Evolutionary and genomic associations of colour and pattern in fire and Alpine salamanders (*Salamandra* spp.)

James David Burgon
B.Sc. (Hons), M.Sc.

Submitted in fulfilment of the requirements for the Degree of Doctor of Philosophy

Institute of Biodiversity, Animal Health and Comparative Medicine

College of Medical, Veterinary and Life Sciences

University of Glasgow

2018

Thesis abstract

Animal colouration is associated with a multitude of ecologically adaptive traits known to drive biological diversification, from predator avoidance to physiological regulation. As such, it is an ideal system in which to study the evolutionary patterns and processes that generate and maintain biological diversity. Within the terrestrial vertebrates, amphibians display some of the greatest complexity and variation in terms of colour patterning, with the salamander genus *Salamandra* particularly renowned for its colour diversity. Typically, *Salamandra* species present bright, highly variable yellow-black patterns consisting of spots and/or stripes, which are thought to hold an aposematic (warning) function related to their toxic secretions. In addition to this, individual species and populations have evolved melanic, fully yellow and fully brown colourations, with gradations seen in-between. Importantly, there are also indications of parallel colour pattern evolution, making *Salamandra* an attractive system for studying the repeated evolution of adaptive phenotypes. However, the genus currently lacks phylogenetic resolution, and the molecular mechanisms underlying amphibian colouration are poorly understood. In this thesis, I aim to fill both of these knowledge gaps through the use of next-generation sequencing (NGS) techniques, which offer both unpredicted opportunities to resolve systematically challenging relationships and allow us to study the genetic basis of ecologically adaptive phenotypes in wild non-model organism. In **Chapter 2** we reconstruct the controversial interspecies phylogeny of *Salamandra* using three largely independent phylogenomic data sets. First, using restriction site associated DNA sequencing (RAD-Seq), I genotyped representatives of all six currently recognised *Salamandra* species (and two outgroup species from its sister genus *Lyciasalamandra*). This was combined with nuclear protein-coding sequences derived from RNA-Seq and full mitochondrial genomes. Analyses of concatenated RNA-Seq and RAD-Seq data retrieved well supported, fully congruent topologies that placed: (1) *S. infraimmaculata* as sister to all other species in the genus; (2) *S. algira* sister to *S. salamandra*; (3) these two species sister to a clade containing *S. atra*, *S. corsica* and *S. lanzai*; and (4) the Alpine species *S. atra* and *S. lanzai* as sister taxa. The phylogeny inferred from mitochondrial genomes differed from this in its placement of *S. corsica*, as did species tree analyses of RNA-Seq and RAD-Seq data. However, the general congruence among topologies recovered from the RNA-Seq and RAD-Seq data sets gives us confidence in our methodologies and results. In **Chapter 3**, I perform more in-depth phylogenomic analyses, using RAD-Seq to genotype 231 salamanders from across the taxonomic and geographic breadth of *Salamandra*. Both Bayesian and maximum

likelihood based analyses of concatenated RAD-loci (comprising 187,080–294,300 nt of sequence data) returned well-supported, largely congruent topologies that supported the monophyly of all six currently recognised species. However, the placement of *S. corsica* was again unclear, and data filtering parameters were found to have a great impact on downstream analyses. Further, I identified undescribed diversity within the North African species (*S. algira*) and find that 43% of *S. salamandra* subspecies do not meet a criterion of monophyly. Following this, I use the phylogenetic hypothesis generated to assess the parallel evolution of reproductive (parity) mode and two colour phenotypes (melanism and stripe formation) through ancestral state reconstruction analyses. I find that pueriparity (giving birth to fully metamorphosed juveniles) has independently arisen in at least four lineages, melanism in at least five, and a striped phenotype in least two, all from a common yellow-black spotted larviparous (larvae depositing) ancestor. Finally, in **Chapter 4**, I leverage and highly colour-variable lineage of the European fire salamander (*S. salamandra bernardezi*) to identify genetic associations with colour, test for selection on colouration, and test the relationship between colour phenotype and toxicity (the functional basis of aposematism). I show that, within a geographically restricted region, colour phenotypes form a gradient of variation, from fully yellow to fully brown, through a yellow-black striped pattern. Population genetic analyses suggest a sympatric evolutionary origin for this colour variation, and I found no association between a salamanders colour pattern and the metabolomic profile of its toxic secretions, which calls into question the adaptive significance attributed to these striking colourations. Following this, I identified 196 significantly differentially expressed genes between skin colours (63 of which are associated with animal colouration) using transcriptomic (RNA-Seq) analyses, and 43 loci associated to representative colour phenotypes (yellow, brown and striped) using genomic (RAD-Seq) approaches. I also found signals of selection on 142 genomic loci between representative colour phenotypes, 19 of which overlap with genomic analyses. Overall, my results provide greater phylogenetic resolution for the genus *Salamandra* than ever before, revealing the need for taxonomic revisions and confirming the convergent (or parallel) evolution of both reproductive and colour phenotypes. My data also represents a significant contribution to our understanding of the genetic basis of amphibian colouration, providing a valuable resource for future comparative research on vertebrate colour evolution.

Table of contents

List of tables

List of figures

Abbreviations

ANOVA (aov)	Analysis of variance
ASR	Ancestral state reconstruction
BIC	Bayesian Information Criterion
BLAST	Basic Local Alignment Search Tool
bp	Base pairs
ddH_2O	Double-distilled water
ddRAD-Seq	Double digest RAD-Seq
DEPC water	Diethyl pyrocarbonate water
ESPCR	European Society for Pigment Cell Research
EtOH	Ethanol
Fst	Fixation index
g	Grams
Gb	Giga bases
GC-MS	Gas chromatography–mass spectrometry
GO	Gene ontology
GTR	General time reversible model of nucleotide evolution
ILS	Incomplete lineage sorting
IMP	Integrative Multi-species Prediction
IUPAC	International Union of Pure and Applied Chemistry
JPEG	Joint Photographic Expert Group
krpm	Kilo revolutions per minute
kU/ml	Thousand units per millilitre
L	Litre
LFMM	Latent Factor Mixed Models
LiCl	Lithium chloride
M	Million
M	Molar
MANOVA	Multivariate analysis of variance
masl	Metres above sea level
MCMC	Markov chain Monte Carlo
mg	Milligrams
ML	Maximum likelihood
ml	Millilitre
m*M*	Millimolar

mRNA	Messenger RNA
mtDNA	Mitochondrial DNA
n	Number/Sample size
NCBI	National Center for Biotechnology Information
nDNA	Nuclear DNA
ng	Nanogram
NGS	Next generation sequencing
NIR	Near infer-red
nm	Nanometre
nt	Nucleotide
PC	Principal component
PCA	Principal component analysis
PCR	Polymerase chain reaction
PE	Paired-end
pg	Picograms
RAD-Seq	Restriction site Associated DNA Sequencing
RF	Random Forest
RGB	red:green:blue pixel values
RNA-Seq	RNA sequencing
rpm	Revolutions per minute
rRNAs	Ribosomal RNA
RT	Room temperature
SD	Standard deviation
sDE	Significantly differentially expressed
SNP	Single nucleotide polymorphism
spp.	*Species pluralis*, the Latin for multiple species
SVL	Snout-to-vent length
TEM	Transmission Electron Microscopy
tRNA	Transfer RNA
TVM	Transversion model of nucleotide evolution
U/µl	Units per microliter
UV	Ultraviolet
VCF	Variant Call Format
WGS	Whole genome sequencing
ZFst	z-transformed *Fst*
µl	Microliter
µ*M*	Micromolar

Author’s declaration

I declare that this thesis is entirely my own composition and that the research described in it is also my own except where otherwise stated.

Some areas were performed in close partnership with study collaborators. Where others have performed or contributed to the work, this has been clearly acknowledged in the text.

Finally, I declare that the work presented in this thesis has not been submitted for any other degree or professional qualification.

Acknowledgements

Before embarking on a research project of this scale, it is hard to appreciate the level of help and collaboration that goes into making it a success. I would like to start by thanking my supervisors, Dr Kathryn R. Elmer and Professor Barbara Mable, who originally proposed the project and gave me the opportunity to pursue and develop it. Their help, support and experience have been invaluable over the last four years. I would also like to thank the Natural Environment Research Council (NERC) for funding my studentship (NE/L501918/1).

During my time at the University of Glasgow, I have had the privilege of working with a diverse and brilliant cohort of early career researchers. In particular, I would thank my lab colleagues Hans Recknagel, Madeleine Carruthers and Arne Jacobs. The four of us started our around the same time, and without your support and insight this project would not have been possible: I envy your future colleagues. I also want to thank more recent additions to the lab: Mel Chen, Kevin Schneider, Marco Crotti and Andrey Yurchenko, it has been a privilege working with you all. The final lab member, Aileen Adam, deserves more thanks than I can give. Aileen patiently helped me throughout my lab work and has provided enough biscuits, chocolate and strong coffee to get me through even my lowest moments.

While all exceptional people, my lab is by no means unique within the Institute of Biodiversity, Animal Health and Comparative Medicine. The highly collegiate and friendly atmosphere means that there are too many people to thank directly for making my time here as enjoyable as it has been. However, a few deserve special mention. First, Lizy Mittell has my thanks for all her help, and apologies for all my rants. I also want to thank Laura Allen for our infrequent, yet highly cathartic, tea breaks, Dr Martin Llewellyn for giving me access to his server, and all those people who helped me run the Institute's blog and podcast Naturally Speaking (particularly Dr Jim Caryl).

I have also had the privilege of working and collaborating with a number of brilliant researchers at other institutions. Chief amongst these are Prof. Miguel Vences and Dr Sebastian Steinfartz from the Technical University of Braunschweig (Germany), who have provided samples, insight and expertise, and Dr David R. Vieites (currently at the University of California, Berkeley), who assisted with field collections and contributed to the conceptual basis of the project. Dr Helen Gunter (Edinburgh Genomics), also provided

aid with RNA sample preparation, and Dr Karl Burgess (Institute of Infection, Immunity & Inflammation, UofG) and Dr Stefan Weidt (Glasgow Polyomics) helped to conduct toxin metabolomic analyses. In addition, Ms Margaret Mullin (Technician, UofG) and Dr Mohsen Nokhbatolfoghahai (IBAHCM) helped me collect and interpret transmission electron microscopy images, for which I am incredibly grateful.

I must also thank my field crew, Vicente Ramos and Rob Williams, and the undergraduate students (Ruth Shepherd, Ryan McIntyre and Amber Mathie), Masters students (Ruth Robinson and Fraser Brydon) and interns (Amelia Cameron and Heather McDevitt), who helped me establish protocols and explore new research avenues.

Further to these individuals, a number of funding bodies have supported this project. The Royal Society provided funding for fieldwork (through a grant awarded to K. Elmer.), The Linnean Society and the Systematics Association aided in the sequencing of phylogenetic samples, and the Glasgow Natural History Society provided funds for the transmission electron microscopy of skin samples. Finally, a Wellcome Trust ISSF catalyst grant was awarded for metabolomics analyses of *Salamandra* toxins (to K. Elmer, through Glasgow Polyomics).

However, my deepest thanks must go my long-suffering family and friends, who have barely seen me in the last year or understood me in the last four. In particular, special thanks go to my parents, who have unconditionally supported and loved me throughout my decade long university career, and my partner Karen, who has borne the brunt of my increasing stress levels and kept our zoo fed and happy.

Finally, I thank the many salamanders I have inconvenienced over the last four years, particularly those few who have died as result of our encounter. You didn’t ask to be involved, and you don’t get to be co-authors, but I couldn’t have done this without you.

Chapter 1: General introduction

1.1. Overview

Animal colouration has long been considered an ideal system in which to study the evolutionary patterns and processes that generate and maintain biological diversity. This is because it is conspicuously affected by natural selection (Caro 2005), and regularly shows striking intra- and interspecific variation (Endler 1990). It also presents a non-lethal way to study genotype-phenotype interactions in ecologically adaptive traits (Hoekstra 2006), with the development of new digital imaging techniques and high-throughput sequencing technologies revolutionising the study of colouration in wild non-model systems. While this has benefited many fields, it is particularly interesting for the study of amphibians, which display some of the most diverse and complex colour patterns of all terrestrial vertebrates (Hoffman and Blouin 2000; Rudh and Qvarnström 2013).

This chromatic diversity is the result of a complex and highly variable cell structure containing up to five kinds of cells called chromatophores, which either contain pigment or reflect light (Bagnara et al. 1968); individually or in combination, these cells can produce almost every known colour (Rudh and Qvarnström 2013). However, the reason that amphibian colouration is of particular interest is its close association to a number of ecologically and physiologically adaptive traits known to drive colour pattern evolution; for example, thermoregulation, predator avoidance and communication (Thayer 1909; Cott 1940; Rudh and Qvarnström 2013). However, while the evolutionary and ecological significance of amphibian colouration has long been recognised, until recently their colour patterns have been challenging to measure objectively, making their quantitative analyses difficult.

Historically, human observers were used to subjectively (qualitatively) rank animal colour patterns. This not only lacked rigor, but also missed biologically significant information outside the human visual system (Endler 1990). However, this changed with the development of spectrophotometry and digital imaging techniques, which have allowed for the quantitative analysis of both colour and shape within powerful computational frameworks, thereby removing one of the most significant barriers to the detailed and objective study of animal colouration (Endler 1990; Stevens et al. 2007; Costa et al. 2009).

Unfortunately, one hurdle remains for evolutionary studies on amphibian colouration: an almost complete lack of robust studies on its underlying genetics.

When compared to other vertebrate taxa, we know very little about amphibian colour genetics, especially in wild populations (Hoffman and Blouin 2000; Rudh and Qvarnström 2013). Amphibians were also the last major vertebrate lineage to have a full reference genome sequenced, with a draft *Xenopus tropicalis* (African clawed frog) assembly only published in 2010 (Hellsten et al. 2010). Unfortunately, while their dorsal colour patterns are highly variable, *X. tropicalis* lacks wild colour polymorphisms, and it is often difficult to understand the ecological significance of genetic variation in lab reared model organisms (Stapley et al. 2010). A similar situation is seen in another model organism: *Ambystoma mexicanum* (the axolotl). While this species has a reference transcriptome assembly (Smith et al. 2005; Keinath et al. 2015), and recent studies have identified important mutations in colour genes between artificially selected colour lines (Woodcock et al. 2017), this does not necessarily aid our understanding of natural colour variation and the selective forces responsible for generating it. However, with new next generation sequencing (NGS) technologies it is now possible to sequence whole genomes, removing our reliance on model-organisms and allowing us to study the genetic basis of pigmentation in wild non-model taxa (Hubbard et al. 2010).

Given the diversity and complex ecological functions of their colour patterns, amphibians make an ideal candidate for studying the interaction between phenotype and genotype in relation to adaptive colouration. This chapter will review our current understanding of amphibian colouration, how we measure colour and our current understanding of the colour genetics, before introducing the Palaearctic salamander genus *Salamandra* (the fire and Alpine salamanders) as a candidate for studying the genetic basis of colour in amphibians. *Salamandra* species are highly colour polymorphic, showing striking inter- and intraspecific variation in dorsal colour patterning (Thorn and Raffaëlli 2001), which is thought to be adaptive for both aposematism (warning colouration associated with unpalatability/toxicity) and potentially thermoregulation at high altitudes (Beukema et al. 2016b). *Salamandra* also shows multiple colour polymorphisms across their range, and their characteristic yellow-black colour patterns are clearly defined and easy to phenotype. Importantly, there are also indications of parallel phenotype evolution (Steinfartz et al. 2000; Vences et al. 2014), which provides 'natural replicates' in which to identify key colour related genes and investigate the relative significance of *de novo* mutation and standing genetic variation in amphibian pattern evolution.

1.2. The cellular basis of amphibian colouration

Epidermal cells called chromatophores determine the colour of amphibian skin. These cells contain either granules of pigment or light reflecting crystalline platelets. Unlike mammals and birds, which only have a single chromatophore type (melanophores; Mills and Patterson 2009), amphibians have up to five different chromatophores (Fig. 1.1). These five colour producing cells all share a common developmental origin, each being derived from pluripotential neural crest stem cells (Bagnara and Ferris 1971; Bagnara 1972; Bagnara et al. 1978). These cells are present during early vertebrate development, and undergo extensive migration throughout the embryo before differentiating into multiple cell types, including neurons, smooth muscle and pigment cells (Mayor and Theveneau 2013).

Differentiated chromatophores are classified based on both their location in the skin and the colour they produce under a white light (Fig. 1.1). The first layer of chromatophores, nearest the surface of the skin, contain either carotenoid (dietary) or pteridine (metabolic) pigments. When yellow these are called xanthophores, and when orange/red they are called erythrophores, although the distinction is vague and not always recognised (see Rudh and Qvarnström 2013). The second layer is composed of chromatophores that create colour through the reflection of light by crystalline purines, particularly guanine (iridophores) and urate (leucophores). This reflectance results in a silvery-iridescent (iridophore) or white (leucophore) colour. Melanophores make up the deepest layer of chromatophores, and contain the black/brown pigment melanin within cellular organelles called melanosomes.

Different chromatophores can also work in tandem to create additional colourations. For example, iridophores and melanophores in combination create blue, and with the addition of xanthophores or erythrophores produce green (Browder 1968; Lyerla and Jameson 1968; Bagnara et al. 2007). This complex cellular structure is known as the "poikilothermic dermal chromatophore unit" (Bagnara et al. 1968; Bagnara and Hadley 1973) and develops during metamorphosis, the transitionary period between the larval and adult life stages. During this time, chromatophores go from being randomly distributed below the subepidermal collagen layer to being highly localised in the dermis, the precise location of each chromatophore the dermis resulting from cell specific migration rates (Yasutomi and Yamada 1998). However, while chromatophores are typically located in the dermal skin layer (Fig. 1.1), during metamorphosis some have been known to migrate into

the epidermis (e.g. xanthophores in *Salamandra salamandra*; Pederzoli et al. 2003) and from the epidermis into the dermis (e.g. melanophores in *Hyla arborea*; Yasutomi and Yamada 1998).

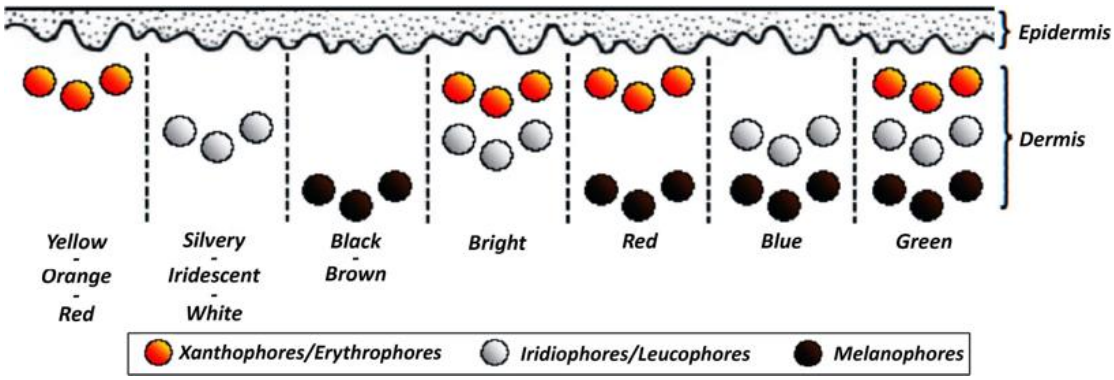


Figure 1.1: Colours produced by different combinations of amphibian chromatophores under a white light. The epidermis is the outermost layer of skin, followed by the dermis. This figure is based on information in: Bagnara et al. 1968, 2007; Browder 1968; Lyerla and Jameson 1968; Bagnara and Hadley 1973.

The occurrence and distribution of chromatophores can also vary within and between individuals and species (Pough et al. 2003). The oriental fire-bellied toad, *Bombina orientalis*, is a classic example of within individual variation, as they have areas of red skin containing xanthophores and melanophores, black skin containing only melanophores, and green skin consisting of the full chromatophore unit (Frost and Robinson 1984). Melanophores are also responsible for the rapid colour change seen in some amphibians, as dendric projections (under neuroendocrine control; Nery and de Lauro Castrucci 1997) are able to expand, masking overlying chromatophores (Bagnara et al. 1968, 1969, 1978). This cellular variation and flexibility is the structural mechanism by which amphibians produce their wealth of complex and ecologically adaptive colour pattern phenotypes.

1.3. Adaptive colouration in amphibians

1.3.1. Amphibian diversity

The taxonomic diversity of amphibians, and their wealth of life history strategies, have influenced the evolution of colour in the group uniquely compared to other vertebrates

(Haddad and Prado 2005). The class Amphibia contains three orders (Frost 2017): Anura (frogs and toads, c. 7728 species), Caudata (newts and salamanders, c. 715 species) and Gymnophiona (caecilians, c. 207 species). Unlike other vertebrates, the 'typical' amphibian life-style is biphasic, with an aquatic (freshwater) egg and larval stage before metamorphosis into a terrestrial adult (Duellman and Trueb 1994). As the visual and landscape properties of aquatic and terrestrial systems differ, it is not surprising that this metamorphosis is associated with ontogenetic changes in colour (e.g. *Pseudacris ornata*; Harkey and Semlitsch 1988) and photoreceptor sensitivity (e.g. *Ambystoma tigrinum*; Chen et al. 2008). Moreover, amphibians have adapted to a wide variety of biomes (IUCN 2008), all of which exert different pressures on colour pattern evolution.

1.3.2. Physiological pressures

Thermoregulation, water loss and ultraviolet light (UV) protection are thought to be important drivers of amphibian colour evolution. This often works in association with behaviourally mediated microhabitat selection, which allows individuals to move along environmental gradients to find more favourable conditions (Rudh and Qvarnström 2013). For example, when Garcia et al. (2004) exposed larval *Ambystoma barbouri* and *A. texanum* salamanders to increased UV radiation, they found not only a chromatic response, but also an increase in selecting a UV-protected microhabitat. However, while this means that amphibian colouration cannot be assumed to be physiologically adaptive, there are several cases of pigmentation providing a direct physiological benefit.

Thermoregulation and the prevention of water loss are often intertwined, and stem from two aspects of amphibian biology: their highly permeable skin and their environmentally regulated body temperature (poikilothermic ectotherms; Duellman and Trueb 1994). The dermal chromatophore unit allows many amphibians (especially anurans) to thermoregulate through a plastic colour change when exposed to high temperatures by rapidly changing from a dark cryptic colour to a very light one: e.g. the desert tree frog, *Litoria rubella* (Withers 1995), the American green tree frog, *Hyla cinerea* (King et al. 1994) and *Bokermannnohyla alvarengai* (Fig. 1.2; Tattersall et al. 2006). This adaptation reduces water loss and helps to prevent over heating as light skin reflects more, and absorbs less, solar radiation (King et al. 1994; Tattersall et al. 2006).

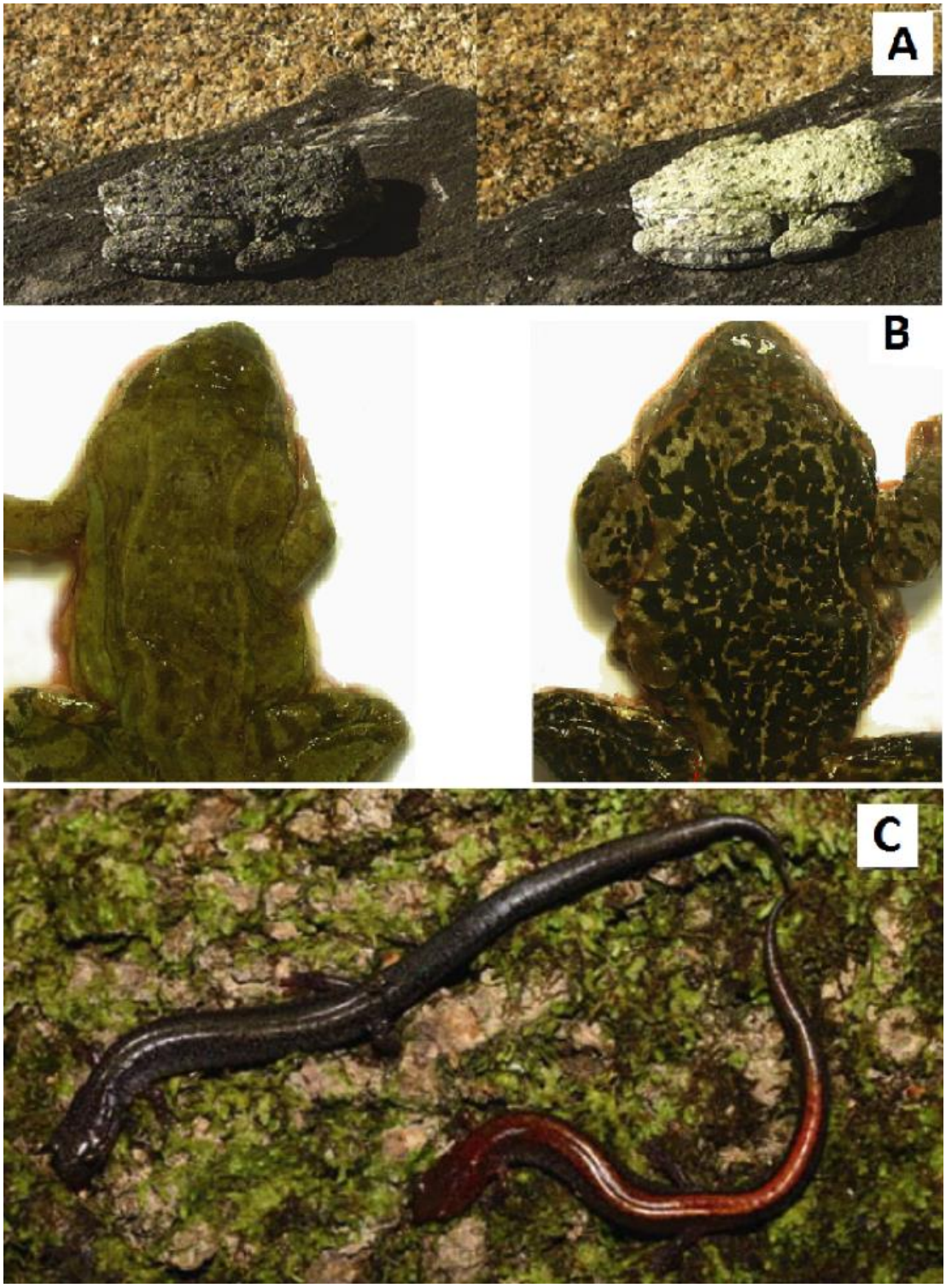


Figure 1.2: Physiological colouration. **A**: Rapid plastic colour change (black to white) in *Bokermannohyla alvarengai* when exposed to direct sunlight (modified from Tattersall et al. 2006); **B**: Increased melanisation in high latitude *Rana temporaria* (right) compared to lower latitude (left; modified from Alho et al. 2010); **C**: Fully melanistic (lead phase) *Plethodon cinereus* (left) next to sympatric red-striped morph (right; modified from Davis and Milanovich 2010).

At the other end of the temperature scale, melanism is thought to be important for thermoregulation in amphibians in colder climates (Clusella-Trullas et al. 2007) and varies along latitudinal (e.g. *Rana temporaria*; Alho et al. 2010; Fig. 1.2) and altitudinal gradients (e.g. *Rana temporaria*; Vences et al. 2002). As black absorbs greater amounts of solar radiation, this leads to an increased body temperature, and therefore higher activity levels in dark compared to light individuals (Clusella-Trullas et al. 2007). Garcia et al. (2004) also found that early stage larval *Ambystoma barbouri* and *A. texanum* (mole salamanders) darkened in response to cold temperatures, though this plasticity was lost with age. An interesting note on melanism comes from Davis and Milanovich (2010) who found that melanic red-backed salamanders (*Plethodon cinereus*) had higher baseline stress levels than red-backed conspecifics (Fig. 1.2), something known to decrease growth and survival in unfavourable environmental conditions and increase disease susceptibility (see Davis and Milanovich 2010). This poses conservation concerns for melanic populations in the wake of climate change and disease-driven amphibian declines (IUCN 2008).

UV protection has also been suggested as a driver of adaptive colouration in amphibians (see Rudh and Qvarnström 2013). High UV exposure has a negative effect on fitness due to molecular damage, and some have suggested that increased melanism may provide protection (e.g. Jablonski 1998). This is supported by studies showing skin darkening in the larvae of seven species of anurans and caudates (*Hyla arborea, Bufo calamita, Triturus cristatus, Rana temporaria, Bufo bufo, Ambystoma barbouri* and *Ambystoma texanum*) when exposed to higher levels of UV radiation compared to controls (Langhelle et al. 1999; Garcia et al. 2004). However, while this trend was also seen in rough skin newt larvae (*Taricha* granulosa) by Belden and Blaustein (2002), they found no increase in fitness or survival between light and dark individuals, leading to uncertainty over the significance of UV radiation as a pressure on amphibian colouration.

1.3.3. Predator avoidance

Predator avoidance is a particularly important selective pressure on amphibians, as they are typically small, soft bodied, and lack hard protective structures (Rudh and Qvarnström 2013). Extensive variation in skin patterns and rapid colour change allow them to cryptically hide from predators (Wente and Phillips 2005), while conspicuous colours are either aposematic (warning colouration signalling unpalatability/toxicity) or used in deimatic (startle/intimidation) displays (Rudh and Qvarnström 2013).

Crypsis is achieved through background matching, disruptive colouration or masquerade (resembling something uninteresting to the predator). Background matching is seen in many ground-dwelling amphibians that are dark green or brown (Rudh and Qvarnström 2013), and leaf-sitting tree frogs that have converged to reflect light in the same wavelengths as photosynthetic leaves (Fig. 1.3; Schwalm and Starret 1977; Emerson et al. 1990). Lightening and darkening *via* rapid colour change is also used, for example in juvenile *Bufo americanus* (Heinen 1986) and *Hyla crucifer* (Kats and Vandragt 1986). In contrast, disruptive colouration breaks up the outline of an individual, making it unrecognisable to predators. Mid-dorsal lines in anurans (e.g. *Eleutherodactylus, Rana* and *Bufo*) and irregular dorsal patterns in many terrestrial salamanders (e.g. *Ambystoma maculatum*; Drake et al. 2009) are often claimed to be examples of disruptive colouration (Duellman and Trueb 1994), but this has rarely been tested (Rudh and Qvarnström 2013). Some species combine aspects of background matching, disruptive colouration and masquerade, such as the mossy frogs (*Theloderma* spp.) or leaf frogs (*Ceratobatrachus* spp.; Fig. 1.3).

