II major histocompatibility complex (MHC) genes (Table D1).

4.2.2.2 - Genome-wide markers

We included an additional 579 loci by running the reference transcriptome through the Marker Development pipeline (available here: https://github.com/CGRL-QB3-UCBerkeley/MarkerDevelopmentPylogenomics; Portik, Smith, and Bi 2016). Orthologs were identified using the *Nanorana parkeri*, *Xenopus laevis* and *Anolis carolinensis* genomes (Sun et al. 2015; Hellsten et al. 2010; Alföldi et al. 2011). We also included a total of 500 bps of mitochondrial sequence for Cytochrome B (CytB) and 500 bps cytochrome oxidase subunit 1 (COI) to reconstruct the maternal lineage (Table D1).

Following a quality control pipeline by Arbor Biosciences (Ann Arbor, MI, USA) to remove repetitive sequences with multiple hits to the *Rana catesbeiana* genome (Hammond et al. 2017), a total of 20,000 tiled baits of 80 bp each were designed across these sequences and ordered by means of a MyBaits Custom 20K kit (Arbor Biosciences, Ann Arbor, MI, USA). Tilling density was dependent on the importance and complexity of the loci (e.g., higher tiling density on the variable copy number MHC loci) and tiling density on the mitochondrial sequences was reduced as mitochondrial molecules are at higher concentrations in most DNA extractions.

4.2.3 - Laboratory methods

4.2.3.1 - Genomic library preparation

We extracted genomic DNA using protein precipitation and eluted all samples in 100ul of EB buffer. Following quantification with the Qubit 2.0 dsDNA HS Assay kit (Invitrogen, Carlsbad, California, USA), we sheared up to 3,000 ng of DNA to a mean size range of around 300 bps using the Q800R Sonicator (Qsonica, Newtown, CT, USA). We prepared DNA libraries from sheared DNA using SureSelect Library Prep Kit (Agilent Technologies, Santa Clara, CA, USA), but using in-house double-indexed 8 bp Nextera-style adapters (Glenn et al. 2019). We amplified libraries in two separate PCR reactions of 50ul of 12-15 cycles using Herculase II Fusion DNA Polymerase (Agilent Technologies, Santa Clara, CA, USA). All reactions and intermediate clean up steps were performed on the Apollo 324 (WaferGen Biosystems, Fremont, CA, USA) in batches of 24 samples. We pooled samples equimolarly in groups of six aiming for a total input of 3,000 ng per sequence capture reaction. A subset of pools and

samples were run on the TapeStation 2100 (Agilent Technologies, Santa Clara, CA, USA) at several stages of the protocol as a quality control using the High Sensitivity Assay.

4.2.3.2 - Sequence capture and sequencing

Pooled libraries and baits were hybridized for 36 hours following the MyBaits v3 protocol (https://arborbiosci.com/wp-content/uploads/2017/10/MYbaits-manual-v3.pdf) but increasing the amount of c0t-1 blocker to 2X. Following stringent washes, we re-amplified the pools in two separate PCR reactions for 9-13 cycles. A final pool was prepared for paired-end 150 bp sequencing on part of two Illumina NovaSeq SP runs at the Oklahoma Medical Research Foundation genomics facility. We pooled samples equimolarly, sourcing DNA from both PCR reactions but with a preference for the reaction with the fewest number of cycles to reduce the number of PCR duplicates in sequencing.

4.2.3.3 - Bd quantification

We analysed Bd infection data for all individuals sampled in winter to ensure that infection status and infection intensity were comparable and relevant to disease outbreaks. We utilized previously generated Bd infection data (Savage, Sredl, and Zamudio 2011; Savage, Becker, and Zamudio 2015) obtained via qPCR, where all values represent averages of at least two replicates and only infection intensities > 1 genome equivalent were considered true positives. Zoospore counts were used as proxy for Bd-intensity for downstream analyses.

4.2.4 - Bioinformatic processing

Because the bait loci were designed from transcriptome cDNA sequences, they did not include introns and thus did not represent an accurate genomic reference for further processing. Thus, to split up the potentially spliced exons in our baits, we applied the IEB-finder pipeline that identifies intron-exon boundaries by means of mapping scores (Deleury et al. 2019). In short, when gDNA reads are mapped against the cDNA reference using the local mapper bwa-mem (Li and Durbin 2009), the parts of the read that represent the intron will be soft-trimmed. IEB-finder scans a bam file to identify regions that have above average soft-trimming compared to the surrounding region and identifies them as putative exon-intron boundaries (Deleury et al.

2019). We used a representative pool of 20 samples to run IEB-finder pipeline (parameters -e 0 -c 10 -x 30) and split up our loci into separate exons for all identified boundaries.

We ran the new exon-based reference through the SECAPR pipeline (Andermann et al. 2018), using the same 20 samples to identify potential paralogs and duplicate loci. In short, reads were quality filtered using default SECAPR parameters and assembled individually using abyss (-k 90), and the resulting assemblies were compared to the reference by means of reciprocal blast using LASTZ (--min-coverage 80, --min-identity 80). We manually examined loci that were found to either contain potential paralogs (multiple contigs hitting the same locus), or duplicate loci (one contig hitting two loci). Duplicate loci typically resulted from short introns/indels that caused IEB-finder to split the locus in two, but in which flanking parts of the sequence reads were long enough to bridge this gap and form one assembled contig. We picked the resulting locus that included the intron to update the reference. We examined all putative paralogs to determine if it was possible to split them into two loci with high confidence. If there was a clear distinction we extracted two separate loci for the updated references. When the putative paralogs were too similar to separate them, we removed the loci from further analyses.

We ran TransDecoder 5.5.0 (Haas and Papanicolaou 2012) on the updated reference to find open reading frames (ORFs) and identify the coding sequence regions (CDS) in our loci, applying 50 consecutive amino acids as our minimum length. The SnpEff pipeline was applied to identify missense (non-synonymous changes resulting in an amino-acid change) and synonymous SNPs across the CDS regions (Cingolani et al. 2012). We annotated the loci by blasting the reference sequence to the *Rana catesbeiana* genome, applying a minimum blast score of 200 using blast 2.7.1+ (Camacho et al. 2009). Loci with no high-quality hits were subsequently blasted against the *Nanorana parkeri* genome (minimum score 200), the non-redundant nucleotide database (minimum score 65) and finally the translated CDS regions were run against the non-redundant protein database using blastp (minimum score 35) to identify additional annotations. To determine the putative locations of the loci along the genome we used the non-annotated reference genome of the closely related *Rana temporaria* to scaffold our loci (unpublished; Daniel Jeffries 2020). We used the chromosome scaffolder RaGOO (Alonge et al. 2019) to identify homologous areas adjusting the minimum overlap length to 300 bp due to small length of our loci, but increasing the grouping confidence score to 0.5 to reduce spurious mapping. Nearly 75% of the current *Rana temporaria* genome assembly is localised to 13 identified chromosomes. We combined the remaining 25% of scaffolds into a pseudo-chromosome 14 and grouped all loci without a hit in pseudo-

chromosome 15. To estimate average distance between loci on both pseudo-chromosomes for plotting purposes we used the average length between loci on the 13 correct chromosomes.

To genotype our individuals, we ran the updated reference through the SECAPR pipeline again, using all 141 individuals but increasing the LASTZ threshold (--min-coverage 90, --min-identity 90). Following the remapping step, we removed all PCR duplicates and ran GATK 3.8.1 (McKenna et al. 2010) across all bam files in order to call high quality SNPs using information from all samples. We used the EMIT-ALL-SITES option in order to keep non-variable sites. This combined strategy allowed us to include all available evidence to determine SNP quality, and additionally identify loci with a heterozygosity excess that were likely paralogs, while still keeping non-variable sites for phylogenetic analyses. Following strict filtering of low-quality (allele balance under 0.25 or over 0.75), low-coverage SNPs (DP < 20), indels, and paralogous loci (ExcessHet > 2.0) alignments were extracted from the vcf file with vcf2phylip (Ortiz 2019) allowing for a maximum of 50% missing data across each site. We concatenated all nuclear loci and analysed the two mitochondrial loci separately. Two SNP datasets were extracted for further analyses using vcftools 0.1.16 (Danecek et al. 2011). An initial 'All-SNPs' dataset included all high-quality SNPs (--min-alleles 2 --max-alleles 2 --minDP 10), and a smaller dataset of putatively 'unlinked SNPs' by removing all but one SNP per locus, keeping the SNP with the least amount of missing data. Individuals with more than 75% missing data across the unlinked SNP dataset were removed from all subsequent analyses.

4.2.5 - Population genetic analyses

4.2.5.1 - Phylogenetic reconstruction

A maximum likelihood tree was constructed using RAxML 8.2.12 (Stamatakis 2014) applying the GTRCAT substitution model on a concatenated alignment of all nuclear loci, requiring a minimum read depth of 5. Analyses were started from 10 distinct starting trees, and bootstrap support was computed on the best scoring tree by means of 100 iterations of rapid bootstrapping (Stamatakis, Hoover, and Rougemont 2008).

4.2.5.2 - Population structure

To identify genetic clusters, we ran Admixture 1.3.0 (Alexander and Lange 2011) from K 1 – 15 and used the cross-validation method to choose the best K. We also ran a principal component analyses (PCA) in R 3.6.3 using the snpgdsPCA function of the SNPRelate 1.20.1 package (Zheng et al. 2012) to visualize the relationships between the different populations. Both analyses were applied to the unlinked SNP dataset to remove potential issues with linkage. All PCA analyses were run using an increased missing data filter of 50% to remove individuals with high levels of missing data (final number of samples after filtering totalling 110). Additional PCA's were applied to six different SNP data subsets (missense only, synonymous only, putative Bd-related markers, non Bd-related markers, TLRs & AMPs), and a heatmap of Nei's distance was generated using the stamppNeisD function of the R package StAMPP 1.61 (Pembleton, Cogan, and Forster 2013).

4.2.5.3 - Mitochondrial haplotyping

We extracted both mitochondrial loci from the dataset and created a consensus sequence for each individual requiring a minimum depth of 12 unique reads to call a base. The loci were concatenated to produce a single alignment and we built a haplotype network using TCS 1.21 (Clement, Posada, and Crandall 2000), and created the network figure with tcsBU (Múrias Dos Santos et al. 2015).

4.2.6 - Measures of genetic diversity and population differentiation

Heterozygosity and allelic richness were calculated from the unlinked SNP dataset and for each population using the basicStats function from R package diversity 1.9.90 (Keenan et al. 2013). We additionally counted the number of private alleles (alleles unique to a population), using the private_alleles function from poppr on all missense SNPs to look at potential functional differences between populations. We calculated pairwise F_{ST} values with the stamppFst function of the StAMPP R package 1.6.1 (Pembleton, Cogan, and Forster 2013), and applied a Mantel test to evaluate isolation by distance. Genetic diversity analyses were only applied to populations with a minimum of nine individuals, thus excluding Dix Creek and Two Mile Tank.

4.2.7 - Signatures of selection

4.2.7.1 - pN/pS analyses

The ratio of synonymous or silent (pN) versus non-synonymous or missense (pS) polymorphisms in gene coding regions can be indicative of the type of selection the gene has experienced. Positive selection on functionally different proteins can increase the ratio of non-synonymous polymorphisms, whereas purifying selection will select against them and reduce this ratio. When looking at within species diversity this ratio is called pN/pS rather than dN/dS, as it does not assume fixed substitutions between sequences but allows for segregating SNPs (Nei and Gojoborit 1986). We calculated pN/pS for every open reading frame larger than 50 amino acids across all loci and individuals using POGENOM (Sjoqvist et al. 2020), and considered all loci with a value higher than 1 to be undergoing positive selection.

4.2.7.2 - Tajima's D

We calculated Tajima's D across all loci using vcftools (Danecek et al. 2011) to quantify pairwise differences relative to the number of segregating sites. Tajima's D is expected to be zero for a neutrally evolving population of constant size. A negative Tajima's D occurs when rare alleles are more abundant than expected and this is indicative of either a population expansion or purifying selection, whereas a positive Tajima's D occurs when rare alleles are scarce and this can be the result of population declines or balancing selection. We calculated the mean Tajima's D for all loci as an indicator of the population size effect and considered loci that were one standard deviation below the mean to be under purifying selection, and one standard deviation above the mean to be under balancing selection.

4.2.7.3 - F_{ST} Outliers

We identified F_{ST} outliers across all populations using BayeScan 2.1 which works on population allele frequencies and corrects significance levels for multiple comparisons (Foll and Gaggiotti 2008). BayeScan was implemented using default parameters and we checked for convergence using the R package coda and Heidelberger and Welch's convergence diagnostic (Plummer et al. 2006). We applied a q-value of 0.05 as our significance threshold,

and used the alpha score to identify balancing selection (alpha < 0) or positive/diversifying selection (alpha > 0).

4.2.7.4 - Genes with multiple lines of evidence

To assess overlap among loci classified as undergoing positive selection based on the three distinct criteria described above (pN/pS, Tajima's D, and F_{ST} outliers), we visualized overlap using a Venn Diagram. We included all loci with a pN/pS over 1, a Tajima's D that was more than one standard deviation over the mean, and all BayeScan loci with a q-value below 0.05 (combining positive and negative alpha values). We combined exons from different genes and used the R package Venn.diagram 1.6.20 (Chen and Boutros 2011) to produce a figure depicting the overlap.

4.2.8 - Association with Bd-intensity

Initial F_{ST} outlier analyses between pairs of susceptible and tolerant populations were complicated by the high F_{ST} values between populations. BayeScan identified many population-specific fixed SNPs as highly significant outliers. Although we cannot refute that these SNPs are related to Bd-resistance, these patterns might also result from genetic drift. To overcome this problem, we grouped samples by individual Bd-intensity as quantified by qPCR. As Bd-intensity fluctuates within populations, these groupings included samples from different populations, thus removing the effects of strong population structure on the BayeScan analyses. We only used samples collected during winter sampling sessions in which Bd was detected in the population (6 populations, 52 individuals), and grouped samples in low, medium and high standardized Bd zoospore counts (<1,000, 1,000 – 10,000, 10,000>), groupings that tracked natural breakpoints in the distribution of all Bd intensity values. We ran BayeScan using default settings and checked for convergence using the Heidelberger and Welch's convergence diagnostic.

We ran the same dataset using SamBada 0.8.3 (Stucki et al. 2017) to include individual Bd-intensity as the environmental variable, using the R package R.SamBada (Duruz et al. 2019) to find associations between Bd-intensity and individual genetic variation. SamBada identifies candidate genes using genotype-environment associations while correcting for population structure and demographic effects. In addition to including Bd-intensity directly, SamBada runs

on individual genotypes and not allele frequencies and thus can detect associations with specific genotypes. We ran SamBada using a minimum allele filter of 0.05 and allowing only 10% missing data.

4.3 - Results

We included 140 individuals sampled from 2007-2017 from localities Willow Creek, Secret Spring, Santa Maria River, Muleshoe Ranch, and Aravaipa Canyon (WC, SS, SM, MR, AC), and from 2009-2017 for Cienega Creek (CIC) and Hassayampa River (HR). For localities Tanque Verde (TV) and Aliso Spring (AS), we only recovered frogs from winter 2007/2008 through winter 2009/2010. Despite repeated annual surveys through 2018, no frogs were observed at either of these localities after January 2010. Because these areas were surveyed throughout the entire range of potential habitat, the lack of observed frogs in summer and winter months for seven consecutive years suggests that Tanque Verde and Aliso Spring are now extirpated. Finally, we included a small number of individuals from localities Two Mile Tank (MT) and Dix Creek (DC) that were sampled in 2011 because they represented distinct population localities and were potentially important for assessing phylogeographic patterns. We focused on individuals sampled in winter months whenever possible because Bd susceptibility can only be assessed during winter months when disease outbreaks occur in Arizona (Savage, Sredl, and Zamudio 2011).

Illumina sequencing resulted in 311,444,510 raw paired-end reads, and after quality filtering 276,115,139 reads were used as input for our pipeline. A total of 133 samples and one outgroup individual (*R. chiricahuensis*) had sufficient coverage and were included for further processing. The final alignment was 884,405 bps, the 'all SNPs' dataset included 10,295 SNPs which was filtered down to 2,248 SNPs for the 'unlinked SNPs' dataset.

4.3.1 - Genetic structure

The mitochondrial alignment was a total of 2,191 bps and did not include any indels. Mitochondrial haplotype diversity was very low with only eight different alleles detected across all *R. yavapaiensis* individuals. Haplotype-network analyses (Figure 4.2B), indicated that there was little mitochondrial structuring across populations. The most mitochondrially diverse populations were Willow Creek (WC) and Hassayampa River (HR) that each had three haplotypes and are both located in the northern part of the sampled range. All other populations had only one or two haplotypes.

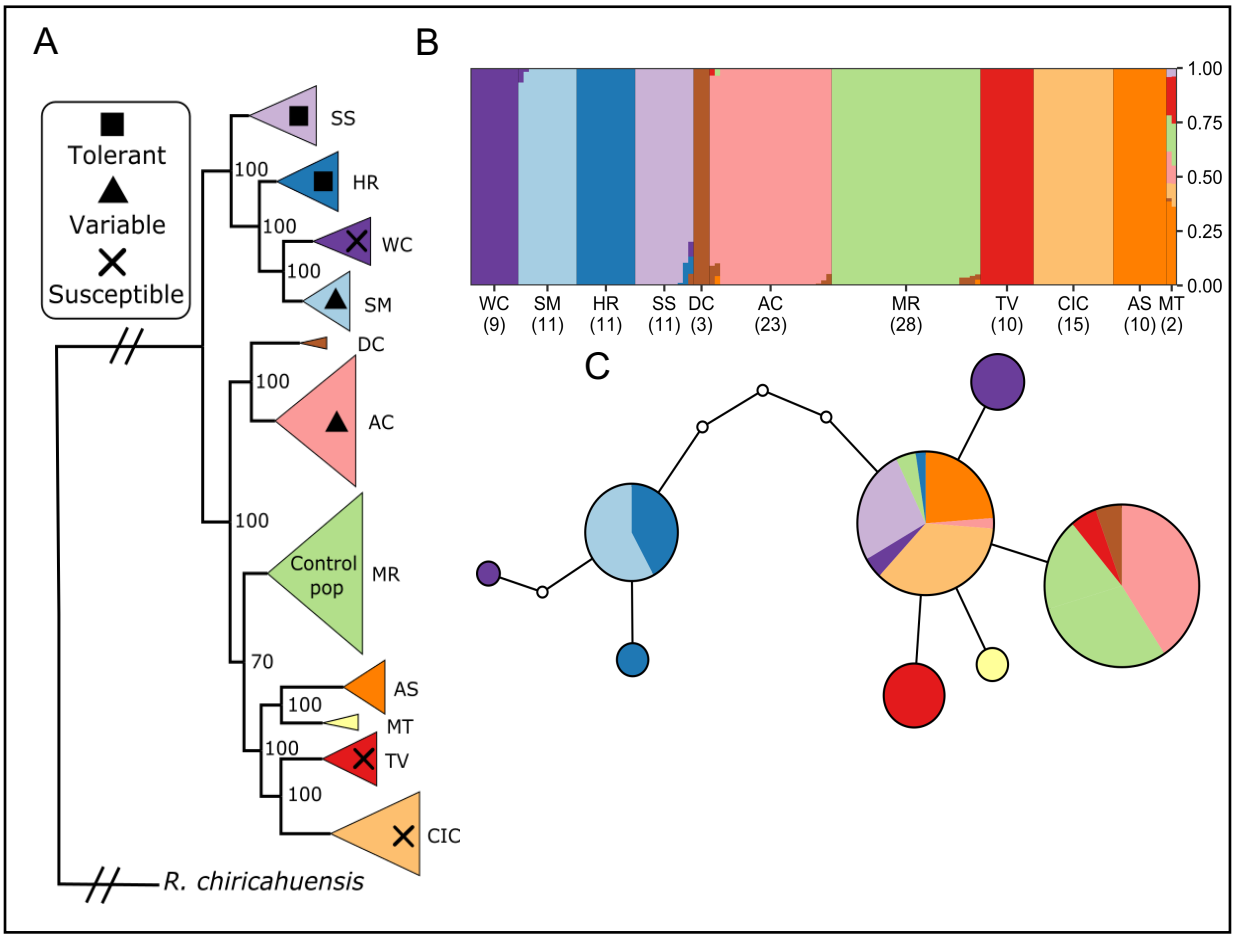
The nuclear maximum likelihood phylogenetic reconstruction based on the total alignment separated all populations into fully-supported, separate clades (Figure 4.2A), and with the exception of the placement of Muleshoe Ranch (MR; bootstrap support of 70), all nodes were fully supported. The deepest divergence was between the four populations found in the north-western part of Arizona (HR, SM, SS and WC) and the southeastern populations (AC, AS, CIC, MR, MT, DC and TV), and populations generally grouped according to their geographic proximity. Interestingly, disease susceptibility varied across the tree and was not associated strongly with any clade. The PCA also showed the same split between northern and southern populations along the first axis and grouped samples by population, generally following geographic proximity with the exception of Aliso Spring (AS) and Two Mile Tank (MT; Figure 4.1B).

Admixture cross validation indicated that K=10 best explained the genetic structure found in the data (Figure 4.2C). All localities formed their own cluster, with the exception of Two Mile Tank which showed mixed ancestry, likely due to the low number of samples from this locality (n=2) which is a known problem for clustering programs (Fogelqvist et al. 2010). The strong population structure found by Admixture was reinforced by the high F_{ST} values found between localities (Table 4.2; mean F_{ST} 0.44), indicating that gene flow between populations is currently low to non-existent. There was a significant pattern of isolation by distance (IBD; $p < 0.01$, see Figure D1). A heatmap of Nei's distance per sample (Figure D2) also followed the same pattern of differentiation as found using phylogenetic and structure-based analyses.

Table 4.2 Table of F_{ST} values by population as calculated with the StAMPP R package in the lower diagonal, and geographic distance in kilometres in the upper triangle. The four top left populations are all found in the north east and the remaining five populations are in the south east. The amount of divergence is visually highlighted by grayscale in steps of 0.10 for genetic distance and 100 km for geographic distance.

	WC	SM	SS	HR	AC	MR	TV	CIC	AS
WC	-	95	201	157	382	427	417	443	457
SM	0.27	-	131	67	308	347	332	357	365
SS	0.35	0.31	-	78	181	226	220	247	273
HR	0.32	0.25	0.26	-	244	282	266	290	300
AC	0.49	0.46	0.39	0.43	-	51	73	99	158
MR	0.50	0.48	0.39	0.45	0.27	-	42	58	124
TV	0.58	0.55	0.46	0.52	0.36	0.35	-	29	86
CIC	0.58	0.56	0.48	0.53	0.38	0.36	0.39	-	66
AS	0.64	0.59	0.52	0.57	0.45	0.46	0.54	0.52	-

Figure 4.2. (a) Maximum likelihood reconstruction of 884,405 bps concatenated alignment as implemented by RaxML, chytridiomycosis susceptibility indicated by symbol if known (b) Admixture plot for 2,248 SNPs and K = 10. All colours match both locality and phylogenetic placement with the exception of Two Mile Tank which shows mixed ancestry. Number of individuals per population indicated below. (c) Mitochondrial haplotype network based on 2,191 bps of sequence from a concatenated COI and CytB alignment.