Deimatic behaviour is where an animal tries to startle or intimidate a predator, usually as a 'last resort' after being detected (Rudh and Qvarnström 2013). A brilliant example of this is the anuran *Physalaemus nattereri*, which has a mottled brown cryptic dorsum, but when startled rears its posterior, inflates its body and presents two large eyespots (Fig. 1.3; Lenzi-Mattos et al. 2005). *Pleurodema brachyops* (the Colombian four-eyed frog) shows a similar display, presenting bright blue eyespots and red marks on their inner legs when attacked (Martins 1989).

Aposematic colouration is an inter-species signal widely seen in amphibians. This is where individuals have bright conspicuous markings, either as a 'warning' to predators that they are toxic or to mimic species that are (Darst and Cummings 2006; Venesky and Anthony 2007; Kuchta et al. 2008). The strong correlation seen between conspicuous colours and potent skin toxins in many amphibians strongly implies a common aposematic function (Rudh and Qvarnström 2013). In some species, brightness has been show to be directly correlated to toxicity (e.g. *Oophaga pumilio*; Maan and Cummings 2012) and toxicity to predation (e.g. *Taricha granulose*; Brodie and Brodie 1990). However, despite many claims, with the exception of a few studies in caudates (e.g. Hurlbert 1970; Hensel and Brodie 1976) and anurans (e.g. Saporito et al. 2007), few have rigorously tested if predators are actually deterred by supposedly aposematic colouration.

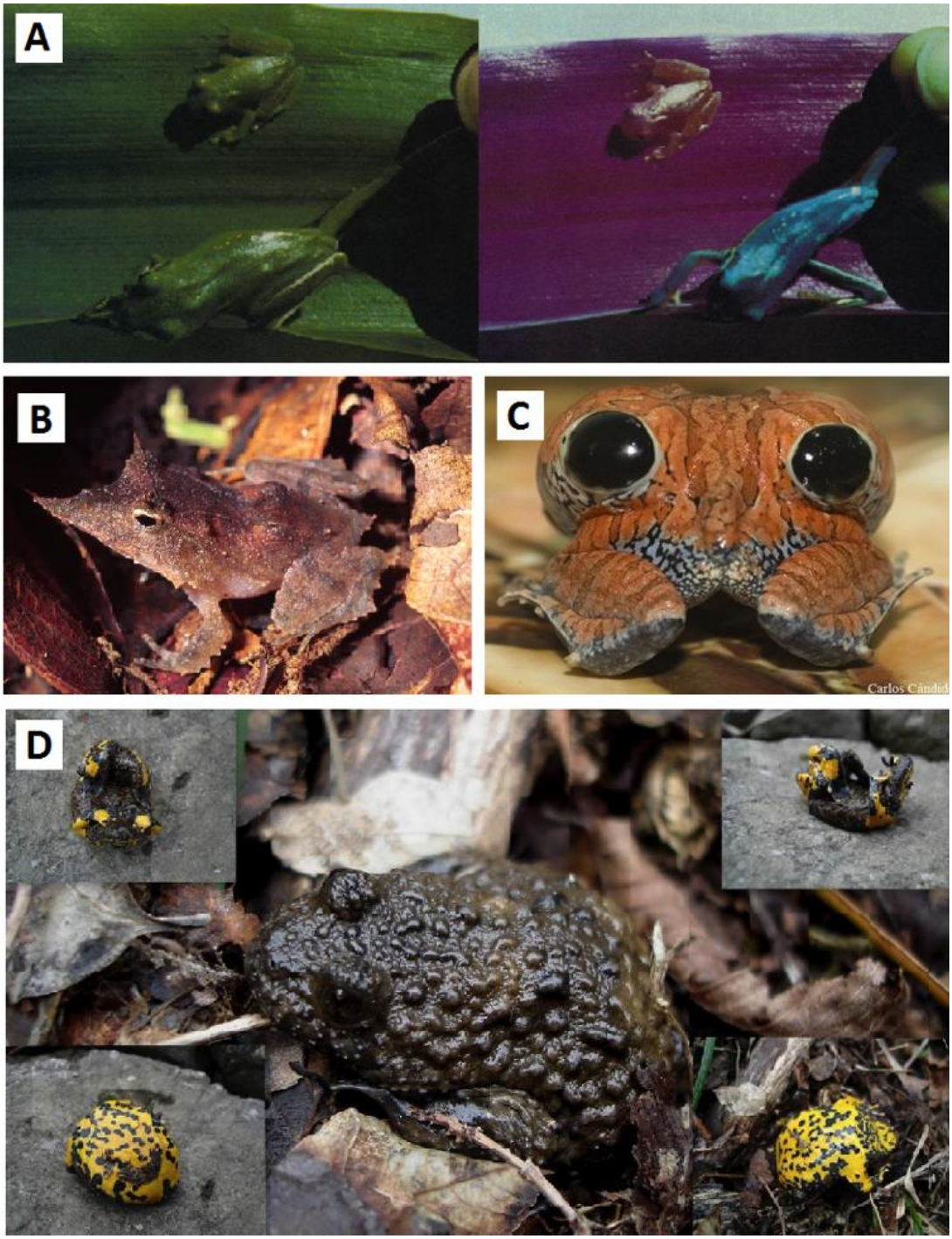


Figure 1.3: Predator avoidance**. A**: Background matching in a centrolenid (top) and a hylid (bottom) tree frog: left, appearance in the visual spectra (~400-700nm); right, appearance in the near-infra red spectra (~700-900nm; modified from Schwalm and Starret 1977); **B**: *Ceratobatrachus guentheri* (Solomon Island Leaf Frog), displaying cryptic masquerade, as it has evolved to look like a leaf (Photocredit: P. Ryan); **C**: Deimatic display by *Physalaemus nattereri* (Photocredit: C. Cándido); **D**: *Bombina variegata* (Yellow-Bellied Toad): middle, cryptic behaviour; surrounding images, unken reflex.

Crypsis, deimatic displays and aposematism are not mutually exclusive. Conspicuous colouration can be cryptic (disruptive colouration) when viewed on a different substrate or from a distance while being aposematic at close range (Stuart-Fox and Moussalli 2009). For example, toxic fire-bellied toads (*Bombina* spp.) have cryptic (black) warty dorsal surfaces, but when threatened display an 'unken reflex', where they flash their bright yellow throat, venter and limbs by flipping onto their backs (a deimatic display). They then remain in this position to display aposematic colours (Fig. 1.3; Bajger 1980). The rough skinned newt (*Taricha granulosa*) displays a similar behaviour (Johnson and Brodie 1975).

1.3.4. Communication

Colouration and colour displays play an important role in intraspecific communication for many vertebrates (Endler 1990; Caro 2005), but have been generally overlooked in amphibians (Rudh and Qvarnström 2013). This is due to the nocturnal nature of many species (Rudh and Qvarnström 2013) and the prevalence of chemical signalling (especially in caudates) or vocal calls (only confirmed in anurans; Wells 2007) for communication within the group. However, sexual dichromatism is more common in amphibians than commonly assumed, with both males and females showing ontogenetic or dynamic differences in colouration (Bell and Zamudio, 2012). The now extinct golden toad (*Bufo periglenes*) showed startling ontogenetic dichromatism, with males being bright orange and females black with scarlet blotches (Fig. 1.4; Jacobson and Vandenberg 1991). While this conspicuousness may have provided an aposematic benefit, sexual selection is a more likely explanation for this difference (Bell and Zamudio 2012).

Figure 1.4: Sexual dichromatism. **Left**: Male (top) and female (bottom) golden toad (*Bufo periglenes;* Photocredit A. Baertschi); **Right**: Male (top) and female (bottom) moor frogs (*Rana arvali*; modified from Rudh and Qvarnström 2013).

Dynamic dichromatism is a seasonal change of colour; for example, male moor frogs (*Rana arvali*; Fig. 1.4) turn bright blue for several weeks during the mating season (Ries et al. 2008). It is hard to attribute this to anything but mate selection, especially when it lasts for only a few hours and disappears soon after mating (e.g. *Incilius luetkenii*; Doucet and Mennill 2010). Studies showing that nocturnal species retain high colour discrimination in low light, and that females prefer conspicuous males, has led to increased interest in the role of colouration in intraspecific amphibian communication (Cummings et al. 2008; Gomez et al. 2009; Rudh and Qvarnström 2013). Sexual selection has even been implicated in the generation and maintenance of aposematic colour polymorphisms in amphibians like *Dendrobates pumilio*, the strawberry poison frog (Rudh et al. 2007).

1.4. Quantifying colour patterns

1.4.1. Subjective ranking

Although colour patterns are conspicuous and often easy to segregate, quantifying them in an objective manner has not always been easy. Colour patterns are complex, multicomponent, three-dimensional structures (Grether et al. 2004), which vary in size, shape, brightness and colour (Endler 1990). Historically, colour was almost universally subjectively ranked based on human perception or colour standards (Endler 1990); for example Harkey and Semlitsch (1988) studied the effect of temperature on developing *Pseudacris ornata* (the ornate chorus frog) and ranked individuals as green, grey, brown or copper by eye; these methods are still used in modern amphibian studies due to their simplicity (e.g. Summers and Clough 2001). The pitfalls of such methods are numerous and extensively reviewed in Endler (1990), but largely come down to the fact that colour is not a physical construct. The colour of an object is dependent on the ambient lighting, surrounding colours and, most importantly, the visual and neural properties of the receiver. This means that colour perception is highly variable both within and between species, and that the human visual system (400-700nm) is inappropriate for many studies. Also, with no quantitative measure, statistical analyses are either impossible or meaningless (Bergman and Beehner 2008).

The inadequacy of human ranking is seen clearly with leaf-sitting tree frogs in the families Centrolenidae and Hylidae. These frogs are predominantly green in the visual spectrum (~400-700nm), providing brilliant background matching and making them almost

undetectable to humans. However, they typically act as heat sinks for near infra-red (NIR) light, so vividly stand out on photosynthetic leaves in the NIR spectra (Schwalm and Starret 1977). While humans cannot see in this spectrum (~700-900nm), many potential frog predators can, including some snakes, birds and potentially even other amphibians (Schwalm and Starret 1977; Emerson et al. 1990; Liebau 2013). In response, several species of leaf-sitting frogs have evolved to reflect light in the NIR spectrum, making them virtually undetectable in both it and the visual spectra (Fig. 1.3; Schwalm and Starret 1977; Emerson et al. 1990). By only focusing on the human visual spectra, we would miss an important predation pressure. As some amphibians can also see in the UV spectrum (~300-400nm) as either larvae (see Chen et al. 2008) or adults (see Przyrembel et al. 1995), some reflect UV spectra (see Ries et al. 2008), and UV vision is seen to varying degrees in all main groups of potential amphibian predators (Jacobs 1992), it is clear that human ranking is inadequate to investigate amphibian colouration with any meaningful evolutionary or ecological perspective.

1.4.2. Spectrophotometry

Colour analysis was revolutionised through the use of spectrophotometry. This is the most accurate and objective method of measuring colour, as it provides spectral data on the wavelengths absorbed or reflected by an organism independent of the receiver's visual system (Stevens et al. 2007). It is used to quantify aspects of pigmentation such as brightness (the amount of light reflected), hue (traditional colour; e.g. red, green, blue) and chroma (intensity or saturation; Endler 1990). Spectrophotometers can measure spectral data from UV, visual and infra-red light wavelengths for use in a wide array of robust statistical analyses.

Spectrophotometric data can also be used in detailed evolutionary and ecological studies through the use of computational models. Where data are available on the photoreceptor sensitivities (and neural processing) of potential receivers, reflectance spectra can be fitted to non-human systems (e.g. Endler et al. 2005). An excellent example of this comes from Maan and Cummings (2012), who used spectral data from ten populations of the highly polymorphic *Oophaga pumilio* (strawberry poison dart frog) to show that brightness (a measure of conspicuousness) was directly correlated to toxicity levels. This was used to infer an aposematic function, and a model was run to evaluate the conspicuousness of individuals to potential predators using representative species with known photoreceptor sensitivities: a trichromatic snake, a dichromatic crab, and two different tetrachromatic

birds. This elegant study accurately measured colour and linked it quantitatively to both its ecological function and selective pressures in a way impossible with subjective human ranking.

Unfortunately, the use of spectrophotometry has several limitations. With increased accuracy comes a decreasing field of view, as readings are only possible for very small areas a couple of millimetres in diameter (e.g. Kraemer et al. 2012). These readings must also be taken flat, leading to issues with the rounded fleshy parts of amphibian anatomy (Villafuerte and Negro 1998). In order to compensate for this, several readings for the same location are often taken and averaged (e.g. Maan and Cummings 2012), but this still treats patterns as solid colours within subjectively defined regions. Standardised anatomical landmarks may be used to try and reduce this bias (e.g. Kraemer et al. 2012), but these are still restrictive and will potentially sacrifice important variation for standardisation.

Another pitfall of spectrophotometry is the risk of collecting 'misleading' data if the questions you are investigating are bound to the visual sensitivity of the receiver. Using spectrophotometry, individuals may show a high reflection at a specific wavelength; however, if the natural ambient light does not contain this wavelength, or the receiver is insensitive to it, this high reflectance is unimportant and could lead to inappropriate analyses (Stevens et al. 2007). Spectrophotometry also carries a large financial burden, as the advanced and highly specialised technology costs thousands to tens of thousands of dollars (US), meaning its application may be beyond the scope of smaller studies. In addition, while field units are available, many are still bulky, fragile and require large external power supplies for the constant and intense light needed (Villafuerte and Negro 1998). Fortunately there is another, cheaper and more versatile technology available for colour analysis: digital photography.

1.4.3. Digital photography: colour

Quantitative digital image analysis techniques allow for the measurement of colour and pattern across a whole specimen, providing the ability to study subtle variations in colour and shape (Davis et al. 2004, 2005; Davis and Maerz 2007). Digital images are a mosaic of pixels, small physical points created from red, green and blue (RGB) brightness values (from 0-225) recorded when a camera's matrix of microscopic photocells scans an image (Villafuerte and Negro 1998; Stevens et al. 2007; Bergman and Beehner 2008). Digital

photography has several advantages over spectrophotometry: it is considerably cheaper, its compact size and internal battery make it much more flexible and field friendly (e.g. Bergman and Beehner 2008), large data sets can be rapidly accumulated, and researchers can analyse entire spatial patterns using powerful computational algorithms, not just point data (Stevens et al. 2007). Digital photography also provides a permanent archive of specimens, which can be re-examined and used for further analysis (Davis et al. 2004), valuable in situations where specimens cannot be removed from the wild or studies with a temporal aspect; for example, ontogenetic studies (e.g. Beukema 2011). However, unlike spectrophotometry, which provides accurate reflectance spectra, digital photography only provides red, green and blue values (RGB colour model) based on the human visual spectra, which requires extensive calibration before analysis.

Images are affected by the object being imaged itself, its geometry, the ambient light, and the characteristics of the camera being used (Davis et al. 2004, 2005; Davis and Maerz 2007). As such, low-end 'point and click' automatic cameras are not sufficient for the digital analysis of colour, as a number of settings require manual control. This is extensively reviewed in Stevens et al. (2007); however, the key requirements needed are: 1) the ability to disable auto white-point balancing, which alters the saturation and can inappropriately weight RGB values; 2) a high resolution so that the pixels are at least half the size of the smallest detail being imaged (the Nyquist frequency) to prevent aliasing (indistinguishable signals) becoming an issue; 3) the ability to save images in an uncompressed file format, e.g. TIFF (Tagged Image File Formats) or RAW, as opposed to the more common JPEG (Joint Photographic Expert Group) format that undergoes 'lossy' compression and irreversibly loses and alters pixel data; 4) manual exposure control, including the aperture size (the opening through which light travels), shutter speed (the length of time the aperture is open) and ISO (the sensitivity of the camera to light) to avoid *a posterior* calibration (which may introduce sources of error)—the smaller/lower these are the more accurately light will converge on the photocells; and 5) good quality optics to reduce chromatic aberration, where different wavelengths focus differently, leading to blurring. However, while these are relatively easy to control, a more challenging aspect of digital photography is accounting for biases in the camera's internal image processing.

Digital cameras are designed to take aesthetically pleasing images, not realistic ones (Stevens et al. 2007). This means that we must be aware of, and compensate for, a number of internal biases. Colour (RGB) values vary from manufacturer to manufacturer and camera to camera (even within a model) and are not linearly related to light intensity, so

even though they can be obtained easily from commercial software (e.g. Photoshop, Adobe Systems Inc.) in their current format they tell us little about the colour of the corresponding point on the real object (Stevens et al. 2007). To linearise and equalise RGB channels, colour standards (e.g. Munsell chips, 1976) must be photographed in every image if using a variable light set up or at the start and end of each session if using a standardised light set up. For example, Stevens et al. (2007) plotted RGB values from images of grey Spectralon diffuse reflectance standards (Labsphere Inc.) of known greyscale value and transformed their data to get a linear relationship. As greys fall on the achromatic locus (Kelber et al. 2003) the three channels should be equal (R=B=G), so by equalising the three channels they also removed the effect of illumination. This is needed as radiance spectra vary dramatically with illumination, but the colour remains relatively constant ('colour consistency'). By linearising and equalising your RGB channels with colour standards you generate a calibration for your data, allowing it to more accurately represent the real colour of the object imaged and make different images comparable. This process can be automated in programs such as MATLAB, Photoshop or open-source software like GIMP and ImageJ. Villafuerte and Negro (1998) developed their own software using Windows Visual Basic, which allowed them to linearise their RGB values across a set of images taken under identical conditions using the first as a standard. Some modern colour standards (e.g. the X-rite ColorChecker Passport) even come with their own calibration software.

While the above issues may sound problematic, they can be dealt with easily and only present a problem if not accounted for in your experimental design; however, digital photography does have some unavoidable drawbacks. Standard digital cameras only operate within the human visual spectra, which may not cover the biologically relevant wavelengths for your study. Though it is possible to image UV (e.g. Kodric-Brown and Johnson 2002; Stevens et al. 2007) and NIR (e.g. Schwalm and Starret 1977) spectra using specialised filters or cameras, this is not common practice and may be cost prohibitive. The data collected are also not wavelength specific, so despite knowing the wavelengths contributing to each RGB channel, it is impossible to tell what wavelengths generated a specific pixel. While this can be approximated using 'multispectral' imaging, by rotating a set of wavelength specific filters in front of the lens, this is time consuming and only practical if detailed waveband data are required (Stevens et al. 2007). The TIFF and RAW file formats needed are also very large, being almost six times the size of JPEGs, potentially creating data storage issues (Stevens et al. 2007). However, if the human visual system is appropriate, exact wavebands are not required and the calibrations can be made,

digital photography can be a powerful tool when researching animal colouration (e.g. Siva-Jothy 1999; Wilson et al. 2001; Thompson et al. 2002; Davis et al. 2004, 2005; Stevens et al. 2007). In addition, digital photography allows the shape of colour patterns to be measured.

1.4.4. Digital photography: shape

Amphibians show a wide diversity of colour patterns and polymorphisms, which differ in both area and shape, including spots, stripes and more complex patterns (Hoffman and Blouin 2000). Subjective ranking by human observers is only adequate when differentiating between distinct polymorphisms: multiple discontinuous phenotypes (Mayr 1963). For example, O'Neill and Beard (2010) used it in their study of colour pattern inheritance in the polymorphic *Eleutherodactylus coqui* (the coqui frog), which displays six distinct phenotypes. However, when more detailed analyses are required, or where colour patterns show a more continuous gradient, this method is less satisfactory. Early attempts at quantification were conducted through measures such as spot number or colour area ratios (e.g. Degani 1986), but as with colour analysis, the large number of powerful, flexible and customisable software programmes now available, coupled with digital imaging, has radically improved our ability to understand and quantify shape.

Digital photography has already been used for a number of quantitative studies on amphibians. Davis and Maerz (2007) used digital image analysis (Fovea Pro, Reindeer Graphics Inc.) to measure the spot number, area and roundness in *Ambystoma maculatum* (the spotted salamander) to investigate Wright and Zamudio's (2002) claim that spot asymmetry increased with environmental stress. A particularly interesting study of the analysis of shape was the application, for the first time, of geometric morphometric techniques to standardise ventral shapes in *Salamandrina perspicillata* (the northern spectacled salamander) by Costa et al. (2009). The colour patterns of *S. perspicillata* are too variable to be classified as distinct polymorphisms, so instead Costa et al. (2009) used a mathematical approach to identify population specific colour patterns. First they used a K Nearest Neighbours (KNN) supervised multivariate clustering method (Belur 1991) in MATLAB (Shakhnarovich et al. 2006) to make their images trichromatic (corresponding to colour patches). The shape of each individual was adjusted using landmark-based geometric morphometry and consensus images for each population generated based on common pixels (TpsSuper; Rohlf 2003). This adjustment ('warping') uses a generalized orthogonal least-squares Procrustes superimposition (translation, scaling, and rotation),

which removes size and orientation as variables, leaving only shape (Rohlf and Slice 1990). A partial least squares (PLS) analysis (ter Braak and de Jong 1993), one of the most powerful statistical methods in multivariate classification (Costa et al. 2009), was then used to show that patterns were population specific. When combined with an ANOSIM (ANalysis Of Similarities; Clarke 1993), they were able to show that the head, neck and fore-girdle are particularly powerful for discriminating populations, suggesting that this region is under selection, perhaps *via* intraspecific communication pressures. A Mantel test was also used to show that population-specific patterns were independent of geographic distance, as even geographically close populations were discernible. As such, this method could conceivably be used to identifying fine-scale populations and appropriate phenotypic characters (units) for genetic investigations.

While the approach used by Costa et al. (2009) is labour intensive and potentially methodologically prohibitive, the accessibility of such methods was recently increased by the release of the R package *patternize* (Van Belleghem et al. 2017). This package allows researchers to quantify variation in colour patterns from digital image data by first detecting homology between pattern positions across images, then categorising the distribution of colours in the image using one of several algorithms. A number of analyses can then be conducted, such as principal component analysis, to conduct population comparisons or accurately quantify colour phenotypes for things like genetic association studies.

1.5. The genetics of colouration

1.5.1. Current synthesis and loci of major effect

Because colour polymorphisms are easily scored and often show simple Mendelian inheritance, species displaying them are ideal for studying microevolutionary processes in nature. While amphibians seem ideal for such studies, there is a lack information on the genetics of amphibian colouration and few appropriate reference genomes (Hoffman and Blouin 2000; Rudh and Qvarnström 2013). Most studies to date have focused on the inheritance of colour polymorphisms, for example, O'Neill and Beard (2010) found that crosses of striped and un-striped coqui frogs (*Eleutherodactylus coqui*) provided an offspring phenotype ratio congruent with a single autosomal locus, 5-allele model in which striped morphs were dominant to un-striped. However, Hoffman and Blouin (2000) found

that the mode of inheritance in colour polymorphic amphibians has only been investigated in 26 species, despite there being 225 polymorphic anuran species described at the time. This is probably due to the difficult captive care needed by, and complex reproductive behaviour seen in, many species (Rudh and Qvarnström 2013), although this can be overcome by reconstructing molecular pedigrees from wild populations (e.g. Richards-Zawacki et al. 2012). However, inheritance based studies do not identify the genetic loci responsible for colour polymorphisms, and it is in this regard that data on amphibians are particularly lacking when compared to other vertebrate taxa.

Most studies on the underlying genetics of vertebrate pigmentation have focused on mammalian model organisms (reviewed in Hoekstra 2006), with over 150 genes identified (see Hubbard et al. 2010) and over 100 of these in mice (Bennett and Lamoreux 2003). This is due to the fact that laboratory mice (*Mus musculus*) have been bred for generations so that strains differ by as little as one gene (knockout/knockin lines). While subsequent candidate gene analyses on wild and domestic populations of other mammals (e.g. Kijas et al. 1998; Eizirik et al. 2003; Hoekstra et al. 2004; Anderson et al. 2009), birds (e.g. Theron et al. 2001; Doucet et al. 2004; Mundy et al. 2004; Poelstra et al. 2013), fish (e.g. Odenthal et al. 1996; Kelsh 2004; Gross et al. 2009; Henning et al. 2013) and even reptiles (e.g. Rosenblum et al. 2004; McLean et al. 2017) has shown that the underlying genetics of vertebrate pigmentation are remarkably conserved, there is a dearth of data from amphibians. A review on the genetics of adaptive pigmentation in vertebrates by Hubbard et al. (2010) failed to mention amphibian taxa once, despite including all other vertebrate groups. This could be due in part to the size of amphibian genomes, which, especially in caudates, tend to be very large (Mohlhenrich and Mueller 2016; Gregory 2017). For example, the genome of *Ambystoma mexicanum* (a model amphibian used in evolutionary, developmental and regeneration studies) is approximately 32Gb—over 10 times the size of the mouse genome (Gregory 2017; Evans et al. 2018). However, data gathered from other taxa have provided candidate colour genes, some of which have been investigated in relation to amphibian pigmentation.

Common colour phenotypes seen across vertebrates include melanism, albinism, stripes and spots. The *melanocortin 1 receptor* (*MC1R*) gene is conserved across vertebrates (see Herczeg et al. 2010) and controls melanin production (see Mills and Patterson 2009). In mammals, birds and reptiles, mutations in *MC1R* lead to either darkening (increased activity, more eumelanin) or lightening (decreased activity, more pheomelanin). However, within amphibians some studies suggest it is associated with increased skin melanisation

(e.g. Posso-Terranova and Andres 2017), and others that it is not (e.g. Alho et al. 2010). In mammals *MC1R* is also closely associated with *Agouti*, a gene responsible for colour banding, and mutations in these two genes are the most common cause of melanism in the group (Mills and Patterson 2009). A potential *Agouti* homolog, melanisation inhibiting factor (*MIF*), has been found in anurans (Fukuzawa and Ide 1988; Fukuzawa and Bagnara 1989; Bagnara and Fukuzawa 1990; Fukuzawa et al. 1995), and may provide insights into the genetic basis of amphibian striping. However, whether this is an *Agouti* homolog has yet to be robustly assessed (Mills and Patterson 2009).

Albinism (the complete lack of melanin), though uncommon, has also been seen in several wild amphibians (Mills and Patterson 2009; Modesti et al. 2011). Mutations in *OCA2* (*ocular and cutaneous albinism 2*) has lead to the parallel evolution of albinism in cave-dwelling tetras (Protas et al. 2006), and mutations in *Tyrosinase* (*TYR*) and *Tyrosinase related protein-1* (*TYRP1*) are known to produce albinism in humans (Oetting *et al*., 2003). Recently, it has also been revealed that mutations in *TYR* are responsible for albinism in laboratory bred *Ambystoma mexicanum* (axolotls; Woodcock et al. 2017), confirming a long suspected causality in amphibians (Smith-Gill et al. 1972).

Stripes and spots are also common amphibian colour pattern components; for example, *Salamandra salamandra* (the European fire salamander) displays both (Veith 1992). Along with fish (e.g. adult zebrafish; *Danio rerio*), amphibians like larval salamanders (e.g. *Ambystoma tigrinum*) are well-known models for structural stripe development (Mills and Patterson 2009). While the genetic basis of stripe formation is unclear in amphibians, recent work has shown that *csf1r* (Parichy and Turner 2003) and *nacre* (Lister et al. 1999) are important for stripe formation in zebrafish (Mills and Patterson 2009). Likewise, *Kit* and *Ednrb* are known to have roles in the creation of spots in mammals and fish (reviewed extensively in Mills and Patterson 2009). However, despite their prevalence, I could find no studies on the genetic basis of spots in amphibians.

1.5.2. Next generation sequencing

Until recently, genome-wide investigations were prohibitively expensive and time consuming, forcing researchers to focus on just a few small fragments of DNA typically hundreds of base pairs in length (Rokas and Abbot 2009). However, with advancing technology, we can now carry out whole genome sequencing (WGS) of wild non-model organisms using next generation sequencing methods (NGS; also known as high-

throughput sequencing); this allows us to study the genomic basis of ecologically-significant characters (Hubbard et al. 2010; Stapley et al. 2010). These NGS approaches can generate gigabases of sequence data and identify tens of thousands of genetic markers (microsatellites, SNPs and indels; Davey et al. 2010; Stapley et al. 2010). They also provide a dramatic reduction in per-base cost and a dramatic increase in the number of bases sequenced compared to earlier methods (Rokas and Abbot 2009; Snyder et al. 2010); sequencing the human genome (~2.85Gb) has fallen from over $18 million in 2004 using automated Sanger sequencing to around $1000 today (Wetterstrand 2017). However, while ultra high-throughput platforms make WGS possible, it is often still impractical, and too expensive, to fully sequence numerous eukaryote genomes over 1000Mbp (Davey and Blaxter 2010; Geneva et al. 2015). This is particularly true of the large genome sized amphibians (Mohlhenrich and Mueller 2016; Gregory 2017).

To date, whole-genome sequence data is only available for five amphibian species: the African clawed frogs *Xenopus tropicalis* (Hellsten et al. 2010) and *Xenopus laevis* (Session et al. 2016; both important model-organisms), the Tibetan frog *Nanorana parkeri* (Sun et al. 2015), the American bullfrog (*Lithobates catesbeianus*; Hammond et al. 2017) and the Iberian ribbed newt (*Pleurodeles waltl*; Elewa et al. 2017). However, extensive genomic resources in the form of transcriptome data are available for at least a further 22 anuran and six caudate species (Kwon 2017). These resources contain data on expressed protein-coding genes, but in order to get markers from non-coding or unexpressed regions of the genome without WGS, a method like restriction site-associated DNA sequencing (RAD-Seq) is required.

RAD-Seq is a reduced complexity short-read method of accurately scoring tens of thousands of genetic markers across a genome (Davey and Blaxter 2010). Developed by Baird et al. (2008), this genome sequencing technique effectively subsamples the genome in a replicable way by using restriction enzymes to digest DNA at specific recognition sites, thereby providing a random but repeatable distribution of short reads that can be analysed for polymorphisms (Davey and Blaxter 2010). This allows for the rapid and cost-effective genotyping of multiple individuals. For example, Baird et al. (2008) used RAD-Seq to identify over 13,000 SNPs and map three ecologically important traits in 96 individuals (two populations) of threespine sticklebacks (*Gasterosteus aculeatus*). Since then, it has been used in countless ecological, evolutionary and conservation genomic studies (reviewed in Andrews et al. 2016).