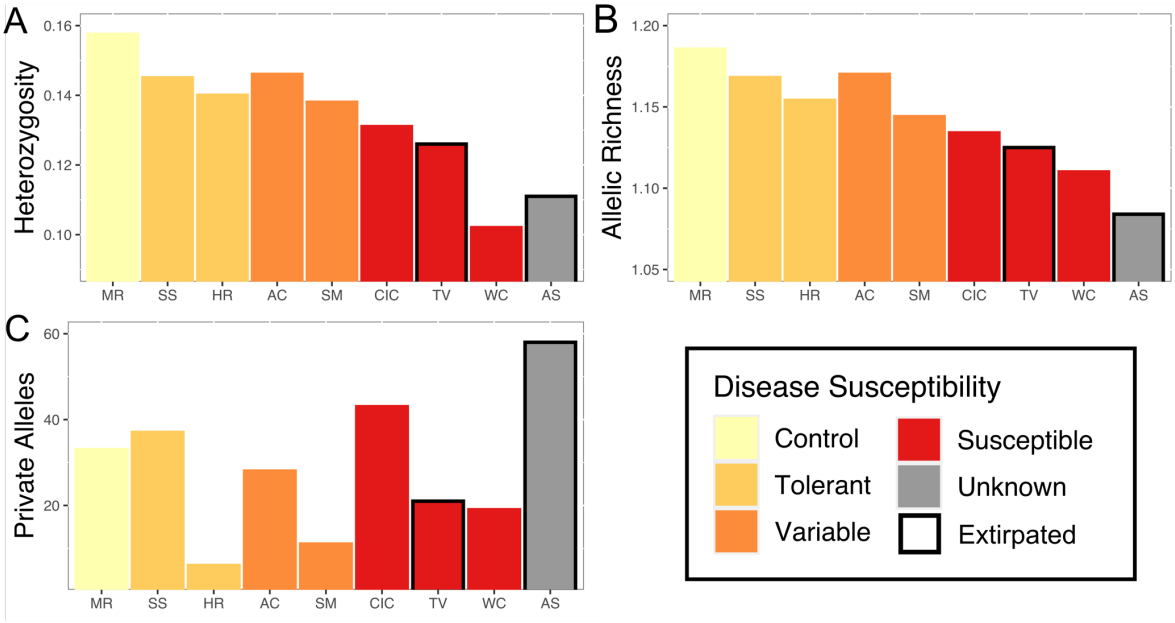


4.3.2 - Measures of genetic diversity

Two measures of genetic diversity (heterozygosity and allelic richness) as calculated across all unlinked SNPs were different between populations and were consistent with our disease susceptibility classifications. Populations with the highest observed Bd susceptibility had the lowest estimates of both heterozygosity (Figure 4.3A) and allelic richness (Figure 4.3B). We do not have susceptibility data for the recently extirpated Aliso Spring population, but it groups with the high susceptibility populations in both measures of genetic diversity. Furthermore, both of the extirpated populations (Tanque Verde and Aliso Spring) had some of the lowest estimates of heterozygosity and allelic richness from our samples collected in the years immediately preceding their extirpations.

The number of functionally unique private alleles did not correspond with allelic richness or heterozygosity (Figure 4.3C). Most notably, the extirpated Aliso Spring population had the lowest allelic richness, but the highest number of private alleles. This held true for both missense SNPs, as well as the complete unlinked SNP dataset (Figure D3). A similar pattern was evident in the different spatial placement of Aliso Spring in the PCAs based on synonymous versus missense SNPs (Figure D4). Overall, the southeastern populations (mean = 41, range 30-64) had more private alleles than the northwestern populations (mean = 13, range 7-22).

Figure 4.3. Three measures of genetic diversity calculated by population and coloured by disease susceptibility (a) Heterozygosity based on all 2,248 unlinked SNPs. (b) Allelic richness based on all 2,248 unlinked SNPs (c) Number of private alleles based on 3,237 missense SNPs, as a measure of uniqueness in functional diversity.



4.3.3 - Signatures of selection

4.3.3.1 - pN/pS analyses

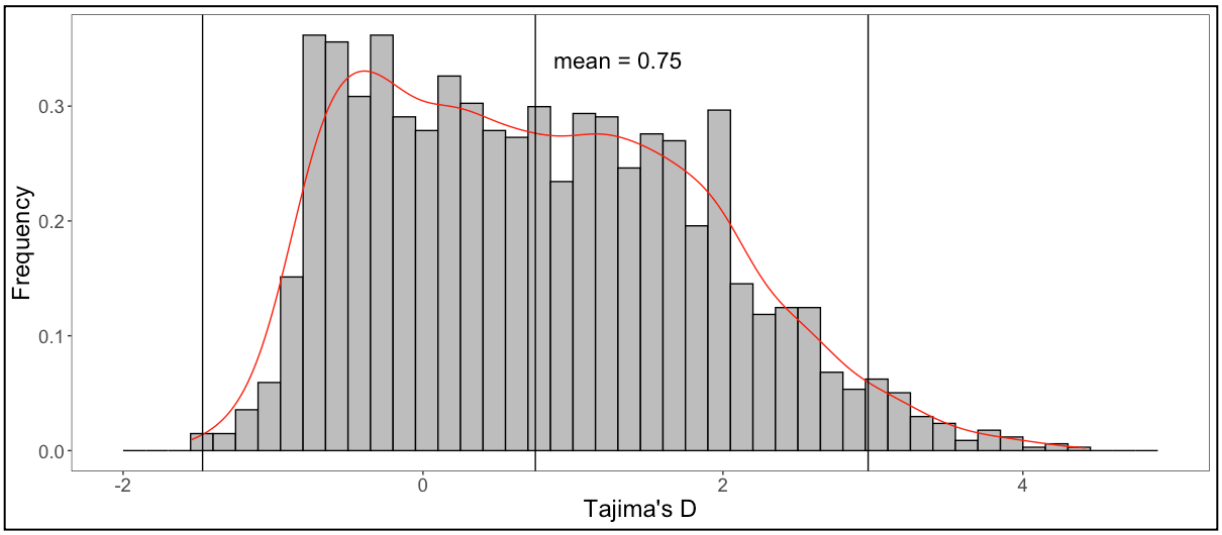
Analyses of within species non-synonymous vs synonymous polymorphisms (pN/pS) detected patterns of purifying and positive selection in our dataset. Of the 874 loci for which we could calculate pN/pS, 780 had a pN/pS below 1, and 94 had a signature of positive/diversifying selection (pN/pS > 1). The top ten genes with the highest pN/pS estimates (all > 1.9) included

five immune system genes (Table 4.3). Nine out of the top ten genes were also found to be differentially expressed in Bd-infection trials (Savage et al. 2020)

4.3.3.2 - Tajima's D

Average Tajima's D across the genome was 0.75 with a short tail on the negative end of the distribution and a longer tail for the positive Tajima's D values (Figure 4.4). Two genes had a Tajima's D that was one standard deviation below the mean, 67 genes were greater than one standard deviation above the mean, and we considered these loci to be potentially under directional and balancing selection, respectively. The top 13 genes with the most extreme Tajima's D values were all on the positive end of the distribution, and included five immune system genes and two others with antimicrobial function (Table 4.3).

Figure 4.4. Histogram of Tajima's D values calculated across all loci. The mean of all values is 0.75, likely indicative of the population decline the species experienced. One standard deviation of the mean on both sides indicated by vertical line at -1.47 and 2.90 and are loci that could be under purifying or balancing selection respectively.



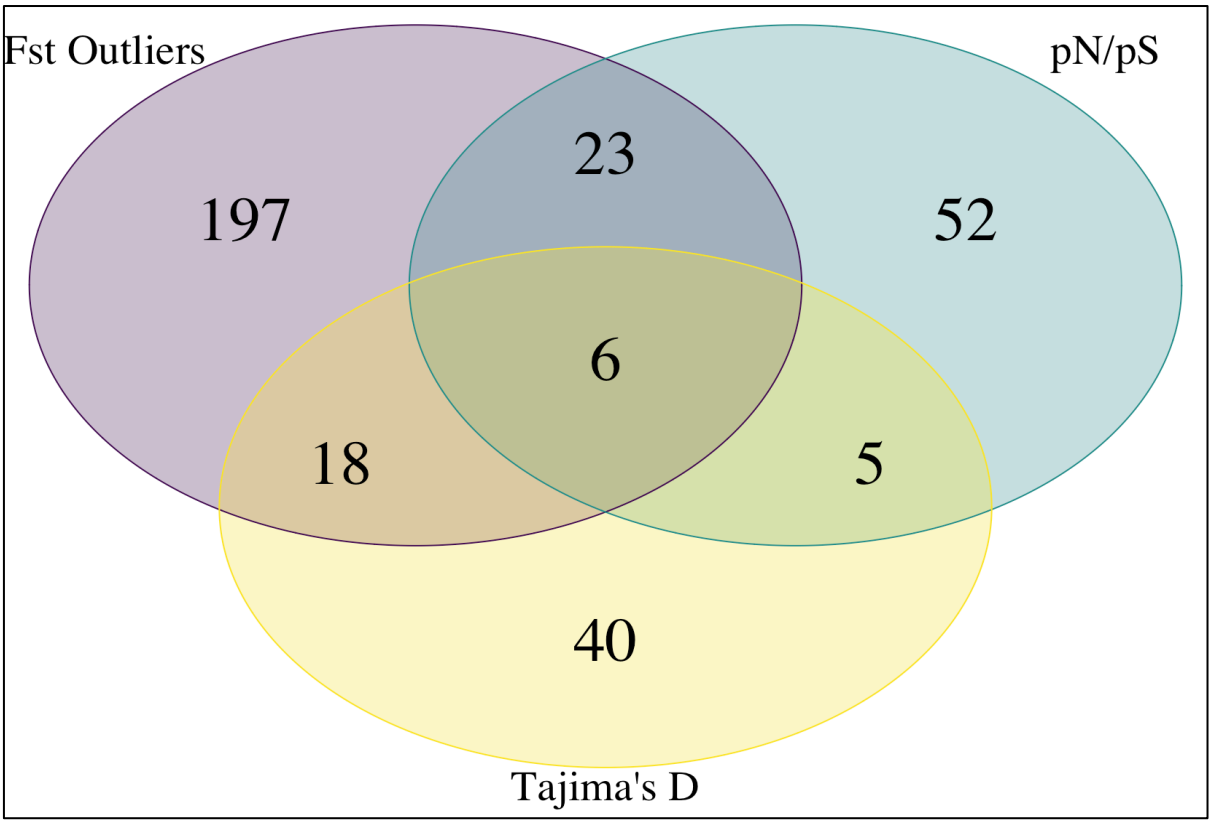
4.3.3.3 - F_{ST} Outliers

BayeScan analyses of F_{ST} outliers across all populations identified 230 SNPs found within 124 genes that had a significant ($q < 0.05$) signature of positive selection, and 444 SNPs within 125 genes with a significant signature of balancing selection. Five genes included SNPs with both balancing and positive selection, thus a total of 244 unique genes contained significant F_{ST} outliers.

4.3.3.4 - Genes with multiple lines of evidence

There was moderate overlap in the genes that were found to be under selection via the three tests of selection we implemented. Six genes were found to be under selection using all three methods, including two genes related to ribosomal function (EF1G and NACA), two genes that are potentially important for amphibian skin integrity (S10A4 and QNR-71), and two immune genes: CD1B4, which is a T-cell surface glycoprotein important for antigen recognition, and a lambda chain-like immunoglobin (LV151) involved in antigen recognition.

Figure 4.5. Venn Diagram of the genes that overlapped between the three methods used to identify signatures of selection (F_{ST} outliers with a q-value below 0.05, pN/pS ratio above 1, Tajima's D more than one standard deviation away from the mean).



3.4 - Genomic associations with Bd

BayeScan analyses between the low, medium and high Bd intensity groupings found two SNPs that were significant F_{ST} outliers with a q-value below 0.05 and with positive alpha values, indicating a signal of positive selection (Figure 4.6A; Table 4.3). The top gene, Interleukin-10 (IL10; q-value: 0.036), was nearly fixed for the alternate allele in the medium

group, when compared to the both other groups (Figure 4.6C). The second gene (q-value: 0.048) was P53 and was fixed for one allele in the medium and high group, but variable in the low group.

In addition to allele frequency differences between the three different Bd-intensity groupings as found by BayeScan, we also found associations between genotypes and individual Bd-intensity using SamBada (Figure 4.6B; Table 4.3). Out of the top three loci with a p-value under 0.0001, two largely followed population structure, but one missense SNP in an unannotated gene (Raya_425) was variable across most populations and showed a strong association with Bd intensity with lower values for heterozygotes (median 112, mean 587) compared to homozygotes (median 2,301, mean 96,418), see Figure 4.6D.

Figure 4.6. (a). Manhattan plot of F_{ST} outliers along the genome as identified by BayeScan when populations are grouped by Bd intensity with on the y-axis the log of the q-value of significance. (a) A Manhattan plot of the log of the p-value for associations with Bd-intensity from SamBada (c) Barplot of the allele frequencies of Interleukin 10 (IL10) by Bd-intensity group. (d) Violin plot of the log-scaled Bd-intensity per individual, split by homozygote or heterozygote for locus Raya_425.

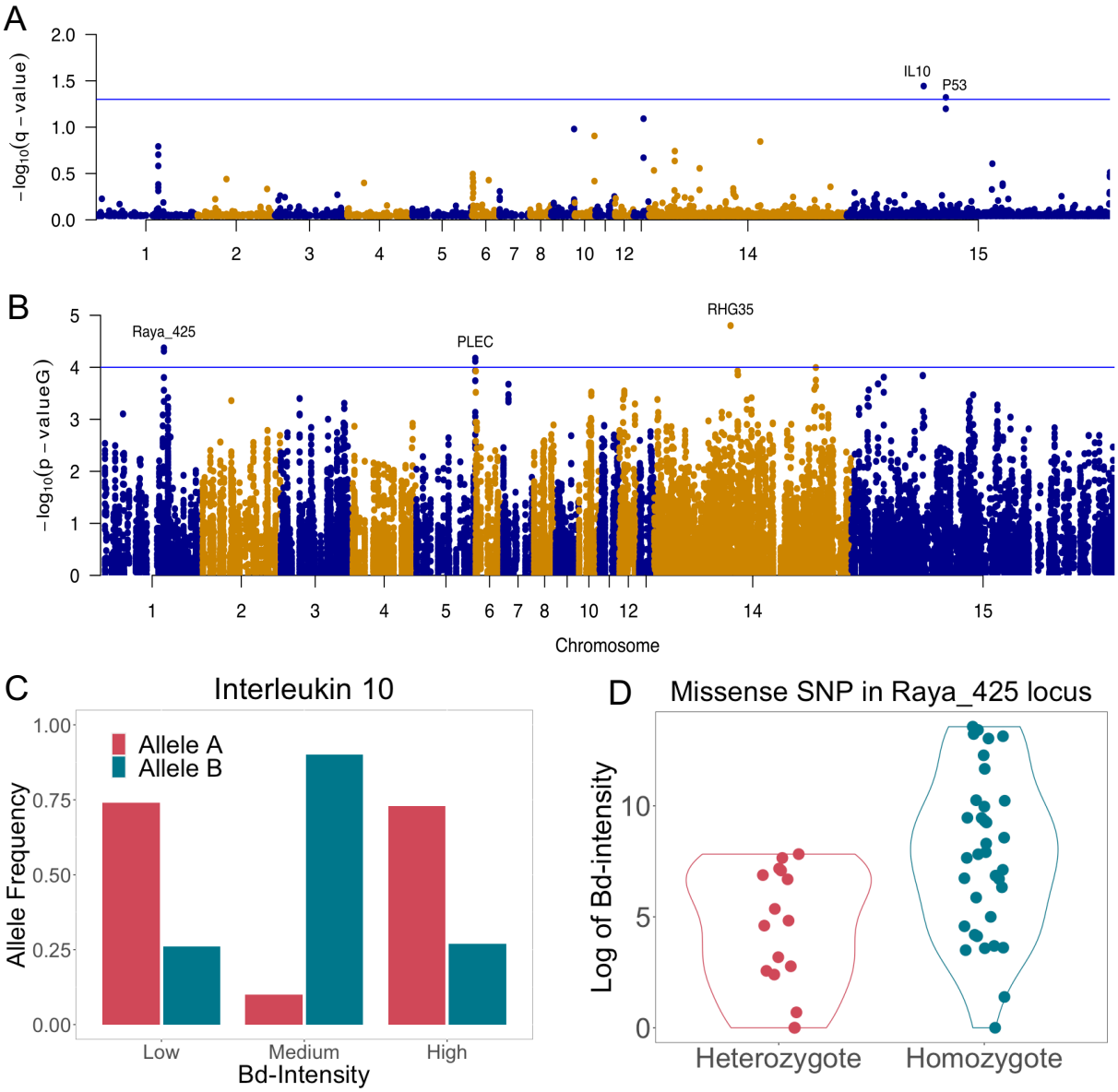


Table 4.3. Overview of the top genes with signatures of selection. Table includes the top 13 loci as ranked by Tajima's D values, the top 10 CDS regions as ranked by pN/pS values and six genes that included signatures of selection for Tajima's D, pN/pS and were also identified as significant F_{ST} outliers. The final five loci were identified as associated with Bd-intensity by either BayeScan or SamBada. Alpha values were included if the q-value was < 0.10. Annotations are based on *Rana catesbeiana* unless noted otherwise: (a) *Nanorana parkeri* genome, (b) non-redundant nucleotide database, (c) non-redundant protein database.

Gene	Blast-score	Tajima's D	pN/pS	BayeScan a	Description
Top loci ranked by Tajima's D					
OXLA	1051	**4.41**	0.06	0.99	L-amino-acid oxidase
OXLA	1225	**4.20**	0.14	NA	L-amino-acid oxidase
Raya_3172	250	**4.16**	0.80	-0.82	Hypothetical protein (*Rana catesbeiana*)
MALL	981	**4.01**	NA	NA	MAL-like protein
LSM3	970	**4.00**	1.90	NA	U6 snRNA-associated Sm-like protein LSm3
MDHM	719	**4.00**	0.00	NA	Malate dehydrogenase, mitochondrial
MK67I	40 (c)	**3.98**	0.00	NA	FHA domain-interacting nucleolar phosphoprotein
BTNL8	128 (b)	**3.89**	0.70	NA	Butyrophilin-like protein 8
PHF1	1061	**3.81**	0.00	NA	PHD finger protein 1
CFAB	1513	**3.80**	NA	NA	Complement factor B
IRF	1048	**3.77**	NA	NA	Interferon regulatory factor
CFAB	706	**3.74**	NA	NA	Complement factor B
IL17F	1098	**3.71**	NA	NA	Interleukin-17F
Top loci tanked by pN/pS values					
PILRA	603	-0.94	**3.40**	NA	Paired immunoglobulin-like type 2 receptor a
NDUA4	66 (b)	-0.66	**3.02**	NA	Cytochrome c oxidase subunit
IFITM1	608	-0.37	**2.97**	NA	Interferon-induced transmembrane protein domain-containing protein
PCOC1	1125	2.74	**2.83**	NA	Procollagen C-endopeptidase enhancer 1
IIGP5	719	0.21	**2.30**	NA	Interferon-inducible GTPase 5
K1C42	1362	1.04	**2.24**	NA	Keratin, type I cytoskeletal 42
FRIH3	446	-0.24	**2.23**	NA	Ferritin heavy chain, oocyte isoform
PGS2	592 (a)	1.22	**2.22**	NA	Decorin
S10A4	505	3.36	**2.06**	1.58	Protein S100-A4
IIGP5	684	0.50	**1.97**	-2.57	Interferon-inducible GTPase 5
Loci overlapping between Tajima's D, pN/pS and F_{ST} outliers					
CD1B4	1112	**3.15**	**1.80**	**-1.01 & 1.32**	T-cell surface glycoprotein CD1b4
EF1G	798	**2.98**	**1.04**	**-0.96**	Elongation factor 1-gamma
LV151	717	**2.99**	**1.86**	**-1.91**	Ig lambda chain V-I region BL2
NACA	673	**3.29**	**1.32**	**-2.34**	Nascent polypeptide-associated complex subunit a
QNR-71	492	**3.22**	**1.28**	**-2.14**	Protein QNR-71
S10A4	861	**3.36**	**1.48**	**1.58**	Protein S100-A4
Loci associated with Bd-Intensity					
IL10	1016	1.14	NA	1.66	Interleukin-10
P53	364	1.89	NA	1.61	Cellular tumor antigen p53
Raya_425	x	0.80	1.21	NA	x
RHG35	1315 (a)	1.04	0.00	1.59	Rho GTPase-activating protein 35
PLEC	3295	1.06	0.06	NA	Plectin

4.4 - Discussion

Sequence capture across 133 individuals and 1,388 genes produced a large genomic dataset that we used to answer questions about population genetics, evolutionary history and adaptation in *Rana yavapaiensis*. Although mitochondrial diversity was low and undifferentiated among the 11 populations sampled, nuclear genetic distances were high and largely correlated with geographic distance. The strong genetic isolation of all populations, together with the lack of association between Bd-susceptibility and phylogenetic placement, means that they may be independently adapting to Bd and other selective pressures. We detected signatures of purifying, balancing and positive selection across the genome and we highlight several candidate genes with important implications for how functional genetic variation is distributed across the species.

4.4.1 - Strong nuclear population structure and loss of genetic diversity

Most biogeographic studies of genetic variation find more structure in mitochondrial than nuclear data, which can occur as a result of introgression or male-biased dispersal (Toews and Brelsford 2012; Mulder et al. 2019). Our data suggests the opposite, with limited to no structuring in the mitochondrial haplotypes, but extreme structuring in nuclear SNPs. The lack of suitable habitat between the localities may explain the strong population structure we find and corresponds with previous work using 14 microsatellite loci (Savage, Becker, and Zamudio 2015). The lack of diversity and structure in the mitochondrial data suggests that this strong structuring occurred quite recently. Recent bottlenecks might also explain the rapid fixation of SNPs between populations (Nei, Maruyama, and Chakraborty 1975), although a caveat is that our choice of loci to include in the array and selecting of fragments with SNPs, might be biased towards more extreme allelic frequencies. Future demographic modelling with strong calibration points might be able to pinpoint the timing of these events. The largest phylogenetic split occurs between the four populations in the north-west and the six southeastern populations, and this split broadly corresponds to the two Arizona refugia that were identified previously using species distribution models for the last glacial maximum (Oláh-Hemmings et al. 2010). Disease susceptibility varied across the tree suggesting that the populations might be adapting independently on an intra-specific level (Richmond 2006).

Across the Arizona range of *R. yavapaiensis* there is evidence of multiple extirpations and range contractions (Witte et al. 2008), and we observed two populations become extirpated during our sampling efforts. Due to the projected increase of anthropogenic habitat destruction (Villarreal et al. 2013), coupled with invasive species such as crayfish and bullfrogs (Witte et al. 2008), and the impact of chytridiomycosis (Bradley et al. 2002; Savage, Sredl, and Zamudio 2011), extirpations of more populations are likely and will greatly reduce the functional genetic variation within the species. Current genetic diversity, as measured by both heterozygosity and allelic richness, is lower in disease susceptible populations, and both extirpated populations (AS and TV) had low genetic diversity before being extirpated (Figure 4.3). Without historical samples we cannot determine whether populations with lower genetic variation were more susceptible to Bd when the pathogen first appeared, or if Bd die-offs caused the observed reduction in genetic variation. Regardless, it is likely that multiple successive die-offs reduced population numbers and genetic diversity and set these populations on an extinction vortex (Fagan and Holmes 2006). Even though each of the populations had low diversity, their impact on genetic diversity for the species was large due to strong genetic differentiation between populations. This was evident in the number of functionally unique private alleles found in extirpated populations (Figure 4.3) that are likely no longer present in the species. Thus, protecting the remaining *R. yavapaiensis* populations is crucial for maintaining functional genetic diversity in this species. Lack of current migration between populations and the lack of re-colonization of both extirpated localities also suggests these extirpation events are likely final without human intervention. Given the important ecological impacts of amphibians in ecosystems (Whiles et al. 2006), the re-introduction of frogs can be considered as a potential last resort, but it will be crucial to consider Bd-susceptibility and how closely related potential source populations are in terms of functional genetic variation that may be adaptive for these localities.

4.4.2 - Multiple signatures of selection across the genome

Measures of molecular evolution including Tajima's D and pN/pS values show that many of the genes in our dataset have been under positive selection. Although Tajima's D is also strongly affected by population expansions and declines, this should mostly be reflected in the average value across all loci. The mean Tajima's D was 0.74, which would represent an overall signature of population bottlenecks, likely corresponding with the large die-offs that were observed across the state in recent decades (Schlaepfer et al. 2007; Clarkson and Rorabaug 1989). By contrast, the tails of the Tajima's D distribution (Figure 4.4) more likely represent

the effect of purifying and balancing selection acting on top of this signal of population decline. These non-neutral signatures can reflect selection events occurring recently or distantly in the evolutionary history of a population, thus we cannot specifically infer Bd is the selective agent. However, the top 13 Tajima's D outlier genes with the strongest signatures of balancing selection were generally highly variable within and across populations, and included several immune function genes that may play a role in population adaptation to Bd or other amphibian pathogens such as Ranavirus or Perkinsea. Two of these genes (Complement Factor B genes; Table 4.3) are part of the complement immune pathway, which shows altered expression in Bd-infected frogs in multiple RNAseq experimental exposure studies (Rosenblum et al. 2009; Ellison, Savage, et al. 2014; Savage et al. 2020).