An extra benefit of RAD-Seq is that reference genomes, while desirable, are not necessary, enabling studies of non-model taxa (Davey and Blaxter 2010; Stapley et al. 2010). This is because reduced complexity genomic studies are cheap and fast enough to be conducted in numerous individuals from closely related populations, or species, which show ecologically important traits over wide phenotypic, geographic and temporal scales, and bioinformatic programmes are sophisticated enough that *de novo* genotyping of loci is possible (Catchen et al. 2011). This allows for *de novo* genome assembly (Catchen et al. 2011), genome-wide linkage mapping and association studies (the statistical identification of markers associated with specific phenotypes), and the identification of outlier loci (areas of the genome showing extreme levels of divergence; Stapley et al. 2010). When integrated with landscape genomic (the integration of environmental and topographic data; Allendorf et al. 2010), phylogeographic, and phylogenomic analyses (McCormack et al. 2013) this will enable us to interpret selection and understand the origin and maintenance of adaptive traits (Stapley et al. 2010). This opens the door to a host of complex ecological and evolutionary applications that were previously impossible in non-model taxa, for example, determining the genomic basis of colouration.

While limited, several studies have used NGS approaches to investigate colour patterns in non-model taxa. By whole genome sequencing three parental genomes and RAD-Seq genotyping 16 offspring at 509,220 SNP-loci, Xu et al. (2013) were able to identify a single mutation in the transporter protein *SLC45A2* as the causative loci for producing the rare white colour morph of the Bengal tiger (*Panthera tigris tigris*). A similar study used a transcriptome profiling technique called EcoP15I-tagged Digital Gene Expression (EDGE) on both domestic cats (*Felis silvestris catus*) and wild cheetahs (*Acinonyx jubatus jubatus*; Kaelin et al. 2012). Through this, the researchers were able to identify mutations in the gene encoding transmembrane aminopeptidase Q (*Taqpep*) and differential expression of Endothelin3 (*Edn3*) as controlling the formation of spots and stripes. Transcriptome profiling (RNA-Seq) was also used by Mallarino et al. (2016) to identify new colour pattern related functions for an already well characterised transcription factor (*ALX3*) in the African striped mouse (*Rhabdomys pumilio*). It has been known for some time that *ALX3* plays an important role in rodent developmental biology; however, only by investigating its expression in wild non-model taxa was its involvement in both stripe pattern formation and reduced melanin production (through the direct repression of *Mitf*, a master regulator of melanocyte differentiation) able to be identified.

Next-generation sequencing approaches can also be used for more exploratory studies, or for the identification of multiple causal or co-expressing pigmentation related genes. For example, Poelstra et al. (2015) looked at gene expression (RNA-Seq) in colour variable corvid species (within and between individuals and species), and Henning et al. (2013) used similar approaches to look at gene expression differences in polychromatic Midas cichlids (*Amphilophus citrinellus*); both identified significantly differentially expressed genes associated with different colour morphs. However, NGS methods are useful even where *a priori* genome information is extremely limited, such as a study by Richards et al. (2013) on colour and banding patterns in *Cepaea* land snails. Using RAD-Seq data they were able to identify a number of putative (unidentified) loci for further study near a supergene known to control colour and banding in the species.

While I could find no studies applying NGS methods to colour loci identification in amphibians, such approaches could easily be applied to investigate colour and pattern formation in the group. Importantly, NGS removes our reliance on captive bred and housed animals, an important consideration as the long generation time and challenging husbandry required by many amphibian species (Ferrie et al. 2014) makes such practices impractical. Also, while the size of amphibian genomes still poses a large bioinformatic task (Metzker 2010), NGS techniques mean that this is no longer a major limiting factor.

1.6. Fire and alpine salamanders (*Salamandra* spp.)

1.6.1. A new colour model

Amphibians clearly represent an understudied group with great potential for studying the genetic basis of adaptive colouration, but one question remains, what taxa do we focus our resources on? While NGS makes it possible to investigate multiple genomes relatively cheaply, it is still impractical and cost prohibitive to do this *ad libitum*. Obvious candidates for colour genomic studies are the traditional amphibian models *Xenopus* spp. and *Ambystoma mexicanum*. However, while both display colour polymorphisms, these tend to be selectively bred not natural, making it hard to interpret the ecological or evolutionary significance of genomic loci identified as being colour related.

Many studies looking at the ecological basis of amphibian colouration have focused on species like *Oophaga pumilio*, the strawberry poison frog (e.g. Saporito et al. 2007; Maan

and Cummings 2012; Dreher et al. 2017). Given its diverse colour polymorphisms, this may seem like an ideal candidate. Unfortunately, most *Oophaga* species have declining endangered populations, are CITES listed, and/or inhabit hard to reach neotropical rainforest canopies (IUCN 2008). The European common frog (*Rana temporaria*) is also considered a good candidate by some researchers (e.g. Herczeg et al. 2010). However, it represents a polyphyletic species (see Frost et al. 2006), and without an accurate phylogeny it is difficult to trace character evolution and therefore make meaningful evolutionary inferences. Many anurans also display physiological colour change based on reproductive state or environmental conditions (Rudh and Qvarnström 2013). In terms of caudates, perhaps the most studied colour polymorphic species is *Plethodon cinereus* (the red-backed salamander). However, this polymorphism manifests as the presence or absence of a single red stripe (Davis and Milanovich 2010), making it less attractive than more variable systems. Given the shortcomings of these species, I instead present the Palaearctic genus *Salamandra* as the ideal candidate for studying the genetic basis of adaptive colouration in amphibians.

This genus of 'true salamanders' (family: Salamandridae) is found throughout most of mainland Europe, on the Mediterranean island of Corsica, and in restricted locations in the Middle East and North Africa (Fig. 1.5; Thiesmeier 2004; Steinfartz et al. 2007). While their taxonomy has been highly controversial (see Thiesmeier 2004; Dubois and Raffaëlli 2009; Speybroeck et al. 2010; Frost 2017), six species are widely recognised: the European fire salamander, *Salamandra salamandra*; the Alpine species *Salamandra atra* and *Salamandra lanzai*; the Corsican fire salamander, *Salamandra corsica*; the North African fire salamander, *Salamandra algira*; and the Near Eastern fire salamander, *Salamandra infraimmaculata* (Fig. 1.6; Speybroeck et al. 2010; Sillero et al. 2014). However, although species distributions are well understood, the biogeographic history of the genus is less clear.

To date, phylogenetic studies have been unable to adequately identify the position of the root, making it difficult to infer the early geographic origin of the genus; although its sister genus, *Lyciasalamandra*, is restricted to south-western Turkey and the Greek Aegean islands (Vences et al. 2014). However, major geological events are likely to have shaped the evolutionary history of *Salamandra*. The desiccation of the Mediterranean during the Messinian salinity crisis (between 5.96 and 5.33 million years ago) produced land bridges between North Africa, the 'island' of Corsica and

mainland Europe (Duggen et al. 2003). The refilling of the Mediterranean approximately 5.33 million years ago may therefore represent a vicariant event between the ancestral populations of *Salamandra* (Vences et al. 2014). Northern Quaternary glaciations are also thought to have had a great impact in the intraspecific diversity of the most geographically widespread species: *S. salamandra*. Like many European amphibians, cyclical glaciation events forced ancestral *S. salamandra* into refugia in the Iberian, Apennine and Balkan peninsulas of Southern Europe (Dufresnes and Perrin 2015). As the ice sheets receded, *S. salamandra* recolonized the continent from these refugia, with Iberian populations proving to be particularly successful (Steinfartz et al. 2000). Such biogeographic events may have played an important role in shaping the colour pattern diversity seen across the genus through the isolation (and occasional secondary contact) of ancient *Salamandra* populations.

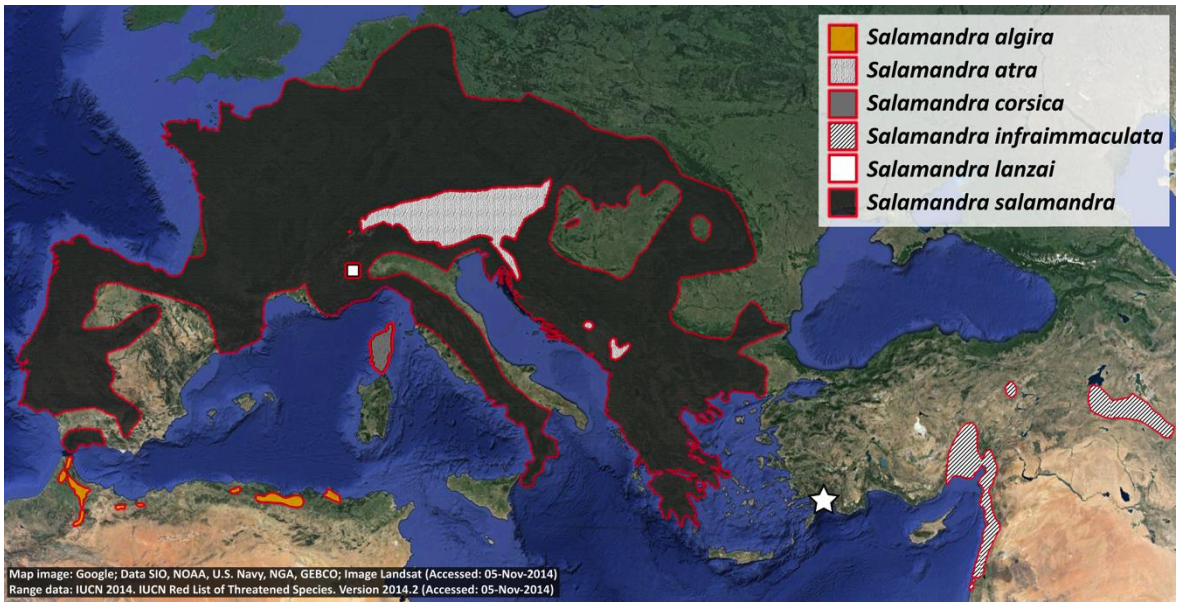


Figure 1.5: The geographic distributions of the six currently recognised species in the genus *Salamandra*. The white star indicates the approximate location of the sister genus *Lyciasalamandra* (south west Turkey).

Salamandra spp. display a striking level of intra- and interspecific variation in dorsal colour patterning (Fig. 1.6), in terms of both colour (yellow, black, orange, red, brown and white) and pattern (spots, stripes and solid; Thiesmeier, 2004; Sparreboom and Arntzen, 2014; Velo-Antón and Buckley, 2015). These colour patterns are putatively adaptive for

both aposematism and thermoregulation (Beukema et al. 2016b), and there are even indications that the same (or similar) colour patterns have evolved independently in more than one lineage (Steinfartz et al. 2000; Vences et al. 2014). This provides natural evolutionary replicates in which to study the contributions of standing genetic variation and *de novo* mutation in the parallel evolution of adaptive colouration. Also, while some studies indicate possible ontogenetic changes during sexual maturation (e.g. Eiselt 1958; Mutz 1992; Donaire-Barroso and Bogaerts 2001; Bogaerts 2002; Beukema 2011), *Salamandra* colour patterns are stable throughout adult life, and can be used to recognise individuals over consecutive years (e.g. Feldmann 1987; Schmidt et al. 2005; Warburg 2006, 2007a,b, 2008a,b). This stability, combined with a lack of plastic colour change, suggests a predominantly genetic origin for skin colouration in this group.

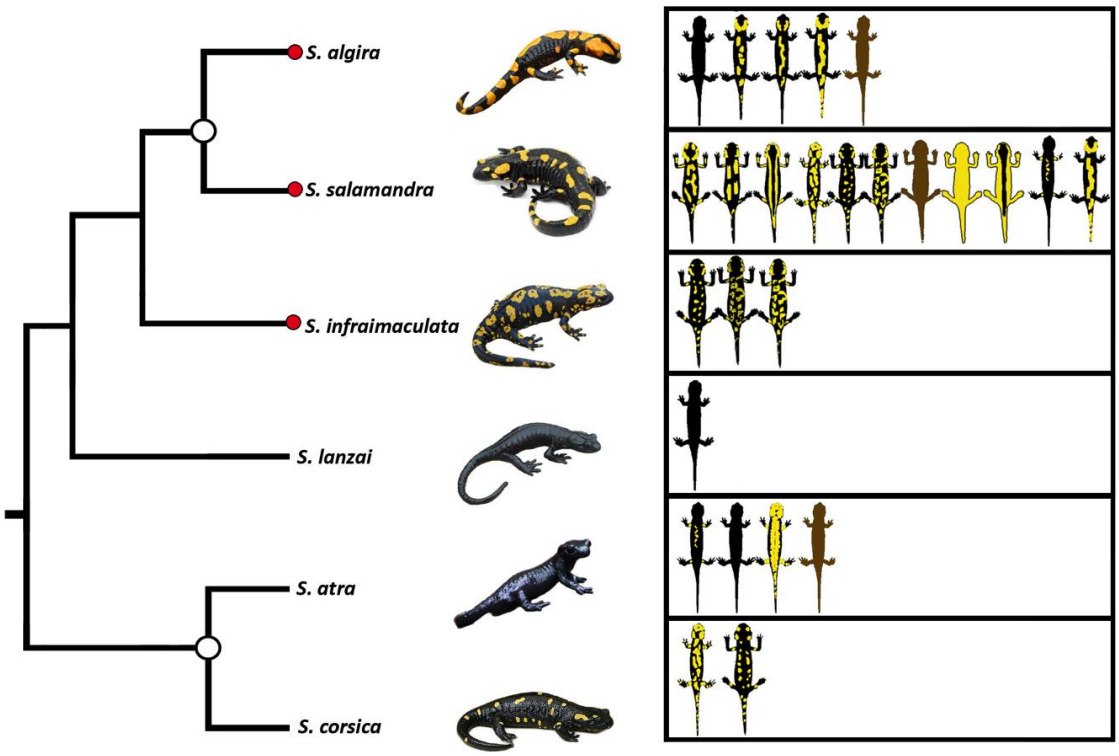


Figure 1.6: Phylogenetic relationships between the six currently recognised *Salamandra* species (data from Vences et al. 2014), along with the 'typical' phenotype for each and simplified representations of intraspecific colour variation (data from Sparreboom and Arntzen 2014). White circles show those nodes consistently recovered across analyses of different molecular datasets in Vences et al (2014). Red circles indicate those species where red pigmentation is commonly, though inconsistently, seen.

1.6.2. Taxonomic and colour diversity within Salamandra

Salamanders in the genus *Salamandra* have fascinated natural historians for centuries due to their remarkable biological traits. For example, they are one of the few vertebrates to show intraspecific variation in reproductive mode, and are one of only two viviparous salamander genera (where the female gives birth to aquatic larvae or fully developed juveniles; García-París et al. 2003; Buckley et al. 2007; Beukema et al. 2010; Buckley 2012). In particular, molecular phylogenetic work indicates that one reproductive mode—pueriparity, the deposition of fully formed terrestrial juveniles—has independently evolved in the *S. lanzai* + *S. atra* clade, *S. algira* and *S. salamandra* (García-París et al. 2003; Steinfartz et al. 2007b; Velo-Antón et al. 2007; Beukema et al. 2010; Vences et al. 2014). However, *Salamandra* is most renowned for its colour diversity.

The fire salamanders (*S. infraimmaculata*, *S. corsica*, *S. algira* and *S. salamandra*), as opposed to the alpine salamanders (*S. atra* and *S. lanzai*), are typically found in broadleaf forests, and are diagnosed based on their characteristic, yet highly variable, yellow-black markings. These patterns may form either spots or stripes, with varying degrees of intraspecific variation (Veith 1992), and are thought to be aposematic in function given their highly conspicuous nature (yellow being particularly conspicuous in a shaded forest habitat; Macedonia et al. 2000) and the fact that these salamanders secrete potent steroidal alkaloid toxins (e.g. García-París et al. 2003; Vences et al. 2014). While these striking colour patterns overlap and cannot be used to robustly categorise *Salamandra* specimens taxonomically (Böhme et al. 2013; Beukema et al. 2016a), they have been used to help inform species and subspecies designations (Thorn and Raffaëlli 2001), and populations/species do display characteristic phenotypes (Thiesmeier 2004; Sparreboom and Arntzen 2014; Velo-Antón and Buckley 2015; Speybroeck et al. 2016).

The most easterly species, *S. infraimmaculata*, is thought to have three distinct lineages. The nominative subspecies, *S. i. infraimmaculata*, ranges from Israel/Palestine to south-western Turkey, while *S. i. orientalis* is found in central and south-central Turkey—both present yellow spots on an otherwise black dorsal surface. The third subspecies, *S. i. semenovi*, is found in neighbouring regions of Iraq, Turkey and Iran (i.e. Kurdistan), and displays yellow rings and semicircles on black, with some populations also having a red/brown tinge. However, colour patterns between subspecies have been known to overlap (Böhme et al. 2013). The Corsican species, *S. corsica*, also displays a yellow-black

spotted pattern; however, its yellow patches are typically larger and more irregular (Beukema 2011).

The North African *S. algira* has a patchy distribution, primarily in northern Morocco and eastern Algeria. The nominative subspecies (*S. a. algira*) is found in eastern Algeria and the Moroccan Mid-Atlas mountains (the later forming a genetically distinct clade, see Chapter 3). As with all *S. algira* subspecies, it typically presents elongated yellow markings on a black dorsum (Martínez-Solano et al. 2005), but some eastern populations also have small red and white dots (Bouzid et al. 2017). Of the remaining three subspecies, two are characterised by more extensive red pigmentation: *S. a. spelaea* (from the Moroccan Central Rif mountains) has small red speckles and *S. a. splendens* (restricted to the Beni Snassen Massif, eastern Morocco) has extensive red patches (Martínez-Solano et al. 2005; Escoriza and Del Mar Comas 2007). Finally, *S. a. tingitana*, from northern Morocco, consistently lacks red colouration; however, hypoluteism (brown colouration) and melanism are common in the northernmost populations (Martínez-Solano et al. 2005).

The final species of fire salamander, the European *S. salamandra*, is the widest ranging and shows the greatest intraspecific colour diversity (Fig. 1.6–1.7). The Iberian peninsula is the most diverse region, where nine of the 13 currently accepted subspecies of *Salamandra salamandra* are found, eight of which are endemic (Salvador 1974; Gasser 1978; Veith 1992; Joger and Steinfartz 1994; Steinfartz et al. 2000; Martínez-Solano et al. 2005). Three of these display a yellow-black spotted phenotype (*S. s. almanzoris*, *S. s. longirostris* and *S. s. terrestris*; Fig. 1.7), with another four showing red pigmentation on top of this (*S. s. crespoi*, *S. s. morenica*, *S. s. bejarae* and *S. s. gallaica*; Fig. 1.7). The remaining two Iberian subspecies (*S. s. fastuosa and S. s. bernardezi*) characteristically have strong yellow-black dorso-lateral stripes (Fig. 1.7). While colour patterns vary considerably within all of these subspecies, two lineages are of particular note. The first is *S. s. gallaica* (found in western Spain and Portugal), which is often considered the most chromatically diverse subspecies, ranging from extensive yellow/red patterning to almost complete melanism (Velo-Antón and Buckley 2015). Also, within the northern Spanish *S. s. bernardezi*, one lineage (formerly known as *S. s. alfredschmidti*) is renowned for having fully brown (hypolutic) and fully yellow (xanthic) individuals (Köhler and Steinfartz 2006; Beukema et al. 2016a).

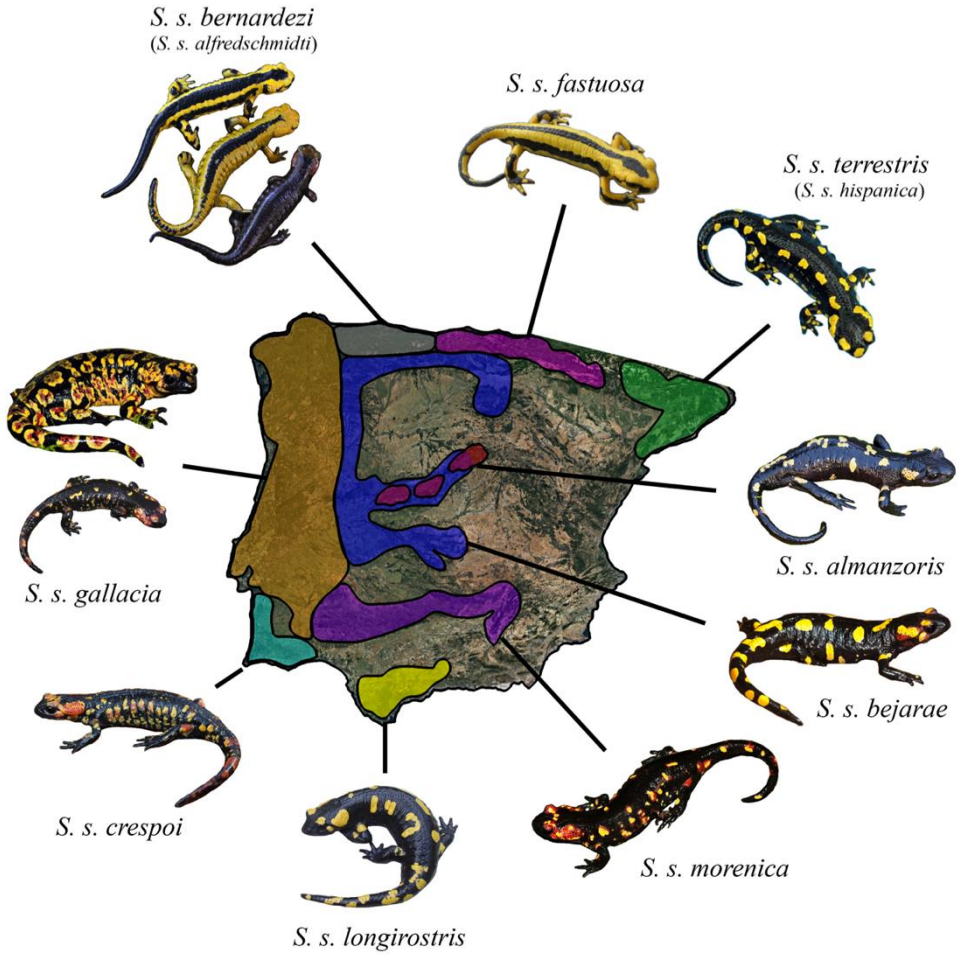


Figure 1.7: Geographic distributions and characteristic colour phenotypes for Iberian *Salamandra salamandra* subspecies. Names in brackets indicate synonyms that are still used by some authors. (Distribution data taken from Pereira et al. 2016.)

Outside of Iberia *S. salamandra* is also chromatically diverse. Distributed in northern and western Europe, *S. s. terrestris* (sometimes known as *S. s. hispanica* in northern Spain) typically presents two continuous or discontinuous dorso-lateral rows of spots (Speybroeck et al. 2016); in most populations these form 'broken' stripes, but in some they are phenotypically similar to *S. s. bernardezi/S. s. fastuosa*. One population of *S. s. terrestris* from the German region of Solling is particularly renowned for its unusually coloured individuals, which include red/orange-patterned, albino and melanic morphs; however, such atypical patterns are rare in the wild (Siedel et al. 2012). In contrast, *S. s. salamandra* (from central, eastern and southern Europe) have an irregular yellow-spotted phenotype. A similar pattern is seen in *S. s. werneri* from southern/central Greece and *S. s. beschkovi,* from the Pirin mountains in Bulgaria, except that *S. s. werneri* also shows red pigmentation and the spots in *S. s. beschkovi* tend to cluster along the spine, creating an irregular stripe (Obst 1981; Biserkov 2007). The final subspecies in *S. salamandra* is the Italian *S. s.*

giglioli, which has an irregular yellow-black pattern, where yellow patches fuse together to cover a large portion (or the entirety) of the dorsal surface (Sparreboom and Arntzen 2014; Speybroeck et al. 2016).

In contrast to the diversity of colour patterns seen in the fire salamanders, the Alpine species *S. lanzai* and *S. atra* are characterised by full melanism, where xanthophore-iridophore complexes fail to develop (Bonato and Steinfartz 2005); as these salamanders inhabit high altitude (over 1200m) microthermic environments, it is possible that their melanism fulfils a thermoregulatory function. The first of these species, *S. lanzai* is a highly sedentary fully melanic species that has the smallest range of any Palaearctic salamander, only being found in a ~10km^2 area of the Cottian Alps (Grossenbacher 1997). They also show an extreme lack of genetic diversity, with just a single mitochondrial lineage (Riberon et al. 2002a).

The second Alpine species, *Salamandra atra*, is more wide-spread and contains at least three lineages corresponding to distinct colour morphs (Bonato and Grossenbacher 2000; Riberon et al. 2002b). The nominative subspecies, *S. atra atra,* is fully melanic (although rare brown individuals are seen; Bonato and Steinfartz 2005) and wide-spread, being found over much of the Alps region in beech woodlands and alpine meadows (Bonato and Grossenbacher 2000). The two other subspecies are geographically restricted to the Venetian Prealps (in northeast Italy), and do display yellow markings despite living in similar habitats to the fully melanic subspecies. The first of these, *S. atra aurorae*, is only found in the northernmost part of the Sette Comuni plateau and is characterised by extensive yellow colouration, which can cover most of its dorsal surface (Grossenbacher 1994). In contrast, *S. a. pasubiensis* (only found in the Pasubio massif) is highly melanistic, possessing only a few yellow spots (Bonato and Steinfartz 2005). A fourth subspecies from the Dinaric Alps (*S. a. prenjensis*) has also been proposed (see Dubois and Raffaëlli 2009) due to its geographic isolation and characteristic hypolutic (brown) colouration. However, this subspecies has not been widely recognised as it is neither genetically or chromatically distinct (see Speybroeck et al. 2010).

Despite the fact that these dorsal patterns have played such a prominent role in the taxonomic description of *Salamandra* species and subspecies, few studies have tried to quantitatively investigate them (but see Degani 1986 and Beukema 2011 for exceptions). However, their bright, well-defined, contrasting colour patterns are ideal for digital image analyses, meaning that within a single genus we see striking variation in putatively

adaptive dorsal colour patterns that is easily quantifiable. When combined with NGS techniques (such as RAD-Seq), *Salamandra* present an ideal system in which to conduct detailed genotype-phenotype association studies. Using such methods, it will be possible to identify regions of the genome significantly correlated with specific phenotypes (e.g. spotted, stripped and melanistic) and identify loci involved with adaptive colouration in amphibians. This will be aided by the observed cases of parallel phenotype evolution, which provide 'natural replicates' in which to test evolutionary hypothesis. However, one obstacle remains to the use of *Salamandra* in such studies: a highly uncertain evolutionary history (Fig 1.8).

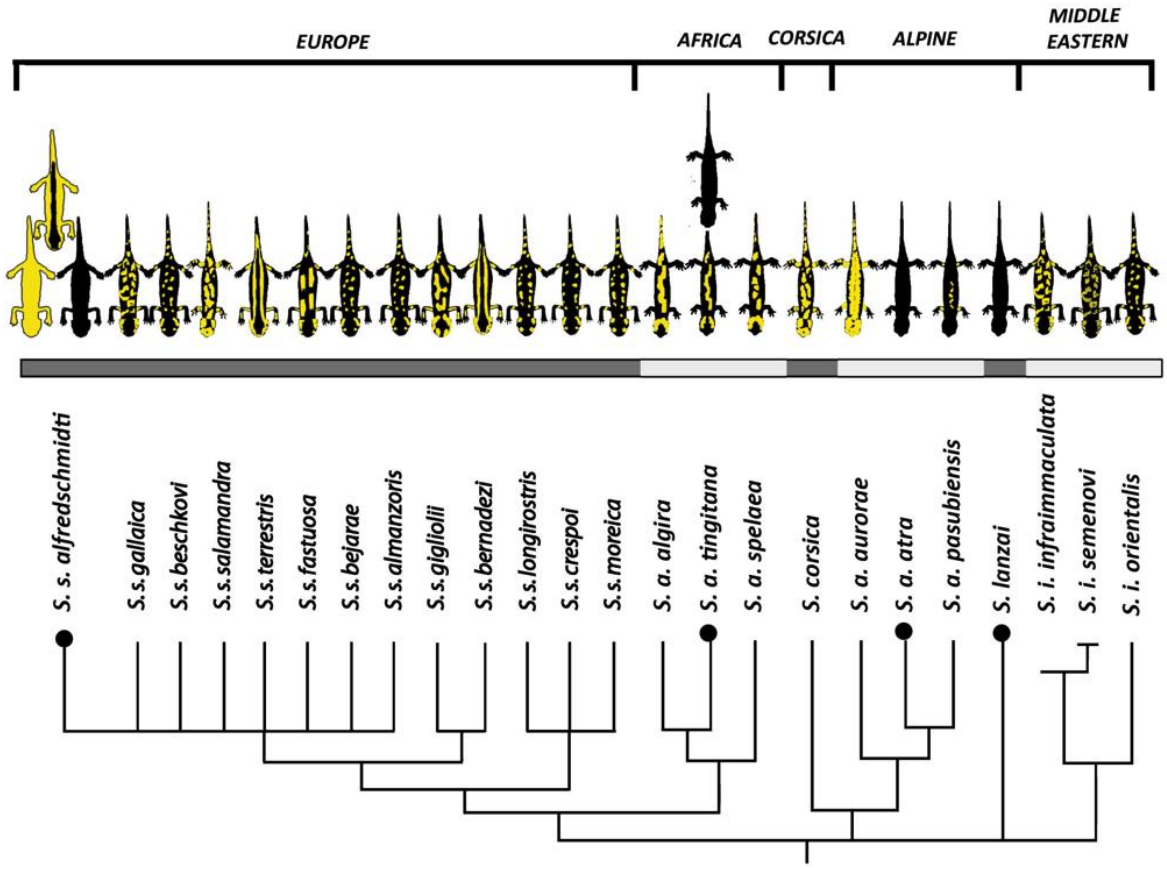


Figure 1.8: Phylogenetic relationships between *Salamandra* species (bars) and described subspecies, along with a simplified representation of the 'typical' colouration for each taxon (colour pattern data taken from Sparreboom and Arntzen 2014). The unresolved phylogenetic relationships show the lack of topological congruence seen across recent phylogenetic analyses (representative phylogeny constructed using data from: Steinfartz et al. 2000; Bonato and Steinfartz 2005; Köhler and Steinfartz 2006; Beukema et al. 2010; Vences et al. 2014).