A low ratio of non-synonymous to synonymous SNPs (pN/pS < 1) is a signal of an evolutionary constraint acting on the amino acid evolution of a locus. On the contrary, a high ratio represents some form of positive selection that is favouring an excess of mutations that change amino acids and alter protein function. Remarkably, a number of loci showed high values of pN/pS across the full coding sequence present in our data. It is likely that additional genes have subsets of codons with exceptionally high or low pN/pS values, reflecting different evolutionary pressures on different parts of the protein. More knowledge of protein structure and function is needed to make sense of these patterns, and in amphibians this has only been done for extensively studied genes such as the MHC (e.g. see Mulder et al. 2017). However, it is noteworthy that half of the genes with the highest overall pN/pS values are primarily involved in immune function based on gene ontology (GO) comparisons (Table 4.3). This suggests that in amphibians, immune genes tend to be primary targets of diversifying selection, a pattern that is long-established in mammalian systems and likely arises from host-pathogen arms races (Ferrer-Admetlla et al. 2008).

A common method to detect positive selection acting on populations is the identification of F_{ST} outliers (Narum and Hess 2011) using programs such as BayeScan and SamBada (Foll and Gaggiotti 2008; Stucki et al. 2017). Positive selection will increase the relative abundance of advantageous SNPs in a population more than expected due to genetic drift, and these SNPs with significantly higher or lower than expected (relative to drift) differentiation values are considered to be F_{ST} outliers. Our study system likely experienced strong selection due to the introduction of chytridiomycosis and the resulting die-offs, yet few outliers were detected when comparing adjacent resistant and susceptible populations. This is likely due to the strength of genetic drift occurring in these small and isolated populations, resulting in many fixed SNPs and high average F_{ST} values between populations. Although some of these SNPs may have

driven to fixation due to recent positive selection, it is not possible to separate this mechanism from alternate SNPs randomly and rapidly fixed by drift. Future studies on systems with high population structure might consider increasing sample sizes in large populations to look at within-population outliers (e.g. between resistant and susceptible frogs), or using temporal datasets to detect selection by means of changes in allele frequencies over time (Alves et al. 2019).

Six genes were considered outliers by all three methods (F_{ST}, Tajima's D and pN/pS outliers) and thus have multiple lines of evidence supporting that they are under selection. Although we cannot determine a definitive selective pressure, they could be related to Bd-immunity or other environmental pressures. Several of these genes are noteworthy in the context of amphibian immune responses. First, the CD1B4 gene holds an important function in the immune system, similar to that of classical MHC loci (Richmond et al. 2009), and in humans has been known to be part of the innate immune system against fungal pathogens (Lionakis and Levitz 2018). Second, LV151 is an immunoglobulin (or antibody) component that functions in pathogen recognition, and thus represents a gene with the sole function of destroying invading pathogens, including in the model frog *Xenopus* (Schwager et al. 1991). Furthermore, two genes (QNR-71 and S10A4) are associated with the skin which is where the Bd infections are restricted in frogs (Brutyn et al. 2012), and several previous studies have identified skin integrity genes as important for Bd susceptibility (Rosenblum et al. 2012, 2009; Ellison, Tunstall, et al. 2014).

4.4.3 - Multiple genes associated with Bd-intensity and potential resistance

Specific associations between Bd infection intensity and genetic variation also highlighted several candidate genes that may be involved with host genetic resistance to *Batrachochytrium dendrobatidis* (Table 4.3). The association of the medium Bd-intensity group with Interleukin-10 is particularly interesting as this gene has been shown to be downregulated in previously Bd-infected frogs compared to naïve frogs (Ellison, Savage, et al. 2014). Overexpression of cytokines such as interleukin in response to an infection can cause damage and actually decrease survival (Sears et al. 2011), and thus the medium Bd-intensity group may be tolerating Bd-growth at non-lethal levels, compared to either tolerant or susceptible frogs. Additionally, for the locus Raya_425 there was significantly lower Bd-intensity for individuals that were heterozygous for a missense SNP in the putative CDS, potentially indicate of heterozygote advantage for this trait.

Chytridiomycosis survival can occur in different ways (i.e. resistance versus tolerance; Medzhitov, Schneider & Soares 2012), and Bd intensity is not necessarily the only indicator of fitness and potential survival. Some amphibians are able to clear Bd-infections (Kriger and Hero 2006) while other species may have high Bd intensities but this confers only sub-lethal fitness costs (Daszak et al. 2004). Indeed, previous studies of *R. yavapaiensis* documented some populations maintaining relatively high winter *Bd* intensities with no apparent mortalities (Savage, Sredl, and Zamudio 2011; Savage, Becker, and Zamudio 2015), suggesting some mechanism of Bd tolerance. In contrast, experimental exposure trials in the lab demonstrated Bd resistance among survivors (Savage and Zamudio 2011). Thus, *R. yavapaiensis* individuals may exhibit Bd tolerance in some environments and disease resistance in others, and this could vary across genotypes. These epidemiological complexities further complicate our ability to tease apart the genetic basis of Bd responses in *R. yavapaiensis*, because multiple mechanisms and gene-by-environment interactions are likely. Future studies on candidate resistance genes in natural populations may address some of these challenges by using data on individual responses to Bd infection for better estimates of chytridiomycosis survival. Following candidate gene identification, further lines of evidence are needed to investigate the functional role of these genes in the host-pathogen relationship between Bd and amphibian hosts. As the number of genomic and transcriptomic studies of amphibian chytridiomycosis continue to grow, it is more important than ever to validate these candidate resistance or susceptibility genes using traditional reverse genetics studies implemented via CRISPR/Cas9 gene knock-out, which has been effectively used to ablate MHC gene expression in the model frog *Xenopus* (Banach, Edholm, and Robert 2017).

4.4.4 - The benefits of sequence capture datasets

Due to the extremely low number of amphibian reference genomes (Koepfli et al. 2015) and the large genome sizes of amphibians (Gregory 2003; Weisrock et al. 2018), there are limited studies assessing amphibian functional genetic variation in a phylogeographic context. Functional genetic variation is interesting to study the adaptation of species and populations to different environments, and it also holds important conservation genetic information with management implications and may differ from neutral genetic variation (Meyer-Lucht et al. 2016). Our sequence capture dataset shows that it is possible to produce population-level and genome-wide datasets based on RNAseq data. These datasets can be applied to both neutral genetic questions on population genetics and demographic history, but also hold information on adaptive differences between populations, and histories of selection acting on them.

Analysing phenotypes of interest in a phylogeographic context helps determine historical population demography and structure as well as local adaptation of phenotypes.

4.5 - Conclusion

The combination of genome wide sequence capture data with population level phenotypic data, show that a phylogeographic context can help explain population history and local adaptation in a frog declining from chytridiomycosis. We show that in the lowland leopard frog species *Rana yavapaiensis* there is strong nuclear DNA population structure, combined with limited and unstructured mitochondrial DNA diversity. This likely reflects a strong population decline and population fragmentation, resulting in rapid allele fixation. Chytridiomycosis susceptibility in the species is related to a reduction in heterozygosity and allelic richness, increasing the risk of extirpation. The number of functional private alleles is not related to susceptibility and highlights the need to protect divergent populations to conserve the species' genetic variation and adaptive potential. The high degree of genetic drift and population structure in the species could be obscuring signals of adaptation, but multiple independent selection analyses identify candidate genes related to population differentiation and potential adaptation to chytridiomycosis.

4.6 - Acknowledgements

We would like to thank Alexa Trujillo, Veronica Urgiles, Dennis Caldwell, Michael Sredl, Lauren Augustine, Zach Rivas, Matt Atkinson, Jacob LaFond, Noah Penney, Rachel Gutner, Karen Tracy, Melissa Lenker, JQ Richmond, Jim Collins, Eric Erb, Mark Haberstitch, Ron and Terrie Day, Jennifer Johnston for help collecting samples and Nancy McInerney, Lilly Parker, Madhvi Venkatraman & Natalia Prezelomska for help and advice with labwork at the Center for Conservation Genomics. Ke Bi and Allison Default shared advice on bait design, Rayna Bell and Stephen Sabatino gave insightful advice on the genetic analyses and Daniel Jeffries shared the assembled *Rana temporaria* genome prior to publication. We thank the Smithsonian Institution National Museum of Natural History's Global Genome initiative (GGI) for funding the majority of the labwork with grant 801-0000-302357-332002-6100-XXXX-4120-33GGI2016GRANTE-FLEISCHERR, in addition to funding from the biology department of the University of Central Florida. The computations performed for this paper were conducted on the Smithsonian High Performance Cluster (SI/HPC), Smithsonian Institution. https://doi.org/10.25572/SIHPC. KPM was supported in part by a doctoral student grant (PD/BD/52604/2014), from the Portuguese ''Fundação para a Ciência e a Tecnologia," funded by Programa Operacional Potencial Humano (POPH)—Quadro de Referência Estratégica Nacional (QREN) from the European Social Fund.

4.7 - References

Albert, Eva M., Saioa Fernández-Beaskoetxea, José A. Godoy, Ursina Tobler, Benedikt R. Schmidt, and Jaime Bosch. 2015. “Genetic Management of an Amphibian Population after a Chytridiomycosis Outbreak.” *Conservation Genetics* 16 (1): 103–11. https://doi.org/10.1007/s10592-014-0644-6.

Alexander, David H., and Kenneth Lange. 2011. “Enhancements to the ADMIXTURE Algorithm for Individual Ancestry Estimation.” *BMC Bioinformatics* 12 (1): 246. https://doi.org/10.1186/1471-2105-12-246.

Alföldi, Jessica, Federica Di Palma, Manfred Grabherr, Christina Williams, Lesheng Kong, Evan Mauceli, Pamela Russell, et al. 2011. “The Genome of the Green Anole Lizard and a Comparative Analysis with Birds and Mammals.” *Nature* 477 (7366): 587–91. https://doi.org/10.1038/nature10390.

Alonge, Michael, Sebastian Soyk, Srividya Ramakrishnan, Xingang Wang, Sara Goodwin, Fritz J. Sedlazeck, Zachary B. Lippman, and Michael C. Schatz. 2019. “RaGOO: Fast and Accurate Reference-Guided Scaffolding of Draft Genomes.” *Genome Biology* 20 (1): 1–17. https://doi.org/10.1186/s13059-019-1829-6.

Alves, Joel M., Miguel Carneiro, Jade Y. Cheng, Ana Lemos de Matos, Masmudur M. Rahman, Liisa Loog, Paula F. Campos, et al. 2019. “Parallel Adaptation of Rabbit Populations to Myxoma Virus.” *Science* 363 (6433): 1319–26. https://doi.org/10.1126/science.aau7285.

Andermann, Tobias, Ángela Cano, Alexander Zizka, Christine Bacon, and Alexandre Antonelli. 2018. “SECAPR - A Bioinformatics Pipeline for the Rapid and User-Friendly Processing of Targeted Enriched Illumina Sequences, from Raw Reads to Alignments.” *PeerJ* 2018 (7): e5175. https://doi.org/10.7717/peerj.5175.

Banach, Maureen, Eva Stina Edholm, and Jacques Robert. 2017. “Exploring the Functions of Nonclassical MHC Class Ib Genes in *Xenopus laevis* by the CRISPR/Cas9 System.” *Developmental Biology* 426 (2): 261–69. https://doi.org/10.1016/j.ydbio.2016.05.023.

Beaumont, Mark A. 2005. “Adaptation and Speciation: What Can F_{ST} Tell Us?” *Trends in Ecology and Evolution* 20 (8): 435–40. https://doi.org/10.1016/j.tree.2005.05.017.

Bradley, Gregory A., Philip C. Rosen, Michael J. Sredl, Thomas R. Jones, and Joyce E. Longcore. 2002. “Chytridiomycosis in Native Arizona Frogs.” *Journal of Wildlife Diseases* 38 (1): 206–12. https://doi.org/10.7589/0090-3558-38.1.206.

Brem, Forrest M.R., and Karen R. Lips. 2008. “*Batrachochytrium dendrobatidis* Infection Patterns among Panamanian Amphibian Species, Habitats and Elevations during

Epizootic and Enzootic Stages." *Diseases of Aquatic Organisms* 81 (3): 189–202. https://doi.org/10.3354/dao01960.

Brennan, Thomas C., and Andrew T. Holycross. 2006. *Field Guide to Amphibians and Reptiles in Arizona*. Arizona Game and Fish Dept.

Brutyn, Melanie, Katharina D'Herde, Maarten Dhaenens, Pascale Van Rooij, Elin Verbrugghe, Alex D. Hyatt, Siska Croubels, et al. 2012. "*Batrachochytrium dendrobatidis* Zoospore Secretions Rapidly Disturb Intercellular Junctions in Frog Skin." *Fungal Genetics and Biology* 49 (10): 830–37. https://doi.org/10.1016/j.fgb.2012.07.002.

Camacho, Christiam, George Coulouris, Vahram Avagyan, Ning Ma, Jason Papadopoulos, Kevin Bealer, and Thomas L. Madden. 2009. "BLAST+: Architecture and Applications." *BMC Bioinformatics* 10: 1–9. https://doi.org/10.1186/1471-2105-10-421.

Campana, Michael G. 2018. "BaitsTools: Software for Hybridization Capture Bait Design." *Molecular Ecology Resources* 18 (2): 356–61. https://doi.org/10.1111/1755-0998.12721.

Cassin-Sackett, Loren, Taylor E. Callicrate, and Robert C. Fleischer. 2019. "Parallel Evolution of Gene Classes, but Not Genes: Evidence from Hawai'ian Honeycreeper Populations Exposed to Avian Malaria." *Molecular Ecology* 28 (3): 568–83. https://doi.org/10.1111/mec.14891.

Chen, Hanbo, and Paul C. Boutros. 2011. "VennDiagram: A Package for the Generation of Highly-Customizable Venn and Euler Diagrams in R." *BMC Bioinformatics* 12 (1): 35. https://doi.org/10.1186/1471-2105-12-35.

Christodoulides, Nicolas, Robert C. Fleischer, Brian Gratwicke, and Anna E. Savage. 2020. "Transcriptome-Wide Immune Function Differs in Natural Frog Populations Based on Chytrid Fungus Susceptibility." *In Prep*.

Cingolani, Pablo, Adrian Platts, Le Lily Wang, Melissa Coon, Tung Nguyen, Luan Wang, Susan J. Land, Xiangyi Lu, and Douglas M. Ruden. 2012. "A Program for Annotating and Predicting the Effects of Single Nucleotide Polymorphisms, SnpEff: SNPs in the Genome of Drosophila Melanogaster Strain W1118; Iso-2; Iso-3." *Fly* 6 (2): 80–92. https://doi.org/10.4161/fly.19695.

Clarkson, Robert W, and James C Rorabaug. 1989. "Status of Leopard Frogs (*Rana pipiens* Complex: Ranidae) in Arizona and Southeastern California." *The Southwestern Naturalist* 34 (4): 531–38. https://doi.org/10.2307/3671513.

Clement, Mark, David Posada, and Keith A. Crandall. 2000. "TCS: A Computer Pragram to Estimate Gene Genealogies." *Molecular Ecology* 9 (10): 1657–60. https://doi.org/10.1046/j.1365-294x.2000.01020.x.

Danecek, Petr, Adam Auton, Goncalo Abecasis, Cornelis A. Albers, Eric Banks, Mark A. DePristo, Robert E. Handsaker, et al. 2011. "The Variant Call Format and VCFtools." *Bioinformatics* 27 (15): 2156–58. https://doi.org/10.1093/bioinformatics/btr330.

Daszak, Peter, Lee Berger, Andrew A. Cunningham, Alex D. Hyatt, D. Earl Green, and Rick Speare. 1999. "Emerging Infectious Diseases and Amphibian Population Declines." *Emerging Infectious Diseases* 5 (6): 735–48. https://doi.org/10.3201/eid0506.990601.

Daszak, Peter, Andrew A. Cunningham, and Alex D. Hyatt. 2001. "Anthropogenic Environmental Change and the Emergence of Infectious Diseases in Wildlife." *Acta Tropica* 78 (2): 103–16. https://doi.org/10.1016/S0001-706X(00)00179-0.

Daszak, Peter, A. Strieby, Andrew A. Cunningham, Joyce E. Longcore, Corrie C. Brown, and David Porter. 2004. "Experimental Evidence That the Bullfrog (*Rana catesbeiana*) Is a Potential Carrier of Chytridiomycosis, an Emerging Fungal Disease of Amphibians." *Herpetological Journal* 14 (4): 201–7.

Deleury, Emeline, Thomas Guillemaud, Aurélie Blin, and Eric Lombaert. 2019. "An Evaluation of Pool-Sequencing Transcriptome-Based Exon Capture for Population Genomics in Non-Model Species." *BioRxiv*, 583534. https://doi.org/10.1101/583534.

Duruz, Solange, Natalia Sevane, Oliver Selmoni, Elia Vajana, Kevin Leempoel, Sylvie Stucki, Pablo Orozco-terWengel, et al. 2019. "Rapid Identification and Interpretation of Gene–Environment Associations Using the New R.SamBada Landscape Genomics Pipeline." *Molecular Ecology Resources* 19 (5): 1355–65. https://doi.org/10.1111/1755-0998.13044.

Ellison, Amy R., Anna E. Savage, Grace V. DiRenzo, Penny Langhammer, Karen R. Lips, and Kelly R. Zamudio. 2014. "Fighting a Losing Battle: Vigorous Immune Response Countered by Pathogen Suppression of Host Defenses in the Chytridiomycosis-Susceptible Frog *Atelopus zeteki*." *G3: Genes, Genomes, Genetics* 4 (7): 1275–89. https://doi.org/10.1534/g3.114.010744.

Ellison, Amy R., Tate Tunstall, Graziella V. Direnzo, Myra C. Hughey, Eria A. Rebollar, Lisa K. Belden, Reid N. Harris, Roberto Ibáñez, Karen R. Lips, and Kelly R. Zamudio. 2014. "More than Skin Deep: Functional Genomic Basis for Resistance to Amphibian Chytridiomycosis." *Genome Biology and Evolution* 7 (1): 286–98. https://doi.org/10.1093/gbe/evu285.

Eskew, Evan A., Barbara C. Shock, Elise E.B. Ladouceur, Kevin Keel, Michael R. Miller, Janet E. Foley, and Brian D. Todd. 2018. "Gene Expression Differs in Susceptible and Resistant Amphibians Exposed to *Batrachochytrium dendrobatidis*." *Royal Society Open Science* 5 (2). https://doi.org/10.1098/rsos.170910.

Fagan, William F., and E. E. Holmes. 2006. "Quantifying the Extinction Vortex." *Ecology*

Letters 9 (1): 51–60. https://doi.org/10.1111/j.1461-0248.2005.00845.x.

Ferrer-Admetlla, Anna, Elena Bosch, Martin Sikora, Tomàs Marquès-Bonet, Anna Ramírez-Soriano, Aura Muntasell, Arcadi Navarro, et al. 2008. "Balancing Selection Is the Main Force Shaping the Evolution of Innate Immunity Genes." *The Journal of Immunology* 181 (2): 1315–22. https://doi.org/10.4049/jimmunol.181.2.1315.

Fogelqvist, Johan, Anne Niittyvuopio, Jon Ågren, Outi Savolainen, and Martin Lascoux. 2010. "Cryptic Population Genetic Structure: The Number of Inferred Clusters Depends on Sample Size." *Molecular Ecology Resources* 10 (2): 314–23. https://doi.org/10.1111/j.1755-0998.2009.02756.x.

Foll, Matthieu, and Oscar Gaggiotti. 2008. "A Genome-Scan Method to Identify Selected Loci Appropriate for Both Dominant and Codominant Markers: A Bayesian Perspective." *Genetics* 180 (2): 977–93. https://doi.org/10.1534/genetics.108.092221.

Forrest, Matthew J., and Martin A. Schlaepfer. 2011. "Nothing a Hot Bath Won't Cure: Infection Rates of Amphibian Chytrid Fungus Correlate Negatively with Water Temperature under Natural Field Settings." *PLoS ONE* 6 (12). https://doi.org/10.1371/journal.pone.0028444.

Gahl, Megan K., Joyce E. Longcore, and Jeff E. Houlahan. 2012. "Varying Responses of Northeastern North American Amphibians to the Chytrid Pathogen *Batrachochytrium dendrobatidis*." *Conservation Biology* 26 (1): 135–41. https://doi.org/10.1111/j.1523-1739.2011.01801.x.

Glenn, Travis C., Roger A. Nilsen, Troy J. Kieran, Jon G. Sanders, Natalia J. Bayona-Vásquez, John W. Finger, Todd W. Pierson, et al. 2019. "Adapterama I: Universal Stubs and Primers for 384 Unique Dual-Indexed or 147,456 Combinatorially-Indexed Illumina Libraries (ITru & INext)." *PeerJ* 2019 (10): 1–31. https://doi.org/10.7717/peerj.7755.

Gregory, T. Ryan. 2003. "Variation across Amphibian Species in the Size of the Nuclear Genome Supports a Pluralistic, Hierarchical Approach to the C-Value Enigma." *Biological Journal of the Linnean Society* 79 (2): 329–39. https://doi.org/10.1046/j.1095-8312.2003.00191.x.

Haas, Brian J., and Alexie Papanicolaou. 2012. "TransDecoder (Finding Coding Regions Within Transcripts)." 2012.

Hammond, S. Austin, René L. Warren, Benjamin P. Vandervalk, Erdi Kucuk, Hamza Khan, Ewan A. Gibb, Pawan Pandoh, et al. 2017. "The North American Bullfrog Draft Genome Provides Insight into Hormonal Regulation of Long Noncoding RNA." *Nature Communications* 8 (1): 1–8. https://doi.org/10.1038/s41467-017-01316-7.

Hellsten, Uffe, Richard M. Harland, Michael J. Gilchrist, David Hendrix, Jerzy Jurka, Vladimir

Kapitonov, Ivan Ovcharenko, et al. 2010. “The Genome of the Western Clawed Frog *Xenopus tropicalis*.” *Science* 328 (5978): 633–36. https://doi.org/10.1126/science.1183670.

Hudson, Michael A., Richard P. Young, Josephine D’Urban Jackson, Pablo Orozco-terWengel, L. Martin, A. James, Machel Sulton, et al. 2016. “Dynamics and Genetics of a Disease-Driven Species Decline to near Extinction: Lessons for Conservation.” *Scientific Reports* 6: 1–13. https://doi.org/10.1038/srep30772.

Keenan, Kevin, Philip Mcginnity, Tom F. Cross, Walter W. Crozier, and Paulo A. Prodöhl. 2013. “DiveRsity: An R Package for the Estimation and Exploration of Population Genetics Parameters and Their Associated Errors.” *Methods in Ecology and Evolution* 4 (8): 782–88. https://doi.org/10.1111/2041-210X.12067.

Koepfli, Klaus-peter, Benedict Paten, Genome 10K Community of Scientists, and Stephen J O’Brien. 2015. “The Genome 10K Project: A Way Forward.” *Annual Review of Animal Biosciences* 3: 57–111. https://doi.org/10.1146/annurev-animal-090414-014900.

Kriger, Kerry M., and Jean Marc Hero. 2006. “Survivorship in Wild Frogs Infected with Chytridiomycosis.” *EcoHealth* 3 (3): 171–77. https://doi.org/10.1007/s10393-006-0027-7.

Kriger, Kerry M., Felicia Pereoglou, and Jean Marc Hero. 2007. “Latitudinal Variation in the Prevalence and Intensity of Chytrid (*Batrachochytrium dendrobatidis*) Infection in Eastern Australia.” *Conservation Biology* 21 (5): 1280–90. https://doi.org/10.1111/j.1523-1739.2007.00777.x.

Lacy, Robert C. 1987. “Loss of Genetic Diversity from Managed Populations: Interacting Effects of Drift, Mutation, Immigration, Selection, and Population Subdivision.” *Conservation Biology* 1 (2): 143–58. https://doi.org/10.1111/j.1523-1739.1987.tb00023.x.

Li, Heng, and Richard Durbin. 2009. “Fast and Accurate Short Read Alignment with Burrows-Wheeler Transform.” *Bioinformatics* 25 (14): 1754–60. https://doi.org/10.1093/bioinformatics/btp324.

Lionakis, Michail S., and Stuart M. Levitz. 2018. “Host Control of Fungal Infections: Lessons from Basic Studies and Human Cohorts.” *Annual Review of Immunology* 36 (1). https://doi.org/10.1146/annurev-immunol-042617-053318.

McKenna, Aaron, Matthew Hanna, Eric Banks, Andrey Sivachenko, Kristian Cibulskis, Andrew Kernytsky, Kiran Garimella, et al. 2010. “The Genome Analysis Toolkit: A MapReduce Framework for Analyzing next-Generation DNA Sequencing Data.” *Genome Research* 20 (9): 1297–1303. https://doi.org/10.1101/gr.107524.110.

Medzhitov, Ruslan, David S. Schneider, and Miguel P. Soares. 2012. “Disease Tolerance as

a Defense Strategy." *Science* 335 (6071): 936–41. https://doi.org/10.1126/science.1214935.

Meyer-Lucht, Yvonne, Kevin P. Mulder, Marianne C. James, Barry J. McMahon, Kieran Buckley, Stuart B. Piertney, and Jacob Höglund. 2016. "Adaptive and Neutral Genetic Differentiation among Scottish and Endangered Irish Red Grouse (*Lagopus lagopus scotica*)." *Conservation Genetics* 17 (3): 615–30. https://doi.org/10.1007/s10592-016-0810-0.

Mulder, Kevin P., Maria Cortazar-Chinarro, D. James Harris, Angelica Crottini, Evan H. Campbell Grant, Robert C. Fleischer, and Anna E. Savage. 2017. "Evolutionary Dynamics of an Expressed MHC Class IIβ Locus in the Ranidae (Anura) Uncovered by Genome Walking and High-Throughput Amplicon Sequencing." *Developmental and Comparative Immunology* 76: 177–88. https://doi.org/10.1016/j.dci.2017.05.022.

Mulder, Kevin P., Nandadevi Cortes-Rodriguez, Evan H. Campbell Grant, Adrianne Brand, and Robert C. Fleischer. 2019. "North-Facing Slopes and Elevation Shape Asymmetric Genetic Structure in the Range-Restricted Salamander *Plethodon shenandoah*." *Ecology and Evolution* 9 (9): 5094–5105. https://doi.org/10.1002/ece3.5064.

Múrias Dos Santos, António, Maria Pilar Cabezas, Ana Isabel Tavares, Raquel Xavier, and Madalena Branco. 2015. "TcsBU: A Tool to Extend TCS Network Layout and Visualization." *Bioinformatics* 32 (4): 627–28. https://doi.org/10.1093/bioinformatics/btv636.

Narum, Shawn R., and Jon E. Hess. 2011. "Comparison of F_{ST} Outlier Tests for SNP Loci under Selection." *Molecular Ecology Resources* 11: 184–94. https://doi.org/10.1111/j.1755-0998.2011.02987.x.

Nei, Masatoshi, and Takashi Gojoborit. 1986. "Simple Methods for Estimating the Numbers of Synonymous and Nonsynonymous Nucleotide Substitutions." *Molecular Biology and Evolution* 3 (5): 418–26. https://doi.org/10.1093/oxfordjournals.molbev.a040410.

Nei, Masatoshi, Takeo Maruyama, and Ranajit Chakraborty. 1975. "The Bottleneck Effect and Genetic Variability in Populations." *Evolution* 29 (1): 1. https://doi.org/10.2307/2407137.

Oláh-Hemmings, Viktoria, Jef R. Jaeger, Michael J. Sredl, Martin A. Schlaepfer, Randy D. Jennings, Charles A. Drost, David F. Bradford, and Brett R. Riddle. 2010. "Phylogeography of Declining Relict and Lowland Leopard Frogs in the Desert Southwest of North America." *Journal of Zoology* 280 (4): 343–54. https://doi.org/10.1111/j.1469-7998.2009.00667.x.

Ortiz, E M. 2019. "Vcf2phylip v2. 0: Convert a VCF Matrix into Several Matrix Formats for Phylogenetic Analysis." *Version V2.*

Painter, Charles W, James N. Stuart, J.Tomasz Giermakowski, and Leland J. S. Pierce. 2017. “Checklist of the Amphibians and Reptiles of New Mexico, USA, with Notes on Taxonomy, Status, and Distribution.” *Western Wildlife* 4 (April): 29–60.

Pauly, Gregory B., Maya C. Shaulsky, Anthony J. Barley, Stevie Kennedy-Gold, Sam C. Stewart, Sharon Keeney, and Robert C. Thomson. 2020. “Morphological Change during Rapid Population Expansion Confounds Leopard Frog Identifications in the Southwestern United States.” *Copeia* 108 (2): 299. https://doi.org/10.1643/ch-19-222.

Pembleton, Luke W., Noel O.I. Cogan, and John W. Forster. 2013. “StAMPP: An R Package for Calculation of Genetic Differentiation and Structure of Mixed-Ploidy Level Populations.” *Molecular Ecology Resources* 13 (5): 946–52. https://doi.org/10.1111/1755-0998.12129.

Plummer, Martyn, Nicky Best, Kate Cowles, and Karen Vines. 2006. “CODA: Convergence Diagnosis and Output Analysis for MCMC.” *R News* 6 (1): 7–11.

Portik, Daniel M., Lydia L. Smith, and Ke Bi. 2016. “An Evaluation of Transcriptome-Based Exon Capture for Frog Phylogenomics across Multiple Scales of Divergence (Class: Amphibia, Order: Anura).” *Molecular Ecology Resources* 16 (5): 1069–83. https://doi.org/10.1111/1755-0998.12541.

Richmond, Jonathan Q. 2006. “Evolutionary Basis of Parallelism in North American Scincid Lizards.” *Evolution and Development* 8 (6): 477–90. https://doi.org/10.1111/j.1525-142X.2006.00121.x.

Richmond, Jonathan Q., Anna E. Savage, Kelly R. Zamudio, and Erica Bree Rosenblum. 2009. “Toward Immunogenetic Studies of Amphibian Chytridiomycosis: Linking Innate and Acquired Immunity.” *BioScience* 59 (4): 311–20. https://doi.org/10.1525/bio.2009.59.4.9.

Rorabaugh, James C., and Julio A. Lemos-Espinal. 2016. “A Field Guide to the Amphibians and Reptiles of Sonora.” Mexico. ECO Herpetological Publishing and Distribution, Rodeo, NM.

Rosenblum, Erica Bree, Thomas J. Poorten, Matthew Settles, and Gordon K. Murdoch. 2012. “Only Skin Deep: Shared Genetic Response to the Deadly Chytrid Fungus in Susceptible Frog Species.” *Molecular Ecology* 21 (13): 3110–20. https://doi.org/10.1111/j.1365-294X.2012.05481.x.

Rosenblum, Erica Bree, Thomas J. Poorten, Matthew Settles, Gordon K. Murdoch, Jacques Robert, Nicole Maddox, and Michael B. Eisen. 2009. “Genome-Wide Transcriptional Response of *Silurana* (*Xenopus*) *tropicalis* to Infection with the Deadly Chytrid Fungus.” *PLoS ONE* 4 (8). https://doi.org/10.1371/journal.pone.0006494.

Savage, Anna E., Carlos G. Becker, and Kelly R. Zamudio. 2015. “Linking Genetic and

Environmental Factors in Amphibian Disease Risk." *Evolutionary Applications* 8 (6): 560–72. https://doi.org/10.1111/eva.12264.

Savage, Anna E., Brian Gratwicke, K. Hope, E. Bronikowski, and R.C. Fleischer. 2020. "Immune Activation Is Associated with Susceptibility to the Amphibian Chytrid Fungus." *Molecular Ecology - Accepted*.

Savage, Anna E., Kevin P. Mulder, Taina Torres, and Stuart Wells. 2018. "Lost but Not Forgotten: MHC Genotypes Predict Overwinter Survival despite Depauperate MHC Diversity in a Declining Frog." *Conservation Genetics* 19 (2): 309–22. https://doi.org/10.1007/s10592-017-1001-3.

Savage, Anna E., Carly R. Muletz-Wolz, Evan H. Campbell Grant, Robert C. Fleischer, and Kevin P. Mulder. 2019. "Functional Variation at an Expressed MHC Class IIβ Locus Associates with Ranavirus Infection Intensity in Larval Anuran Populations." *Immunogenetics* 71 (4): 335–46. https://doi.org/10.1007/s00251-019-01104-1.

Savage, Anna E., Michael J. Sredl, and Kelly R. Zamudio. 2011. "Disease Dynamics Vary Spatially and Temporally in a North American Amphibian." *Biological Conservation* 144 (6): 1910–15. https://doi.org/10.1016/j.biocon.2011.03.018.

Savage, Anna E., and K. R. Zamudio. 2011. "MHC Genotypes Associate with Resistance to a Frog-Killing Fungus." *Proceedings of the National Academy of Sciences* 108 (40): 16705–10. https://doi.org/10.1073/pnas.1106893108.

Savage, Anna E., and Kelly R. Zamudio. 2016. "Adaptive Tolerance to a Pathogenic Fungus Drives Major Histocompatibility Complex Evolution in Natural Amphibian Populations." *Proceedings of the Royal Society B: Biological Sciences* 283 (1827): 20153115. https://doi.org/10.1098/rspb.2015.3115.

Scheele, Ben C., Frank Pasmans, Lee F. Skerratt, Lee Berger, An Martel, Wouter Beukema, Aldemar A. Acevedo, et al. 2019. "Amphibian Fungal Panzootic Causes Catastrophic and Ongoing Loss of Biodiversity." *Science* 363 (6434): 1459–63. https://doi.org/10.1126/science.aav0379.

Schlaepfer, Martin A., Michael J. Sredl, Phil C. Rosen, and Michael J. Ryan. 2007. "High Prevalence of *Batrachochytrium dendrobatidis* in Wild Populations of Lowland Leopard Frogs *Rana yavapaiensis* in Arizona." *EcoHealth* 4 (4): 421–27. https://doi.org/10.1007/s10393-007-0136-y.

Schwager, Joseph, Niels Bürckert, Martine Schwager, and Melanie Wilson. 1991. "Evolution of Immunoglobulin Light Chain Genes: Analysis of *Xenopus* IgL Isotypes and Their Contribution to Antibody Diversity." *The EMBO Journal* 10 (3): 505–11. https://doi.org/10.1002/j.1460-2075.1991.tb07976.x.

Sears, Brittany F., Jason R. Rohr, Judith E. Allen, and Lynn B. Martin. 2011. "The Economy

of Inflammation: When Is Less More?" *Trends in Parasitology* 27 (9): 382–87. https://doi.org/10.1016/j.pt.2011.05.004.

Sjoqvist, Conny, Luis Fernando Delgado Zambrano, Johannes Alneberg, and Anders F Andersson. 2020. "Revealing Ecologically Coherent Population Structure of Uncultivated Bacterioplankton with POGENOM." *BioRxiv*, 2020.03.25.999755.

Smith, Katherine F., Dov F. Sax, and Kevin D. Lafferty. 2006. "Evidence for the Role of Infectious Disease in Species Extinction and Endangerment." *Conservation Biology* 20 (5): 1349–57. https://doi.org/10.1111/j.1523-1739.2006.00524.x.

Sredl, Michael J. 1997. "Ranid Frog Conservation and Management."

Stamatakis, Alexandros. 2014. "RAxML Version 8: A Tool for Phylogenetic Analysis and Post-Analysis of Large Phylogenies." *Bioinformatics* 30 (9): 1312–13. https://doi.org/10.1093/bioinformatics/btu033.

Stamatakis, Alexandros, Paul Hoover, and Jacques Rougemont. 2008. "A Rapid Bootstrap Algorithm for the RAxML Web Servers." *Systematic Biology* 57 (5): 758–71. https://doi.org/10.1080/10635150802429642.

Stucki, Sylvie, Pablo Orozco-terWengel, Brenna R. Forester, Solange Duruz, Licia Colli, Charles Masembe, Riccardo Negrini, et al. 2017. "High Performance Computation of Landscape Genomic Models Including Local Indicators of Spatial Association." *Molecular Ecology Resources* 17 (5): 1072–89. https://doi.org/10.1111/1755-0998.12629.

Sun, Yan Bo, Zi Jun Xiong, Xue Yan Xiang, Shi Ping Liu, Wei Wei Zhou, Xiao Long Tu, Li Zhong, et al. 2015. "Whole-Genome Sequence of the Tibetan Frog *Nanorana parkeri* and the Comparative Evolution of Tetrapod Genomes." *Proceedings of the National Academy of Sciences* 112 (11): E1257–62. https://doi.org/10.1073/pnas.1501764112.

Tobler, Ursina, and Benedikt R. Schmidt. 2010. "Within- and among-Population Variation in Chytridiomycosis-Induced Mortality in the Toad *Alytes obstetricans*." *PLoS ONE* 5 (6): e10927. https://doi.org/10.1371/journal.pone.0010927.

Toews, David P.L., and Alan Brelsford. 2012. "The Biogeography of Mitochondrial and Nuclear Discordance in Animals." *Molecular Ecology* 21 (16): 3907–30. https://doi.org/10.1111/j.1365-294X.2012.05664.x.

Villarreal, Miguel L., Laura M. Norman, Kenneth G. Boykin, and Cynthia S.A. Wallace. 2013. "Biodiversity Losses and Conservation Trade-Offs: Assessing Future Urban Growth Scenarios for a North American Trade Corridor." *International Journal of Biodiversity Science, Ecosystem Services and Management* 9 (2): 90–103. https://doi.org/10.1080/21513732.2013.770800.

Voyles, Jamie, Douglas C. Woodhams, Veronica Saenz, Allison Q Byrne, Rachel Perez,

Gabriela Rios-Sotelo, Mason J. Ryan, et al. 2018. “Shifts in Disease Dynamics in a Tropical Amphibian Assemblage Are Not Due to Pathogen Attenuation.” *Science* 359 (6383): 1517–19. https://doi.org/10.1126/science.aao4806.

Weisrock, David W., Paul M. Hime, Schyler O. Nunziata, Kara S. Jones, Mason O. Murphy, Scott Hotaling, and Justin D. Kratovil. 2018. “Surmounting the Large-Genome ‘Problem’ for Genomic Data Generation in Salamanders.” In *Population Genomics: Wildlife*, 1–28. Springer. https://doi.org/10.1007/13836_2018_36.

Whiles, Matt R., Karen R. Lips, Cathy M. Pringle, Susan S. Kilham, Rebecca J. Bixby, Roberto Brenes, Scott Connelly, et al. 2006. “The Effects of Amphibian Population Declines on the Structure and Function of Neotropical Stream Ecosystems.” *Frontiers in Ecology and the Environment* 4 (1): 27–34. https://doi.org/10.1890/1540-9295(2006)004[0027:TEOAPD]2.0.CO;2.

Witte, Carmel L., Michael J. Sredl, Andrew S. Kane, and Laura L. Hungerford. 2008. “Epidemiologic Analysis of Factors Associated with Local Disappearances of Native Ranid Frogs in Arizona.” *Conservation Biology* 22 (2): 375–83. https://doi.org/10.1111/j.1523-1739.2007.00878.x.

Zamudio, Kelly R., Rayna C. Bell, and Nicholas A. Mason. 2016. “Phenotypes in Phylogeography: Species’ Traits, Environmental Variation, and Vertebrate Diversification.” *Proceedings of the National Academy of Sciences* 113 (29): 8041–48. https://doi.org/10.1073/pnas.1602237113.

Zheng, Xiuwen, David Levine, Jess Shen, Stephanie M. Gogarten, Cathy Laurie, and Bruce S. Weir. 2012. “A High-Performance Computing Toolset for Relatedness and Principal Component Analysis of SNP Data.” *Bioinformatics* 28 (24): 3326–28. https://doi.org/10.1093/bioinformatics/bts606.

CHAPTER 5: GENERAL DISCUSSION

5.1 - Concluding remarks on the application of genomic tools to questions of local adaptation in amphibians

Advances in genomic sequencing and the study of adaptive traits were initially focused on humans and model organisms, but subsequently spread to many other lineages in the tree of life. In chapter 1, I explained why the field of amphibian genomics was slow to start, mainly due their exceptionally large genomes. This meant higher costs for sequencing, and additional computational challenges with assembly (Keinath et al. 2015), SNP calling (Mulder et al. 2019), and annotation (Kwon 2017). I discussed several of the new and updated library preparation and analyses methods that can alleviate some of these challenges, including ddRAD, RNA-seq and exome capture, specifically reviewing how these methods can generate data for studies of adaptation.

At the start of this doctoral thesis, many of these techniques had been developed quite recently and were starting to be applied at wider scales. In amphibians specifically, studies applying genome-wide markers were still exceedingly rare, and most were focused on phylogenomics and/or species with relatively small genomes. In this thesis, I showed that exon-capture can generate genome-wide population level datasets in amphibians and that they can be applied to questions of phylogenomics and ancestral state reconstruction (chapter 2), and population genomics and adaptation (chapter 4). I also applied RNAseq sequencing to a variable adaptive trait by using natural populations (chapter 3). Both sequence capture studies were completed using affordable in-house library methods and limited sequencing efforts. Baits were designed based on species-specific transcriptome data (chapter 3 and Savage et al.

2020), showing that developing a reference transcriptome can also be an efficient tool for functional marker development. Enrichment rates were relatively high (over 40%) and resulted in largely complete alignments across individuals. The produced transcriptome data based on seven different tissues was also of high quality, and resulted in a near complete transcriptome assembly (BUSCO score of 91), and successful differential expression analyses between uterus and oviduct tissues (chapter 3).

The sequence capture data for both chapter 2 and 4 had to be strictly filtered and conservatively analysed to produce high quality results. Filtering and quality control are important steps in outlier detection to reduce the potential for false positives (Weale 2010). For example, identifying loci with excess heterozygosity was an important step to reduce the number of putative paralogs in our dataset. Identifying exon-intron boundaries was also crucial to improve mapping and alignments, and to increase the quality and scale of the data. Including intronic sequences helped identify paralogs, as duplicate loci often diverged more in the flanking introns, and increased the number of mapped reads and their mapping score. The bioinformatic analyses of these datasets was greatly helped by the development of recently published software pipelines such as SECAPR (Andermann et al. 2018) and IEB-finder (Deleury et al. 2019).

Transcriptome assembly and differential expression analyses in chapter 3 largely followed general bioinformatic pipelines and did not have to be tweaked for amphibians. However, annotation of the assembled transcripts was troublesome due to the lack of reference material for amphibians available in the UniProt database. Annotation is extremely important for high quality analyses of gene-level differential expression (Mudge and Harrow 2016). By using the annotated *Pleurodeles waltl* genome and associated transcriptome (Elewa et al. 2017), we improved our annotation substantially. Both the *Rana catesbeiana* and *R. temporaria* genome assemblies were also important for annotation and scaffolding of the lowland leopard frog data. The recent increase in amphibian genomes and transcriptomes should help increase the quality of annotation, mapping and gene expression analyses in the coming years.

The level of population structure found in chapter 4, and the effect it had on finding F_{ST} outliers, is a warning note for future studies on local adaptation in amphibians. As amphibians on average have more limited dispersal and strong site fidelity (Smith and Green 2005), population structure is likely to be exceptionally high in many systems. Strong population structure and associated high average F_{ST} values may make F_{ST} outliers more difficult to detect, and conceal potential signatures of selection.

Specific (micro) habitat requirements such as waterbodies can also result in more disjunct ranges, especially with regards to breeding (Duellman and Trueb 1994). The range of *Rana yavapaiensis*, for example, is currently quite patchy across Arizona (although it may have been more wide-ranging in the past). Studies of amphibians often sample numerous individuals per breeding habitat, increasing the amount of genetic data, but limiting the number of different environmental data points available for the species. This will reduce the effectiveness of genotype-environment associations (GEA), as we found during exploratory analyses. One possible solution to this challenge is to collect micro-environmental data at a fine scale where amphibians are sampled, precluding the need to rely on global databases with environmental data at much larger spatial scales.

Strong knowledge of the evolutionary history of a species is important for any study of adaptation, as was evident in the results of both chapter 2 and 4. Within *Salamandra* it is now clear that there is likely only one transition to pueriparity within *S. algira*, and that both pueriparous islands of *S. s. gallaica* are not each other's closest relative. Furthermore, it appears *S. s. bernardezi* is pueriparous across both geographic and phylogenetic space. Within *R. yavapaiensis* there is a strong phylogenetic split between the populations in the north and the south (chapter 4), which was not evident based on previously collected mitochondrial and microsatellite data (Oláh-Hemmings et al. 2010; Savage, Becker, and Zamudio 2015). The increased knowledge of the evolutionary history of both these study systems is crucial for future comparative studies of adaptation.

Both study systems investigated in this thesis also highlighted the importance of both range-wide sampling, and good phenotypic characterization of traits of interest. Even with the increase in the availability of genomic data, continued effort in the collection of both genetic and phenotypic data are crucial for studies of adaptation. Within *Salamandra* increased sampling of both genetic and natural history data in *S. s. fastuosa* is needed to fully understand the evolution of pueriparity in this clade. The high number of private alleles in the Aliso Spring population of *R. yavapaiensis* in southern Arizona also indicates that increased sampling of the species in its Mexican range is crucial. Although very difficult to collect, increased knowledge of Bd-susceptibility across space and time in the lowland leopard frog will also improve our understanding of the evolution of resistance to Bd.

5.2 - *Salamandra salamandra* and implications for the study of pueriparity

Using exon capture data and ancestral state reconstruction we showed that there are at least five independent transitions from larviparity to pueriparity in the genus, and that they happened at different evolutionary timescales. Four of the transitions are intra-specific, three of which are within *S. salamandra*. This data provides the evolutionary framework for comparative analyses, and is crucial for future work on the environmental, ecological and physiological determinants of the shift to pueriparity,

Although there is clearly a genetic component to the differences in larviparity and pueriparity (Velo-Antón et al. 2015), the number of (recent) transitions in this clade shows that the trait has evolved both often and fast. This appears contradictory to the strong phenotypic differences found between both modes (Buckley et al. 2007), and the relative rarity of pueriparity across both salamanders and frogs. Two possible explanations for this apparent contradiction are exaptation and genetic assimilation.

It is possible that the common ancestor of both *Salamandra* and *Lyciasalamandra* already displayed certain traits or characters that made it easier for pueriparity to develop. This term was previously called pre-adaptation, but the preferred wording is now exaptation (Gould and Vrba 1982). Major knowledge gaps exist about reproduction and embryonic development in amphibians, and it is not clear what these characters could be. More work on the biology of pueriparity in frogs and caecilians might improve our understanding of common patterns of pueriparity and the physiological traits needed for its development.

Although *Salamandra* are not phenotypically plastic right now, there is some variation in the number and stage of larval development upon delivery (Velo-Antón et al. 2015). It is possible that some phenotypic plasticity and epigenetic changes may initially promote the development of pueriparity in water-limited environment, after which genetic assimilation occurs and the trait becomes non-plastic. The process and extent of genetic assimilation in explaining biological variation is still largely unknown and an expanding area of research.

RNA-seq analyses of two of the transitions to pueriparity identified convergent changes in gene expression, indicative of potential candidate genes important in explaining this shift, and the associated physiological and developmental changes. Highlighted genes included

numerous with putative functions in embryonic development. Although this was a correlative study and we cannot determine causation, these genes are strong candidates that merit further study. Although heritable gene expression differences can also be due to allelic variation in other parts of the genome, follow up studies may investigate how genetic variation in these candidate genes varies across the landscape, and across the five identified transitions. Admixture zones between both reproductive modes can also be used to apply admixture mapping and genome-wide association studies to determine if there are genetic variants across the genome that correspond with the gene expression differences.

5.3 - *Rana yavapaiensis* and implications for disease resistance

The amphibian extinction crisis, and the continued spread of diseases such as *Batrachochytrium dendrobatidis* (Scheele et al. 2019), *Batrachochytrium salamandrivorans* (Martel et al. 2013), Ranavirus (Miller, Gray, and Storfer 2011; Gray, Miller, and Hoverman 2009) and Perkinsea (Isidoro-Ayza et al. 2017), demonstrate the importance of increased knowledge of amphibian immunity for conservation. Most knowledge of both the innate and adaptive immune system in frogs is still based on *Xenopus*, a genus of frog that is phylogenetically distinct from the Neobatrachia which holds over 95% of the species diversity in anurans. Studies of adaptation to disease in currently declining frogs such as *R. yavapaiensis* (chapter 4) can inform questions of local adaptation and direct conservation efforts in this species, but also increase our general understanding of immunity in a large clade of at-risk frogs.

A challenge with studies of disease immunity is assessing susceptibility. Data on survival in the field are hard to collect and laboratory conditions are not representative of natural settings. For diseases like Bd, data on the prevalence and intensity of fungal zoospores on the skin of amphibians are regularly collected, and relatively easy to quantify using qPCR (Hyatt et al. 2007). Infection intensity, however, is not always a good representative of disease susceptibility as there are different mechanisms that organisms employ to survive diseases. Some species and populations of frogs appear to tolerate medium to high levels of Bd on their skin without showing apparent symptoms or a decrease in survival (Reeder, Pessier, and Vredenburg 2012; Savage and Zamudio 2016). Others are more resistant and actively manage to reduce or remove Bd from their skin after exposure (Márquez et al. 2010). In chapter 4 we classify susceptibility using numerous metrics (Table 4.1) in this well studied species, but assessing phenotypic diversity of complex traits like susceptibility remains more challenging than simpler traits like morphological differentiation.