Although *Salamandra* has an apparent wealth of taxonomic diversity, the evolutionary relationships between these species and subspecies have been challenging to elucidate. The phylogenetic analyses of Sanger sequence generated mitochondrial (mtDNA) and nuclear (nDNA) sequences has returned unresolved, weakly supported or topologically conflicting trees (Veith et al. 1998; Steinfartz et al. 2000; García-París et al. 2003; Vences et al. 2014). Even the most extensive analyses to date, which combined 10 mtDNA and 13 nDNA genes, was unable to adequately resolve inter-species relationships or even infer the correct position of the root (Vences et al. 2014). With sub-specific taxa largely delimited based on chromatic patterns (e.g. Veith 1992; Thorn and Raffaëlli 2001; Köhler and Steinfartz 2006; Beukema et al. 2010), and the known unreliability of this for taxonomic classifications in the group (Böhme et al. 2013; Beukema et al. 2016a), it is not currently possible to trace the evolutionary history of colour phenotypes, casting doubt on the parallel evolution of certain morphs. However, again, NGS methods may be of use, as they are offering unprecedented opportunities to study the phylogenomic relationships of systematically challenging taxa (Philippe et al. 2005; Leaché et al. 2015a; Prum et al. 2015; Wen et al. 2015; Massatti et al. 2016; Posada 2016; Irisarri et al. 2017). As Vences et al. (2014) speculate: "it is likely that only phylogenomic methods will be able to provide a robustly resolved phylogenetic tree for true salamanders. Such a phylogenomic tree will reveal the evolution of the many remarkable phenotypic characters".

1.7. Thesis overview

Amphibian colouration presents an ideal system in which to investigate genotype-phenotype relationships in the evolution of ecologically, physiologically and behaviourally adaptive traits. However, such research is hampered by an almost complete lack of information regarding its underlying genetics. In this thesis, I introduce the genus *Salamandra* as a candidate for elucidating the genetic basis of amphibian colouration. This genus displays striking inter- and intraspecific polymorphisms in dorsal colour patterns, which are putatively adaptive for both aposematism and thermoregulation. Its wide distribution also allows for the integration of a range of environmental data, including latitude, longitude, altitude and habitat. The suspected parallel evolution of colour phenotypes is particularly important as it will allow us to determine the importance of *de novo* mutation and standing genetic variation in the evolution and maintenance of adaptive colour phenotypes. However, first we must accurately reconstruct the evolutionary history of *Salamandra* and identify important colour genes/loci. In order to achieve this, I applied

new NGS technologies, bioinformatics methods, and quantitative colour analysis. The objectives of each data chapter are:

Chapter 2: We aimed to accurately reconstruct the controversial shallow phylogenetic relationships between *Salamandra* species through the analysis of three largely independent molecular data sets generated using multiple phylogenomic approaches (RNA-Seq, RAD-Seq and full mitochondrial genome sequencing). The data included samples from all six recognised species—*Salamandra atra, S. algira, S. corsica, S. infraimmaculata, S. lanzai and S. salamandra*—as well as outgroup samples from the genus *Lyciasalamandra*. This work has been published as a collaborative paper (https://doi.org/10.1016/j.ympev.2017.07.009), on which I am second author. My main contribution was the generation and phylogenetic analysis of the RAD-Seq data set. I also assessed the independence of the RNA-Seq and RAD-Seq datasets, and contributed to the conception, writing and editing of the manuscript. It is presented here as published with only minor corrections.

Chapter 3: Having validated the use of RAD-Seq in **Chapter 2**, I aimed to confirm species monophyly and inter-species relationships, and resolve subspecific relationships within the two most diverse taxa: *Salamandra algira* and *Salamandra salamandra*. For this, I genotyped 231 salamanders from across the taxonomic and geographic breadth of *Salamandra*; samples included representatives of all six species and almost all currently recognised subspecies. I also assessed the impact of a range of SNP filtering parameters on downstream phylogenetic reconstructions. Concatenated RAD-loci were then analysed using both maximum likelihood and Bayesian approaches. Following this, having constructed a robust phylogenetic hypothesis, I aimed to identify cases of parallel/convergent evolution in reproductive mode and two colour phenotypes (melanism and stripes) through ancestral character state estimation analyses.

Chapter 4: Using quantitative colour analysis and NGS approaches in a colour-variable population of *S. salamandra*, I aimed to: 1) identify genetic associations with colour and pattern; 2) test for selection on colouration; and 3) test the relationship between colour phenotype and toxicity (the functional basis of aposematism). For this, I utilised a contact zone within a rare, highly colour-

variable lineage in northern Spain (*S. s. bernardezi*) to investigate the genetic associations of colour and pattern. This was based on detailed phenotyping (using image analysis, spectrophotometry, and electron microscopy), differential gene expression analyses of skin tissue (using RNA-Seq data), and both machine learning and outlier based genotype-phenotype association studies (using RAD-Seq data). I also tested the supposed aposematic function of their colouration by investigating the metabolomic profiles of these salamanders' toxic secretions.

Finally, in **Chapter 5**, I discuss the broader implications of my work. Specifically, I outline my key findings, highlight study limitations, discuss potential future directions, and draw broad conclusions regarding the research presented in this thesis.

Chapter 2: Inferring the shallow phylogeny of true salamanders (*Salamandra*) by multiple phylogenomic approaches

2.1. Preface

A version of the work presented in this chapter has been previously published (open access) in the journal *Molecular Phylogenetics and Evolution* (Rodríguez et al. 2017); it is freely available at http://www.sciencedirect.com/science/article/pii/S1055790317301148. This was a collaborative project, comparing the phylogenomic analyses of three different high-throughput data sets: RAD-Seq, RNA-Seq and full mitochondrial genomes. I generated and conducted all analyses of the RAD-Seq dataset (except the 'gene' jackknifing, see section 2.4.2), compared the independence of the RNA-Seq and RAD-Seq data sets (see section 2.4.4), and contributed to the conception, writing and editing of the manuscript. It is presented here with only minor corrections.

2.2. Abstract

The rise of high-throughput sequencing techniques provides the unprecedented opportunity to analyse controversial phylogenetic relationships in great depth, but also introduces a risk of being misinterpreted by high node support values influenced by unevenly distributed missing data or unrealistic model assumptions. Here, we use three largely independent phylogenomic data sets to reconstruct the controversial phylogeny of true salamanders of the genus *Salamandra*, a group of amphibians providing an intriguing model to study the evolution of aposematism and viviparity. For all six species of the genus *Salamandra*, and two outgroup species from its sister genus *Lyciasalamandra*, we used RNA sequencing (RNA-Seq) and restriction site associated DNA sequencing (RAD-Seq) to obtain data for: (1) 3070 nuclear protein-coding genes from RNA-Seq; (2) 7440 loci obtained by RAD-Seq; and (3) full mitochondrial genomes. The RNA-Seq and RAD-Seq data sets retrieved fully congruent topologies when each of them was analysed in a concatenation approach, with high support for: (1) *S. infraimmaculata* being sister group to all other *Salamandra* species; (2) *S. algira* being sister to *S. salamandra*; (3) these two species being the sister

group to a clade containing *S. atra*, *S. corsica* and *S. lanzai*; and (4) the alpine species *S. atra* and *S. lanzai* being sister taxa. The phylogeny inferred from the mitochondrial genome sequences differed from these results, most notably by strongly supporting a clade containing *S. atra* and *S. corsica* as sister taxa. A different placement of *S. corsica* was also retrieved when analysing the RNA-Seq and RAD-Seq data under species tree approaches. Closer examination of gene trees derived from RNA-Seq revealed that only a low number of them supported each of the alternative placements of *S. atra*. Furthermore, gene jackknife support for the *S. atra* + *S. lanzai* node stabilized only with very large concatenated data sets. The phylogeny of true salamanders thus provides a compelling example of how classical node support metrics such as bootstrap and Bayesian posterior probability can provide high confidence values in a phylogenomic topology even if the phylogenetic signal for some nodes is spurious, highlighting the importance of complementary approaches such as gene jackknifing. Yet, the general congruence among the topologies recovered from the RNA-Seq and RAD-Seq data sets increases our confidence in the results.

2.3. Introduction

The rise of high-throughput sequencing techniques has provided molecular systematists with unprecedented opportunity to analyse controversial phylogenetic relationships in great depth (da Fonseca et al. 2016). In most organisms the sequencing of entire genomes is technologically within reach, but complexity reduction approaches, such as restriction site associated DNA sequencing (RAD-Seq), anchored hybrid enrichment, or sequencing of transcriptomes (i.e., the transcribed RNA; RNA-Seq), are more affordable and are increasingly being used to obtain markers that are representative of the nucleotide diversity across the genome (e.g., Emerson et al. 2010; Lemmon et al. 2012; Prum et al. 2015; Wen et al. 2015). Typically, phylogenomic approaches based on single nucleotide polymorphisms (SNPs) have been applied to inferences of population-level differentiation, phylogeography, and phylogenetic relationships among closely related species (Davey and Blaxter 2010; Peterson et al. 2012; Rubin et al. 2012; Darwell et al. 2016), whereas those based on sequences of protein-coding genes derived from RNA-Seq or full genomes have been used for inferring deep nodes in the tree of life, often analysed at the amino acid level (Bapteste et al. 2002; Chiari et al. 2012; Jarvis et al. 2014; Wickett et al. 2014; Chen et al. 2015; Irisarri and Meyer 2016). More recently, phylotranscriptomic analysis of RNA-Seq

derived markers has proven to also provide valuable insights into shallow relationships between species (Wen et al. 2015; Wang et al. 2017)

The ever-increasing amount of data available for phylogenomic analyses provides unprecedented opportunities to resolve the tree of life (Philippe et al. 2005), but also introduces novel risks of drawing misleading conclusions. In particular, current approaches of assessing node stability, such as non-parametric bootstrapping and Bayesian posterior probabilities, tend to provide very high node support with large amounts of data, but these might reflect artefacts such as unevenly distributed missing data rather than real phylogenetic signal (e.g., Dell'Ampio *et al.*, 2014). Also, if model assumptions are not realistic, incorrect phylogenetic relationships can be supported by high values thus representing systematic error. Accordingly, just adding more sequence information is not necessarily a route to increasing the quality of phylogenomic inference (Philippe et al. 2011). The scale of the problem might also differ by the type of data used. For example, SNP-based analyses cannot incorporate codon-based nucleotide substitution models into the phylogenetic inference process. Similarly, concatenating across genes in an RNA-Seq analysis precludes gene-specific estimation of transition matrices or accurate estimation of rate heterogeneity, because current software to estimate partition schemes (e.g., Lanfear *et al.*, 2012) is still in its infancy when it comes to efficiently handling thousands of genes. It is also poorly understood how sensitive phylogenomic resolution is to combining different types of data in the same analyses. While these issues are not new (e.g. ‘‘total evidence" debate from the 1990s; e.g. Bull *et al.*, 1993), the ‘‘big data" typical of modern approaches could require new strategies for assessing confidence. For example, resampling genes or loci rather than individual nucleotide sites (e.g. gene jackknifing) in a concatenated analysis could be more informative than traditional bootstrap or posterior probability analyses (Irisarri et al. 2017). However, direct empirical comparisons between gene-based and SNP-based data obtained from the same set of samples, necessary to determine their relative sensitivity to phylogenetic error and to assess their performance in resolving shallow phylogenetic relationships, are scarce.

Here we empirically compare the use of RNA-Seq, RAD-Seq, and whole mitochondrial genomes to test evolutionary hypotheses about relationships in a prominent group of amphibians, the 'true salamanders' of the genus *Salamandra* (family Salamandridae). This genus includes six recognised species (Speybroeck et al. 2010; Sillero et al. 2014) that vary in both colour patterns and reproductive modes. Some species have a conspicuous yellow-black coloration, thought to be of aposematic function (*S. algira* from North Africa; *S.*

corsica from Corsica; *S. infraimmaculata* from the Near East; and *S. salamandra* from Europe), whereas others have a uniformly black coloration (*S. atra* and *S. lanzai*; both distributed in the European Alps). They are also one of the few groups of vertebrates to vary in viviparity across their range, including instances of deposition of aquatic larvae or terrestrial juveniles reared on yolk nutrition, vs. release of fully metamorphosed young reared on maternal nutrition (Wake 1993; Greven and Guex 1994; Greven 2003a; Buckley et al. 2007; Caspers et al. 2014). Clarifying the phylogenetic relationships among these species is thus of interest for studies of biogeography, colouration and toxicity (function of aposematism), and the evolution of different reproductive modes.

Previous molecular phylogenetic studies based on DNA sequences of mitochondrial and nuclear genes, and complete mitochondrial genomes, placed the Asian *Lyciasalamandra*, another clade of viviparous salamanders, as sister taxon to *Salamandra*, and a clade comprising *Chioglossa* and *Mertensiella* sister to the *Salamandra*/*Lyciasalamandra* clade (Titus and Larson 1995; Veith et al. 1998; Weisrocka et al. 2001; Veith and Steinfartz 2004; Frost et al. 2006; Weisrock et al. 2006; Steinfartz et al. 2007a; Zhang et al. 2008; Pyron 2014). However, despite combining DNA sequences of 10 mitochondrial and 13 nuclear genes (Vences et al. 2014), the relationships among species of *Salamandra* have remained poorly resolved. Several relationships were supported by both types of markers, such as a clade containing the black-coloured alpine species (*S. atra* and *S. lanzai*) plus *S. corsica*, but most other relationships did not receive strong support.

In the present study, we newly sequenced and assembled three phylogenomic data sets to resolve the phylogenetic relationships among species of *Salamandra*: (1) 3070 protein-coding nuclear genes obtained from transcriptomes (RNA-Seq); (2) 7440 anonymous nuclear markers obtained *via* double-digest Restriction Site Associated DNA sequencing (ddRAD-Seq); and (3) full mitochondrial genomes. We analyse these data using concatenation and 'species tree' approaches to assess the phylogeny of true salamanders. Further, we scrutinize the congruence of the different molecular data sets and analytical approaches for phylogenetic resolution.

2.4. Materials and methods

2.4.1. RNA-Seq analyses

Transcriptomic data from one individual each of *Salamandra salamandra* from Germany (Kottenforst near Bonn) and of *S. infraimmaculata* from Israel were available from the study of Czypionka et al. (2015). A further transcriptomic data set was available for *S. salamandra* from France (Banyuls; geographical coordinates 42.479183, 3.101555) from the study of Figuet et al. (2014). New transcriptomic data were generated for single individuals of the other four species of *Salamandra* (*S. algira*, *S. atra*, *S. corsica* and *S. lanzai*), along with two species of *Lyciasalamandra* as an outgroup. We used pooled samples of different organs (skin, muscle and liver) preserved in RNAlater and frozen at -80°C. RNA extraction from 100mg of tissue of each salamander was carried out using a trizol protocol (see Appendix 3). RNA was prepared for sequencing following the Illumina TruSeq mRNA protocol. Sequencing was carried out on the Illumina MiSeq (2 x 250 bp paired-end) platform. Illumina reads were quality trimmed and filtered using Trimmomatic v.0.32 (Bolger et al. 2014) with default settings. Filtered reads, paired and unpaired, were used for *de novo* transcriptome assembly using Trinity v.2.1.0 (Grabherr et al. 2011) following published protocols (Haas et al. 2013). Candidate coding regions within transcript sequences from the final assembly were identified and translated using Transdecoder 2.1.0 (Haas et al. 2013). Raw reads were submitted to the NCBI Short Read Archive database (Bioproject PRJNA385088).

As a basis for selecting nuclear protein-coding genes for analysis, we used a previously compiled alignment from Irisarri et al. (2017), in the following called reference alignment. For detailed methods of obtaining this reference alignment, see Irisarri et al. (2017). In brief, the reference alignment was assembled by first grouping 20 vertebrate proteomes into putative orthology using USEARCH (Edgar 2010) and OrthoMCL (Li et al. 2003) and discarding orthogroups with missing data for major clades of jawed vertebrates. After aligning and custom paralog-splitting, the resulting protein clusters were complemented with 80 additional published genomic and transcriptomic data sets of vertebrates using the software Forty-Two (or '42'; D. Baurain; https://bitbucket.org/dbaurain/42/) that controls for orthology using several proteomes in strict three-way reciprocal best BLAST hit tests. Subsequently, following the methods outlined in Irisarri et al. (2017), the reference alignment went through several decontamination steps to remove: (1) all human and non-vertebrate sequences; (2) crosscontaminations; (3) highly incomplete genes; (4) genes with

poor alignment or frame shifts; (5) genes resulting in extremely long branches in some taxa suggesting the possibility of contamination or undetected paralogy. The final vertebrate reference data set as used for the analysis of Irisarri et al. (2017) contained 4593 genes.

Sequences of *Salamandra* transcriptomes were aligned to this reference data set using the software Forty-Two. The resulting alignment was then submitted to a pipeline of thorough filtering and decontamination, composed of the following eight steps: (1) Sequences from non-vertebrate sources (e.g. Bacteria or Platyhelminthes) were detected by BLAST searches and 21,265 sequences were removed (almost exclusively from the previously published *S. infraimmaculata* and *S. salamandra* sequences). (2) To remove redundant and/or divergent sequences, for each sample represented by at least two sequences for a given gene (e.g., multiple short transcripts that could not be assembled together), every sequence was compared against all the other sequences in the alignment by BLAST. A total of 16,295 sequences were eliminated if their average bit score was at least 10% lower than the best average bit score of the redundant set and if there was a length overlap of ≥95% between the two sequences. (3) We then excluded genes providing unrealistic phylogenetic resolution in our target group (salamanders) by excluding such genes for which the genus *Salamandra* or the family Salamandridae were not recovered as monophyletic; for this analysis, phylogenetic trees were inferred using RAxML v8 (Stamatakis 2014) and an LG + C model (Le and Gascuel 2008) using only amphibian sequences. (4) To remove from the remaining genes those that might be affected by undetected ancient paralogy, we split the respective alignments by looking for the branch that maximizes taxonomic diversity (see Amemiya et al. 2013). Phylogenetic trees were inferred again on these split alignments and we again retained only genes for which the genus *Salamandra* or the family Salamandridae were monophyletic. In total, 3105 genes recovered a monophyletic genus *Salamandra* and among the remaining genes retained (i.e. with nonmonophyletic *Salamandra*), 508 recovered Salamandridae as monophyletic. (5) We then retrieved nucleotide sequences for these genes from the original transcriptomic data; genes for which sequences were available for fewer than five *Salamandra* species or without *Lycisalamandra* data were discarded. The corresponding retained nucleotide sequences were both recovered and aligned according to the amino acid alignments using the software Leel (or '1331'; D. Baurain; https://bitbucket.org/dbaurain/42/). All subsequent analyses were based on these nucleotide alignments. (6) We checked for remaining contaminating or paralogous sequences by comparing the branch lengths in gene trees and in the concatenation tree as in Irisarri et al. (2017), and removed 445 sequences having a branch length ratio (gene tree vs. concatenation tree) >7. (7) To reduce stochastic

error in gene trees, we removed all the codons that were present in less than 50% of the species and any sequence having less than 30 nucleotides. (8) The resulting nucleotide data set (available as Mendeley Research Data) was then concatenated using SCaFoS (Roure et al. 2007) and the 3070 genes with fewer than three species missing were retained and used for phylogenetic analysis.

Maximum likelihood (ML) phylogenetic inference was conducted on the concatenated nucleotide matrix using RAxML, partitioning the data by genes. For each gene, we estimated a separate general time reversible (GTR) model of nucleotide evolution (i.e. six substitution rates), with rate heterogeneity modelled according to a gamma distribution (shape parameter alpha) with four rate categories. We assessed node support with 1000 non-parametric bootstrap replicates. As conflicting genealogical histories often exist in different genes throughout the genome, concatenation methods can result in incorrect trees with high support (Kubatko et al. 2007; Degnan and Rosenberg 2009). We thus took several strategies for assessing potential conflicts.

First, node support was assessed using a gene jackknife approach (Delsuc et al. 2008) to determine what proportion of the data would need to be sampled to resolve the maximal number of nodes: one hundred alignment replicates were generated by randomly sampling genes up to ca. 10,000, 50,000, 100,000, 500,000, 1,000,000 and 3,000,000 nucleotide positions, respectively. For each replicate, unpartitioned ML trees were estimated using RAxML with a GTR + C model defined for the whole data set, and gene jackknife proportions estimated for each node.

Second, we also used ASTRAL II (Mirarab and Warnow 2015), a statistically consistent algorithm to estimate the species tree topology under the multi-species coalescent model (Mirarab and Warnow 2015). Clade support was evaluated by computing the local posterior probability, a feature of ASTRAL II that has shown high precision compared with multi-locus bootstrapping on a wide set of simulated and biological datasets (Sayyari and Mirarab 2016). As species tree analyses do not require outgroups (Heled and Drummond 2010) the ASTRAL II analyses was carried out with ingroup sequences (genus *Salamandra*) only, but an additional exploratory analysis including the outgroup was also performed.

2.4.2. RAD-Seq analyses

Tissue samples were collected from two individuals of *Salamandra algira, S. atra, S. corsica, S. infraimmaculata* and *S. salamandra*, and one individual of *S. lanzai*. Tissue was also collected from one individual each of *Lyciasalamandra billae* and *L. flavimembris* to provide an outgroup. Genomic DNA was extracted using the Macherey-Nagel NucleoSpin® Tissue kit following the manufacturer's instructions. We applied double-digest Restriction Site Associated DNA sequencing (ddRAD-Seq; Peterson et al. 2012) and for simplification hereafter refer to the resulting sequences as RAD-Seq data set. The library was prepared as follows (per Recknagel et al. (2015) with modification of Illumina adapters): 1mg of DNA from each individual was double-digested using the PstI-HF® and AclI restriction enzymes (New England Biolabs); modified Illumina adaptors with unique barcodes for each individual were ligated onto this fragmented DNA; samples were multiplexed (pooled); and a PippinPrep used to size select fragments around a tight range of 383 bp, based on the fragment length distribution identified using a 2200 TapeStation instrument (Agilent Technologies). Finally, enrichment PCR was performed to amplify the library using forward and reverse RAD primers. Sequencing was conducted on an Illumina Next-Seq® 550 machine at Glasgow Polyomics to generate paired-end reads 75 bp in length. Raw reads were submitted to the NCBI Short Read Archive database (Bioproject PRJNA386146).

Sequence reads were de-multiplexed, Illumina adaptors and barcodes removed, and reads truncated to 60 nucleotides using Stacks v.1.35 (Catchen et al. 2013). Processed reads per sample ranged from 4.9 to 16.8 million, compared to 5–17 million raw reads per sample. Reads were assembled *de novo* into loci using pyRAD v.3.0.6 (Eaton 2014). Reads were first clustered within an individual at a minimum depth of 10 with a clustering threshold of 85%. The same clustering threshold was then used to assemble *de novo* loci across samples; final RAD-loci ranged in length from 109 to 144 nucleotides, with an average of 111 nucleotides. As the performance of *de novo* assembled RAD-Seq data matrices in phylogenetic reconstructions depends on the sample coverage and potential intra-locus paralogy (Huang and Knowles 2014; Takahashi et al. 2014), we explored a range of thresholds for loci coverage between samples (4, 6, 8 10 and 11 individuals; equivalent to 31–100% of the in-group) and maximum number of SNPs per RAD locus (2, 4, 6, 8 and 10). RAxML, with a GTRGAMMA model and 100 rapid bootstraps, was used to explore the resulting concatenated sequence matrices in order to choose the filters that produced

the most reliable trees (based on node resolution and support). All trees agreed in almost all aspects (except the relationships among *S. atra*, *S. lanzai*, and *S. corsica* that were left unresolved in some analyses), and we eventually chose a between sample coverage of 6 with a maximum number of 2 SNPs per locus.

Phylogenetic analyses of the RAD-Seq data set were conducted in BEAST 2.4.2 (Bouckaert et al. 2014). For the concatenated analysis of loci, a BEAST xml file was generated using BEAUTi 2.4.2. The best fitting evolutionary model inferred by jModeltest 2.1.10 was the transversion model (TVM; based on the Bayesian information criterion). As BEAST2 only has four base substitution models, which do not include TVM, a GTR substitution model with the alpha gamma rate parameter fixed at one was selected to simulate it. A relaxed clock (log normal) was used, with all other parameters left on default settings. A MCMC chain of 10 million generations was run (10% burn-in) with tree and parameter estimates sampled every 1000 MCMC generations. Tracer 1.6.0 (Drummond and Rambaut 2007) was used to assess chain convergence. Eight prior operators (tree Scaler; Subtree Slide; Rate AG Scaler; Rate AT Scaler; Yule Model Tree Scaler; Yule Model Subtree Slide; Fix Mean Mutation Rates Operator; and ucldStdevScaler) were optimised based on the output of this trial, and the analysis re-run. A maximum clade credibility tree was then generated from the output of the optimised analysis using TreeAnnotator 2.4.2. 'Gene' jackknifing (i.e., jackknifing of RAD-loci) was carried out as described for the RNA-Seq data, with replicates of ca. 10,000, 50,000, 100,000, 500,000 and 800,000 nucleotide positions.

In addition to the concatenation approach, we also used the coalescent-based program SNAPP (Bryant et al. 2012) to infer the species tree under a finite-sites model of mutation from unlinked biallelic SNPs extracted from the RAD-Seq data set. The *Lyciasalamandra* outgroup was removed as no outgroup is required in species tree analyses (Heled and Drummond 2010) and loci were refiltered in pyRAD to extract a single SNP per locus, giving a final data set consisting of 3586 loci from across the 11 *Salamandra* samples. Using the SNAPP template in BEAUTi, 2.4.2 a BEAST xml file was generated. Given a lack of reliable prior information, mutation rates were sampled and a uniform distribution was used for the lambda parameter of the Yule prior; all other priors were left at default. BEAST2 was run with 10 million generations, 10% burn-in, and tree and parameter estimates sampled every 1000 MCMC generations. Convergence was assessed with Tracer 1.6.0 and the maximum clade credibility tree generated using TreeAnnotator 2.4.2.

2.4.3. Mitogenome analyses

We assembled mitogenomes of all *Salamandra* species from the quality-trimmed RNA-Seq data. We first randomly sampled 20% of the raw data and subsequently retrieved and assembled mitochondrial sequences with MIRA v4.0 (Chevreux et al. 1999) and MITObim v1.8 (Hahn et al. 2013) following Machado, Lyra and Grant (2016) and using default parameters. We used the complete mitochondrial sequence of *Salamandra infraimmaculata* (EU880331) as reference genome in the first MIRA step. Assemblies in CAF format were manually verified in Geneious software v.6 (Biomatters) to evaluate the coverage and quality of each mtDNA element. All positions with coverage lower than 4 were coded as ambiguous ('N'). Preliminary annotation of each sequence was done using the mitochondrial genome annotation server MITOS (Bernt et al. 2013) with default parameters. Validation of tRNA sequences were performed using tRNAscan-SE (Lowe and Chan 2016). The resulting automatic annotation was confirmed and edited manually by comparison to *Salamandra infraimmaculata* EU880331. All newly determined sequences were submitted to Genbank (accession numbers MF043386–MF043393). We also included complete or almost complete mitochondrial genome sequences of *Chioglossa lusitanica* (EU880308) and *Mertensiella caucasica* (EU880319) as outgroups, added species of *Lyciasalamandra* (EU880318, AF154053) as hierarchical outgroups, and furthermore added one species of *Salamandra* for which a full mitogenome sequence was available from Genbank (EU880331). The latter sample was originally analyzed as *S. salamandra* (Zhang et al. 2008), but corresponds to a sample of *S. infraimmaculata* from Turkey.

We aligned mitochondrial sequences using MAFFT v.7 (Katoh and Standley 2013) and determined the optimal among-gene partitioning scheme and model choice for dataset in PartitionFinder (Lanfear et al. 2012) under the Bayesian Information Criterion (BIC). Bayesian phylogenetic inference was performed with MrBayes v.3.2.6 (Ronquist et al. 2012) using two independent runs of eight chains. Chains were started from random trees and run for 10 million generations each, being sampled every 1000 generations. Twenty-five percent of the trees were discarded as 'burn-in' before generating a consensus tree. The full mitogenomic data set was also analysed under the ML optimality criterion in RAxML v. 8. (Stamatakis 2014), using the GTRGAMMA model of nucleotide substitution and a partitioned approach, with partitions and substitution models as defined by Partitionfinder (Lanfear et al. 2012). Node support was assessed using 1000 bootstrap replicates.

2.4.4. Independence of data sets

To understand whether the RNA-Seq and RAD-Seq datasets were independent, we calculated sequence overlap between them. For each individual, sequences representing the RAD-loci were aligned to the RNA-Seq data set using Bowtie2 v.2.2.9. An overall alignment rate of 1.23% was found: of 56,987 paired reads that mapped, none aligned concordantly, 46 (0.08%) aligned discordantly once, and when single-end reads were aligned independently (113,882 in total), 817 (0.72%) aligned one time and 497 (0.44%) aligned >1 time. This confirms that the loci used for the RNA-Seq and RAD-Seq analyses were almost completely non-overlapping and that the two analyses can be considered independent subsamples of the same underlying genomes. We also confirmed that no genes encoding mitochondrial proteins were present in the final RNA-Seq alignment used for analysis.

2.4.5. Gene Ontology analyses

Given that mitochondrial and nuclear DNA sequences gave a contradicting signal regarding the monophyly of black or alpine salamanders (i.e. a grouping of *S. atra* + *S. corsica* vs. *S. atra* + *S. lanzai*; see Results) we tested whether this discordance could be explained by differences in the functional categories of genes. We first counted the total number of genes supporting a certain topology using Phylosort, v.1.3 (Moustafa and Bhattacharya 2008). We then tested whether incongruencies among analyses were due to particular nuclear genes coding for proteins whose functions interact with mitochondria and might have thus coevolved with mitochondrial genes (Hill 2016). We first created a consensus protein sequence for each of the 3070 RNA-Seq genes and used BLAST to compare them to the UniProtKB/Swiss-Prot database (http:// www.uniprot.org/). We selected the most similar sequence to represent the gene ontology term for the given protein. From the total list of genes, we selected those for which the phylogenetic tree supported either the *S. atra* + *S. corsica* or *S. atra* + *S. lanzai* sister group relationship. We then used the UniProt Retrieve/ID mapping web server (http://www.uniprot.org/uploadlists/) to classify genes into Gene Ontology domains (Supp. Table A3.1).