We found a pattern of lower levels of genetic variation in metrics like allelic richness and heterozygosity correlating with higher Bd-susceptibility (Figure 4.3). It is unknown if population level Bd-susceptibility was caused by lower levels of genetic variation, or if populations that were more susceptible had previously lost more genetic variation. This question would be interesting to investigate using museum specimens collected before Bd arrived into the system. As sequence capture works on fragmented DNA and can be applied to DNA extracted from formalin-fixed samples (Ruane and Austin 2017), this is a promising application of exon

capture. By combining historical and modern samples it is possible track genomic signatures of adaptation over time (Alves et al. 2019).

Nonetheless, we did find some consistent patterns through our different outlier and selection analyses that suggest that immune system genes (such as CFAB and CD1B4), and genes important for skin integrity, may be important for how lowland leopard frog populations and individuals are adapting to Bd. Several overlap with previous RNAseq studies that have identified these genes as differentially expressed during Bd infections. This list of candidate genes reinforces how useful the sequence capture approach can be, especially when combined with RNAseq data, even for complex traits like disease susceptibility.

5.4 - The future of amphibian adaptation genomics

In this doctoral thesis I successfully applied genomic tools and genome-wide markers to two amphibian study systems showing intra-specific variation in adaptive traits. This demonstrates the promise of techniques such as exon capture and RNA-seq in amphibian genomics and studies of adaptation. Using analyses of gene expression and signatures of selection I identified candidate genes for the genetic basis of two traits of interest across two distinct amphibian study systems. Future work can use functional analyses to determine how these genes maybe influencing the phenotype. Two promising techniques that might help determine the functionality of these genes are chromosome level genome assemblies and CRISPR/Cas9 gene editing.

Although population level whole-genome resequencing is likely still prohibitively expensive for the large genomes of amphibians, new techniques such as HiC, PacBio and Bionano have produced high quality reference genomes for amphibians (chapter 1). Annotated reference genomes can be used to identify and design probes for both the coding sequence and surrounding regions of previously identified candidate genes, and more directly target potential genetic polymorphisms responsible for the phenotype. This includes important non-coding transcription regulators such as promotors, silencers, and enhancers.

Historically, knockout mice have been crucial in determining the function of many genes but this technique has not been commonly applied in non-model organisms. The recent development of CRISPR/Cas9 gene editing has made gene editing and functional analyses of candidate genes achievable. It has already been successfully applied to *Xenopus laevis* (Banach, Edholm, and Robert 2017), and conceptually it can be applied to any species. Gene editing holds great promise to confirm the functional genetic basis of the identified candidate genes.

Overall the increase in both genomic tools, resources and bioinformatic pipelines developed these past years has substantially helped amphibian genomics, and will likely continue to improve with time. The variety of different techniques and analyses in this thesis, applied to two different amphibian systems shows the promise of studies of adaptation genomics in amphibians.

5.5 - References

Alves, Joel M., Miguel Carneiro, Jade Y. Cheng, Ana Lemos de Matos, Masmudur M. Rahman, Liisa Loog, Paula F. Campos, et al. 2019. "Parallel Adaptation of Rabbit Populations to Myxoma Virus." *Science* 363 (6433): 1319–26. https://doi.org/10.1126/science.aau7285.

Andermann, Tobias, Ángela Cano, Alexander Zizka, Christine Bacon, and Alexandre Antonelli. 2018. "SECAPR - A Bioinformatics Pipeline for the Rapid and User-Friendly Processing of Targeted Enriched Illumina Sequences, from Raw Reads to Alignments." *PeerJ* 2018 (7): e5175. https://doi.org/10.7717/peerj.5175.

Banach, Maureen, Eva Stina Edholm, and Jacques Robert. 2017. "Exploring the Functions of Nonclassical MHC Class Ib Genes in *Xenopus laevis* by the CRISPR/Cas9 System." *Developmental Biology* 426 (2): 261–69. https://doi.org/10.1016/j.ydbio.2016.05.023.

Buckley, David, Marina Alcobendas, Mario Garcia-Paris, and Marvalee H. Wake. 2007. "Heterochrony, Cannibalism, and the Evolution of Viviparity in *Salamandra Salamandra*." *Evolution and Development* 9 (1): 105–15. https://doi.org/10.1111/j.1525-142X.2006.00141.x.

Deleury, Emeline, Thomas Guillemaud, Aurélie Blin, and Eric Lombaert. 2019. "An Evaluation of Pool-Sequencing Transcriptome-Based Exon Capture for Population Genomics in Non-Model Species." *BioRxiv*, 583534. https://doi.org/10.1101/583534.

Duellman, William E., and Linda Trueb. 1994. *Biology of Amphibians*. JHU Press.

Elewa, Ahmed, Heng Wang, Carlos Talavera-López, Alberto Joven, Gonçalo Brito, Anoop Kumar, L. Shahul Hameed, et al. 2017. "Reading and Editing the *Pleurodeles waltl* Genome Reveals Novel Features of Tetrapod Regeneration." *Nature Communications* 8 (1): 1–9. https://doi.org/10.1038/s41467-017-01964-9.

Gould, Stephen Jay, and Elisabeth S. Vrba. 1982. "Exaptation - A Missing Term in the Science of Form." *Society* 8 (1): 4–15.

Gray, Matthew J., Debra L. Miller, and Jason T. Hoverman. 2009. "Ecology and Pathology of Amphibian Ranaviruses." *Diseases of Aquatic Organisms* 87 (3): 243–66. https://doi.org/10.3354/dao02138.

Hyatt, Alex D., D. G. Boyle, Veronica Olsen, David B. Boyle, Lee Berger, David Obendorf, A. Dalton, et al. 2007. "Diagnostic Assays and Sampling Protocols for the Detection of *Batrachochytrium dendrobatidis*." *Diseases of Aquatic Organisms* 73 (3): 175–92. https://doi.org/10.3354/dao073175.

Isidoro-Ayza, Marcos, Jeffrey M. Lorch, Daniel A. Grear, Megan Winzeler, Daniel L.

Calhoun, and William J. Barichivich. 2017. “Pathogenic Lineage of Perkinsea Associated with Mass Mortality of Frogs across the United States.” *Scientific Reports* 7 (1): 10288. https://doi.org/10.1038/s41598-017-10456-1.

Keinath, Melissa C., Vladimir A. Timoshevskiy, Nataliya Y. Timoshevskaya, Panagiotis A. Tsonis, S. Randal Voss, and Jeramiah J. Smith. 2015. “Initial Characterization of the Large Genome of the Salamander *Ambystoma mexicanum* Using Shotgun and Laser Capture Chromosome Sequencing.” *Scientific Reports* 5 (1): 16413. https://doi.org/10.1038/srep16413.

Kwon, Taejoon. 2017. “AmphiBase: A New Genomic Resource for Non-Model Amphibian Species.” *Genesis* 55 (1–2): e23010. https://doi.org/10.1002/dvg.23010.

Márquez, Mariella, Francisco Nava-González, Dinora Sánchez, Marina Calcagno, and Margarita Lampo. 2010. “Immunological Clearance of *Batrachochytrium dendrobatidis* Infection at a Pathogen-Optimal Temperature in the Hylid Frog *Hypsiboas crepitans*.” *EcoHealth* 7 (3): 380–88. https://doi.org/10.1007/s10393-010-0350-x.

Martel, An, Annemarieke Spitzen-Van Der Sluijs, Mark Blooi, Wim Bert, Richard Ducatelle, Matthew C. Fisher, Antonius Woeltjes, et al. 2013. “*Batrachochytrium salamandrivorans* Sp. Nov. Causes Lethal Chytridiomycosis in Amphibians.” *Proceedings of the National Academy of Sciences* 110 (38): 15325–29. https://doi.org/10.1073/pnas.1307356110.

Miller, Debra, Matthew Gray, and Andrew Storfer. 2011. “Ecopathology of Ranaviruses Infecting Amphibians.” *Viruses* 3 (11): 2351–73. https://doi.org/10.3390/v3112351.

Mudge, Jonathan M., and Jennifer Harrow. 2016. “The State of Play in Higher Eukaryote Gene Annotation.” *Nature Reviews Genetics* 17 (12): 758–72. https://doi.org/10.1038/nrg.2016.119.

Mulder, Kevin P., Nandadevi Cortes-Rodriguez, Evan H. Campbell Grant, Adrianne Brand, and Robert C. Fleischer. 2019. “North-Facing Slopes and Elevation Shape Asymmetric Genetic Structure in the Range-Restricted Salamander *Plethodon shenandoah*.” *Ecology and Evolution* 9 (9): 5094–5105. https://doi.org/10.1002/ece3.5064.

Oláh-Hemmings, Viktoria, Jef R. Jaeger, Michael J. Sredl, Martin A. Schlaepfer, Randy D. Jennings, Charles A. Drost, David F. Bradford, and Brett R. Riddle. 2010. “Phylogeography of Declining Relict and Lowland Leopard Frogs in the Desert Southwest of North America.” *Journal of Zoology* 280 (4): 343–54. https://doi.org/10.1111/j.1469-7998.2009.00667.x.

Reeder, Natalie M.M., Allan P. Pessier, and Vance T. Vredenburg. 2012. “A Reservoir Species for the Emerging Amphibian Pathogen *Batrachochytrium dendrobatidis* Thrives in a Landscape Decimated by Disease.” *PLoS ONE* 7 (3): 1–7. https://doi.org/10.1371/journal.pone.0033567.

Ruane, Sara, and Christopher C. Austin. 2017. “Phylogenomics Using Formalin-Fixed and 100+ Year-Old Intractable Natural History Specimens.” *Molecular Ecology Resources* 17 (5): 1003–8. https://doi.org/10.1111/1755-0998.12655.

Savage, Anna E., Carlos G. Becker, and Kelly R. Zamudio. 2015. “Linking Genetic and Environmental Factors in Amphibian Disease Risk.” *Evolutionary Applications* 8 (6): 560–72. https://doi.org/10.1111/eva.12264.

Savage, Anna E., Brian Gratwicke, K. Hope, E. Bronikowski, and R.C. Fleischer. 2020. “Immune Activation Is Associated with Susceptibility to the Amphibian Chytrid Fungus.” *Molecular Ecology - Accepted*.

Savage, Anna E., and Kelly R. Zamudio. 2016. “Adaptive Tolerance to a Pathogenic Fungus Drives Major Histocompatibility Complex Evolution in Natural Amphibian Populations.” *Proceedings of the Royal Society B: Biological Sciences* 283 (1827): 20153115. https://doi.org/10.1098/rspb.2015.3115.

Scheele, Ben C., Frank Pasmans, Lee F. Skerratt, Lee Berger, An Martel, Wouter Beukema, Aldemar A. Acevedo, et al. 2019. “Amphibian Fungal Panzootic Causes Catastrophic and Ongoing Loss of Biodiversity.” *Science* 363 (6434): 1459–63. https://doi.org/10.1126/science.aav0379.

Smith, M. Alex, and David M. Green. 2005. “Dispersal and the Metapopulation Paradigm in Amphibian Ecology and Conservation: Are All Amphibian Populations Metapopulations?” *Ecography* 28 (1): 110–28. https://doi.org/10.1111/j.0906-7590.2005.04042.x.

Velo-Antón, Guillermo, Xavier Santos, Iago Sanmartín-Villar, Adolfo Cordero-Rivera, and David Buckley. 2015. “Intraspecific Variation in Clutch Size and Maternal Investment in Pueriparous and Larviparous *Salamandra salamandra* Females.” *Evolutionary Ecology* 29 (1): 185–204. https://doi.org/10.1007/s10682-014-9720-0.

Weale, Michael E. 2010. “Quality Control for Genome-Wide Association Studies.” In *Genetic Variation*, 341–72. Springer.

APPENDIX A: LIST OF ADDITIONAL PUBLICATIONS

List of publications that are within the scope of this thesis and that I led or participated in as a co-author.

Mulder, Kevin P., André Lourenço, Miguel Carneiro, and Guillermo Velo-Antón. 2016. "The Complete Mitochondrial Genome of *Salamandra salamandra* (Amphibia: Urodela: Salamandridae)." *Mitochondrial DNA Part B* 1 (1): 880–82. https://doi.org/10.1080/23802359.2016.1253042.

Mulder, Kevin P., Nandadevi Cortes-Rodriguez, Evan H. Campbell Grant, Adrianne Brand, and Robert C. Fleischer. 2019. "North-Facing Slopes and Elevation Shape Asymmetric Genetic Structure in the Range-Restricted Salamander *Plethodon shenandoah*." *Ecology and Evolution* 9 (9): 5094–5105. https://doi.org/10.1002/ece3.5064.

Savage, Anna E., Carly R. Muletz-Wolz, Evan H. Campbell Grant, Robert C. Fleischer, and **Kevin P. Mulder**. 2019. "Functional Variation at an Expressed MHC Class IIβ Locus Associates with Ranavirus Infection Intensity in Larval Anuran Populations." *Immunogenetics* 71 (4): 335–46. https://doi.org/10.1007/s00251-019-01104-1.

APPENDIX B: SUPPLEMENTAL INFORMATION FOR CHAPTER 2

Table B1: Table of genetic samples used for the *S. salamandra* sequence capture. Information included are initial species identification, location and reproductive mode. Samples included in the SNAPP species-tree analyses are indicated in bold.

Species	Locality	Latitude	Longitude	Mode	Sample ID
S. a. algira	Algeria: Tamalous	36.89	6.72	L	**KHASAA**
S. a. atlantica	Morocco: Taffert	33.65	-4.16	L	**GVA4172**
S. a. spelaea	Morocco: Berkane	34.84	-2.28	L	**SAL1513**
S. a. splendens	Morocco: Fifi	35.00	-5.19	L	GVA6175
S. a. splendens	Morocco: Talassemtane	35.18	-5.23	L	**GVA5801**
S. a. splendens	Morocco: Tetouan	35.52	-5.40	P	GVA4156
S. a. tingitana	Morocco: Adayourha	35.26	-5.42	L	GVA5773
S. a. tingitana	Morocco: Amsa	35.53	-5.21	P	GVA6103
S. a. tingitana	Morocco: Dar Chaoui	35.57	-5.77	L	GVA1941
S. a. tingitana	Morocco: Jebel Suna	35.13	-5.43	L	**GVA6183**
S. a. tingitana	Morocco: Spirada	35.52	-5.55	L	GVA5793
S. a. tingitana	Morocco. Chefchaouen	35.27	-5.49	L	GVA5762
S. a. tingitana	Spain: Ceuta	35.89	-5.36	P	GVA5748
S. a. tingitana	Spain: Ceuta	35.89	-5.36	P	GVA5749
S. atra atra	Austria: Turracherhohe	46.91	13.87	P	**GVA7606**
S. atra pasubiensis	Italy: Monte Pasubio	45.79	11.17	P	**SAL0337**
S. corsica	France: Corsica	42.10	9.11	L	**MVTS6241**
S. infraimmaculata	Turkey: Fevcipasa	37.10	36.64	L	**SAL0009**
S. infraimmaculata	Lebanon: Bater	33.59	35.61	L	**SAL0027**
S. lanzai	Italy: Pian del Re. Monte Viso	44.70	7.09	P	**SAL0243**
S. s. almanzoris	Spain: Circo de Gredos	40.26	-5.27	L	**GVA6474**
S. s. bejarae	Spain: Boñar	42.86	-5.29	L	**GVA4115**
S. s. bejarae	Spain: Cabeza de Campo	42.54	-6.87	L	GVA5002
S. s. bejarae	Spain: Candelario	40.36	-5.74	L	GVA8017
S. s. bejarae	Spain: Cistierna	42.81	-5.14	L	GVA6537
S. s. bejarae	Spain: Fuentes Carrionas	42.88	-4.55	L	GVA6525
S. s. bejarae	Spain: Sedano	42.68	-3.74	L	GVA4248
S. s. bernardezi	Spain: Baselgas	43.31	-6.06	P	**GVA8621**
S. s. bernardezi	Spain: Bolgues	43.40	-6.03	P	GVA8525
S. s. bernardezi	Spain: Brañes	43.41	-5.92	P	GVA8400
S. s. bernardezi	Spain: Cabañaquinta	43.15	-5.60	P	GVA4937
S. s. bernardezi	Spain: Castropol	43.46	-6.92	P	GVA8370
S. s. bernardezi	Spain: Coaña	43.51	-6.75	P	GVA8398
S. s. bernardezi	Spain: Gijón	43.55	-5.66	P	GVA4980
S. s. bernardezi	Spain: Inguanzo	43.31	-4.86	P	GVA5029
S. s. bernardezi	Spain: La Castañar	43.20	-5.87	P	GVA3627
S. s. bernardezi	Spain: La Cueta	43.00	-6.19	P	GVA3600
S. s. bernardezi	Spain: La Espina	43.42	-6.34	P	GVA4532
S. s. bernardezi	Spain: Marea	43.29	-5.42	P	GVA4951
S. s. bernardezi	Spain: Mondoñedo	43.39	-7.31	P	GVA3674
S. s. bernardezi	Spain: O Vicedo	43.67	-7.71	P	GVA8306
S. s. bernardezi	Spain: Oviedo	43.36	-5.87	P	GVA3733
S. s. bernardezi	Spain: Pigueña	43.16	-6.32	P	GVA3618
S. s. bernardezi	Spain: Posada	43.46	-5.86	P	GVA7263
S. s. bernardezi	Spain: Restriello	43.30	-6.19	P	GVA8588
S. s. bernardezi	Spain: Ribadesella	43.44	-5.05	P	GVA3798
S. s. bernardezi	Spain: Somiedo	43.10	-6.26	P	GVA8748
S. s. bernardezi	Spain: Tendi	43.20	-5.15	P	GVA3546
S. s. bernardezi	Spain: Trubia	43.33	-5.95	P	GVA8502

S. s. bernardezi	Spain: Tuiza	43.03	-5.91	P	GVA3533
S. s. bernardezi	Spain: Villamar	43.37	-5.91	P	GVA6264
S. s. beschovi	Bulgaria: Zelenigrad	42.84	22.55	L	**GVA5735**
S. s. crespoi	Portugal: Odemira	37.73	-8.64	L	GVA5527
S. s. crespoi	Portugal: Tavira	37.42	-7.80	L	**GVA5521**
S. s. fastuosa	Spain: Hijas	43.26	-3.98	L	GVA9190
S. s. fastuosa	Spain: Beasain	43.07	-2.24	L	**GVA4097**
S. s. fastuosa	Spain: Larrau	43.00	-1.07	L	GVA6099
S. s. gallaica	Spain: Bárcena de la Abadía	42.79	-6.63	L	GVA3031
S. s. gallaica	Spain: Coiro	42.30	-8.77	L	GVA1619
S. s. gallaica	Spain: Cotorredondo	42.36	-8.68	L	GVA1076
S. s. gallaica	Spain: Melide	42.25	-8.86	L	GVA1683
S. s. gallaica	Spain: Melide	42.25	-8.86	L	**GVA1684**
S. s. gallaica	Spain: Melide	42.25	-8.86	L	GVA1688
S. s. gallaica	Spain: Monteferrro	42.15	-8.84	L	GVA0720
S. s. gallaica	Spain: Monteferrro	42.15	-8.84	L	GVA0721
S. s. gallaica	Spain: Nerga	42.26	-8.82	L	GVA1610
S. s. gallaica	Spain: Nerga	42.26	-8.82	L	GVA1612
S. s. gallaica	Spain: O Grove	42.47	-8.89	L	GVA0596
S. s. gallaica	Spain: O Grove	42.47	-8.89	L	GVA0597
S. s. gallaica	Spain: O Grove	42.47	-8.89	L	GVA0630
S. s. gallaica	Spain: Ons	42.37	-8.93	P	GVA_ONS25
S. s. gallaica	Spain: Ons	42.37	-8.93	P	GVA_ONS26
S. s. gallaica	Spain: Ons	42.37	-8.93	P	GVA8752
S. s. gallaica	Spain: Ons	42.37	-8.93	P	GVA8753
S. s. gallaica	Spain: Ons	42.37	-8.93	P	GVA8754
S. s. gallaica	Spain: Ons	42.37	-8.93	P	GVA8755
S. s. gallaica	Spain: Ouriz	42.39	-7.80	L	GVA4035
S. s. gallaica	Spain: San Martiño	42.20	-8.91	P	GVA_CIES26
S. s. gallaica	Spain: San Martiño	42.20	-8.91	P	GVA5719
S. s. gallaica	Spain: San Martiño	42.20	-8.91	P	GVA5720
S. s. gallaica	Spain: San Martiño	42.20	-8.91	P	GVA5721
S. s. gallaica	Spain: San Martiño	42.20	-8.91	P	GVA5722
S. s. gallaica	Spain: San Martiño	42.20	-8.91	P	GVA5723
S. s. gallaica	Spain: Saramagal	42.13	-8.69	L	GVA0766
S. s. gallaica	Spain: Vigo	42.23	-8.71	L	GVA0071
S. s. gigliolii	Italy: La Spezia	44.09	9.78	L	**SAL0255**
S. s. gigliolii	Italy: Serra San Bruno	38.55	16.31	L	SAL0114
S. s. longirostris	Spain: Medina Siotonia	36.46	-5.89	L	GVA5639
S. s. longirostris	Spain: Villanueva del Rosario	36.98	-4.35	L	**GVA6248**
S. s. morenica	Spain: Sierra del Relumbar	38.61	-2.69	L	GVA5471
S. s. salamandra	Poland: Czarmorzeki	49.75	21.80	L	**GVA6006**
S. s. terrestris	Deutschland: Schwarzwald	47.80	7.95	L	**GVA6582**
S. s. terrestris	France: Jublains	48.24	-0.56	L	GVA5376
S. s. werneri	Greece: Prespes	40.78	21.27	L	**GVA9206**

Table B2: List of pregnant females and known offspring. List includes known data from the literature and new data collected for this study.

Subspecies	Locality	Latitude	Longitude	Mode	Reference/Sample ID
S. a. tingitana	Amsa (N=4)	35.53	-5.20	P	Dinis & Velo-Antón 2017
S. a. tingitana	Yebel Musa (N=1)	35.90	-5.40	P	Donaire-Barroso and Bogaerts, 2000
S. a. tingitana	Tleta Taghramet (N=1)	35.80	-5.50	P	Donaire-Barroso and Bogaerts, 2001
S. s. bernardezi	San Julián de Bimenes (N=1)	43.33	-5.56	P	Buckley et al. 2007; Velo-Antón et al. 2015
S. s. bernardezi	Monasterio de Hermo (N=1)	42.97	-6.54	P	Buckley et al. 2007; Velo-Antón et al. 2015
S. s. bernardezi	Brañes, Oviedo (N=1)	43.41	-5.91	P	Buckley et al. 2007; Velo-Antón et al. 2015
S. s. bernardezi	San Miguel de Lillo (N=1)	43.38	-5.86	P	Buckley et al. 2007; Velo-Antón et al. 2015
S. s. bernardezi	Oviedo (N=10)	43.36	-5.85	P	Buckley et al. 2007; Velo-Antón et al. 2015
S. s. bernardezi	Pico Cuadrazales (N=1)	43.23	-5.15	P	Buckley et al. 2007; Velo-Antón et al. 2015
S. s. fastuosa	Les Cauterets (N=1)	42.88	-0.11	P	Joly 1968
S. s. bernardezi	Mondoñedo (N=1)	43.42	-7.36	P	Gasser 1978
S. s. fastuosa	Urgull (N=8)	43.32	-1.99	P	Uotila et al. 2013
S. s. fastuosa	Aiako Harria (N=2)	43.28	-1.80	L	Uotila et al. 2013
S. s. fastuosa	Asteasu (N=4)	43.19	-2.10	L	Uotila et al. 2013
S. s. fastuosa	Landarbaso (N=3)	43.25	-1.89	L/P	Uotila et al. 2013
S. s. bernardezi	Lugo: Serra do Xistral	43.49	-7.56	P	Galán et al. 2007
S. s. gallaica	Ons (N=12)	42.37	-8.94	P	Velo-Antón et al. 2015
S. s. gallaica	San Martiño (N=2)	42.20	-8.90	P	Velo-Antón et al. 2015
S. s. bernardezi	Pie de Sierra	43.38	-4.64	P	GVA9311
S. s. bernardezi	Tendi	43.31	-5.25	P	GVA9312
S. s. bernardezi	Tendi	43.31	-5.25	P	GVA9313
S. s. bernardezi	Tendi	43.31	-5.25	P	GVA9314
S. s. bernardezi	Pie de Sierra	43.38	-4.64	P	GVA9315
S. s. bernardezi	Piedra	43.35	-4.92	P	GVA9316
S. s. bernardezi	Fuensanta	43.35	-5.48	P	GVA9317
S. s. bernardezi	Fito	43.44	-5.20	P	GVA9318
S. s. bernardezi	Fito	43.44	-5.20	P	GVA9319
S. s. bernardezi	Fario	43.43	-5.57	P	GVA9320
S. s. bernardezi	Piedra	43.35	-4.92	P	GVA9321
S. s. bernardezi	Piedra	43.35	-4.92	P	GVA9322
S. s. bernardezi	Tiñana	43.37	-5.75	P	GVA9323
S. s. bernardezi	La Pesanca	43.27	-5.34	P	GVA9324
S. s. bernardezi	Jolagua (Mueñegru)	43.32	-4.93	P	GVA9325
S. s. bernardezi	Piedra	43.35	-4.92	P	GVA9326
S. s. bernardezi	Color	43.31	-5.26	P	GVA9328
S. s. bernardezi	Buferrera	43.28	-4.99	P	GVA9329
S. s. bernardezi	Buferrera	43.28	-4.99	P	GVA9330
S. s. bernardezi	Buferrera	43.28	-4.99	P	GVA9331
S. s. bernardezi	Andara	43.22	-4.72	P	GVA9332
S. s. bernardezi	Andara	43.22	-4.72	P	GVA9333
S. s. bernardezi	Andara	43.22	-4.72	P	GVA9334
S. s. bernardezi	Ramoniel	43.42	-4.95	P	GVA9335
S. s. bernardezi	Ramoniel	43.42	-4.95	P	GVA9336
S. s. bernardezi	Brañes	43.41	-5.92	P	GVA9337
S. s. bernardezi	O Vicedo	43.64	-7.74	P	GVA8950
S. s. bernardezi	Brañes	43.41	-5.92	P	GVA9338
S. s. bernardezi	Brañes	43.41	-5.92	P	GVA9339
S. s. bernardezi	Llerandi	43.31	-5.21	P	GVA9340
S. s. bernardezi	Llerandi	43.31	-5.21	P	GVA9341

S. s. bernardezi	Llerandi	43.31	-5.21	P	GVA9342
S. s. bernardezi	Pimiango	43.40	-4.53	P	GVA9343
S. s. bernardezi	Piedra	43.35	-4.92	P	GVA9344
S. s. fastuosa	Bárcena Mayor	43.13	-4.17	L	GVA9327
S. s. bernardezi	Somiedo	43.11	-6.32	P	x
S. s. fastuosa	Cantabria, La Cotera	43.27	-4.41	L	GVA9184

Table B3: List of genes included in the sequence array. The majority of loci (1,287) stem from Rodriguez et al. 2017, see first line in bold.