2.5. Results

2.5.1. RNA-Seq analyses

The final concatenated alignment derived from RNA-Seq contained 3,255,534 bp including 74,801 variable (2.30%) and 28,125 parsimony-informative (0.87%) positions. This corresponded to the nucleotide sequences of 3070 genes with mean alignment lengths of 1060 ± 566 bp SD (min–max = 228–7068 bp). Taxonomic coverage varied from 6 to 9 per gene, and the percentage of missing data per taxon ranged from 0 to 35.8%. The ML tree calculated from this concatenated matrix (Fig. 2.1A) provided a fully resolved tree of *Salamandra* species, with *S. infraimmaculata* sister to all remaining species, and a clade of *S. algira* and *S. salamandra* sister to a clade of *S. atra*, *S. corsica* and *S. lanzai*, with the two alpine species *S. atra* and *S. lanzai* forming a monophyletic group sister to *S. corsica*.

Gene jackknife proportions of the RNA-Seq data set revealed that up to 3 million nucleotide positions are necessary to recover all final-tree bipartitions with high support (>75%; Fig. 2.2A). With 10,000 nucleotide positions, only one out of five nodes of interspecific relationships were recovered with high support (corresponding to the placement of the two *Lyciasalamandra* species as sister group); 50,000 nucleotide positions were sufficient to recover the close affinity of *S. algira* and S. *Salamandra*; whereas the remaining three nodes required replicates of more than one million nucleotide positions to be resolved with high support.

The tree obtained from the ASTRAL II species tree analysis (Fig. 2.3A) partly agreed with the trees obtained by the analyses of the concatenated data set (Fig. 2.1) but placed *S. corsica* apart from the *atra-lanzai* clade. Repeating the same analysis with outgroup sequences resulted in yet another topology, where *S. corsica* was placed sister to *S. atra* (Supp. Fig. A3.2).

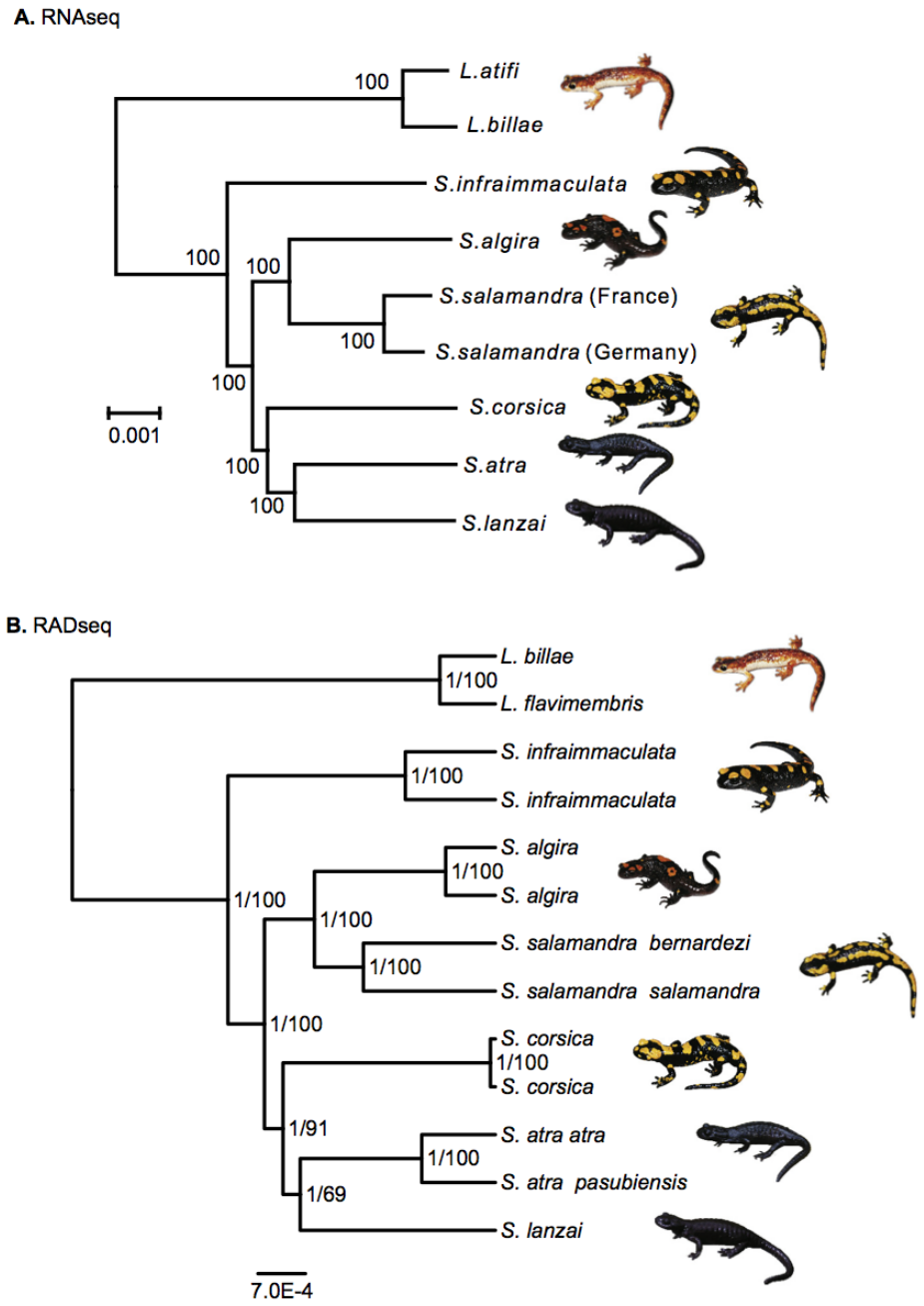


Figure 2.1: (A) Phylogenetic tree resulting from the analysis of 3070 orthologous loci (3,256,500 bp) obtained from transcriptomes (RNA-Seq) of *Salamandra* species partitioned by genes and analyzed under a GTR + C model in RAxML; branch support was estimated with 1000 rapid bootstraps. (B) Phylogenetic tree based on a BEAST2 analysis of 7440 concatenated RAD-Seq loci (17,985 SNPs) with a minimum number of 6 samples per locus and a maximum number of 2 SNPs per locus; branch support is based on Bayesian posterior probabilities (first number at nodes) and ML bootstrap analyses (second number at nodes; RaxML rapid bootstrapping, 100 replicates).

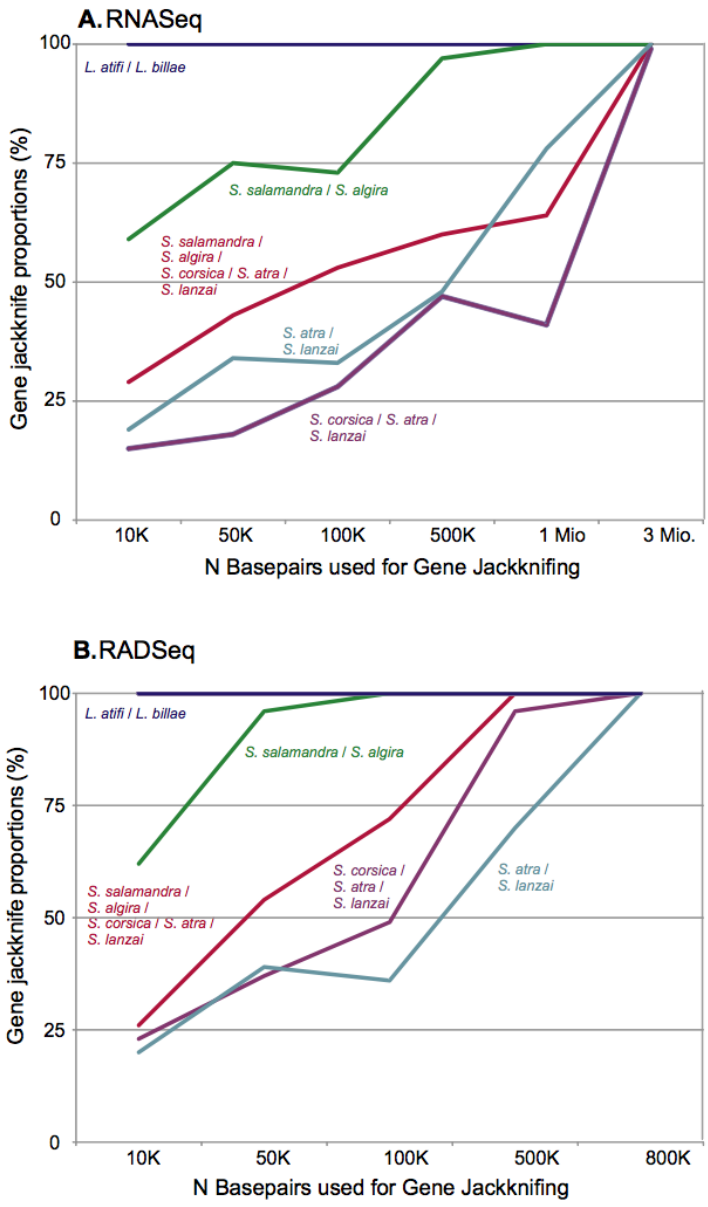


Figure 2.2: Gene jackknifing results for (A) the RNA-Seq and (B) the RAD-Seq data. All results including those of the largest data sets are based on gene resampling without replacement. For RAD-Seq, the procedure treated each locus as ‘‘gene” for jackknifing purposes.

2.5.2. RAD-Seq analyses

From the RAD-Seq data, we assembled an alignment of 7440 loci present in at least 6 samples (33.4% missing data), with a maximum number of 2 SNPs per locus (single end), and containing 822,917 nucleotide positions, 17,985 SNPs, and 7189 parsimony-informative sites (0.87%). Bayesian inference analysis of this concatenated RAD-Seq data set yielded a tree identical to that obtained with the concatenated RNA-Seq data (Fig.

2.1B), with all nodes showing posterior probabilities of 1. Exploratory analysis of a more stringently selected RAD-Seq data set, with 1541 SNPs from 586 loci present in at least 11 samples and with at most 2 SNPs per locus (missing data 8.0%), recovered an identical topology, but with lower support values for several nodes. A very similar jackknife pattern was observed for the RNA-Seq data (Fig. 2.2A) as compared to the RAD-Seq data (Fig. 2.2B). Up to 500,000 nucleotide positions were needed to recover all nodes representing interspecific relationships with support values of 75% or higher. The SNAPP analysis of the RAD-Seq data led to a tree in agreement with the ASTRAL II tree of the RNA-Seq data, failing to group *S. corsica* with the clade containing *S. atra* and *S. lanzai* (Fig. 2.3B).

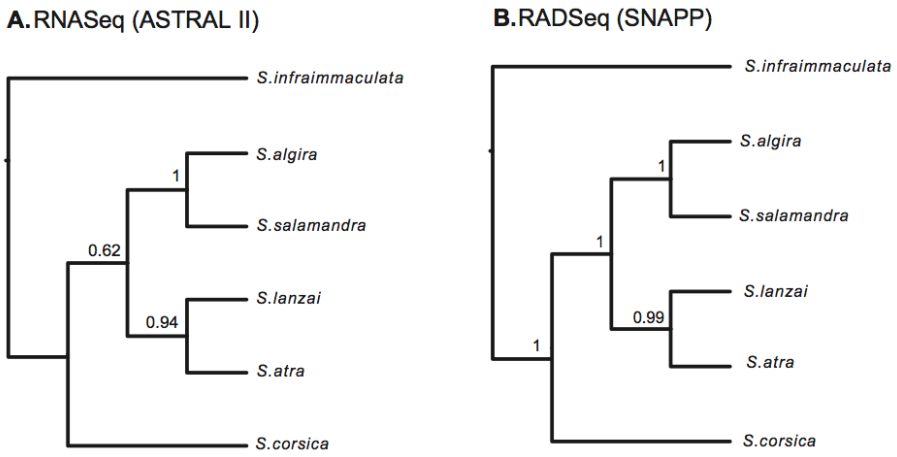


Figure 2.3: Results of species tree analyses of *Salamandra*. (A) Species tree obtained from ML gene trees of each of the 3070 orthologous loci from the RNA-Seq analysis, summarized with ASTRAL II. Branch support was estimated by computing the local posterior probability (not calculated by ASTRAL II for the basal most node, which is strongly supported in an analysis including *Lyciasalamandra* as outgroup; see Supp. Fig. A3.2). (B) Maximum clade credibility tree (cladogram representation) obtained from a SNAPP analysis of 3586 unlinked SNP loci identified from RAD-Seq data.

2.5.3. Mitogenome analyses

We recovered almost complete mitochondrial genomes for all samples analyzed (coding genes, rRNAs and tRNAs) from the RNA-Seq raw sequence reads. The phylogenetic inference of the mitogenomic alignments (Fig. 2.4) was largely congruent with the nuclear data, with two exceptions. The mitogenome tree showed strong bootstrap support for

different relationships among the three species in the clade including the two alpine salamanders (*S. atra* and *S. lanzai*) plus *S. corsica*; *S. atra* and *S. corsica* were identified as being sister groups, as compared to *S. atra* and *S. lanzai* in the RNA-Seq and RAD-Seq analyses. The placement of *S. infraimmaculata* also differed: it was inferred to be sister to *S. algira* and *S. salamandra* in the mitogenome tree (although this node was poorly resolved) but ancestral to all of the other species in the nuclear data analyses. The relationships among these three species varied according to partitioning scheme used for the mitogenomic sequences (Supp. Fig. A3.1).

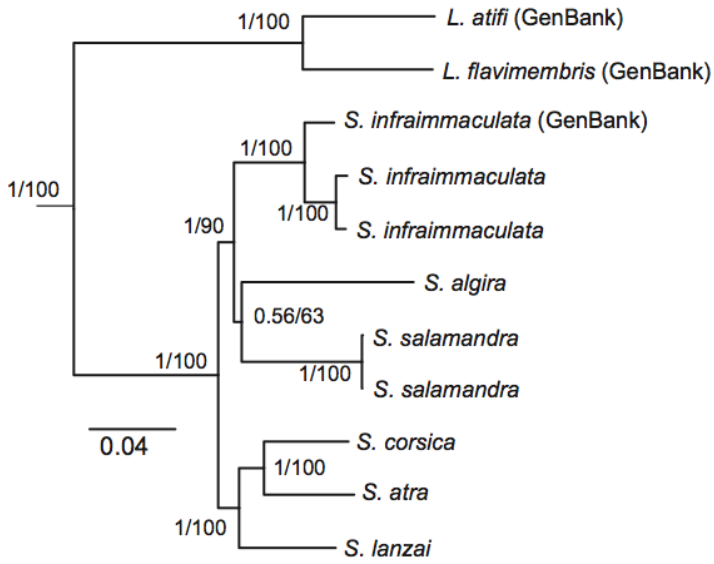


Figure 2.4: Majority-Rule consensus tree obtained by partitioned Bayesian Inference from complete or almost complete mitochondrial genomes (sequences of proteincoding genes only; see Supp. Fig. A3.1 for trees based on data sets including also non-coding rRNA and tRNA genes). Numbers at nodes are Bayesian posterior probabilities, followed by bootstrap proportions in percent from a ML (1000 replicates).

2.5.4. Gene Ontology analyses

To investigate the origin of the discordance observed between the concatenated nuclear gene analyses compared to the mitogenome analyses, we specifically analyzed which sets of genes supported the two alternative topologies within the clade containing *S. atra*, *S. corsica* and *S. lanzai*. Altogether, out of a total of 3070 genes, only 680 genes supported the *atra-lanzai* clade and a similar number, 665 genes, supported the alternative *atra-*

corsica clade. The clade of *algira-Salamandra-atra-lanzai* to the exclusion of *corsica* (as recovered by the species tree analyses; Fig. 2.3) was supported by 279 genes.

A comparison of the ontology of genes supporting either the *atra-corsica* or the *atra-lanzai* clade revealed no clear pattern; for different functional properties, a similar proportion of gene trees supporting either topology were found (Supp. Table A3.1). Given the conflicting phylogenetic resolution among concatenation and species tree approaches of the nuclear data, we tested whether the nuclear phylogenetic signal might have been influenced by genes functionally coupled to mitochondrial genes, e.g. in the respiratory chain. However, these genes again supported the *atra-corsica* vs. the *atra-lanzai* clade in similar proportions (4 vs. 8 genes tightly connected to mitochondrial functions, and 33 vs. 39 genes weakly connected to mitochondrial functions; Supp. Table A3.2).

2.6. Discussion

2.6.1. Phylotranscriptomic analysis of shallow phylogenetic relationships

In this study, we used nucleotide sequences of nuclear protein-coding genes derived from various phylogenomic data sets to reconstruct shallow phylogenetic relationships among closely related species of amphibians. Among these were sequences of nuclear protein-coding genes obtained by RNA-Seq, a kind of data set typically used to resolve deep phylogenies, with amino acid sequences as phylogenetic characters and taxa often separated for hundreds of millions of years (e.g., Jarvis et al. 2014; Misof et al. 2014; Chen et al. 2015). Here we followed an approach that recovered the nucleotide sequences of those transcripts that passed a stringent decontamination pipeline at the amino acid level (to remove sequences and genes potentially affected by sample contamination, paralogy, sequencing errors, or other artefacts; see Methods), and used the resulting concatenated nucleotide alignment of expressed genes for phylogenetic reconstruction. The results fully agreed with those derived from a concatenated alignment of RAD-Seq-derived SNPs (Fig. 2.1), which represented a largely independent subsample of the salamander's genomes and is a kind of data set typically used for shallow phylogenetic inferences. This validates our usage of a phylotranscriptomic approach, and confirms that such RNA-Seq-derived data hold promise for reconstructing not only deep nodes of the tree of life but also shallow phylogenetic relationships among taxa probably characterized by recent gene flow.

It is remarkable that with double-digest RAD sequencing it was possible to obtain a substantial number of homologous RAD-Seq loci, despite the enormous genome size of the target species, with haploid C-values estimated between 27 and 41 pg (Gregory 2016). The overall number of nucleotide positions and phylogenetically informative sites in the RNA-Seq data set was almost fourfold that of the RAD-Seq data set, but the RAD-Seq data sets still led to a highly resolved tree and most nodes stabilized at shorter gene jackknife replicates as compared to the RNA-Seq dataset (Fig. 2.2). RAD-Seq loci represent a relatively random subsample of the entire genome, potentially capturing a wider range of evolutionary signals than protein coding genes; i.e. it should contain fast as well as slowly evolving loci that might broaden the phylogenetic spectrum covered. By contrast, the RNA-Seq transcript sequences used in our analysis were restricted to those loci that are consistently expressed and conserved across vertebrates, and hence potentially more limited in phylogenetic resolution. Another potential shortcoming is that alleles at heterozygous positions are not called in the RNA-Seq analysis pipeline and a considerable amount of nucleotide variation is therefore neglected, with possible influences on the RNA-Seq-derived species tree. Interestingly, despite these very different characteristics of the two data sets, the proportion of phylogenetically informative sites was identical in both of them (0.87%) and they resolved the same relationships among the ingroup taxa. Together these results emphasise that both types of data might be useful for phylogenomics of closely related species and congruence analyses comparing the two can increase confidence in relationships resolved.

2.6.2. Conflict between concatenation and species tree approaches

Massive phylogenomic data sets, such as those obtained from RNA-Seq, certainly have the potential to lead to improved phylogenetic inference. However, simply adding more sequences to the data set is not enough (Philippe et al. 2011). Our analysis clearly exemplifies the limitations of large amounts of sequence data, as different analysis methods can result in opposing phylogenetic hypotheses, each with strong support using classical statistical metrics such as non-parametric bootstrap or Bayesian posterior probabilities. For the RAD-Seq data, the species tree analysis with SNAPP (Fig. 2.3) placed *S. corsica* away from the *atra-lanzai* clade, with maximum posterior probability, conflicting with the analysis of the concatenated alignment that placed these three taxa in one clade (Fig. 2.1). For the ASTRAL II analysis of RNA-Seq data, the same species tree topology was found, albeit with partly weaker support.

These differences between the species tree analyses vs. concatenated analyses are surprising, given that in studies on other organisms, congruent results were obtained from the two approaches (e.g., Herrera and Shank, 2016; Tucker et al., 2016). The incongruence in our study could be caused by shortcomings of one of the approaches in (1) dealing with a clade of closely related species, probably affected by incomplete lineage sorting (ILS) and introgression, or (2) dealing with the kind of data, i.e. long protein-coding sequences derived from RNA-Seq vs. SNPs derived from RAD-Seq.

It has been hypothesized that in the presence of introgression or incomplete lineage sorting, as it can be expected in the case of the closely related *Salamandra* species, multispecies coalescent species tree analyses should provide a more realistic phylogenetic resolution than concatenation (Liu et al. 2009; Leaché and Rannala 2011; Mirarab et al. 2016). On the other hand, strong advocates for concatenation approaches in phylogenomics are often concerned with resolving deep nodes in the tree of life where ILS should be less of an issue (Gatesy and Springer 2014). Still, concatenated data sets also appear to perform well for shallower nodes (Wang et al. 2017) and have correctly recovered relationships in studies with simulated sequence data sets (Rubin et al. 2012; Cariou et al. 2013; Tonini et al. 2015; Rivers et al. 2016). Based on these previous studies, we assume that in principle, both coalescence and concatenation approaches should be effective in reconstructing *Salamandra* relationships, given sets of DNA sequences appropriate for the respective method.

However, it is questionable whether the *Salamandra* data sets are equally appropriate for being analysed with the two methods. It has been contended that coalescence methods should not be applied to complete protein-coding loci because they amalgamate potentially recombining genomic regions with different evolutionary histories, therefore violating important assumptions of the multispecies coalescent model (Springer and Gatesy 2016). In our analyses, it is obvious that most of the *Salamandra* RNA-Seq derived loci corresponded to multiple exons that are distant from each other in the genome, separated by long intronic sequences that were not included in our data set (as they are not translated into mRNA).

Furthermore, coalescence methods can be less accurate than concatenation when the gene trees have poor phylogenetic signal (Mirarab et al. 2016). The example of *S. atra* relationships indicates that only a minority of all gene trees supports the favoured placement of the species, in agreement with other examples of lacking phylogenetic

congruence among single-gene trees (e.g., Dikow and Smith, 2013). This indicates that the protein-coding nuclear loci derived from RNA-Seq might not be sufficiently informative for reconstructing single-gene trees. Also the RAD-Seq loci correspond to very short sequences, each with few SNPs, which might not be suitable for calculating single-locus trees with adequate phylogenetic resolution among the various species. For these loci we therefore used a SNP-based species tree approach; however, this has not yet been extensively tested and therefore might require methodological refinement.

Given that our data sets might not be optimal for being analysed with species tree approaches, we consider the phylogenies obtained by the concatenated analyses (Fig. 2.1) to be more reliable. As a further cause for the conflicting topologies, we emphasize that in *Salamandra*, the phylogenetic signal supporting relationships within the *atra-corsica-lanzai* clade is at best very weak, given that in RNA-Seq gene jackknifing, over 1 million base pairs are needed to stabilize the preferred concatenated topology with bootstrap support values >60%.

2.6.3. Evolutionary history and biogeography of Salamandra

We hypothesize that the phylogenies placing *S. atra* sister to *S. lanzai* (e.g., Fig. 2.1) represent most accurately the evolutionary history of the genus *Salamandra*. We base this hypothesis on: (1) the congruence of the trees obtained from separate concatenated analyses of the RNA-Seq and RAD-Seq datasets (Fig. 2.1); and (2) the phenotypic similarity of *atra* and *lanzai* in many key traits. Both species occur in the Alps, are entirely black-coloured (except the subspecies *S. atra aurorae* and *S. atra pasubiensis*), and are pueriparous, i.e., giving birth to fully metamorphosed juveniles (Fig. 2.5). In contrast, *S. corsica* has a geographic distribution restricted to the island of Corsica, is larviparous and is yellow-black coloured. If our preferred phylogenetic hypothesis is correct, then the alternative clade (*atra-corsica*) as strongly supported by the mitogenomic data probably reflects ancient hybridization, with introgression of the mitochondrial genome of an ancestral *S. atra* population into the ancestor of *S. corsica* and replacement of the original mitochondrial genome of that species.

As pointed out by Vences et al. (2014), one of the main sources of disagreement in previous molecular studies of *Salamandra* phylogeny was the placement of the root: while the unrooted topology was almost fully congruent among all analyses published to date, the outgroup (*Lyciasalamandra*) in previous studies was connected alternatively to almost

every branch in the *Salamandra* tree (see Fig. 2.3 in Vences et al. 2014), leading to radically different phylogenetic scenarios. Here, we provide rather strong evidence that the position of *S. infraimmaculata* as the sister group of all other *Salamandra* species is most likely the one correctly reflecting the evolution of these salamanders. This topology is stable across all of the nuclear phylogenomic trees we resolved (Figs. 1 and 3) and was previously recovered in a phylogeny based on a small set of nuclear genes (Vences et al. 2014). In gene jackknifing this topology (i.e., the clade grouping all *Salamandra* to the exclusion of *S. infraimmaculata*) received support of 60% with a data set of 500,000 bp (RNA-Seq) and of 72% with 100,000 bp (RAD-Seq), confirming it is highly supported by the data. Accepting this relationship of *S. infraimmaculata* as the sister species of all other *Salamandra*, the consensus topology of the nuclear gene data (Fig. 2.5) also suggests an origin of the genus in the Near East, considering that the two earliest branching clades (*Lyciasalamandra* and *S. infraimmaculata*) are restricted to this region.

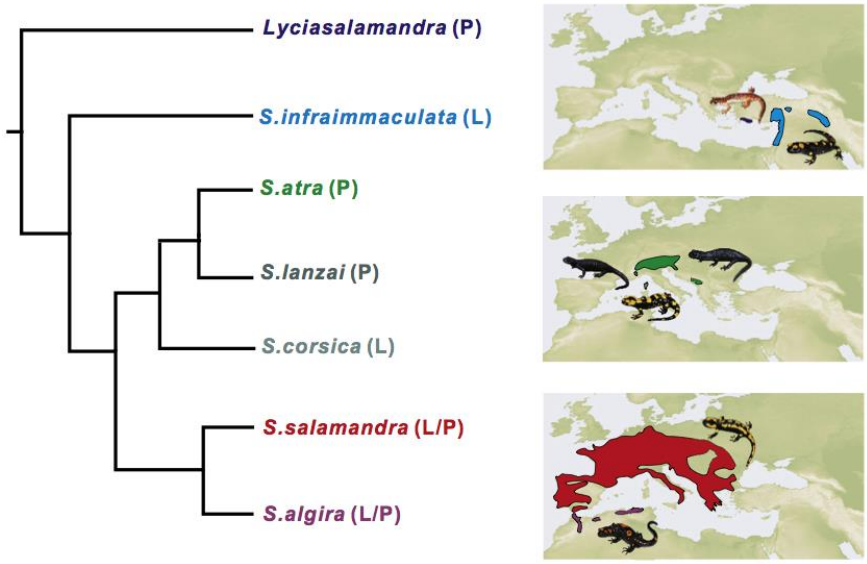


Figure 2.5: Phylogenetic hypothesis of the genus *Salamandra*, based on concatenation analyses of nuclear gene data (RNA-Seq and RAD-Seq), and maps showing approximate distribution of each species (or genus in the case of *Lyciasalamandra*). L and P after species names indicate whether the respective species are larviparous (L; giving birth to larvae), pueriparous (P; giving birth to fully metamorphosed juveniles), or comprise populations with both reproductive modes (L/P). Inset pictures are representative individuals of each taxon.

2.7. Conclusion

The data presented here have shown the potential of phylogenomic data sets to elucidate shallow relationships among closely related taxa, even if these have probably been characterized by past episodes of introgression. It is encouraging that different commonly used phylogenomic approaches, such as RNA-Seq and RAD-Seq, result in data sets that yield congruent results, despite having very distinctive properties. Yet, our results also confirm the need for caution in interpreting high bootstrap proportions or Bayesian posterior probability values: with an increase in quantity of phylogenomic data, high values of these classical support metrics can be misleading as they do not necessarily reflect a strong phylogenetic signal for a certain branch.

We are convinced that further improvement of analytical tools is of highest importance to deal with phylogenomic and phylotranscriptomic data sets, because model violations are to some degree inherent to all the methods used herein. For example, identification of the best fitting partition and substitution models is currently a computational hurdle for such large data sets and additional analytical tools are needed to better unravel, and critically assess, node support in both concatenation and species tree methods.

Although the massive data sets discussed herein provide a well-founded evolutionary and biogeographic hypothesis for the genus *Salamandra*, some doubts still remain on the relationships within this interesting group of terrestrial salamanders. Full genome sequences, currently prohibitive in costs, would allow a more conclusive understanding of past demography and possible episodes of introgressive hybridization among species of *Salamandra*, by identifying contiguous parts of genomic sequences affected by introgression. However, this might not necessarily result in a stronger phylogenetic signal. By the analysis of different comprehensive molecular data sets, we have definitely approached a limit to resolve the phylogenetic relations of these amphibians.

Chapter 3: Phylogenomic relationships and convergent phenotype evolution in the Palaearctic fire and Alpine salamanders (*Salamandra* spp.)

3.1. Abstract

The genus *Salamandra* is renowned for its biological diversity, with many researchers considering it an ideal model for studying the evolution of aposematic (warning) colouration and viviparity. This is because it shows striking inter- and intraspecific variation in colour patterning and parity mode, with indications of parallel phenotype evolution in these two traits. However, the phylogenetic relationships within the genus have been hard to elucidate, leaving its true diversity and evolutionary history a topic of debate and contention. In this study, we used double digest restriction site associated DNA sequencing (ddRAD-Seq) to genotype salamanders from across the taxonomic and geographic breadth of *Salamandra* in order to resolve species and subspecies relationships within the genus. The resulting phylogenetic hypothesis was then used to assess the parallel evolution of reproductive (parity) mode and two colour phenotypes (melanism and stripe formation) through ancestral state reconstruction analyses. Both maximum likelihood and Bayesian based phylogenetic analyses of concatenated RAD-loci returned well-supported, largely congruent topologies that supported the monophyly of all six currently recognised species (*Salamandra algira*, *Salamandra atra*, *Salamandra corsica*, *Salamandra infraimmaculata*, *Salamandra lanzai*, and *Salamandra salamandra*). However, uncertainty remains regarding the phylogenetic relationships between these species, particularly in the placement of *S. corsica*. Phylogenetic analyses also highlighted undescribed diversity within the North African *S. algira* and, for the first time, robustly revolved the relationships between all *S. salamandra* subspecies—43% of which do not meet a criterion of monophyly. Finally, through ancestral state reconstruction analyses we found that pueriparity (giving birth to fully metamorphosed juveniles) has independently arisen in at least four lineages, melanism in at least five, and a striped phenotype in at least two, all from a common yellow-black spotted larviparous (larvae depositing) ancestor. These results confirm the convergent (or parallel) evolution of both reproductive and colour phenotypes in *Salamandra* spp., and highlight important considerations for the application of RAD-Seq based phylogenomic methods in taxonomically challenging groups.

3.2. Introduction

In order to study the evolutionary patterns and processes that generate and maintain biological diversity, we must have a clear understanding of how organisms relate to one another on the tree of life. Yet, over three decades on from the PCR revolution, the phylogenetic histories of many lineages evade adequate resolution. While the reasons for this are numerous and well documented (see Philippe et al. 2011), the development of new high-throughput sequencing technologies and phylogenomic methods has provided an unprecedented opportunity to investigate evolutionarily challenging relationships (Philippe et al. 2005; Leaché et al. 2015a). From the sequencing of whole genomes, to the use of reduced complexity approaches like restriction site associated DNA sequencing (RAD-seq; Baird et al. 2008), transcriptome sequencing (RNA-seq; Wang et al. 2009), and anchored hybrid enrichment (Lemmon et al. 2012), molecular systematists can now obtain thousands of genome wide molecular markers to test evolutionary hypotheses (Prum et al. 2015; Wen et al. 2015; Massatti et al. 2016; Posada 2016; Irisarri et al. 2017).

One lineage where such phylogenomic approaches offer the opportunity to robustly resolve shallow phylogenetic relationships is the Palaearctic salamander genus *Salamandra*. This genus currently contains six recognised species: the European fire salamander, *Salamandra salamandra*; the Alpine species *Salamandra atra* and *Salamandra lanzai*; the Corsican fire salamander, *Salamandra corsica*; the North African fire salamander, *Salamandra algira*; and the Near Eastern fire salamander, *Salamandra infraimmaculata* (Speybroeck et al. 2010; Sillero et al. 2014). However, the evolutionary relationships between these species have been challenging to elucidate, with the analyses of few traditional mitochondrial (mtDNA) and nuclear (nDNA) sequences returning unresolved, weakly supported or topologically conflicting trees (Veith et al. 1998; Steinfartz et al. 2000; García-París et al. 2003; Vences et al. 2014). It required the phylogenomic analyses of multiple high-throughput sequencing data sets (RNA-Seq, RAD-Seq and full mitochondrial genome sequencing) to robustly infer the inter-species relationships within the genus (Rodríguez et al. 2017/Chapter 2). However, while a validation of such phylogenomic approaches, this study was taxonomically limited, leaving much of the subspecific diversity within the genus phylogenetically unresolved.

While the taxonomic rank of subspecies is often controversial (reviewed in Phillimore and Owens 2006), it has been applied extensively across *Salamandra*. Within the genus, the greatest intraspecific diversity is seen within *S. salamandra*, where 13–14 subspecies are

currently recognised, largely based on external morphological characters and colour patterns (Thorn and Raffaëlli 2001; Thiesmeier 2004; Sparreboom and Arntzen 2014; Velo-Antón and Buckley 2015). Of these, nine are found in the Iberian Peninsula, with seven endemic to the region (García-París et al. 2003; Martínez-Solano et al. 2005). Subspecific taxa have also been described within *S. atra* (three subspecies; see Bonato and Steinfartz 2005), *S. algira* (four subspecies; see Beukema et al. 2013) and *S. infraimmaculata* (three subspecies; see Böhme et al. 2013). However, the recognition of species and subspecies across the genus has often been contentions, largely due to an overreliance on unstable or uninformative phenotypic characters and distributional data (see Dubois & Raffaëlli 2009; Speybroeck et al. 2010; Frost 2017). For example, detailed genetic studies have revealed that colouration is unsuitable for the taxonomic categorisation of *S. salamandra* specimens in northern Spain (Beukema et al. 2016a), and colour patterns have been shown to overlap between *S. infraimmaculata* subspecies (Böhme et al. 2013). However, with the emergence of reduced complexity high-throughput genome sequencing techniques like ddRAD-Seq (Davey and Blaxter 2010; Peterson et al. 2012), it may now be possible to accurately infer the relationships between these subspecific taxa, thereby allowing us to study the evolution of complex, ecologically adaptive traits within the group, like colouration and reproductive mode.

The striking colour patterns displayed by these toxic amphibians have fascinated scientists for centuries due to their supposed aposematic (warning) function (Beukema et al. 2016b). Typically, *Salamandra* display yellow spots or stripes on a black dorsal surface, although patterns vary dramatically within and between taxa (see Sparreboom and Arntzen 2014). Some populations also possess red colouration, and the emergence of melanic, brown (hypolutic) and yellow (xanthic) populations has been noted (Thiesmeier 2004; Sparreboom and Arntzen 2014; Velo-Antón and Buckley 2015). Importantly, there are also indications of parallel colour phenotype evolution, particularly in terms of melanism (Bonato and Steinfartz 2005; Vences et al. 2014). However, greater phylogenetic resolution is required to identify the repeated loss or emergence of colour patterns.

Salamandra are also renowned for their reproductive diversity, and are thought to be an ideal system for studying the evolution of viviparity (live birth). Within caudate amphibians (the newts and salamanders), *Salamandra* is one of only two genera to display this reproductive mode (the other being its sister genus *Lyciasalamandra*; Veith et al. 1998; Weisrocka et al. 2001; Veith and Steinfartz 2004; Frost et al. 2006; Weisrock et al. 2006; Steinfartz et al. 2007a; Zhang et al. 2008; Buckley 2012; Pyron 2014). *Salamandra*

are particularly fascinating due to the diversity of viviparity strategies they display, from the deposition of yolk fed larvae (ovoviviparity/larviparity) to the birth of fully metamorphosed terrestrial juveniles (pueriparity) fed on yolk or maternal nutrition (Wake 1993; Greven and Guex 1994; García-París et al. 2003; Greven 2003b; Buckley et al. 2007; Caspers et al. 2014). Fascinatingly, the species *S. salamandra* and *S. algira* display both larviparity and pueriparity, making them two of the few vertebrate taxa to possess intraspecific variation in reproductive mode (García-París et al. 2003; Beukema et al. 2010; Buckley 2012). However, while pueriparity is suspected to have evolved independently in several lineages (see Buckley et al. 2007), a lack of phylogenetic resolution again prevents the robust assessment of parallel/convergent evolution in this trait in *Salamandra*.

In this study, I conducted the most taxonomically comprehensive phylogenomic analyses of the genus *Salamandra* to date. This was done in order to 1) assess species level diversity within the genus; 2) resolve subspecies relationships within *S. algira* and *S. salamandra*; and 3) assess the parallel (or convergent) evolution of parity mode and colour pattern phenotypes. For this, I ddRAD-Seq genotyped 231 salamanders from across the taxonomic and geographic breadth of *Salamandra* and conducted both maximum likelihood and Bayesian based phylogenetic analyses. I also explored the impact of two data filtering parameters on downstream phylogenetic analyses: per locus missing data and per locus variable site number. Finally, I conducted ancestral state reconstructing analyses to trace the evolutionary history of parity, melanism and stripe formation within the genus.

3.3. Methods

3.3.1. Taxonomic sampling

Tissue samples (toe or tail clips; n=272) were collected from across the taxonomic and geographic breadth of *Salamandra* and two samples from its sister genus, *Lyciasalamandra* (Table 3.1; Fig. 3.1; for full collection data see Appendix 1, Table A1.2). These samples included almost all currently recognised *Salamandra* species and subspecies, with the exception of *Salamandra atra aurorae*, a locally protected yellow-black patterned subspecies of *S. atra* endemic to the Venetian Prealps (north-eastern Italy), and *Salamandra infraimmaculata orientalis*, from the Adana region of southern Turkey, for which tissue samples were not available.

Table 3.1: Taxonomic sampling for phylogenomic analyses, showing the total number of samples collected per taxon and the number that passed sequence quality filtering for phylogenetic analyses (see sections *3.3.2.* to *3.3.4.*). * = *S. s. alfredschmidti* has recently been synonymised with *S. s. bernardezi* (Beukema et al. 2016a).

Species	Subspecies	Authority	Total samples	231 Indiv. data set	89 Indiv. data set
Salamandra algira	*algira*	Bedriaga, 1883	23	15	6
	tingitana	Donaire-Barroso and Bogaerts, 2003	18	16	3
	spelaea	Escoriza and Comas, 2007	3	3	3
	splendens	Beukema, de Pous, Donaire-Barroso, Bogaerts, Garcia-Porta, et al. 2013	10	10	3
Salamandra atra	*atra*	Laurenti, 1768	2	2	2
	aurorae	Trevisan, 1982	0	0	0
	pasubiensis	Bonato and Steinfartz, 2005	3	3	3
Salamandra corsica	-	Savi, 1838	3	3	3
Salamandra infraimmaculata	*infraimmaculata*	Martens, 1885	26	25	6
	orientalis	Wolterstorff, 1925	0	0	0
	semenovi	Nesterov, 1916	1	0	0
Salamandra lanzai	-	Nascetti, Andreone, Capula and Bullini, 1988	1	1	1
Salamandra salamandra	*alfredschmidti**	Köhler and Steinfartz, 2006	12	12	3
	almanzoris	Müller and Hellmich, 1935	2	2	2
	bejarae	Wolterstorff, 1934	3	2	2
	bernardezi	Wolterstorff, 1928	10	9	6
	beschkovi	Obst, 1981	2	2	2
	crespoi	Malkmus, 1983	2	2	2
	fastuosa	Schreiber, 1912	6	6	4
	gallaica	López-Seoane, 1885	17	15	10
	gigliolii	Eiselt and Lanza, 1956	6	6	3
	hispanica	Wolterstorff, 1937	2	2	2
	longirostris	Joger and Steinfartz, 1994	11	9	3
	morenica	Joger and Steinfartz, 1994	11	11	3
	salamandra	(Linnaeus, 1758)	32	29	9
	terrestris	Lacépède, 1788	63	41	3
	werneri	Sochurek and Gayda, 1941	3	3	3
Lyciasalamandra billae	-	(Franzen and Klewen, 1987)	1	1	1
Lyciasalamandra flavimembris	-	(Mutz and Steinfartz, 1995)	1	1	1

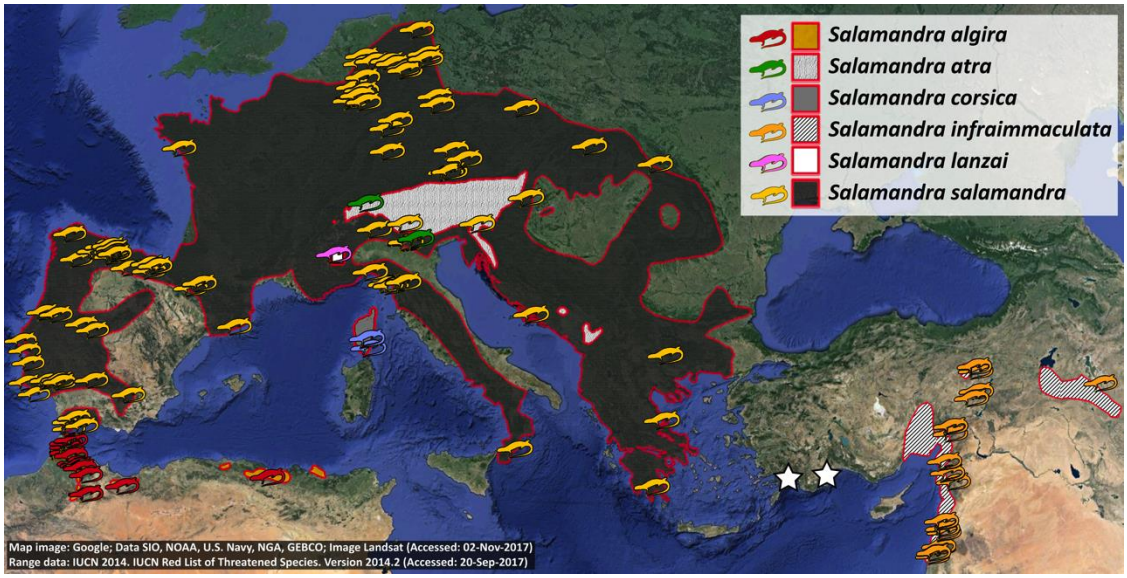


Figure 3.1: Geographic distributions for the six *Salamandra* species (shaded areas), with approximate tissue-sampling localities (salamander cartoons). Stars indicate *Lyciasalamandra* outgroup sample sites.

3.3.2. Sequencing and de novo assembly

Genomic DNA was extracted from tissue samples using the Macherey-Nagel NucleoSpin® Tissue kit following manufacturer's instructions. The ddRAD-Seq library preparation followed Recknagel et al. (2015), with modification of Illumina adapters (see Appendix 2 for a detailed protocol). Briefly, 1µg of DNA from each individual was digested using the *PstI-HF*® and *AclI* restriction enzymes (New England Biolabs, Ipswich); seventeen technical replicates were also run to quantify genotyping error rates (n=291 total samples). Modified Illumina adaptors with unique combinatorial barcodes for each sample were ligated onto this fragmented DNA (see Appendix 2, Table A2.1 for barcode information); samples were multiplexed; and a PippinPrep (Sage Science, Beverly) used to size select fragments around a tight selection of 383 bp (range: 350–416 bp), based on the fragment length distribution identified using a 2200 TapeStation instrument (Agilent Technologies, Santa Clara). Finally, enrichment PCR was performed to amplify the library using forward and reverse RAD primers. Samples were split across five libraries (49–87 individuals per library; see Appendix 2, Table A2.1), each of which was sequenced on the Illumina NextSeq™ 500 platform at Glasgow Polyomics to generate 3853.3M paired-end reads 75bp in length.

Raw sequence reads were first quality checked using FastQC v.0.11.3 (Andrews 2010). Samples were then de-multiplexed, Illumina adaptors and barcodes removed, and reads truncated to 60 nt using Stacks v.1.44 *process_radtags* (Catchen et al. 2011), removing reads with uncalled bases or low quality scores and rescuing barcodes and RAD-Tags. Processed PE reads per sample varied considerably, ranging between 7078 and 37.8M (average: 12.6M); those samples having less than 150,000 reads were immediately excluded from further analyses (n=4). As no reference genome exists for *Salamandra*, reads were assembled *de novo* in Stacks: the minimum number of identical raw reads required to create a stack was set to six, with other settings left on default (two mismatches were allowed between loci when processing a single individual, one mismatch was allowed between loci when building the catalogue, and highly repetitive RAD-Tags were removed, or broken up, in the ustacks program).

Preliminary data filtering was then carried out to identify low coverage individuals. For this, a whitelist of all loci containing 1–5 SNPs was generated from the Stacks catalogue using custom scripts in R (R Core Team 2013; see Appendix 6). This was used to filter the samples through the Stacks *populations* pipeline (as a single population). Whitelisted loci were retained if they were present in ≥75% of samples, had a minimum individual stack depth of six, a maximum observed heterozygosity of 0.5, and a minimum minor allele frequency of 0.05. Data were exported in Variant Call Format (VCF; one SNP per locus). This identified 4130 RAD-loci shared across samples, with any showing >75% missing data excluded from further analyses (n=10, including one technical replicate).

3.3.3. Genotyping error

The technical replicates (n=16) were then assessed to calculate genotyping error rates. After excluding the low coverage individuals identified above (n=14), data for the remaining 231 individuals were again filtered through Stacks *populations* (using the same settings as before) with data were exported in Phylip format, including all variable sites encoded using IUPAC notation (9107 loci; 30,742 SNPs). Using custom scripts in R (see Appendix 6), an average genotyping error rate of 0.12% was calculated between replicates (range: 0.02–0.54%), lower than that reported in published protocols (e.g. Recknagel *et al*. 2015). A RAxML analysis (raxmlHPC; Stamatakis 2014) was then run to confirm that technical replicates clustered together in an analysis of the full dataset, using a GTRCAT model and 100 rapid bootstrap replicates. The resulting best tree was visualised in FigTree 1.4.0 (http://tree.bio.ed.ac.uk/software/figtree/). Following this, the replicate with the most

missing data of the pair was removed from further analyses. As a result of these analyses, five mislabelled samples putatively assigned to *S. algira* were also identified and excluded (as their collection data could not be confirmed), and one population of *S. s. terrestris* was down sampled from 23 to four individuals (retaining those individuals with the greatest number of loci) as samples clustered together. This left a total 237 individuals (Table 3.1).

3.3.4. SNP calling optimization

Exploratory phylogenetic analyses were carried out to optimise SNP calling parameters, as sample coverage can effect the performance of *de novo* assembled RAD-Seq data matrices in phylogenetic reconstructions (Huang and Knowles 2014; Takahashi et al. 2014). I used the species-level tree generated using our comparative phylogenomics analyses as a guide for choosing the most appropriate parameter set (Rodríguez et al. 2017/Chapter 2). Initially, the impact of the maximum number of SNPs allowed per locus on species-level relationships was assessed. Custom scripts in R (see Appendix 6) were used to create whitelists for all loci in the Stacks catalogue containing a maximum of 2, 3, 4, 5, 6, 7, 8, 9 and 10 SNPs. Samples (n=237) were filtered through the Stacks *populations* pipeline as a single population (as before, with the exception of the minimum stack depth, which was increased to 10) using each of these whitelists. The number of loci retained ranged between 1218 and 10,086, with an average increase of 1109 loci per additional SNP (although this varied considerably; Appendix 4 Fig. A4.1). RAxML analyses were then carried out on each of the resulting data sets as above.

The impact of missing data was then assessed. Using the whitelist allowing up to 5 SNPs per locus, data were filtered through Stacks as before, but this time allowing for different levels of per locus missing data, from 10% to 50% (in 5% increments). The number of retained loci ranged between 977 and 19,375, with an average increase of 2422 loci per 5% increase in missing data (Appendix 4 Fig. A4.1). As before, RAxML analyses were carried out on each of the resulting data sets.

From these analyses, a further six samples were identified for exclusion. These included one sample of *S. infraimmaculata* from Iran (the only one corresponding to the subspecies *S. i. semenovi*) and one sample of *S. s. bernardezi*, both of which returned fewer than 10 loci when using the max-SNPs 5 whitelist and allowing for 25% per locus missing data. Four samples with unconfirmed collection details were found not to cluster with other samples from the same subspecies, and as such were also removed: one sample of *S. s.*

bejarae, one sample of *S. s. salamandra*, and two samples of *S. s. longirostris.* The remaining 231 individuals (including the outgroup) were then filtered back through Stacks *populations* allowing 1–5 SNPs and up to 25% missing data per locus, resulting in a data set of 3475 RAD-loci for phylogenetic analyses of variation within and between subspecies (12,750 SNPs; 7280 parsimony-informative sites; full concatenated sequence length per individual = 187,080 nucleotides (nt); total missing data = 13.8%).

To further clarify species and subspecies relationships, data were collapsed to 89 individuals for Bayesian inference (see Table 3.1 and Fig 3.3–4). For this, one to three individuals per species/subspecies (those with the least missing data in the 231 individual data set) were retained, depending on sample availability. However, where RAxML analyses of the 231 individual data set indicated additional sub-clades or potential paraphyly, extra samples were included. This was done for *S. a. algira* (n=3), *S. s. salamandra* (n=6), *S. s. gallaica* (n=7), *S. s. bernardezi* (n=6), *S. s. fastuosa* (n=1), and *S. infraimmaculata* (n=3). Data for these 89 individuals were re-filtered in Stacks *populations* as above, generating a data set of 4905 RAD-loci (16,875 SNPs; 15,731 parsimony-informative sites; full concatenated sequence length per individual = 294,300 nt; total missing data = 13.2%).

3.3.5. Genetic variation within and between species and subspecies

Using the 231 individual dataset, p-distances between the six recognised species were calculated using MEGA 7 (Kumar et al. 2016). This is the raw proportion of nucleotide sites at which two sequences differ from one another. P-distances were also calculated between subspecies in *S. algira* and *S. salamandra*. Further, to assess intraspecific sequence variation, within group p-distance was also calculated for all taxa with more than one sample. To look at genetic clustering between and within species (excluding the outgroup), principal component analyses (PCAs) were then conducted using the *dudi.pca* function in the *ade4* R package (Dray and Dufour 2007).

3.3.6. Phylogenetic analyses

Phylogenetic analysis of concatenated RAD-loci for 231 individuals was carried out in RAxML, using a GTRCAT nucleotide substitution model and 1000 bootstrap replicates. Following this, the RAxML analysis was repeated on the 89 individual data set (to check for congruence between data sets) and Bayesian inference conducted on the 89 individual

data set in BEAST 2.4.2 (Bouckaert et al. 2014). For the latter, jModeltest 2.1.10 (Guindon et al. 2003; Darriba et al. 2012) was used to infer the best fitting evolutionary model of nucleotide substitution from 11 substitution schemes; based on the lowest Bayesian information criterion (BIC), this was determined to be the transversion model (TVM). A BEAST xml file was then generated using BEAUTi 2.4.2, and the TVM substitution model approximated by selecting a GTR substitution model and fixing the alpha gamma rate parameter at one (following Bagley 2016). A relaxed clock (log normal) was used, with all other parameters left as default. BEAST2 analyses were then carried out using the CIPRES Science Gateway server (Miller et al. 2010).

An optimisation MCMC chain of 100M generations was run, with tree and parameter estimates sampled every 1000 MCMC generations. Based on the output of this, several prior operators were adjusted based on BEAST2 operator estimations: the scale factor for the Yule model tree scaler was set to 0.944, the tree root scaler was set to 0.945 and the size of the subtree slide was decreased to 0.039. Rate scalars for the mutation model were also adjusted, with the AC, AT, CG, and GT scale factors set to 0.924, 0.922, 0.907 and 0.926, respectively. Finally, the scale factor for the standard deviation of the uncorrelated log-normal relaxed clock was set to 0.964. The analysis was then run for 300M generations (sampling every 30,000 generations and discarding the first 10% as burn-in). Tracer 1.6.0 (Drummond and Rambaut 2007) was used to assess chain convergence and a maximum clade credibility tree was generated using TreeAnnotator 2.4.2 (Bouckaert et al. 2014).

3.3.7. Ancestral state reconstructions of reproductive mode, melanism and colour pattern

To identify the independent, repeated evolution of colour and reproductive phenotypes, I conducted ancestral character state reconstruction analyses. First, the literature was searched to determine the phenotype of contemporary species and subspecies. For parity mode, three character states were identified: strict pueriparity, strict larviparity, or both. Pueriparity, where salamanders give birth to fully metamorphosed terrestrial juveniles (as opposed to depositing aquatic larvae; larviparity), is found in the Alpine species *S. lanzai* and *S. atra* (Buckley 2012). It is also seen as one of two reproductive strategies in *S. algira tingitana* (Beukema et al. 2010), and three subspecies of *S. salamandra*: Atlantic island populations *of S. s. gallaica* (closly related to mainland populations in Galicia; Velo-Antón et al. 2007), *S. s. bernardezi* and *S. s. fastuosa* (García-París et al. 2003). All other *Salamandra* species and subspecies are larviparous, while the sister genus (*Lyciasalamandra*) displays strict pueriparity (see Buckley 2012).

For colouration, given the range of intraspecific variation seen, the evolvability of two phenotypes was assessed: melanism (coded as either brown/black or yellow-black patterned), and gross colour pattern (coded as striped, spotted or without pattern). Melanism is seen within five lineages, including the fully melanic (black) species *S. lanzai*, the predominantly melanic *S. atra* (*S. a. atra*; Bonato and Steinfartz 2005), and two subspecies where it is seen as a common colour variant: *S. algira tingitana* (Martínez-Solano et al. 2005) and *S. salamandra gallaica* (Velo-Antón and Buckley 2015). In addition, brown (or hypolutic) colouration, which also results from an increase in melanin pigments, is common in *S. s. bernardezi* (Köhler and Steinfartz 2006) and *S. algira tingitana* (Beukema et al. 2010), and is occasionally seen in *S. a. atra* (Bonato and Steinfartz 2005). All other lineages show variable yellow-black colourations (see Sparreboom and Arntzen 2014; Velo-Antón and Buckley 2015; Speybroeck et al. 2016).

In terms of gross colour patters, three broad phenotypes were identified: yellow-black striped (*S. s. bernardezi*, *S. s. fastuosa* and some populations of *S. s. terrestris*; Siedel et al. 2012; Sparreboom and Arntzen 2014; Speybroeck et al. 2016), single colours (i.e. no pattern; *S. a. atra* and *S. lanzai*; Bonato and Steinfartz 2005) and yellow-black spotted (all other species and subspecies; see Sparreboom and Arntzen 2014; Velo-Antón and Buckley 2015; Speybroeck et al. 2016). While these colourations show a great deal of variability, for ASR analyses the most common phenotype per taxon was selected in all cases except *S. s. terrestris*. While this subspecies is typically spotted, these spots form discontinuous dorso-later stripes, and in some populations have fused to form continuous stripes (Siedel et al. 2012; Sparreboom and Arntzen 2014). As such, it was coded as "striped" to assess the potential independent evolution of this phenotype.

To reconstruct ancestral states, I used the maximum clade credibility tree generated through BEAST2 analysis of the 89 individual dataset (294,300 nt) as a phylogenetic hypothesis. This tree was selected due to its greater level of support and its placement of *S. corsica* as sister to the *S. lanzai* + *S. atra* clade, which is congruent with comparative phylogenomic analyses in **Chapter 2**. Using the R package *phytools* (Revell 2012), I trimmed taxa to a single representative per species/subspecies (except *S. s. gallaica*, for which a single representative was retained for each of the three separate clades identified). Following this, we used the *ace* function in *phytools* to reconstruct ancestral states through maximum likelihood estimation, assuming equal rates of transition between discrete character states (see Appendix 4 Table A4.1 for input character states).

3.4. Results

3.4.1. The influence of data filtering parameters on inter-species relationships

The phylogenetic relationships recovered between the six currently recognised *Salamandra* species were found to be highly sensitive to data filtering parameters. When allowing different maximum numbers of SNPs per locus (from two to 10), RAxML analyses recovered seven different inter-species topologies, with only the sister placement of *S. salamandra* and *S. algira* consistently found (Appendix 4 Fig. A4.2). Notably, only analyses of the data generated when allowing for a maximum of five and six SNPs per locus resulted in topologies congruent with earlier comparative phylogenomic analyses (Rodríguez et al. 2017/Chapter 2). While inter-species relationships were more stable when allowing for different levels of per locus missing data, they still became highly variable when allowing ≤20% missing data; *S. salamandra* was even found to be paraphyletic with regards to *S. algira* when allowing for 1–5 SNPs and ≤15% missing data per locus (Appendix 4 Fig. A4.3). However, when allowing for ≥25% per locus missing data inter-species relationships stabilised. As a result of this, parameters allowing for 1–5 SNPs and up to 25% missing data per locus were chosen for ddRAD-Seq data filtering.

3.4.2. Diversity and relationships between Salamandra species

Based on the 231 individual dataset, p-distances between the six currently recognised species averaged 2.52% (see Appendix 4 Table A4.2 for full comparisons). The greatest distance was found between *S. infraimmaculata* and both *S. algira* and *S. salamandra* (2.9%), and the lowest (2.1%) was seen in two comparisons: *S. algira vs*. *S. salamandra* and *S. atra vs*. *S. lanzai*. Intraspecific sequence variation was considerably lower, averaging 1.1% (range: 0.4–1.6%). This contrasts sharply to comparisons with the outgroup, *Lyciasalamandra*, from which all *Salamandra* species differed by 3.5–4%.

Principle component analyses of the same data revealed four genetic clusters (Fig. 3.2). Three of these correspond to the species *S. infraimmaculata*, *S. algira* and *S. salamandra*, with the last cluster comprising the three remaining species (*S. corsica*, *S. lanzai* and *S. atra*). However, this structuring only explains 20.9% of the variation in the data.

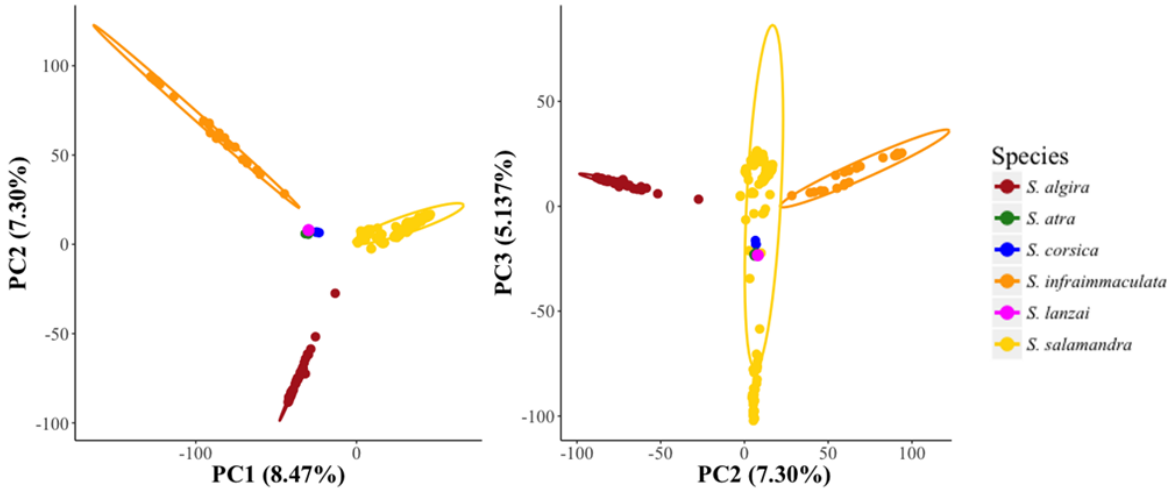


Figure 3.2: PCA of the 231 individual data set, with samples clustered by species. Ellipses show the 95% confidence interval for each cluster.