Gene	Source	GenBank	length (bps)	Full gene name
1287 cDNA transcripts	**Rodríguez et al. 2017**	**-**	**109-338**	**-**
H3	GenBank	DQ284416	301	histone H3a
POMC	GenBank	KF645798	312	proopiomelanocortin
RHO	GenBank	DQ347354	316	rhodopsin
BDNF	GenBank	EF453369	367	brain-derived neurotrophic factor
PDGFRa	GenBank	KF645656	498	platelet-derived growth factor receptor alpha
SLC8A3	GenBank	KU295307	513	solute carrier family 8 member 3
RAG-2	GenBank	KF645724	585	recombination activating protein 2
CXCR4	GenBank	KU295299	601	chemokine receptor 4
NCX1	GenBank	KU295326	607	sodium/calcium exchanger 1
Bfib	GenBank	KU295332	664	beta-fibrinogen
SACS	GenBank	KF645438	682	spastic ataxia of Charlevoix-Saguenay
TTN	GenBank	KF645452	706	titin
KIAA1239	GenBank	KF645474	785	KIAA1239 protein
RAG-1	GenBank	AY650135	1540	recombination activating protein 1
CytB	Mulder et al. 2016	KX094979	130	Cytochrome B
COI	Mulder et al. 2016	KX094979	130	cytochrome oxidase subunit 1
ENC1	Shen et al. 2013	KC165218	255	ectodermal-neural cortex 1
FAT4	Shen et al. 2013	KC165261	300	FAT tumor suppressor-like protein 4
FICD	Shen et al. 2013	KC165283	300	FIC domain-containing protein
PANX2	Shen et al. 2013	KC165541	300	pannexin 2
KIAA2013	Shen et al. 2013	KC165412	323	KIAA2013-like protein
BPTF	Shen et al. 2013	KC165070	376	bromodomain PHD finger transcription factor
CAND1	Shen et al. 2013	KC165090	400	cullin-associated and neddylation-dissociated
DET1	Shen et al. 2013	KC165112	400	de-etiolated-like protein 1
DISP1	Shen et al. 2013	KC165134	400	dispatched-like protein 1
DNAH3	Shen et al. 2013	KC165156	400	dynein axonemal heavy chain 3
DOLK	Shen et al. 2013	KC165174	400	dolichol kinase
DSEL	Shen et al. 2013	KC165196	400	dermatan sulfate epimerase-like protein
GRM2	Shen et al. 2013	KC165303	400	glutamate receptor metabotropic 2
KCNF1	Shen et al. 2013	KC165370	400	potassium voltage-gated channel, subfamily F member 1
LIG4	Shen et al. 2013	KC165432	400	ATP-dependent ligase IV
LRRN1	Shen et al. 2013	KC165474	400	leucine rich repeat neuronal 1
PDP1	Shen et al. 2013	KC165559	400	pyruvate dehyrogenase phosphatase catalytic subunit 1
POZ	Shen et al. 2013	KC165348	400	kelch repeat and BTB domain-containing 2 protein
MGAT4C	Shen et al. 2013	KC165495	442	mannosyl alpha-1,3--glycoprotein beta-1,4-N-acetylglucosaminyltransferase isozyme C
LPHN2	Shen et al. 2013	KC165454	506	latrophilin 2
-	Shen et al. 2013	KC165325	586	hypothetical protein gene
MIOS	Shen et al. 2013	KC165516	651	missing oocyte meiosis regulator-like protein
ZBED4	Shen et al. 2013	KC165682	704	zinc finger BED-type containing 4
EXTL3	Shen et al. 2013	KC165241	1000	exostoses multiple-like 3 protein
PPL	Shen et al. 2013	KC165581	1169	periplakin

Figure B1: Dated Bayesian inference based on a concatenated dataset of 574k bps using BEAST 2.6.0 applying a strict molecular clock under the constant coalescent population model. Timescale in million years before present.

Figure B2: DensiTree plot of 45000 generated SNAPP trees of the 6 individuals found closest to the type locality of the respective species. The most common topology is in dark grey (34%), and the second most common in yellow (31%). Remaining topologies have been removed for clarity. On the right is the consensus tree as generated by tree-annotator and with posterior probabilities indicated on the nodes.

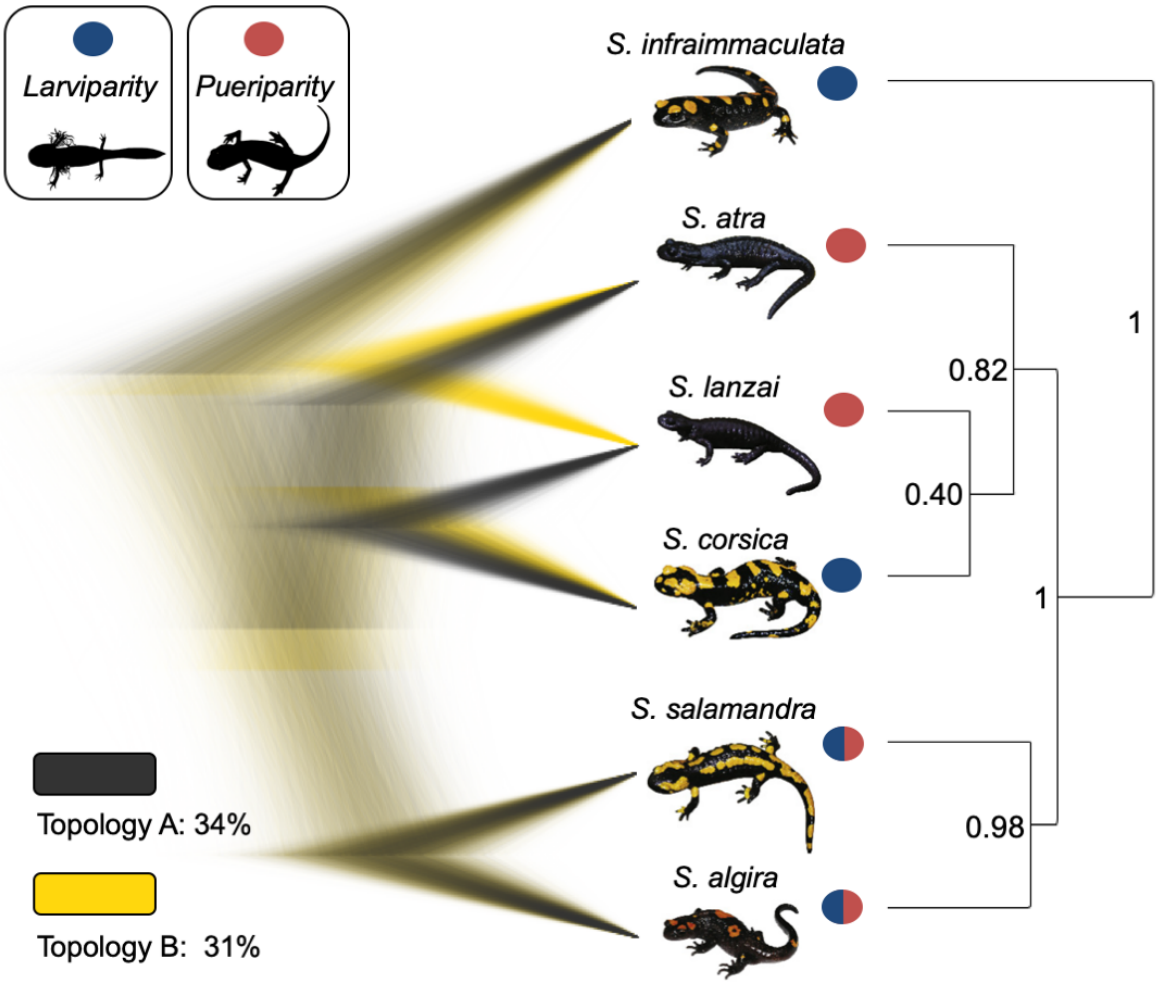


Figure B3: Maximum likelihood phylogenetic reconstruction based on a concatenated dataset of 574k bps in RAxML 8.1.12. Black dots indicate >99 bootstrap support, dark grey >90 and light grey >80.

Figure B4: Bayesian inference of 802 bps of the Cytochrome B alignment. Samples include all 94 samples included in the phylogenetic reconstructions and the 37 females assessed for reproductive strategy. In blue are the females grouping in the larger bernardezi clade, in red are three females showing mitochondrial introgression, but morphology and locality show clear *S. s. bernardezi* ancestry. The figure shows that pueriparity is widespread across the mitochondrial lineages and is likely the prevailing strategy in this clade.

APPENDIX C: SUPPLEMENTAL INFORMATION FOR CHAPTER 3

Table C1: List of samples included in Chapter 3. The first 14 samples (GVA6688 and GVA6082) were only used to generate the reference transcriptome.

Sample	Individual	Collection Date	Tissue	Location	Latitude	Longitude	SVL (mm)	Uterus & Oviduct content	RIN	Conc (ng/ul)	Instrument + Readlength	Raw Reads
140815Heart	GVA6688	April 9, 2015	Heart	Mindelo	41.327	-8.729	116	-	9.7	38	HiSeq1500, 2 x 125	42,159,169
140815Kidney	GVA6688	April 9, 2015	Kidney	Mindelo	41.327	-8.729	116	-	9.7	115	HiSeq1500, 2 x 125	31,827,814
140815Liver	GVA6688	April 9, 2015	Liver	Mindelo	41.327	-8.729	116	-	9.6	38	HiSeq1500, 2 x 125	27,727,629
140815Lung	GVA6688	April 9, 2015	Lung	Mindelo	41.327	-8.729	116	-	8.8	32	HiSeq1500, 2 x 125	36,502,912
140815Muscle	GVA6688	April 9, 2015	Muscle	Mindelo	41.327	-8.729	116	-	9.3	5	HiSeq1500, 2 x 125	25,700,526
140815Oviduct	GVA6688	April 9, 2015	Oviduct	Mindelo	41.327	-8.729	116	-	8.5	11	HiSeq1500, 2 x 125	35,753,721
140815Uterus	GVA6688	April 9, 2015	Uterus	Mindelo	41.327	-8.729	116	eggs	9.3	21	HiSeq1500, 2 x 125	25,735,217
IAGVA001A	GVA6082	April 11, 2016	Heart	Ons	42.370	-8.930	107	-	8.6	82	HiSeq4000, 2 x 150	50,799,067
IAGVA001B	GVA6082	April 11, 2016	Kidney	Ons	42.370	-8.930	107	-	7.3	73	HiSeq4000, 2 x 150	62,553,010
IAGVA001C	GVA6082	April 11, 2016	Liver	Ons	42.370	-8.930	107	-	9.5	253	HiSeq4000, 2 x 150	36,941,075
IAGVA001D	GVA6082	April 11, 2016	Lung	Ons	42.370	-8.930	107	-	9.0	168	HiSeq4000, 2 x 150	46,887,526
IAGVA001E	GVA6082	April 11, 2016	Muscle	Ons	42.370	-8.930	107	-	9.8	22	HiSeq4000, 2 x 150	58,846,533
IAGVA001F	GVA6082	April 11, 2016	Oviduct	Ons	42.370	-8.930	107	-	9.0	178	HiSeq4000, 2 x 150	38,642,245
IAGVA001G	GVA6082	April 11, 2016	Uterus	Ons	42.370	-8.930	107	empty	8.8	72	HiSeq4000, 2 x 150	53,829,469
Coiro2_LO	GVA6713	Oct 14, 2016	Oviduct	Coiro	42.296	-8.769	120	-	8.0	107	HiSeq4000, 2 x 100	20,031,562
Coiro2_LU	GVA6713	Oct 14, 2016	Uterus	Coiro	42.296	-8.769	120	10 larvae	9.3	127	HiSeq4000, 2 x 100	22,629,139
Coiro2_RU	GVA6713	Oct 14, 2016	Uterus	Coiro	42.296	-8.769	120	13 larvae	8.4	114	HiSeq4000, 2 x 100	15,211,224
Coiro3_LO	GVA6729	Oct 14, 2016	Oviduct	Coiro	42.296	-8.769	113	-	7.6	50	HiSeq4000, 2 x 100	4,067,397
Coiro3_LU	GVA6729	Oct 14, 2016	Uterus	Coiro	42.296	-8.769	113	10 larvae	9.4	132	HiSeq4000, 2 x 100	19,250,836
Coiro3_RU	GVA6729	Oct 14, 2016	Uterus	Coiro	42.296	-8.769	113	13 larvae	9.3	155	HiSeq4000, 2 x 100	12,109,634
MF4_LO	GVA6716	Oct 14, 2016	Oviduct	Monteferro	42.154	-8.841	125	-	8.7	110	HiSeq4000, 2 x 100	21,804,863
MF4_LU	GVA6716	Oct 14, 2016	Uterus	Monteferro	42.154	-8.841	125	many eggs	9.6	76	HiSeq4000, 2 x 100	19,429,013
MF4_RU	GVA6716	Oct 14, 2016	Uterus	Monteferro	42.154	-8.841	125	7 larvae	8.1	189	HiSeq4000, 2 x 100	25,799,233
MF5_LO	GVA6719	Oct 14, 2016	Oviduct	Monteferro	42.154	-8.841	137	-	9.0	102	HiSeq4000, 2 x 100	18,304,316
MF5_LU	GVA6719	Oct 14, 2016	Uterus	Monteferro	42.154	-8.841	137	12 eggs	9.1	101	HiSeq4000, 2 x 100	13,326,298
MF5_RU	GVA6719	Oct 14, 2016	Uterus	Monteferro	42.154	-8.841	137	5 larvae + 20 eggs	9.0	69	HiSeq4000, 2 x 100	24,807,115

Ons4_LO	GVA6715	Oct 14, 2016	Oviduct	Ons	42.370	-8.930	108	-	8.1	63	HiSeq4000, 2 x 100	15,976,935
Ons4_LU	GVA6715	Oct 14, 2016	Uterus	Ons	42.370	-8.930	108	5juv + 1larvae	9.4	124	HiSeq4000, 2 x 100	13,128,118
Ons4_RU	GVA6715	Oct 14, 2016	Uterus	Ons	42.370	-8.930	108	5juv + 1larvae	8.0	218	HiSeq4000, 2 x 100	24,928,197
Ons6_LO	GVA6714	Oct 14, 2016	Oviduct	Ons	42.370	-8.930	102	-	8.0	153	HiSeq4000, 2 x 100	28,639,561
Ons6_LU	GVA6714	Oct 14, 2016	Uterus	Ons	42.370	-8.930	102	9 juv	9.5	155	HiSeq4000, 2 x 100	16,959,061
Ons6_RU	GVA6714	Oct 14, 2016	Uterus	Ons	42.370	-8.930	102	10 juv	9.7	136	HiSeq4000, 2 x 100	21,440,160
Ons9_LO	GVA6717	Oct 14, 2016	Oviduct	Ons	42.370	-8.930	114	-	9.6	250	HiSeq4000, 2 x 100	22,209,776
Ons9_LU	GVA6717	Oct 14, 2016	Uterus	Ons	42.370	-8.930	114	9 juv	9.6	122	HiSeq4000, 2 x 100	15,464,891
Ons9_RU	GVA6717	Oct 14, 2016	Uterus	Ons	42.370	-8.930	114	3 juv + 2 tiny larvae	9.0	200	HiSeq4000, 2 x 100	15,199,510
Ora2_LO	GVA6818	Oct 26, 2016	Oviduct	Villablino	42.963	-6.349	102	-	8.9	220	HiSeq4000, 2 x 100	15,736,669
Ora2_LU	GVA6818	Oct 26, 2016	Uterus	Villablino	42.963	-6.349	102	23 small larvae	8.5	253	HiSeq4000, 2 x 100	20,618,464
Ora2_RU	GVA6818	Oct 26, 2016	Uterus	Villablino	42.963	-6.349	102	26 larvae	8.0	108	HiSeq4000, 2 x 100	13,822,428
Ora3_LO	GVA6816	Oct 26, 2016	Oviduct	Villablino	42.963	-6.349	116	-	9.4	101	HiSeq4000, 2 x 100	22,278,969
Ora3_LU	GVA6816	Oct 26, 2016	Uterus	Villablino	42.963	-6.349	116	21 larvae + 2 eggs	8.2	215	HiSeq4000, 2 x 100	24,999,415
Ora3_RU	GVA6816	Oct 26, 2016	Uterus	Villablino	42.963	-6.349	116	13 larvae	9.5	89	HiSeq4000, 2 x 100	15,502,357
Ora5_LO	GVA6821	Oct 26, 2016	Oviduct	Villablino	42.963	-6.349	100	-	9.0	200	HiSeq4000, 2 x 100	14,106,992
Ora5_LU	GVA6821	Oct 26, 2016	Uterus	Villablino	42.963	-6.349	100	13 larvae (1 small)	8.8	171	HiSeq4000, 2 x 100	18,033,373
Ora5_RU	GVA6821	Oct 26, 2016	Uterus	Villablino	42.963	-6.349	100	8 larvae (2 small)	7.8	116	HiSeq4000, 2 x 100	13,542,145
Som1_LO	GVA6815	Oct 26, 2016	Oviduct	Somiedo	43.100	-6.256	112	-	8.5	82	HiSeq4000, 2 x 100	13,887,283
Som1_LU	GVA6815	Oct 26, 2016	Uterus	Somiedo	43.100	-6.256	112	empty	9.0	54	HiSeq4000, 2 x 100	23,309,980
Som1_RU	GVA6815	Oct 26, 2016	Uterus	Somiedo	43.100	-6.256	112	empty	9.3	85	HiSeq4000, 2 x 100	13,626,228
Som3_LO	GVA6813	Oct 26, 2016	Oviduct	Somiedo	43.100	-6.256	103	-	8.7	168	HiSeq4000, 2 x 100	19,372,814
Som3_LU	GVA6813	Oct 26, 2016	Uterus	Somiedo	43.100	-6.256	103	7 larvae	9.1	142	HiSeq4000, 2 x 100	13,438,548
Som3_RU	GVA6813	Oct 26, 2016	Uterus	Somiedo	43.100	-6.256	103	5 larvae	9.5	81	HiSeq4000, 2 x 100	13,718,067
Som5_LO	GVA6817	Oct 26, 2016	Oviduct	Somiedo	43.100	-6.256	103	-	9.3	171	HiSeq4000, 2 x 100	14,457,645
Som5_LU	GVA6817	Oct 26, 2016	Uterus	Somiedo	43.100	-6.256	103	2 advanced larvae	7.5	95	HiSeq4000, 2 x 100	26,369,036
Som5_RU	GVA6817	Oct 26, 2016	Uterus	Somiedo	43.100	-6.256	103	4 larvae	9.3	160	HiSeq4000, 2 x 100	15,724,372

Table C2: EdgeR results for all uterus samples, only the top 100 indicated. LogFC is the Log of the Fold-Change, LogCPM is the Log of the Counts Per Million, p-value is the exact p-value for differential expression and the FDR is the significance value after correcting for multiple testing by means of the Benjamini & Hochberg false discovery rate.

n	Gene	logFC	logCPM	p-value	FDR
1	MCM6	2.249	5.645	2.66E-10	3.47E-06
2	PDGFD	-1.453	4.302	1.81E-07	1.18E-03
3	MTNR1A	-3.818	1.502	2.80E-07	1.22E-03
4	NFU1	-1.142	6.494	5.00E-07	1.52E-03
5	B9D1	0.895	4.794	5.83E-07	1.52E-03
6	PNLIPRP1	-6.075	4.405	8.31E-07	1.67E-03
7	DDX19B	-0.923	6.776	1.02E-06	1.67E-03
8	TMEM56	-6.886	2.340	1.12E-06	1.67E-03
9	STPG4	-4.707	1.706	1.35E-06	1.67E-03
10	MTUS1	1.035	5.630	1.41E-06	1.67E-03
11	NMUR3	-7.602	1.677	1.42E-06	1.67E-03
12	TPPP3	1.919	8.482	1.54E-06	1.67E-03
13	CDKL1	3.754	3.189	1.69E-06	1.69E-03
14	THTPA	-1.758	2.837	2.14E-06	1.78E-03
15	BBL030307	-3.212	4.639	2.19E-06	1.78E-03
16	GIMAP4	-3.195	4.195	2.22E-06	1.78E-03
17	CLDN23	-1.629	4.848	2.32E-06	1.78E-03
18	CCDC12	-1.453	4.359	2.54E-06	1.84E-03
19	PSMC3	-0.871	7.879	3.33E-06	2.29E-03
20	EFCAB7	0.755	4.727	3.85E-06	2.42E-03
21	ZC2HC1C	1.885	3.348	3.90E-06	2.42E-03
22	FAAH2	-2.020	4.348	4.20E-06	2.49E-03
23	STK33	1.940	2.712	5.31E-06	3.01E-03
24	TOR3A	1.626	3.758	8.17E-06	4.44E-03
25	SCN4B	3.864	-0.566	9.57E-06	4.94E-03
26	TMEM121	-1.432	5.903	9.85E-06	4.94E-03
27	EAPP	-1.049	5.533	1.08E-05	5.00E-03
28	CCDC175	4.143	5.224	1.09E-05	5.00E-03
29	MCHR2	9.873	4.105	1.18E-05	5.00E-03
30	LAMTOR2	-0.991	7.372	1.18E-05	5.00E-03
31	RHCG	-1.432	10.534	1.19E-05	5.00E-03
32	EDAR	2.355	0.913	1.24E-05	5.03E-03
33	TIGAR	1.084	5.089	1.27E-05	5.03E-03
34	DCN	1.449	7.643	1.32E-05	5.06E-03
35	TGFB3	-1.184	4.308	1.42E-05	5.26E-03
36	SNX15	-1.261	4.425	1.45E-05	5.26E-03
37	NMB	5.879	2.169	1.57E-05	5.49E-03
38	MFSD7	-0.920	4.254	1.60E-05	5.49E-03
39	GDPD5	-0.958	5.516	1.65E-05	5.51E-03
40	CKBA	1.164	7.907	1.71E-05	5.59E-03
41	A2ML1	-1.830	13.428	2.02E-05	6.41E-03
42	GLRA1	-2.817	1.035	2.08E-05	6.43E-03
43	TMA16	-1.797	5.150	2.14E-05	6.43E-03
44	BBL013153	-0.881	8.826	2.17E-05	6.43E-03
45	ZFYVE1	-0.579	6.178	2.29E-05	6.63E-03
46	GATA2	-1.128	9.345	2.43E-05	6.89E-03
47	GCK	5.259	3.056	2.57E-05	7.12E-03