RAxML analysis of the 231 individual (187,080 nt) data set returned a moderately well supported tree (Fig. 3.3–4). In this analysis, the eastern *Salamandra infraimmaculata* was found to be the sister to all other species, after which there was a clade containing the Alpine species *S. lanzai* and *S. atra*. Following this, the Mediterranean *S. corsica* was found to be sister to a clade containing the European *S. salamandra* and the North African *S. algira*. Most interspecies nodes had a bootstrap support <85%, except those placing *S. infraimmaculata* as sister to all other species and *S. algira* and sister to *S. salamandra*, which both had 100% support. However, this analysis did not strongly support the monophyly of *S. salamandra*, as it found low support (41%) for the sister placement of two major clades within the species, suggesting their relationship with regard to the North African *S. algira* remains unresolved. In contrast, Bayesian phylogenetic analyses of the 89 individual concatenated data set (294,300 nt) returned a highly resolved phylogeny, which strongly supports the monophyly of all six currently recognised species (posterior probabilities = 1–0.98 for most nodes; Fig. 3.5). This analysis conflicts with the 231 individual data set only in its placement of *S. corsica*, which it places as sister to the *S. atra* + *S. lazai* clade. RAxML analysis of the 89 individual data set returned the same interspecies topology as the Bayesian analysis (Appendix 4 Fig. A4.4).

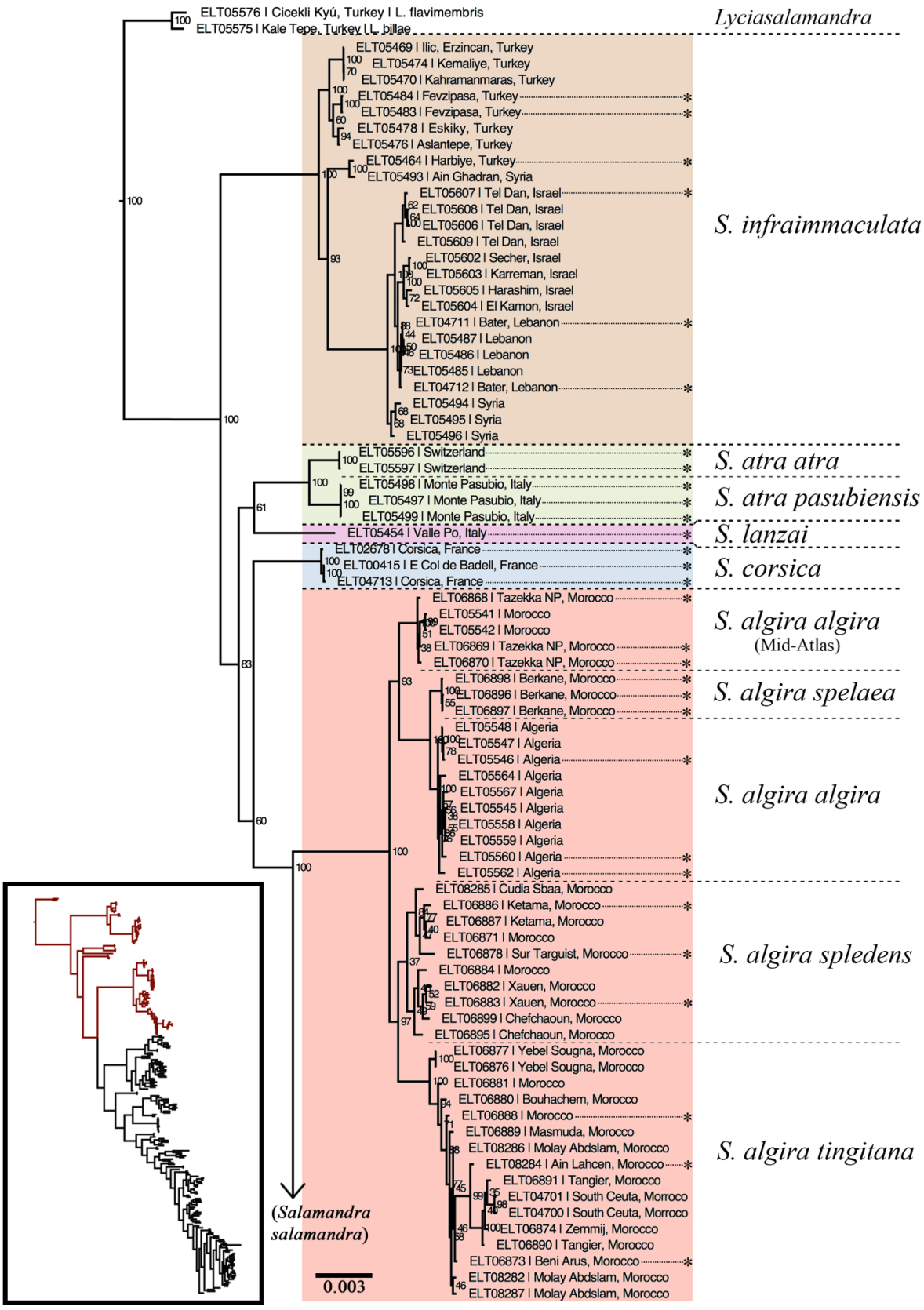


Figure 3.3: RAxML phylogenetic tree of 231 individuals based on 3475 RAD-loci (187,080 nt); continued in Figure 3.4. Coloured blocks highlight currently recognised species. Node values show bootstrap support. Asterisks (*) indicate those samples selected for the 89 individual data set. Inset: full RAxML tree with the portion shown in detail highlighted in red.

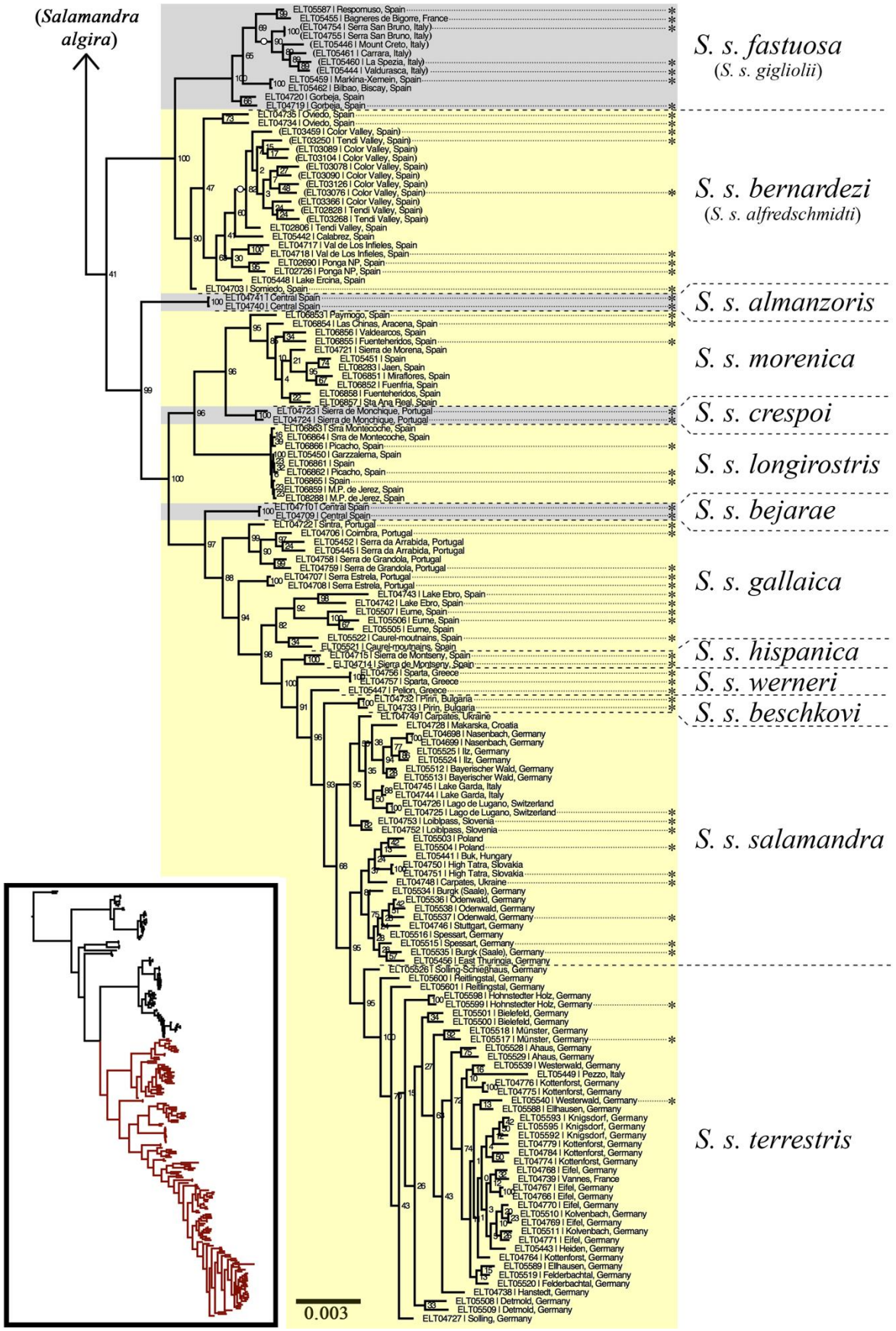


Figure 3.4: RAxML phylogenetic tree continued from Figure 3.3. Shading indicates monophyletic groupings in relation to descried *S. salamandra* subspecies. Node values show bootstrap support. Asterisks (*) indicate those samples selected for the 89 individual data set. Inset: full RAxML tree with the portion shown in detail highlighted in red.

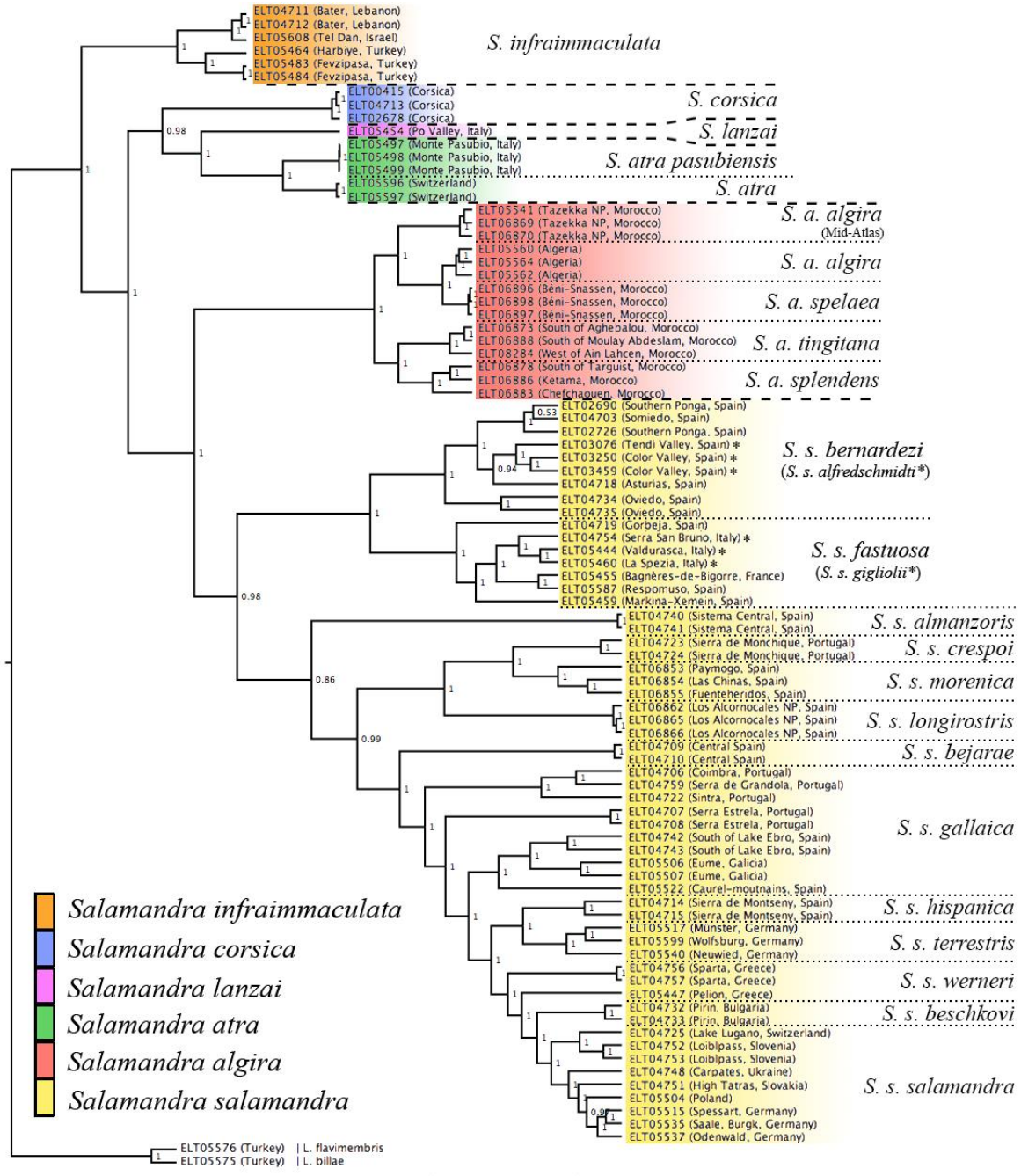


Figure 3.5: BEAST2 phylogeny of concatenated RAD-loci (294,300 nt) for 89 individuals. Coloured blocks highlight currently recognised species. Dashed lines separate species; dotted lines separate subspecies. Node values display posterior probabilities.

3.4.3. Diversity and relationships within Salamandra algira

Within *Salamandra algira*, all phylogenetic (Fig. 3.3; Fig. 3.5; Appendix 4 Fig. A4.4) and principal component (Fig. 3.6) analyses identified five clear clusters. These corresponded to the subspecies *S. a. tingitana*, *S. a. spelaea*, *S. a. spledens* and two further lineages both currently recognised as *S. a. algira*: one from the Algerian type locality and one from the Moroccan Mid-Atlas mountains (making *S. algira algira* paraphyletic). The genetic distance between these clusters averaged 1.58% (range: 1.4–1.8%; see Appendix 4 Table A4.3), lower than between species comparisons (Appendix 4 Table A4.2). Interestingly, while the Moroccan Mid-Atlas *S. a. algira* population is geographically closer to the *S. a. tingitana* and *S. a. spelaea* populations, it clustered as sister to the more eastern subspecies, *S. a. spledens* and the Algerian *S. a. algira* (Appendix 4 Fig. A4.5).

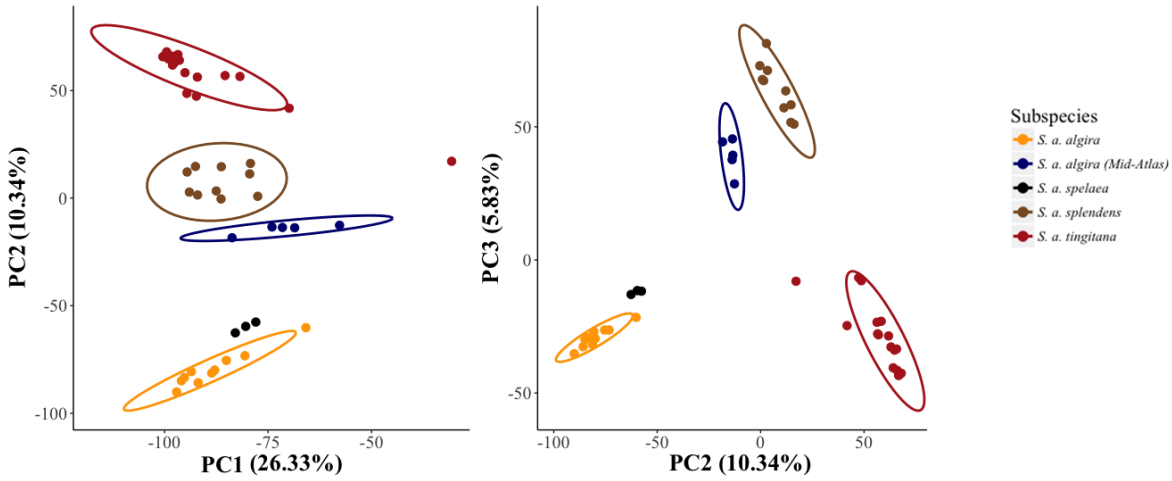


Figure 3.6: PCA of 44 *S. algira* samples sub-set from the 231 individual dataset (coloured by phylogenetic clade). Ellipses show the 95% confidence interval for each cluster.

3.4.4. Diversity and relationships within Salamandra salamandra

Within *Salamandra salamandra*, all phylogenetic analyses split samples into two major clades (Fig. 3.4; Fig. 3.5; Appendix 4 Fig. A4.4). Interestingly, while p-distances between the 14 subspecies included in these analyses varied greatly (from 1–2.5%), the average distance between the two major clades was just 1.9% (Appendix 4 Table 4.3). One of these clades contained the northern Spanish *S. s. bernardezi* and *S. s. fastuosa*, and the Italian *S. s. gigliolii*, *S. s. fastuosa* being paraphyletic with regards to *S. s. gigliolii*. The atypically

colour diverse population formerly known as *S. s. alfredschmidti* (from the Rio Color and Rio Tendi valleys in Asturias, northern Spain) was found to nest within the more widespread *S. s. bernardezi*. The other major *S. salamandra* clade contained all 11 remaining subspecies. In the 231 individual RAxML analysis (Fig 3.3), the first lineage to diverge within this clade was the central Spanish *S. s. almanzoris*, which is sister to all other subspecies. Next, the southern Spanish *S. s. longirostris* and *S. s. morenica* and the southern Portuguese *S. s. crespoi* clustered, all being monophyletic and the later two forming sister taxa. The central Spanish subspecies *S. s. bejarae* was then resolved as sister to all remaining taxa. The next lineage to diverge was the Portuguese/western Spanish subspecies *S. s. gallaica*, but this was recovered as paraphyletic, containing at least three distinct lineages. Following this, there was a monophyletic grouping of the northern Spanish *S. s. hispanica* samples, followed by the Greek *S. s. werneri*, which was found to be paraphyletic. Finally, the Bulgarian *S. s. beschkovi* was found to be sister to a clade containing the Central, Southern and Eastern European *S. s. salamandra* and the Western European *S. s. terrestris*; *S. s. salamandra* being paraphyletic with regard to *S. s. terrestris*. While phylogenetic analyses of the 89 individual data set were largely congruent with this, they did conflict in their placement of *S. s. terrestris*. In the 89 individual RAxML analysis *S. s. terrestris* was placed as sister to a clade containing *S. s. werneri* + *S. s. beschkovi* + *S. s. salamandra* (Appendix 4 Fig. A4.4), while Bayesian analysis of the same data found *S. s. terrestris* and *S. s. hispanica* to form a sister clade to the *S. s. werneri* + *S. s. beschkovi* + *S. s. salamandra* clade (Fig. 3.5). As such, these phylogenetic analyses only resolve eight of the fourteen subspecies included as monophyletic: *S. s. bernardezi*, *S. s. fastuosa*, *S. s. almanzoris*, *S. s. crespoi*, *S. s. morenica*, *S. s. longirostris*, *S. s. bejarae* and *S. s. salamandra*.

A PCA of 151 *S. salamandra* samples (data extracted from the 231 individual data set) showed similar relationships to the phylogenetic analyses (Fig. 3.7). First, the two major *S. salamandra* clades clearly separated along PC1, although this structure was lost in PCs 2 and 3. Along PC2, the southern Iberian *S. s. longirostris* + *S. s. morenica* + *S. s. crespoi* clade, along with the central Spanish *S. s. almanzoris*, show a clear separation from all other subspecies, with *S. s. almanzoris* further separating from all other subspecies along PC3. Finally, the clusters for *S. s. fastuosa* + *S. s. gigliolii*, *S. s. bernardezi* + *S. s. alfredschmidti*, and *S. s. beschkovi* + *S. s. hispanica* + *S. s. terrestris* + *S. s. salamandra* + *S. s. werneri* were found to greatly overlap.

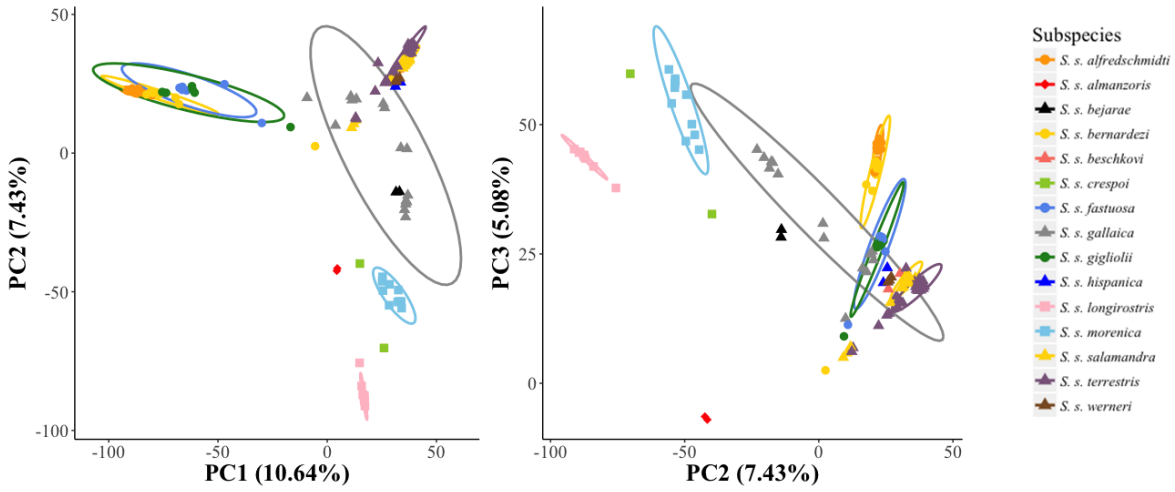


Figure 3.7: PCA of 151 *S. salamandra* samples sub-set from the 231 individual dataset (coloured by *a priori* subspecies designation). Ellipses show the 95% confidence interval for each cluster.

3.4.5. Ancestral state reconstructions of reproductive mode, melanism and colour pattern

Ancestral state reconstruction (ASR) analyses were performed to test hypotheses on the independent evolution of reproductive mode (parity), melanism and colour pattern. The parity of the *Salamandra* common ancestor was predicted to be larviparity, from which pueriparity has independently evolved at least four times: once in the common ancestor of *S. lanzai* and *S. atra*, once in *S. a. tingitana*, once in *S. s. gallaica* and at least once in the clade containing *S. s. bernardezi* and *S. s. fastuosa* (Fig. 3.7). The common ancestor of *Salamandra* was also predicted to have had a yellow-black colour pattern, from which melanism has arisen independently five times, once in *S. lanzai*, once in *S. atra*, once in *S. algira tingitana*, once in *S. s. gallaica* and once in *S. s. bernardezi*. Finally, the ancestral *Salamandra* colour pattern is predicted to be spotted, from which a striped phenotype has evolved at least twice, once in the *S. s. bernardezi* + *S. s. fastuosa* clade and once in *S. s. terrestris*. The phylogenetic placement of the yellow-black blotched larviparous *S. s. gigliolii* makes the colour pattern phenotype and reproductive mode of the *S. s. bernardezi* + *S. s. fastuosa* + *S. s. gigliolii* common ancestor unclear.

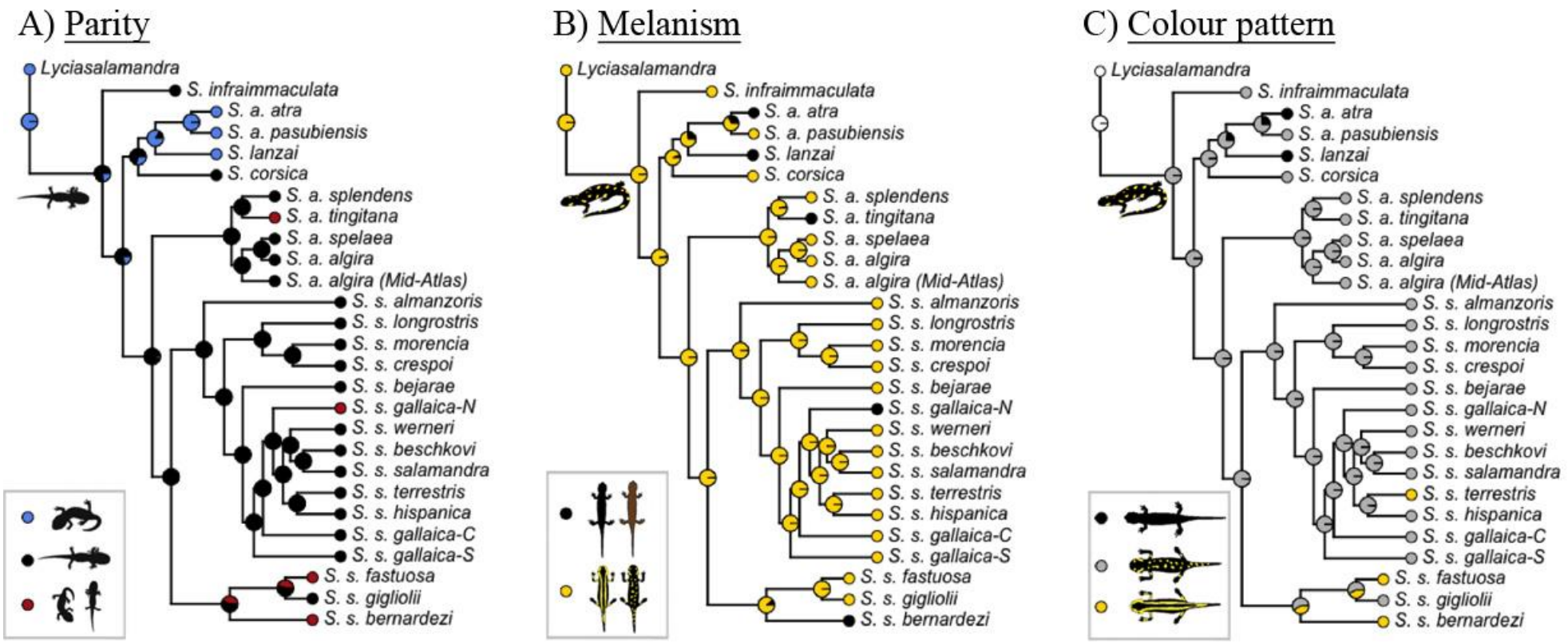


Figure 3.8: Ancestral state reconstructions for parity and colour phenotype in the genus *Salamandra*, indicating a larviparous yellow-black spotted common ancestor. **A)** Evolutionary history of parity mode: blue = strict pueriparity; black = strict larviparity; red = both viviparity modes observed. **B)** Melanism seen as a common (though not always exclusive) phenotype: black = loss of pattern (either brown or black); yellow = yellow-black patterned. **C)** Evolution of gross colour pattern: black = no pattern; grey = spotted; yellow = striped (*Lyciasalamandra* has been coded as a fourth character state). The N, C and S after *S. s. gallaica* indicate the clades approximate geographic position in western Iberia (northern, central and southern respectively).

3.5. Discussion

In this study I present the most taxonomically comprehensive phylogenomic analyses of so-called 'true-salamanders' in the genus *Salamandra* to date. My data reveal previously unknown complexities of subspecific diversity and confirm the repeated, independent evolution of both colour phenotypes and reproductive modes. Most notably, two genetically distinct lineages were resolved within the most widespread species (*Salamandra salamandra*) and almost fully elucidate its intraspecific relationships, which have been challenging to resolve using mtDNA and nDNA markers (Veith et al. 1998; Steinfartz et al. 2000; García-París et al. 2003; Vences et al. 2014). However, while my data confirm the monophyly of all six currently recognised species, the relationships between them remain less certain.

3.5.1. Inter-species relationships within Salamandra

Only recently have the phylogenetic relationships between the six recognised species of *Salamandra* been considered robustly resolved (Rodríguez et al. 2017/Chapter 2). This required the analyses of multiple phylogenomic datasets, including RNA-Seq, RAD-Seq and full mitochondrial genome data. However, uncertainty remained over the placement of one species: *S. corsica*. While all analyses placed *S. algira* as sister to *S. salamandra*, *S. atra* as sister to *S. lanzai*, and *S. infraimmaculata* as sister to all other *Salamandra* species, *S. corsica* was alternately placed as either sister to the *S. atra* + *S. lanzai* clade (concatenated sequence analyses), or sister to a clade containing *S. algira* + *S. salamandra* and *S. atra* + *S. lanzai* (species tree analyses). This ambiguity in regard to *S. corsica* was also observed in the analyses presented in this chapter, based on broader sampling within species. Using optimised filtering parameters and Bayesian inference, *S. corsica* was recovered as sister to the Alpine *S. lanzai* + *S. atra* clade (Fig. 3.5). However, RAxML analyses either returned this relationship with weak support (89 individual data set; Appendix 4 Fig. A4.4) or placed *S. corsica* as sister to the *S. algira* + *S. salamandra* clade (231 individual data set; Fig. 3.3). This suggests a complex history of introgression or incomplete lineage sorting that may be difficult or impossible to resolve, even with thousands of molecular markers. Such relationships may be resolvable using coalescent-based phylogenetic and demographic modelling, and F_3, F_4, and Patterson's D (or ABBA-BABA) statistics can be used to assess levels of current or historical introgression (Eaton and Ree 2013; Rheindt et al. 2014; Hou et al. 2015; Skoglund et al. 2015; Barbato et al.

2017). However, such methods remain difficult to implement without whole genome sequence data (Nater et al. 2015), which is currently unavailable for *Salamandra*.

My analyses also highlight the importance of exploring the impact of data filtering parameters in RAD-Seq based phylogenomics. This is due to the fact that the amount of per locus missing data, and the maximum number of variable sites allowed per locus, had a strong effect on resulting phylogenetic reconstructions. This is less of a concern for missing data, as inter-species relationships stabilised when allowing at least 20% per locus missing data. Moreover, more conservative filtering (e.g. requiring a locus be present in all individuals) disproportionately excludes those loci with high mutation rates or low coverage, which negatively impacts downstream phylogenetic analyses (Huang and Knowles 2014; Eaton et al. 2017). However, less studied is the impact of the maximum number of variable sites allowed per locus. For example, recent phylogenomic studies on Phrynosomatid lizards (Leaché et al. 2015b), yucca moths (Darwell et al. 2016) and tunas (Díaz-Arce et al. 2016) all explored the effect of missing data on *de novo* assembled RAD-Seq data, but either did not consider per locus variable site number or do not report it. This omission is important, as our analyses returned different topologies when allowing for different numbers of SNPs per locus. While our comparative phylogenetic analyses (Rodríguez et al. 2017/Chapter 2) give us confidence that the parameter settings chosen here have returned the best supported inter-species topology (Fig. 3.5), the impact of 'SNPs per locus' on RAD-Seq based phylogenetic reconstructions is an area warranting further investigation.