48	KLF4	2.385	5.671	2.68E-05	7.17E-03
49	LUM	1.318	8.152	2.69E-05	7.17E-03
50	BBL015221	-3.643	2.919	2.89E-05	7.43E-03
51	KIF19	2.078	7.682	2.90E-05	7.43E-03
52	SEC16B	2.757	3.037	3.02E-05	7.58E-03
53	RPGRIP1L	0.765	3.566	3.28E-05	7.70E-03
54	TDO2	5.448	0.152	3.31E-05	7.70E-03
55	HGF	2.120	3.584	3.31E-05	7.70E-03
56	HIF3A	-1.461	8.829	3.35E-05	7.70E-03
57	SULFOTRANSFERASE	-2.460	6.943	3.37E-05	7.70E-03
58	DISP2	-1.018	5.595	3.48E-05	7.82E-03
59	DDX6	-0.554	7.419	3.54E-05	7.82E-03
60	LRRC3CA	-1.812	3.185	3.63E-05	7.90E-03
61	GALC	-1.140	7.913	4.02E-05	8.24E-03
62	RNF182	-2.353	3.695	4.02E-05	8.24E-03
63	FGFBP1	3.255	4.826	4.05E-05	8.24E-03
64	BBL013387	-1.032	4.376	4.17E-05	8.24E-03
65	IFT46	0.752	5.681	4.17E-05	8.24E-03
66	PTGER4	1.213	4.677	4.20E-05	8.24E-03
67	CAR15	7.213	2.332	4.27E-05	8.24E-03
68	ZC2HC1A	0.641	4.931	4.30E-05	8.24E-03
69	FOXQ1A	-1.849	4.554	4.36E-05	8.24E-03
70	TRNAU1AP	-0.622	5.226	4.52E-05	8.42E-03
71	TBC1D22A	-0.506	9.160	5.01E-05	9.15E-03
72	TNFAIP6	2.262	1.259	5.05E-05	9.15E-03
73	TMEM62	-1.106	5.143	5.20E-05	9.29E-03
74	CAMK4	1.829	2.621	5.35E-05	9.43E-03
75	POP5	-1.057	3.988	5.54E-05	9.57E-03
76	SLC9A3	6.284	0.745	5.58E-05	9.57E-03
77	NOV	1.931	5.826	5.80E-05	9.82E-03
78	KIF22	2.065	3.797	5.93E-05	9.92E-03
79	INSIG2	-0.639	6.056	6.03E-05	9.95E-03
80	WBP1	-0.592	8.564	6.20E-05	1.01E-02
81	VWA8	-0.781	4.224	6.47E-05	1.04E-02
82	PHYHIPL	1.924	3.017	6.54E-05	1.04E-02
83	ZNF367	2.114	2.170	6.79E-05	1.05E-02
84	TMEM147	-0.698	6.033	6.87E-05	1.05E-02
85	CLDN10	5.594	0.666	6.95E-05	1.05E-02
86	CETN1	-0.764	4.420	7.06E-05	1.05E-02
87	NFIL3-6	-1.508	6.760	7.12E-05	1.05E-02
88	RAB40B	-1.360	5.574	7.14E-05	1.05E-02
89	BPIFCL	-1.766	8.535	7.18E-05	1.05E-02
90	BBL026939	2.117	3.279	7.37E-05	1.07E-02
91	ANGPTL1	1.867	7.267	7.58E-05	1.09E-02
92	MAP9	0.701	3.270	7.82E-05	1.11E-02
93	ALKBH6	-1.273	8.055	7.88E-05	1.11E-02
94	NELL2	-3.730	4.030	8.42E-05	1.17E-02
95	CAPN10	-0.989	2.880	8.65E-05	1.19E-02
96	SPEF1	3.207	4.794	8.75E-05	1.19E-02
97	EMP3	1.855	5.451	8.99E-05	1.21E-02
98	CFP	2.321	3.594	9.31E-05	1.21E-02
99	PSMG3	-0.729	5.999	9.37E-05	1.21E-02
100	HS3ST2	4.347	-0.079	9.38E-05	1.21E-02

Table C3: EdgeR results for all oviduct samples, only the top 100 indicated. LogFC is the Log of the Fold-Change, LogCPM is the Log of the Counts Per Million, p-value is the exact p-value for differential expression and the FDR is the significance value after correcting for multiple testing by means of the Benjamini & Hochberg false discovery rate.

n	Gene	logFC	logCPM	p-value	FDR
1	SNCB	2.696	1.849	1.29E-04	5.94E-01
2	MCM6	2.109	5.229	2.02E-04	5.94E-01
3	CEACAM8	4.551	3.649	2.13E-04	5.94E-01
4	NMUR3	-7.206	2.698	2.43E-04	5.94E-01
5	GP5	3.504	2.711	2.81E-04	5.94E-01
6	GP9	4.194	2.400	3.19E-04	5.94E-01
7	GIMAP4	-4.412	3.332	3.97E-04	5.94E-01
8	MCHR2	7.480	3.022	4.44E-04	5.94E-01
9	GCK	6.401	3.089	4.78E-04	5.94E-01
10	SLC1A2	5.189	-0.062	5.15E-04	5.94E-01
11	VWA5B2	-2.313	3.190	5.54E-04	5.94E-01
12	PRL2	-5.960	1.836	5.83E-04	5.94E-01
13	BBL048454	-5.387	-0.374	5.92E-04	5.94E-01
14	NFU1	-0.907	6.398	7.66E-04	6.51E-01
15	BBL026937	6.406	1.447	9.31E-04	6.51E-01
16	FARSB	0.838	8.018	1.01E-03	6.51E-01
17	CDKL2	-3.424	0.822	1.06E-03	6.51E-01
18	TMEM232	-4.299	5.848	1.14E-03	6.51E-01
19	HAAO	-5.545	1.529	1.16E-03	6.51E-01
20	CHRNB2	5.904	1.206	1.22E-03	6.51E-01
21	PARP3	-0.855	8.622	1.23E-03	6.51E-01
22	PDE2A	2.645	4.761	1.25E-03	6.51E-01
23	PRIM2	-1.192	3.761	1.26E-03	6.51E-01
24	MFNG	4.870	0.110	1.36E-03	6.51E-01
25	SUN2	2.236	4.107	1.42E-03	6.51E-01
26	PTTG1	-2.007	3.948	1.47E-03	6.51E-01
27	KLF17	2.025	3.951	1.51E-03	6.51E-01
28	AKAP8L	1.016	9.872	1.52E-03	6.51E-01
29	FCN1	5.345	2.248	1.62E-03	6.51E-01
30	COL12A1	-1.973	3.980	1.64E-03	6.51E-01
31	BBL027510	3.478	0.007	1.68E-03	6.51E-01
32	HSPB1	3.820	1.534	1.74E-03	6.51E-01
33	PTPRJ	-5.856	-0.037	1.75E-03	6.51E-01
34	ATR	0.872	4.131	1.83E-03	6.51E-01
35	TNFRSF11B	3.676	0.200	1.84E-03	6.51E-01
36	SERPINC1	-8.253	3.003	1.88E-03	6.51E-01
37	OC90	-7.139	2.644	1.92E-03	6.51E-01
38	VPREB1	5.561	1.660	1.92E-03	6.51E-01
39	SGCG	-3.343	3.850	1.99E-03	6.51E-01
40	NKX2-5	1.642	2.258	2.07E-03	6.51E-01
41	SERPINF1	-3.198	7.932	2.10E-03	6.51E-01
42	BBL024721	2.539	2.317	2.18E-03	6.51E-01
43	SGIP1	-1.558	4.438	2.23E-03	6.51E-01
44	HPGD	-1.567	5.188	2.24E-03	6.51E-01
45	GP1BB	2.920	1.762	2.26E-03	6.51E-01
46	STMN3	3.427	1.099	2.33E-03	6.51E-01
47	ZFC3H1	0.770	4.724	2.46E-03	6.51E-01

48	PTGIR	3.208	-1.035	2.50E-03	6.51E-01
49	BBL061588	2.709	4.395	2.50E-03	6.51E-01
50	BBL024513	-6.906	2.545	2.55E-03	6.51E-01
51	FAM219A	2.755	-1.237	2.69E-03	6.51E-01
52	THTPA	-1.992	2.831	2.71E-03	6.51E-01
53	SCX	5.124	0.676	2.73E-03	6.51E-01
54	XB-GENE-5817456	4.327	1.666	2.82E-03	6.51E-01
55	CNTF	1.711	5.295	2.87E-03	6.51E-01
56	AHNAK	1.649	3.090	2.88E-03	6.51E-01
57	CAAP1	1.844	4.441	2.89E-03	6.51E-01
58	CBARP	3.223	1.863	2.90E-03	6.51E-01
59	CABZ01085857.1	2.991	2.509	2.97E-03	6.51E-01
60	DDX19B	-0.936	6.594	3.05E-03	6.51E-01
61	RRN3	2.108	4.242	3.13E-03	6.51E-01
62	DOC2A	2.649	3.424	3.15E-03	6.51E-01
63	HEBP1	-1.360	4.522	3.18E-03	6.51E-01
64	KIF22	3.025	3.307	3.21E-03	6.51E-01
65	DHRS7	-0.778	6.191	3.25E-03	6.51E-01
66	CISD3	0.652	4.901	3.35E-03	6.62E-01
67	TTC9	1.606	4.819	3.51E-03	6.83E-01
68	IRF1	-2.982	5.519	3.67E-03	6.89E-01
69	SMTNL1	4.124	-0.606	3.71E-03	6.89E-01
70	MPL	6.666	0.242	3.73E-03	6.89E-01
71	SNX15	-1.419	3.818	3.76E-03	6.89E-01
72	CBLN2	3.989	1.403	3.83E-03	6.89E-01
73	CPA5	-5.531	4.174	3.89E-03	6.89E-01
74	ABAT	-2.729	4.571	3.91E-03	6.89E-01
75	HOXB3	0.876	4.598	4.02E-03	6.90E-01
76	FAM20B	1.616	2.911	4.02E-03	6.90E-01
77	BBL019666	7.512	0.802	4.12E-03	6.98E-01
78	TMEM121	-1.967	5.621	4.25E-03	7.11E-01
79	FTR04	1.448	1.511	4.43E-03	7.32E-01
80	BBL045744	-6.910	0.758	4.55E-03	7.41E-01
81	KIAA1468	-0.668	5.718	4.60E-03	7.41E-01
82	BBL024722	5.569	0.601	4.91E-03	7.55E-01
83	FAM78B	2.788	-1.111	4.91E-03	7.55E-01
84	ESPN	2.937	-0.919	4.92E-03	7.55E-01
85	D630003M21RIK	2.284	1.603	4.92E-03	7.55E-01
86	HSPB7	2.388	4.170	5.14E-03	7.71E-01
87	B4GALNT2	-4.759	-0.700	5.14E-03	7.71E-01
88	BBL018241	2.793	0.936	5.29E-03	7.79E-01
89	TRAF3IP3	1.344	3.668	5.32E-03	7.79E-01
90	BBL008029	-3.105	-0.544	5.42E-03	7.85E-01
91	NPSN	1.767	2.787	5.57E-03	7.89E-01
92	STPG4	-3.258	3.153	5.68E-03	7.89E-01
93	BNC2	0.675	6.578	5.69E-03	7.89E-01
94	EAPP	-0.964	5.904	5.69E-03	7.89E-01
95	WNT16	5.322	0.455	5.77E-03	7.93E-01
96	NYAP2	-5.247	1.667	6.07E-03	7.97E-01
97	CCDC175	2.556	5.559	6.07E-03	7.97E-01
98	GPR176	3.584	0.178	6.12E-03	7.97E-01
99	SLC52A3	3.565	0.498	6.19E-03	7.97E-01
100	AOX6	-1.087	5.513	6.22E-03	7.97E-01

Table C4: EdgeR results for the uterus samples across the island transition, only the top 100 indicated. LogFC is the Log of the Fold-Change, LogCPM is the Log of the Counts Per Million, p-value is the exact p-value for differential expression and the FDR is the significance value after correcting for multiple testing by means of the Benjamini & Hochberg false discovery rate.

n	Gene	logFC	logCPM	p-value	FDR
1	SCEL	-8.144	3.877	2.50E-07	3.26E-03
2	TIGAR	1.616	5.456	6.63E-07	4.32E-03
3	RFWD3	-1.617	5.532	1.30E-06	5.18E-03
4	BPI	-2.738	4.807	1.66E-06	5.18E-03
5	HDAC10	1.737	4.413	2.26E-06	5.18E-03
6	MCM6	2.017	5.742	2.38E-06	5.18E-03
7	PCBD2	1.774	8.006	3.94E-06	7.27E-03
8	AKAP8L	1.442	9.839	4.46E-06	7.27E-03
9	CACNA1C	-2.106	4.993	6.93E-06	1.00E-02
10	FGL2	-2.209	8.218	7.71E-06	1.01E-02
11	FRMPD1	-5.089	1.101	9.14E-06	1.08E-02
12	BTG2	-1.196	8.929	1.16E-05	1.13E-02
13	SUN2	1.813	4.761	1.17E-05	1.13E-02
14	URGCP	4.090	2.718	1.21E-05	1.13E-02
15	FARSB	1.252	8.109	1.41E-05	1.23E-02
16	SNX15	-1.586	4.501	1.55E-05	1.26E-02
17	TMEM25	5.383	2.705	1.79E-05	1.38E-02
18	CAAP1	2.338	5.284	2.03E-05	1.42E-02
19	CLN6	-1.115	5.918	2.07E-05	1.42E-02
20	FOSL2	-1.627	5.370	2.24E-05	1.44E-02
21	TMEM56	-7.497	2.720	2.37E-05	1.44E-02
22	MRI1	1.197	5.640	2.76E-05	1.44E-02
23	LRGUK	1.551	3.986	2.76E-05	1.44E-02
24	FOXQ1A	-2.944	4.543	2.77E-05	1.44E-02
25	EXTL2	-2.057	3.280	2.77E-05	1.44E-02
26	GPR161	-1.797	3.043	3.71E-05	1.60E-02
27	IFT81	1.891	5.169	3.73E-05	1.60E-02
28	RHEBL1	1.470	3.561	3.75E-05	1.60E-02
29	DOHH	0.948	5.884	3.79E-05	1.60E-02
30	DPM1	0.901	6.032	3.82E-05	1.60E-02
31	EIF3C	0.983	8.525	3.83E-05	1.60E-02
32	ATXN7L1	-1.400	4.621	3.93E-05	1.60E-02
33	CKBA	1.682	8.325	4.60E-05	1.82E-02
34	MRPS14	-1.092	4.794	4.95E-05	1.90E-02
35	BBL022435	8.113	2.604	5.40E-05	2.01E-02
36	BBL031754	-4.136	1.367	7.52E-05	2.72E-02
37	PDGFD	-1.444	4.144	8.02E-05	2.78E-02
38	RSPH4A	2.264	4.394	8.32E-05	2.78E-02
39	FREM1	-0.798	6.639	8.90E-05	2.78E-02
40	STK33	2.872	2.959	8.92E-05	2.78E-02
41	PASK	-1.624	3.802	9.08E-05	2.78E-02
42	PGK1	0.987	8.556	9.09E-05	2.78E-02
43	PHYHIPL	2.382	3.603	9.18E-05	2.78E-02
44	TACR1	2.394	2.879	9.63E-05	2.82E-02
45	TTC12	1.129	5.143	9.88E-05	2.82E-02
46	SCGN	5.360	5.219	9.97E-05	2.82E-02

47	SULT1C2	2.870	4.608	1.05E-04	2.91E-02
48	FGFR1OP2	-1.277	4.652	1.10E-04	2.94E-02
49	UDP-GLCNAC	2.780	2.651	1.10E-04	2.94E-02
50	RRN3	2.426	4.965	1.13E-04	2.95E-02
51	USF3	-1.685	4.115	1.16E-04	2.95E-02
52	MTNR1A	-4.394	1.811	1.30E-04	3.24E-02
53	NTRK3	-2.235	2.594	1.32E-04	3.24E-02
54	PITPNBL	-0.951	6.171	1.43E-04	3.43E-02
55	BHLHE40	-1.790	5.578	1.45E-04	3.43E-02
56	HS3ST2	4.567	0.700	1.49E-04	3.43E-02
57	TDRD5	-6.178	0.005	1.50E-04	3.43E-02
58	YPEL3	-1.166	5.547	1.57E-04	3.51E-02
59	APEH	1.327	6.690	1.59E-04	3.51E-02
60	NFU1	-1.204	6.474	1.63E-04	3.54E-02
61	CARS	1.043	5.515	1.70E-04	3.62E-02
62	UCP1	1.959	9.602	1.72E-04	3.62E-02
63	IRS2	-0.950	7.561	1.76E-04	3.65E-02
64	TTC36	1.891	6.041	1.79E-04	3.65E-02
65	C8ORF82	1.307	3.693	1.86E-04	3.68E-02
66	C15ORF57	-6.304	0.053	1.86E-04	3.68E-02
67	CCDC12	-1.177	3.790	1.94E-04	3.75E-02
68	GCK	7.007	4.032	1.96E-04	3.75E-02
69	FO834829.1	5.459	5.267	2.00E-04	3.75E-02
70	THEMIS	-4.122	1.481	2.01E-04	3.75E-02
71	BBL020833	-3.275	3.475	2.04E-04	3.75E-02
72	C19ORF53	1.547	7.574	2.36E-04	4.23E-02
73	RASGEF1A	-1.604	4.513	2.37E-04	4.23E-02
74	CXCR3	2.147	2.037	2.44E-04	4.23E-02
75	FAM20B	1.619	2.917	2.47E-04	4.23E-02
76	PPP1R3E	-6.030	-0.105	2.50E-04	4.23E-02
77	RBM33	-1.001	4.655	2.54E-04	4.23E-02
78	COL20A1	-1.478	7.188	2.56E-04	4.23E-02
79	FOXA4	2.770	1.471	2.56E-04	4.23E-02
80	UCHL3	1.079	6.136	2.79E-04	4.42E-02
81	ERI2	-1.114	4.223	2.79E-04	4.42E-02
82	SMOC1	3.205	4.220	2.80E-04	4.42E-02
83	SDHAF4	0.889	5.232	2.81E-04	4.42E-02
84	SCAMP1	1.228	4.910	2.92E-04	4.48E-02
85	CETN1	-1.141	4.462	2.92E-04	4.48E-02
86	SPEF1	4.353	5.745	2.96E-04	4.48E-02
87	BBL022284	3.435	2.790	3.30E-04	4.75E-02
88	TDP1	1.133	4.319	3.32E-04	4.75E-02
89	SULF1	1.192	9.892	3.33E-04	4.75E-02
90	SGCA	2.762	1.246	3.34E-04	4.75E-02
91	GNMT	0.935	5.029	3.39E-04	4.75E-02
92	LCMT2	-1.725	6.159	3.39E-04	4.75E-02
93	BBL024016	-1.623	8.012	3.42E-04	4.75E-02
94	MED20	1.563	4.472	3.42E-04	4.75E-02
95	PPIL6	3.665	4.147	3.53E-04	4.81E-02
96	GPR176	5.161	1.899	3.56E-04	4.81E-02
97	NFIL3-6	-0.941	6.995	3.58E-04	4.81E-02
98	FABP3	3.034	1.669	3.67E-04	4.89E-02
99	WDR95	2.674	2.482	3.80E-04	5.00E-02
100	ZC2HC1C	1.441	3.334	3.91E-04	5.06E-02

Table C5: EdgeR results for the uterus samples across the mountain transition, only the top 100 indicated. LogFC is the Log of the Fold-Change, LogCPM is the Log of the Counts Per Million, p-value is the exact p-value for differential expression and the FDR is the significance value after correcting for multiple testing by means of the Benjamini & Hochberg false discovery rate.

n	Gene	logFC	logCPM	p-value	FDR
1	CCDC175	4.759	6.046	6.79E-11	8.86E-07
2	PRSS33	-2.369	4.177	1.67E-09	1.09E-05
3	NMB	7.599	2.962	7.13E-09	3.10E-05
4	LMNA	2.307	8.564	3.03E-08	8.49E-05
5	PSMC3	-1.387	7.813	3.25E-08	8.49E-05
6	SHISA2	-2.414	5.162	4.34E-08	9.43E-05
7	FABP7	-6.387	3.745	6.82E-08	9.86E-05
8	UBP1	-2.080	7.847	6.89E-08	9.86E-05
9	SCNN1A	-2.507	8.475	7.15E-08	9.86E-05
10	KIAA1841	5.527	2.235	7.56E-08	9.86E-05
11	BBL016091	-2.473	3.696	1.05E-07	1.24E-04
12	RHCG	-1.810	10.632	1.29E-07	1.41E-04
13	TXNDC11	-1.645	7.145	1.86E-07	1.86E-04
14	PDE7A	-1.491	7.611	2.20E-07	2.05E-04
15	MTUS1	1.347	5.796	2.90E-07	2.52E-04
16	SAMD12	-1.687	4.357	3.22E-07	2.56E-04
17	GALC	-1.639	8.049	3.49E-07	2.56E-04
18	NTRK3	4.023	3.818	3.69E-07	2.56E-04
19	TMEM62	-1.624	5.285	3.73E-07	2.56E-04
20	FGFBP1	3.653	5.561	4.16E-07	2.71E-04
21	KLF4	3.130	6.222	5.55E-07	3.34E-04
22	LIPA	2.994	6.203	5.63E-07	3.34E-04
23	NIPAL4	-2.891	7.185	6.59E-07	3.53E-04
24	BBL019666	7.864	1.434	7.10E-07	3.53E-04
25	MTHFS	-2.272	3.884	7.22E-07	3.53E-04
26	BBL015290	-2.079	4.813	7.24E-07	3.53E-04
27	COQ3	-1.329	5.962	7.32E-07	3.53E-04
28	CERS2	-2.005	9.993	9.18E-07	4.20E-04
29	TMC7	1.435	4.876	9.49E-07	4.20E-04
30	ATP1B3	-1.281	7.732	9.90E-07	4.20E-04
31	SLC24A4	-1.700	5.167	1.03E-06	4.20E-04
32	BBL014344	-5.565	5.698	1.03E-06	4.20E-04
33	DRD4	-3.163	3.783	1.06E-06	4.20E-04
34	HIP1R	-1.202	6.894	1.20E-06	4.52E-04
35	NOV	2.857	5.899	1.27E-06	4.52E-04
36	TPPP3	2.521	8.754	1.31E-06	4.52E-04
37	BBL000845	1.301	9.933	1.36E-06	4.52E-04
38	PER3	-2.155	6.061	1.46E-06	4.52E-04
39	NMUR3	-7.958	2.491	1.47E-06	4.52E-04
40	CBSL	1.520	5.919	1.52E-06	4.52E-04
41	KCNJ1	-2.628	6.178	1.53E-06	4.52E-04
42	AGRN	2.297	5.357	1.54E-06	4.52E-04
43	PTGER4	1.787	4.611	1.55E-06	4.52E-04
44	CCDC12	-1.458	4.710	1.57E-06	4.52E-04
45	DCLK2	2.046	4.540	1.57E-06	4.52E-04
46	PTDSS2	-1.279	7.793	1.59E-06	4.52E-04

47	GPER1	-6.860	2.718	1.72E-06	4.78E-04
48	SLCO1C1	5.617	0.566	1.77E-06	4.80E-04
49	DISP2	-1.494	5.697	1.82E-06	4.84E-04
50	ORC1	2.390	2.837	1.92E-06	4.96E-04
51	ATP1B1	-1.361	12.172	1.95E-06	4.96E-04
52	ALDH3A1	2.071	8.050	2.03E-06	4.96E-04
53	GPX4	-1.861	8.236	2.03E-06	4.96E-04
54	RNF112	-1.787	8.222	2.05E-06	4.96E-04
55	IDH1	-1.598	10.682	2.26E-06	5.36E-04
56	FAM83B	1.955	3.349	2.32E-06	5.41E-04
57	ECM1	1.665	10.125	2.52E-06	5.69E-04
58	BBL011948	1.294	6.036	2.53E-06	5.69E-04
59	ATP1A1	-1.629	12.871	2.73E-06	6.04E-04
60	CPN2	4.370	3.537	2.86E-06	6.13E-04
61	GABRR1	3.502	1.673	2.87E-06	6.13E-04
62	STRAP	-1.191	7.429	3.01E-06	6.34E-04
63	BRI3	-1.165	7.825	3.29E-06	6.55E-04
64	GLUL	-1.590	9.811	3.31E-06	6.55E-04
65	MCM6	2.456	5.558	3.32E-06	6.55E-04
66	MICALL2	-0.974	7.119	3.38E-06	6.55E-04
67	RAP2C	1.089	6.122	3.40E-06	6.55E-04
68	IL22RA1	3.134	3.463	3.44E-06	6.55E-04
69	ANKRD28	-0.931	6.634	3.47E-06	6.55E-04
70	TFPI	2.365	7.672	3.66E-06	6.82E-04
71	GLOD4	-0.989	7.171	3.75E-06	6.90E-04
72	SCARA3	3.408	3.786	3.86E-06	6.99E-04
73	SOCS2	-2.390	1.912	3.97E-06	7.09E-04
74	TTL	1.505	5.193	4.20E-06	7.40E-04
75	TSKU	2.666	6.665	4.27E-06	7.43E-04
76	N4BP3	-1.733	7.329	4.54E-06	7.79E-04
77	CST3	-1.110	8.155	4.64E-06	7.85E-04
78	LHX1	-1.404	6.091	5.00E-06	8.33E-04
79	GLB1L2	-1.643	4.659	5.05E-06	8.33E-04
80	FAM46A	2.003	4.588	5.15E-06	8.35E-04
81	GLDC	-2.305	7.898	5.23E-06	8.35E-04
82	PAPSS2	-1.253	9.467	5.25E-06	8.35E-04
83	TTC6	3.848	1.388	5.34E-06	8.39E-04
84	OTUD7A	-1.843	5.773	5.46E-06	8.48E-04
85	LAMTOR2	-1.411	7.446	5.70E-06	8.69E-04
86	ZFP36L2	1.473	5.461	5.73E-06	8.69E-04
87	SH3KBP1	2.026	5.604	6.14E-06	9.05E-04
88	C19ORF53	-2.301	7.138	6.20E-06	9.05E-04
89	ANGPTL1	2.756	7.586	6.21E-06	9.05E-04
90	SERPINA1	1.012	6.854	6.28E-06	9.05E-04
91	LAMP1	-0.975	9.327	6.33E-06	9.05E-04
92	MAT2A	1.489	6.018	6.39E-06	9.05E-04
93	MARVELD2	-1.567	6.251	6.62E-06	9.28E-04
94	PNLIPRP1	-4.206	1.901	6.74E-06	9.34E-04
95	DDX49	-1.339	5.911	6.85E-06	9.41E-04
96	UPK1B	-1.413	7.768	7.34E-06	9.97E-04
97	ATP1B2A	-3.033	4.267	7.51E-06	1.01E-03
98	GALT	-0.925	5.446	7.61E-06	1.01E-03
99	TMEM150C	-1.184	5.777	7.79E-06	1.02E-03
100	BLOC1S4	-1.625	6.103	7.92E-06	1.02E-03

Figure C1: Tissue specific expression by tissue. Blue bars show absolute specific genes (Tau score of 1) and the green bars show highly specific genes (Tau > 0.85).

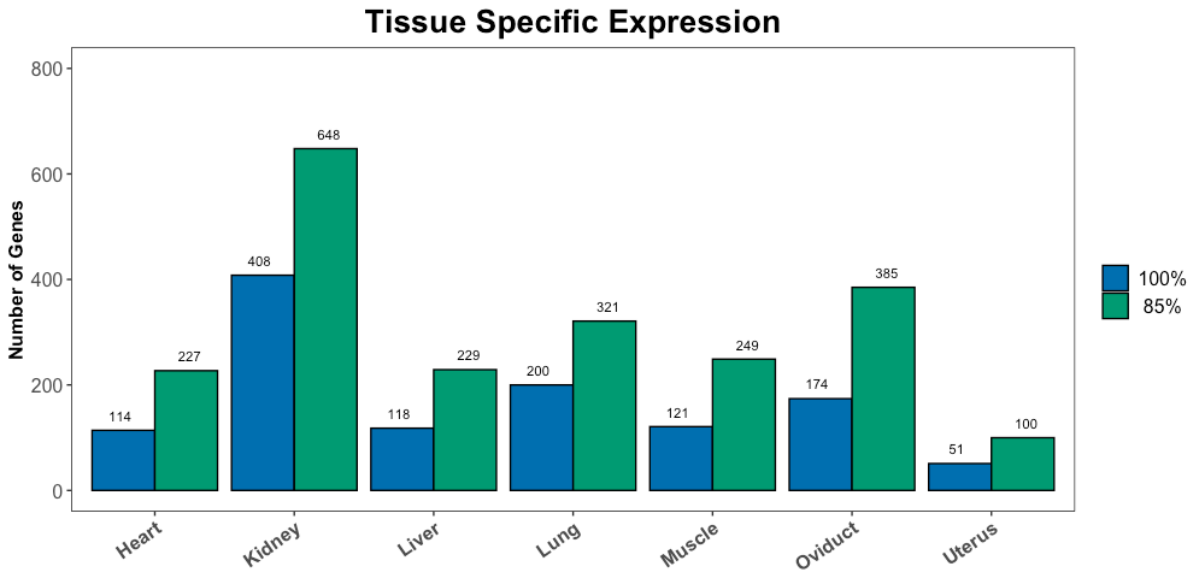


Figure C2: The top 14 genes found by edgeR when comparing the uterus tissues for the island transition (top 7 are also shown in Table 3.2)

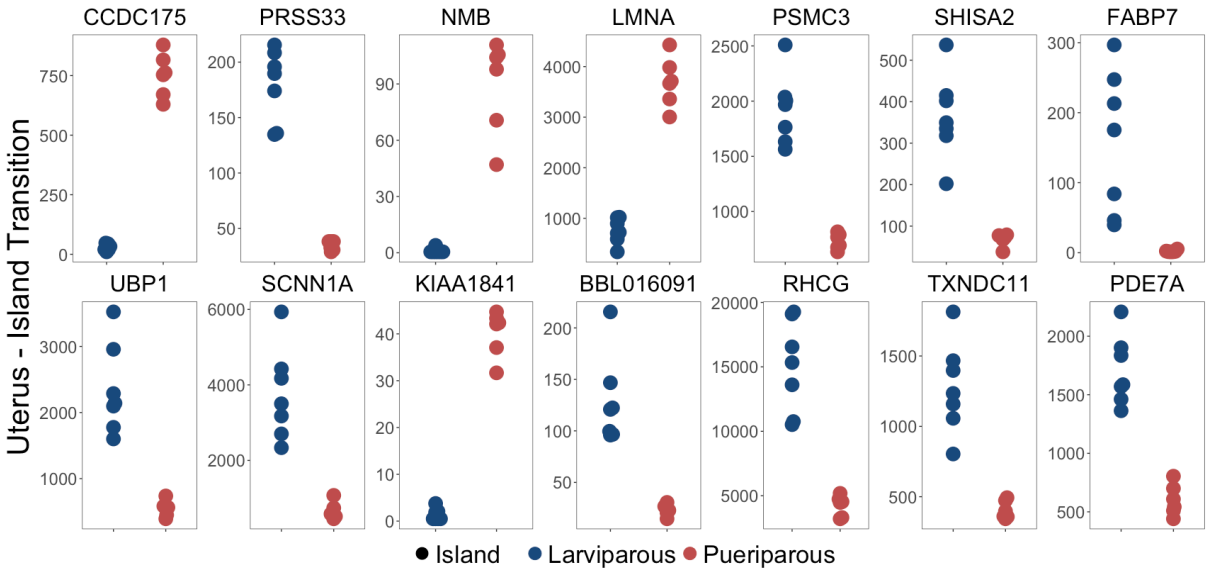


Figure C3: The top 14 genes found by edgeR when comparing the uterus tissues for the mountain transition (top 7 are also shown in Table 3.2)

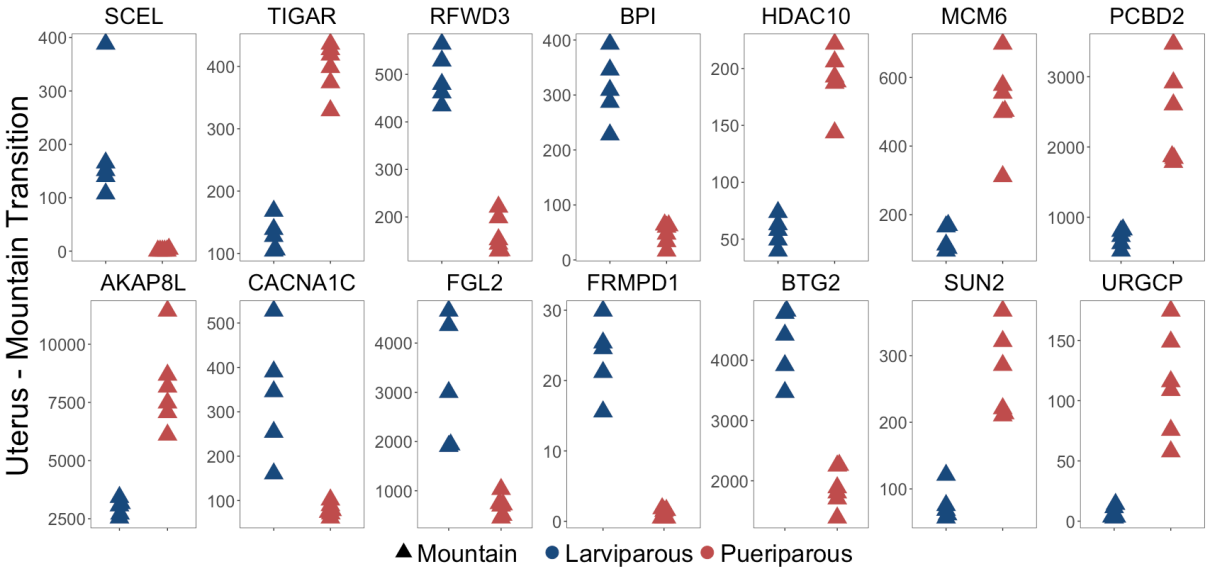


Figure C4: The top 28 genes found by edgeR when comparing the uterus tissues across both transitions (top 7 are displayed in Figure 3.4C).

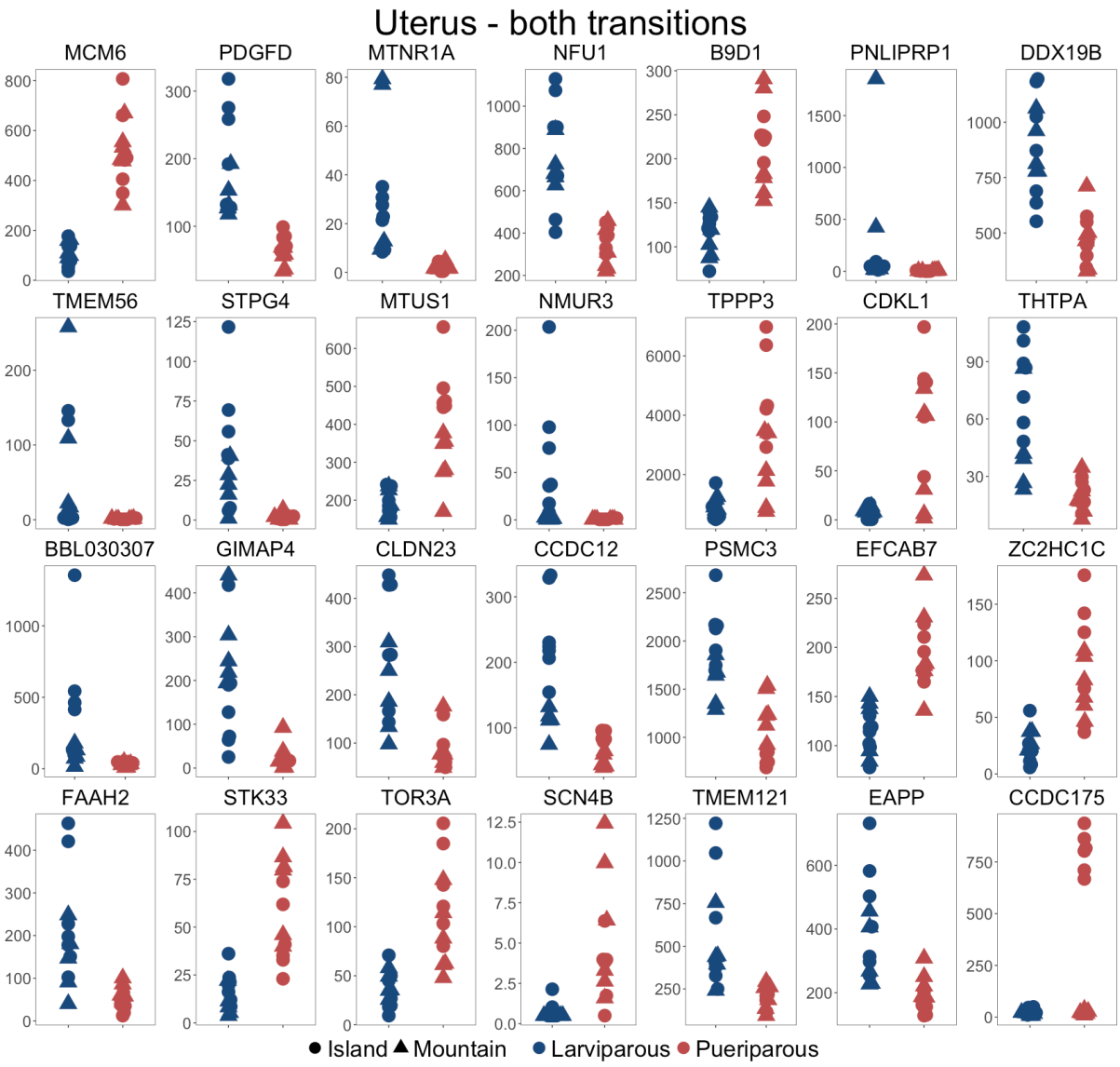
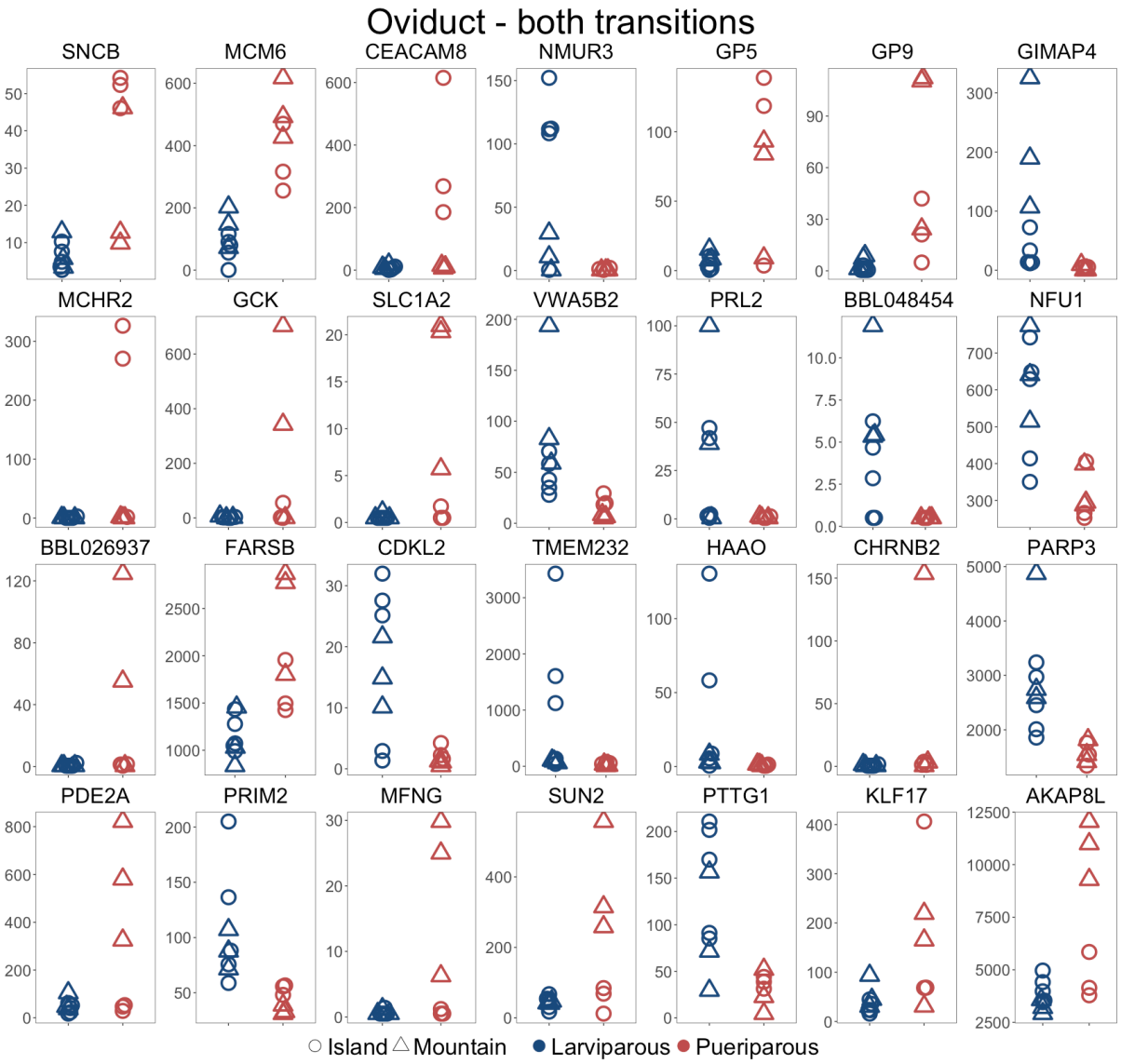


Figure C5: The top 28 genes found by edgeR when comparing the oviduct tissues across both transitions (top 7 are displayed in Figure 3.4D), all genes were non-significant.



APPENDIX D: SUPPLEMENTAL INFORMATION FOR CHAPTER 4

Figure D1: Relationship between genetic distance as measured by F_{ST} and geographic distance.

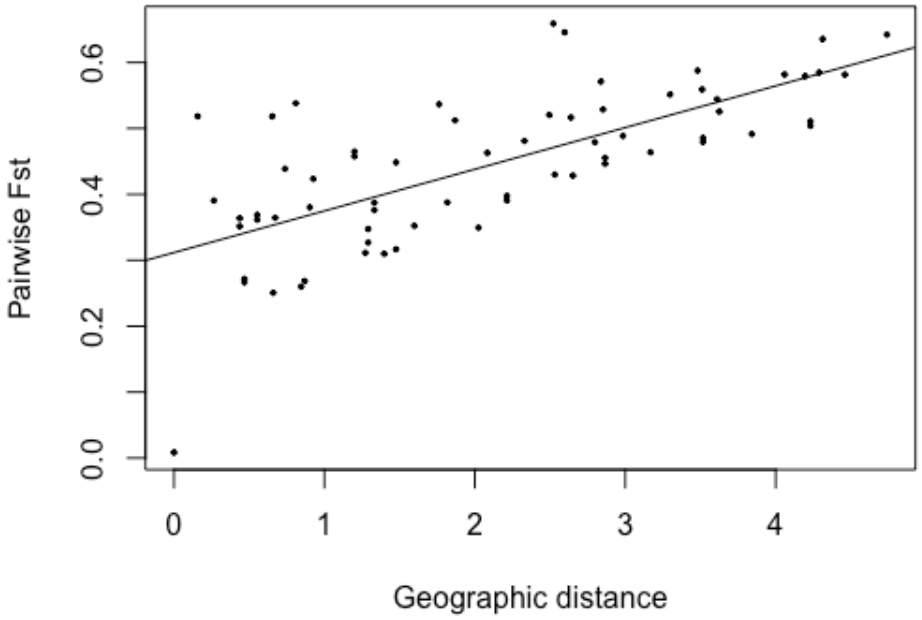


Figure D2: Heatmap of Nei's distance by individual.

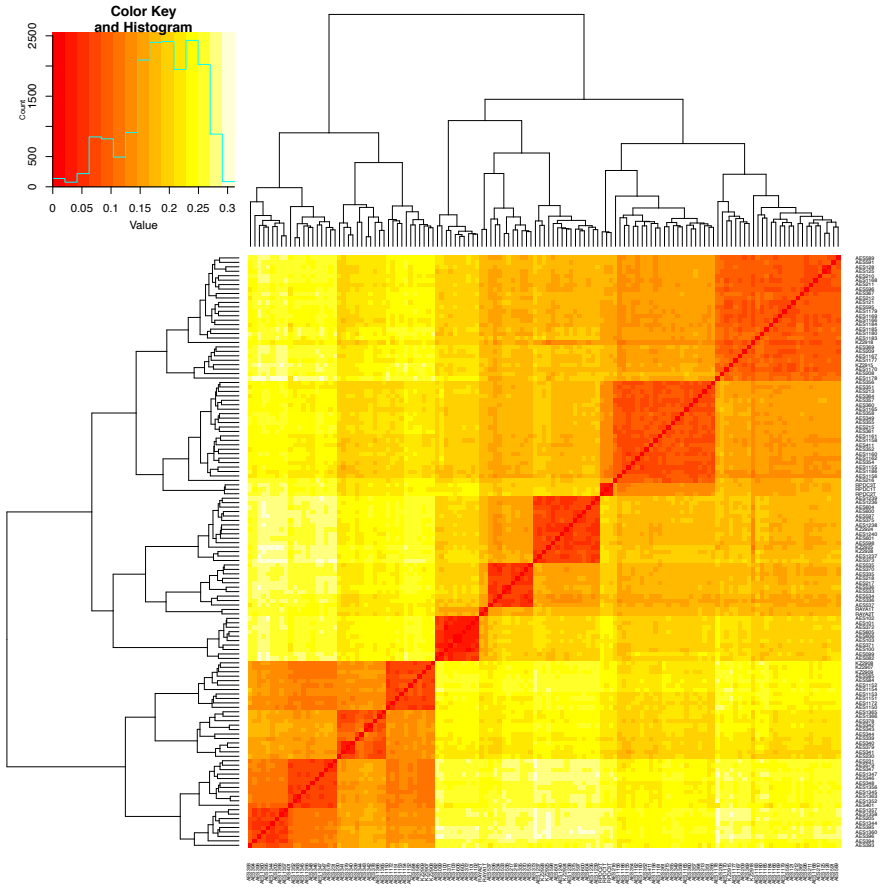


Figure D3: Number of Private alleles per population plotted in the same order Figure 4.3C. Data split for all missense SNPs and by unlinked SNPs. Currently extirpated populations indicated by black outlines.

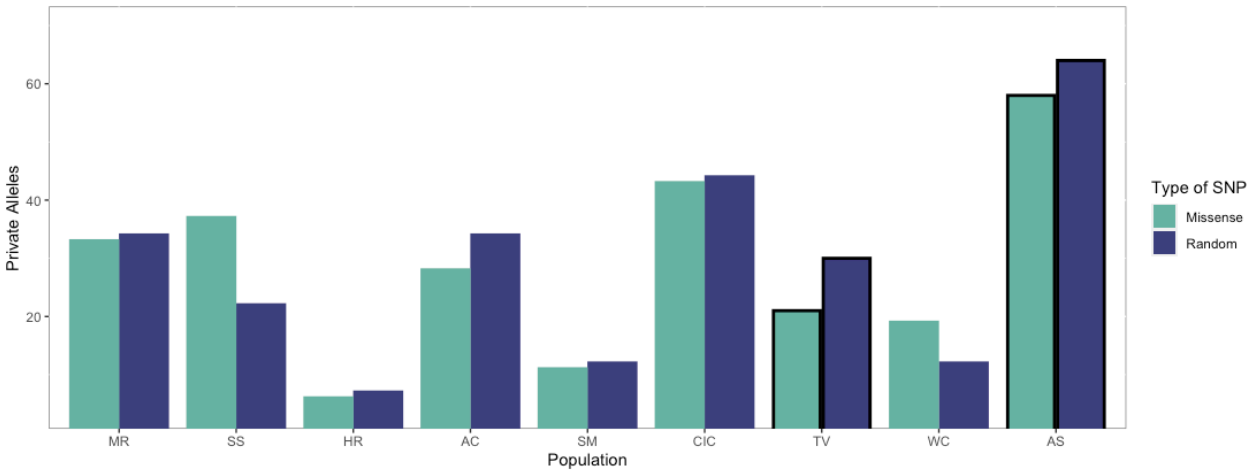


Figure D4: Principal component analyses of different subsets of SNPs.

Table D1: Table of loci included in the capture array

Market-set	Number of Loci	Base pairs
Putative Bd-related transcripts	746	287,028
Genome-wide transcripts	579	77,793
Major Histocompatibility Complex (MHC)	11	4,575
Antimicrobial Peptides (AMPs)	37	4,100
Toll-Like receptors (TLRs)	13	3,300
Mitochondrial	2	1,000
Total	**1,388**	**377,796**