3.5.2 Diversity and relationships within subspecies

While some ambiguity remains over the inter-species relationships within *Salamandra*, my analyses do confirm the monophyly of all six currently recognised species. However, they also indicate inconsistencies between phylogenetic clusters and subspecies groupings within *S. algira* and *S. salamandra*. This would suggest that some taxonomic designations are inconsistent with a phylogenetic species/subspecies concept (Nixon and Wheeler 1990), and may require some reassessment.

3.5.2.1. Taxonomic implications for Salamandra algira

Within *S. algira*, previous studies have identified four distinct groups based mainly on mtDNA, phenotypic and distributional data (Steinfartz et al. 2000; Donaire and Bogaerts

2003; Escoriza and Del Mar Comas 2007). These correspond to the subspecies *S. a. tingitana* in Northern Morocco; *S. a. spledens* from the Moroccan central Rif mountains; *S. a. spelaea* from eastern Morocco/western Algeria; and *S. a. algira*, found in both eastern Algeria and northern Morocco. However, my analyses indicate that *S. a. algira* as currently recognised (e.g. Escoriza, 2016) is paraphyletic, with salamanders from the Moroccan Mid-Atlas mountains being genetically differentiated from those in eastern Algeria. In addition, despite its geographic proximity to *S. a. tingitana* and *S. a. spledens*, the Mid-Atlas population did not cluster with them, but instead came out as sister to the eastern clade containing *S. a. spelaea* and the Algerian *S. a. algira*. Given that it is as genetically distinctive as other currently recognised subspecies (Fig. 3.5; Appendix 4 Table A4.2), and the mismatch seen between phylogenetic clustering and geographic distance, the Moroccan Mid-Atlas population may warrant promotion to subspecies level. Otherwise, *S. a. spelaea* would need to be collapsed into *S. a. algira* in order to meet a criterion of monophyly.

Some authors have also suggested that certain *S. algira* subspecies be promoted to species level rank. For example, Dubois and Raffaëlli (2009) proposed splitting *S. algira* into three species: *Salamandra tingitana, Salamandra algira* (containing the subspecies *S. a. algira* and *S. a. spelaea*), and a final unnamed species, which would now correspond to *Salamandra spledens* (the subspecies having been named in 2013; Beukema et al. 2013). While this proposal has some support from the perspective of life history (e.g. parity mode) and molecular data, and is broadly congruent with the phylogenetic clusters in my analyses, it relied on published studies where the authors questioned the ability of their data to delimit species, and has therefore not been widely accepted (see Moali et al. 2015 and Merabet et al. 2016). The taxonomic status of *S. algira* will be the focus of future publications, combining phylogenomics with environmental niche modelling (Dinis, et al. *in prep*).

3.5.2.2. Taxonomic implications for Salamandra salamandra

Salamandra salamandra poses a more complex taxonomic challenge than *S. algira*, as it contains the greatest subspecific diversity and geographic spread in the genus (Thorn and Raffaëlli 2001; García-París et al. 2003). My data suggest two main clades within *S. salamandra*, with one comprising the subspecies *S. s. fastuosa, S. s. bernardezi,* and *S. s. gigliolii*, and the other all remaining subspecies: *S. s. almanzoris, S. s. morenica, S. s. crespoi, S. s. longirostris, S. s. bejarae, S. s. gallaica, S. s. hispanica, S. s. terrestris, S. s. werneri, S. s. beschkovi*, and *S. s. salamandra*. However, my analyses found that only eight

of these fourteen subspecies meet a criterion of phylogenetic monophyly (following Nixon and Wheeler 1990).

The most unexpected finding in the first of these two major clades was the paraphyly of *S. s. fastuosa* with respect to *S. s. gigliolii*, which makes these subspecies designations questionable. This is different to previous studies using mtDNA and nDNA markers, which placed *S. s. gigliolii* as sister to *S. s. fastuosa* (Steinfartz et al. 2000; García-París et al. 2003; Vences et al. 2014). The nested position of *S. s. gigliolii* within *S. s. fastuosa* in this study is particularly interesting given the discordance it shows with geographic (Appendix 4 Fig. A4.6) and phenotypic data. While *S. s. fastuosa* and *S. s. bernardezi* share a common striped colour pattern, display pueriparity and have abutting distributions in northern Spain, *S. s. gigliolii* is larviparous, yellow blotched/spotted and restricted to western Italy (Sparreboom and Arntzen 2014). Therefore, these relationships indicate a complex evolutionary and biogeographic history within this lineage.

Also unexpected was the placement of the *S. s. bernardezi* + *S. s. fastuosa* + *S. s. gigliolii* clade as sister to all other *S. salamandra* subspecies. This relationship has not been found in any previous molecular analyses, largely due to a lack of phylogenetic resolution when using mtDNA and nDNA markers (Veith et al. 1998; Steinfartz et al. 2000; García-París et al. 2003; Vences et al. 2014). However, while their strong phylogenetic separation and characteristic (though not fixed) differences in reproductive mode and colouration (Fig. 3.7) suggest these two clades may warrant recognition as separate species, sequence divergence between them (~1.9%) falls within the range of intraspecific variation seen within the larger of the two (1.5–2.2%; Appendix 4 Table A4.4).

Within the other major *S. salamandra* clade, the subspecies sister to all other groups is the geographically restricted *S. s. almanzoris*. Occasionally, it has been proposed that this central Spanish subspecies be elevated to species level rank (e.g. Dubois and Raffaëlli, 2009) due to its morphological distinctiveness, extremely long larval period (>1yr; Guerrero et al. 1990) and genetic differentiation. However, this has been disputed (see Speybroeck et al. 2010) because allozyme and genetic differentiation between *S. s. almanzoris* and geographically proximate *S. s. bejarae* fall within accepted intraspecific limits (Martínez-Solano et al. 2005). A similar result was found in this study, with sequence divergence between the two subspecies being just 1.1%, one of the lowest seen (Appendix 4 Table 4.4). Previous phylogenetic analyses using mtDNA have also placed *S. s. almanzoris* within the larger *S. salamandra* clade (García-París et al. 2003; Ainhoa

Iraola and Garcia-Paris 2004), and hybridization has been noted between *S. s. almanzoris* and neighbouring *S. s. bejarae* populations (Martínez-Solano et al. 2005; Pereira et al. 2016). As I also found *S. s. almanzoris* to fall within the wider *S. salamandra* clade (Fig. 3.4; Fig. 3.5), I find little support for recognising it as a distinct and separate species.

The other subspecies often considered a candidate for promotion to species-level status by some authors is the southern Spanish *S. s. longirostris*. For example, Dubois and Raffaëlli (2009) suggested this based on distributional, morphological and molecular grounds. While Speybroeck, Beukema and Crochet (2010) also refute this promotion, they consider it more likely than Dubois and Raffaëlli's (2009) other taxonomic suggestions, largely due to previous studies placing *S. s. longirostris* as sister to all other *S. salamandra*, either on its own or as sister to the *S. s. morenica* + *S. s. crespoi* clade (Steinfartz et al. 2000; García-París et al. 2003; Vences et al. 2014). My analyses support the placement of *S. s. longirostris* as sister to *S. s. morenica* + *S. s. crespoi*, but do not place this clade as sister to all other *S. salamandra* lineages. Therefore, recognising it as a separate species would result in *S. salamandra* being paraphyletic.

The status of the subspecies *S. s. bejarae* is less clear. It formed a monophyletic grouping sister to *S. s. gallaica, S. s. hispanica, S. s. terrestris, S. s. werneri, S. s. beschkovi, and S. s. salamandra*. However, several samples originally attributed to this subspecies were actually collected within the range of *S. s. gallaica* (and clustered with *S. s. gallaica* in phylogenetic analyses). This resulted in only one population of 'true' *S. s. bejarae* (from the approximate type locality) being included in my phylogenetic analyses. Given its wide range in central and northern Spain, greater sampling is required to confirm the monophyly of *S. s. bejarae* as currently recognised.

The phylogenetic finding with the greatest taxonomic impact in this study is the paraphyly of *S. s. gallaica*. While this has been previously indicated (e.g. García-París *et al.*, 2003), we confirm that *S. s. gallaica* does not meet a criterion of monophyly, instead constituting at least three genetically distinct clusters, broadly corresponding to southern Portugal, central Portugal, and northern Portugal/Spain. The position of '*S. s. gallaica*' as sister to *S. s. hispanica*, *S. s. terrestris*, *S. s. werneri*, *S. s. beschkovi* and *S. s. salamandra* therefore calls into question the validity of these subspecies designations. In addition, principal component analyses showed no strong separation between these lineages (with the exception of *S. s. gallaica*; Fig. 3.7), *S. s. werneri* was found to be paraphyletic in phylogenetic analyses, and the position of *S. s. terrestris* was topologically unstable.

Questioning the subspecies distinctiveness of *S. s. hispanica*, *S. s. terrestris*, *S. s. werneri*, *S. s. beschkovi*, and *S. s. salamandra* is also supported by phenotypic data. The primary phenotypic difference between *S. s. salamandra* and *S. s. werneri* is the presence of red spots in the later (Speybroeck et al. 2016), and while *S. s. beschkovi* is slightly smaller and stockier than *S. s. salamandra*, it has a highly restricted distribution in the Bulgarian Pirin Mountains (Obst 1981). As such, the validity of both *S. s. werneri* and *S. s. beschkovi* have already been questioned (e.g. Speybroeck et al. 2016). In addition, the main phenotypic difference between *S. s. salamandra* and *S. s. terrestris* is the orientation of their yellow spots/blotches, which are irregular in *S. s. salamandra* and typically form two discontinuous dorso-lateral striped in *S. s. terrestris* (Sparreboom and Arntzen 2014; Speybroeck et al. 2016). Finally, a lack of phenotypic difference has also resulted in some authors regarding *S. s. hispanica* as indistinct from *S. s. terrestris* (e.g. Velo-Antón and Buckley 2015; Pereira et al. 2016; Speybroeck et al. 2016). Combined, this suggests that these five subspecies—*S. s. gallaica*, *S. s. hispanica*, *S. s. terrestris*, *S. s. werneri*, *S. s. beschkovi*, and *S. s. salamandra*—likely constitute a single clade ranging across most of mainland Europe that is divided along regional differences in minor phenotypic traits.

This incongruence between *S. salamandra* subspecies descriptions and phylogenetic clades is not entirely unexpected. While many have been delimited based on traditional subspecies concepts, like discontinuous geographical distributions and phenotypic traits (following Mayr and Ashlock 1991), recent studies in other taxonomic groups have revealed such data to be unreliable. For example, Phillimore and Owens (2006) found that ~64% of avian subspecies failed to form phylogenetically distinct lineages, and a phylogenomic analysis of *Storeria* snakes by Pyron et al. (2016) found a lack of support for subspecies described using distributional and colour pattern data. However, while phylogenomic analyses call into question many *Salamandra* subspecies, when combined with ancestral state reconstruction (ASR) analyses they confirm the independent evolution of both reproductive and colour pattern phenotypes.

3.5.3. The evolutionary history of pueriparity in Salamandra

While viviparity has evolved in all three orders of amphibians, within caudates, it is only found in the sister genera *Salamandra* and *Lyciasalamandra* (Buckley 2012; Wake 2015). However, while all *Lyciasalamandra* species give birth to fully formed terrestrial juveniles (pueriparity), *Salamandra* spp. typically deposited aquatic larvae (larviparity), with pueriparity only found in certain lineages (Buckley 2012). Through a combination of

phylogenomic and ASR analyses, I have identified at least four independent transitions from larviparity to pueriparity in *Salamandra*. These occurred once in the common ancestor of *S. lanzai* and *S. atra*, once in *S. a. tingitana*, once in *S. s. gallaica* and once in the *S. s. bernardezi* + *S. s. fastuosa* clade (Fig. 3.7); whether *S. s. gigliolii*, which was found to phylogenetically nest within *S. s. fastuosa*, represents a transition back to larviparity, or *S. s. fastuosa* and *S. s. bernardezi* represent independent transitions to pueriparity, remains unclear. However, determining whether these transitions to pueriparity constitute cases of parallel or convergent evolution (i.e. the same or different phenotypic and/or molecular routes to a similar contemporary trait; reviewed in Elmer and Meyer 2011) will require detailed phenotypic and genetic characterisation.

Unfortunately, the genetic basis of pueriparity *Salamandra* is unknown, although ecological genomics advances provide the possibility to study this in wild populations (Elmer and Meyer 2011). However, detailed reproductive and developmental information is available for two pueriparous lineages: *S. atra* and *S. s. fastuosa*/*S. s. bernardezi*. In *S. atra*, two offspring are retained for 3–4 years gestation, being fed successively on yolk provision, unfertilized eggs (oophagy) and enlarged maternal epithelial uterine cells (epitheliophagy); this last stage is associated with specialised tooth structures (Greven and Guex 1994; Greven 1998). The reproductive biology of *S. lanzai* is unclear but thought to be similar to that of *S. atra* (Miaud et al. 2001; Buckley 2012). In contrast, *S. s. bernardezi* and *S. s. fastuosa* typically give birth to 1–15 fully metamorphosed juveniles after 80–90 days of gestation, having been fed on yolk provision, unfertilized eggs, and smaller siblings (adelphophagy; Buckley et al. 2007). As they present highly different phenotypes, these constitute a convergence on pueriparity through different evolutionary trajectories. Whether the evolution of pueriparity in some populations of *S. a. tingitana* (Beukema et al. 2010) and two insular island populations of *S. s. gallaica* (Velo-Antón et al. 2007, 2012) display greater parallelism with the *S. s. bernardezi/S. s. fastuosa* phenotype is unclear, as their exact biology is not fully understood. However, the two island populations of *S. s. gallaica* are particularly interesting, as they became isolated from a larviperous ancestor by rising sea levels within the last 9000 years and may each represent an independent, recent transition to pueriparity (Velo-Antón et al. 2012).

3.5.4. The evolutionary history of Salamandra colour patterns

Salamandra are renowned for their chromatic diversity, which shows striking inter- and intraspecific variation (Thiesmeier 2004; Sparreboom and Arntzen 2014; Velo-Antón and

Buckley 2015). In my study, the most recent common ancestor of all *Salamandra* species was predicted to have had a black dorsum covered with yellow spots, which is still the typical pattern for *S. infraimmaculata*, *S. corsica*, *S. algira* and *S. salamandra*. However, the repeated, independent evolution of two derived colour phenotypes was also confirmed. One of these is the evolution of a striped pattern in two lineages: *S. s. fastuosa* + *S. s. bernardezi* and *S. s. terrestris*. Whether *S. s. gigliolii*, which nests within *S. s. fastuosa*, represents a loss of stripes, or *S. s. bernardezi* and *S. s. fastuosa* represent two independent evolutionary transitions to a striped phenotype, cannot currently be resolved. However, it is worth noting that *S. s. gigliolii* colour patterns display a high level of yellow pigmentation (Sparreboom and Arntzen 2014; Speybroeck et al. 2016), more closely resembling individuals from an atypically colour diverse lineage of *S. s. bernardezi* (Köhler and Steinfartz 2006; Beukema et al. 2016a) than the inferred ancestral phenotype.

Perhaps more interesting is the finding that melanism has evolved independently in at least five *Salamandra* lineages. It has been known for some time that the two melanic taxa, *S. atra atra* and *S. lanzai*, do not share a common melanic ancestor due to the phylogenetic placement of the subspecies *S. a. aurorae* and *S. a. pasubiensis*, which are both yellow-black patterned (Bonato and Steinfartz 2005). However, melanism is also found in certain populations of *S. algira tingitana* (Martínez-Solano et al. 2005) and *S. s. gallaica* (Velo-Antón and Buckley 2015), with the ASR analyses presented in this study confirming the independent evolutionary origins of these phenotypes. A similar brown (or hypolutic) phenotype also appears to have evolved independently in three lineages: *S. s. bernardezi* (Köhler and Steinfartz 2006), *S. a. atra* (Bonato and Steinfartz 2005) and *S. a. tingitana* (Martínez-Solano et al. 2005). While phenotypically different, both melanism and hypolutism involve the loss of the two colour producing cell types: yellow-red pigment containing xanthophores and light reflecting iridophores (Bonato and Steinfartz 2005). This leaves only dark pigment containing melanophores. As a result, it is likely that both brown and black skin are produced *via* similar structural and molecular pathways. The co-occurrence of melanic and hypolutic individuals in *S. a. atra* and *S. a. tingitana* is also indicative of a shared evolutionary origin. However, determining whether these melanic populations represent convergent or parallel phenotype evolution will require more detailed information the underlying genetic basis of amphibian colouration, which is currently poorly understood (Hoffman and Blouin 2000; Rudh and Qvarnström 2013).

3.5.5. A shared evolutionary origin of pueriparity and melanism in Salamandra?

An unexpected result from the ASR analyses in this thesis is an apparent correlation between pueriparity and melanism. These two characters display almost completely overlapping evolutionary histories, with all of the melanic and/or hypolutic *Salamandra* lineages also displaying pueriparity. While these characters are not necessarily associated with one another, given that *S. s. fastuosa* displays pueriparity but not melanism/hypolutism and melanism is only the dominant colour phenotype in the Alpine species, some authors have speculated that they may be derived from similar ecological drivers. For example, Beukema et al. (2010) hypothesised that the tendency of northern *S. a. tingitana* populations to be both pueriparous and hypolutic may be related to low precipitation and a lack of standing water resulting from the underlying karstic geology. A lack of water in karstic systems has also been hypothesised as a driver of viviparity in the *S. s. bernardezi* + *S. s. fastuosa* clade (García-París et al. 2003). This is notable, as *S. s. bernardezi* also contains hypolutic individuals (Köhler and Steinfartz 2006; Beukema et al. 2016a) and is almost fully pueriparous, unlike the more reproductively variable, yellow-black patterned *S. s. fastuosa* (Greven and Guex 1994; García-París et al. 2003). As such, melanic *Salamandra* could potentially represent a vertebrate example of the so-called melanisation-desiccation hypothesis.

The melanisation-desiccation hypothesis was originally formulated through studies on wild *Drosophila* spp. in India, which found a correlation between melanisation and desiccation resistance (Rajpurohit et al. 2008, 2013; Rajpurohit and Nedved 2013). This was supported by experimental evolution studies on *Drosophila melanogaster* (Ramniwas et al. 2013) and experiments on mosquito eggs (Farnesi et al. 2017), although the hypothesis remains controversial (see Rajpurohit et al. 2016). Mechanistically, it is hypothesised that this is the result of hydrophobic melanin pigments thickening or decreasing the permeability of the cuticle (Rajpurohit et al. 2016). While the exact mechanism would likely be different in vertebrates, the melanin pigment responsible for this increased insect melanisation is eumelanin (Farnesi et al. 2017), the same one responsible for creating melanic (black and brown) colouration in vertebrates, including amphibians (Bagnara et al. 1978). Given that melanic and pueriparous phenotypes show a propensity to co-occur, and a lack of standing water has been hypothesised to be involved in the evolution of both phenotypes (García-París et al. 2003; Velo-Antón et al. 2007; Beukema et al. 2010), it is possible that they are adaptations to more xeric habitats than those occupied by yellow-black patterned salamanders. However, testing this is beyond the scope of the current study.

3.6. Conclusion

The phylogenetic analyses presented in this study, based on a comprehensive subspecies sampling and 187,080–294,300 nt of sequence, constitutes one of the most in-depth assessments of *Salamandra* diversity to date. Strong support was found for the monophyly of all currently recognised species: *Salamandra atra*, *Salamandra algira*, *Salamandra corsica*, *Salamandra infraimmaculata*, *Salamandra lanzai*, and *Salamandra salamandra*. I also find strong support for the sister placement of *S. algira* + *S. salamandra*, *S. atra* + *S. lanzai*, and *S. infraimmaculata* to all other species in the genus. While the topological position of *S. corsica* remains uncertain, I find it most likely sister to the *S. atra* + *S. lanzai* clade. However, the placement of an upper bound on the number of variable sites allowed per locus during data filtering had an unexpectedly large impact on downstream analyses, highlighting important considerations for RAD-Seq based phylogenetic reconstructions. Further to this, I also explored subspecific diversity in the two most diverse species: *S. algira* and *S. salamandra*. Within *S. algira*, I confirm the monopyly of three out of four currently recognised subspecies, but also identify previously undescribed diversity in the Moroccan Mid-Atlas Mountains. This makes the nominative subspecies (*S. algira algira*) paraphyletic, and perhaps warrants the description of a fifth *S. algira* subspecies. In contrast, within *S. salamandra*, I only find that eight of the 14 currently recognised subspecies meet a criterion of monophyly (*S. s. bernardezi*, *S. s. fastuosa*, *S. s. almanzoris*, *S. s. crespoi*, *S. s. morenica*, *S. s. longirostris*, *S. s. bejarae* and *S. s. salamandra*), all others potentially representing over-splitting based on regional variations in phenotypic traits like colour pattern. I also confirm the independent evolution of phenotypic characters through a combination of phylogenomic and ancestral state reconstruction analyses. I find that a transition from larviparity to pueriparity has occurred at least four times across the genus, and can infer the independent evolution of a striped phenotype in two lineages and melanic colouration (black or brown) in five, all from a common yellow-black spotted ancestor. Most notably, I find a striking correlation in the evolution of pueriparity and melanism, potentially indicating shared adaptive ecological basis for these two complex, derived phenotypes, which warrants further study. Overall, the results presented here lay the groundwork for detailed studies on the parallel evolution of complex, adaptive phenotypes like reproductive mode and colour pattern within *Salamandra*. Further, through comparative study, the convergence of such phenotypes across vertebrate lineages will allow us to look at the repeatability and predictability of, as well as the constraints on, evolution by natural selection.

Chapter 4: Genetic associations of colour and pattern in the European fire salamander (*Salamandra salamandra*)

4.1. Abstract

Animal colouration is associated with a multitude of ecologically adaptive traits known to drive biological diversification, which broadly fall under aspects of: intraspecific communication, predator avoidance and physiological processes. However, despite the fact that amphibians present some of the most striking colour patterns among the terrestrial vertebrates, the molecular mechanisms underlying this variation have remained poorly understood. The European fire salamander (*Salamandra salamandra*) in particular displays a remarkable level of color variation, including: yellow, brown, black, and yellow-black patterned morphs. These patterns are hypothesized to be adaptive for either aposematism or thermoregulation, although this has not been robustly tested. Here, I leverage a highly colour-variable lineage from northern Spain (*S. s. bernardezi*) to identify genetic associations with colour using a suite of approaches, test for selection on colouration, and test the relationship between colour phenotype and toxicity (the functional basis of aposematism). Through this, I aim to, 1) resolve the evolutionary origins of sympatric *S. s. bernardezi* colour morphs, 2) assess the ecological drivers of colour diversification within *S. s. bernardezi*, and 3) identify the underlying genetic basis of these colour patterns. Through quantitative colour pattern analyses I show that, within a geographically restricted region, colour phenotypes form a gradient of variation from fully yellow to fully brown, through a yellow-black striped pattern. I also show that brown and black skin are near equivalent in both cellular structure and spectral reflectance. Population genetic analyses show no indications of assortative mating by colour phenotype and suggest a sympatric evolutionary origin for divergent colour morphs. In addition, I found no association between a salamanders colour pattern and the metabolomic profile of its toxic secretions, which calls into question the adaptive significance attributed to these striking phenotypes. Following this, using transcriptomic data (RNA-Seq) I identified 196 significantly differentially expressed genes between yellow, brown and black skin. Of these, 63 are known or suspected to be involved in animal colouration. For the remaining 133 genes, this study likely represents their first association to vertebrate pigmentation. Further,

through genotype-phenotype association analyses of RAD-Seq data I identified 43 genomic loci able to discriminate between colour pattern phenotypes. These results contribute greatly to our understanding of the molecular and evolutionary basis of amphibian pigmentation.

4.2. Introduction

Animal coloration provides an ideal system in which to study the evolutionary processes that generate and maintain biological diversity. This is because it is both conspicuously affected by natural selection (Caro 2005) and is quantifiable non-invasively (Endler 1990), which provides a non-lethal way to study genotype-phenotype interactions in ecologically adaptive traits (Hoekstra 2006). Typically, such traits fall into three broad functional categories: 1) intraspecific signalling, such as mate attraction and intra-sexual competition, 2) predator avoidance, and 3) physiological regulation (reviewed in Hubbard et al. 2010). However, the molecular, structural and ecological basis of animal pigmentation is highly variable, with some taxonomic groups showing greater levels of diversity than others.

While striking colour patterns abound throughout the tree of life, amphibians present some of the most diverse and complex of all vertebrates (Hoffman and Blouin 2000; Rudh and Qvarnström 2013). Generated by way of a variable cell structure, called the dermal chromatophore unit (Bagnara et al. 1968), amphibian colouration is associated with a multitude of ecologically and physiologically adaptive traits known to drive colour pattern evolution (Thayer 1909; Cott 1940). For example, numerous colour-based strategies for avoiding predation have evolved, such as crypsis (concealment) and aposematism (warning colouration that communicates toxicity or unpalatability). Thermoregulation, preventing water loss, and mate attraction are also important (Duellman and Trueb 1994; Rudh and Qvarnström 2013). However, to fully understand the evolution and persistence of an adaptive phenotype, or its convergent evolution, simply understanding its ecological function is not enough, we must also have detailed information on its underlying genetic basis (Orr 2005; Wiener and Wilkinson 2011; Roulin and Ducrest 2013).

Compared to other vertebrate taxa, little is known about the molecular mechanisms involved in amphibian skin pigmentation, especially in wild populations (Hoffman and Blouin 2000; Hubbard et al. 2010; Rudh and Qvarnström 2013; Posso-Terranova and Andres 2017). To date, most studies have focused on either the inheritance of

polymorphisms (Hoffman and Blouin 2000; O'Neill and Beard 2010) or the role of candidate genes identified in other vertebrate lineages, for example melanocortin 1 receptor (*Mc1r*; e.g. Alho et al. 2010; Posso-Terranova and Andres 2017) and tyrosinase (*TYR*; e.g. Woodcock et al. 2017). The lack of robust molecular studies on amphibian colouration could result from the difficult captive care required by many species, which makes laboratory based studies and/or experimental crosses impractical (Rudh and Qvarnström 2013). However, the emergence of new 'omic technologies may change this, as they are allowing us to study colour genetics in wild, non-model organisms.

A combination of whole genome sequencing (WGS) and restriction site associated DNA sequencing (RAD-Seq) was used to identify mutations in the transporter protein *SLC45A2* of white Bengal tigers (*Panthera tigris tigris*) as the molecular basis of their rare phenotype (Xu et al. 2013). A different technique, transcriptome profiling, was used in cheetahs (*Acinonyx jubatus jubatus*) and domestic cats (*Felis catus*) to show that mutations in *Taqpep* and differential expression of *Edn3* results in increased blotching and striping (the so-called king or tabby phenotype; Kaelin et al. 2012). Similarly, Mallarino et al. (2016) used transcriptomics, in the form of RNA-Seq, to look at colour pattern formation in rodents. They discovered that the well-characterised transcription factor *ALX3* plays a vital, and previously unknown, role in stripe formation and melanin production. RNA-Seq has also been used to look at the gene expression differences underlying colour variable corvids (within and between individuals and species; Poelstra et al. 2015) and Midas cichlids (*Amphilophus citrinellus*; Henning et al. 2013), and RAD-Seq has been used to identify candidate colour loci in polymorphic *Cepaea* land snails (Richards et al. 2013). Such studies have been slow to take place with amphibian taxa, in part due to the large size of their genomes compared to other vertebrates (Gregory 2017; Montero-Mendieta et al. 2017). However, with improving technology and bioinformatics techniques, this barrier is being eroded.

In this study, I use 'omic techniques to investigate genotype-phenotype associations in a colour variable lineage of the European fire salamander (*Salamandra salamandra*). This species displays a remarkable level of intraspecific colour variation, including: yellow, brown, black, and yellow-black spotted or striped patterns (Sparreboom and Arntzen 2014; Beukema et al. 2016b); these patterns are thought to be adaptive for both aposematism and thermoregulation, with the parallel evolution of melanic phenotypes hypothesised (Vences et al. 2014; Beukema et al. 2016b; Rodríguez et al. 2017). Of particular note is a colour diverse lineage in Asturias, northern Spain, which shows a striking level of intra-

population variation. Previously described as a separate subspecies (*S. s. alfredschmidti*; Köhler and Steinfartz 2006), but now recognised to be a part of the more widespread *S. s. bernardezi* (Beukema et al. 2016a), this lineage shows a gradation of phenotypes between yellow-black striped (ancestral), fully yellow (xanthic), and fully brown (hypolutic), with data from previous studies using only a few loci indicating no genetic differences between colour morphs (Robinson 2014; Beukema et al. 2016a). This level of sympatric diversity is not seen in other populations, enabling a unique assessment of the evolutionary genetic drivers of such colour variation.

Here, I leverage this admixture zone to uncover the genetic basis of different colour phenotypes and the evolutionary processes giving rise to them. To better understand the phenotypic diversity within the region I apply quantitative colour analyses, and use gene expression data (RNA-Seq) to identify genes and molecular pathways associated with yellow, black and brown *Salamandra* skin. In order to investigate assortative mating by colour phenotype (and potential signs of incipient speciation associated with colour) I conduct population genomic analyses for the colour-variable population. Following this, I utilise machine-learning algorithms and outlier detection methods to conduct genotype-phenotype association studies, thereby allowing for the prediction of genomic loci associated with different colour phenotypes, and use a range of approaches to identify signals of selection on colour pattern. Finally, I investigate the supposed aposematic function of this colouration by profiling the metabolomic content of these salamanders' toxic secretions.

4.3. Methods

4.3.1. Field sampling

Adult *S. s. bernardezi* were sampled from seven localities in Asturias (northern Spain) during fieldwork in 2013, 2014 and 2015 (n=781; Fig 4.1). Sample localities ranged in altitude from 78–1312 meters above sea level (masl), with striped morphs found across this range and hypolutic (brown) and xanthic (yellow) morphs only found below 680 masl. All salamanders were photographed (dorsal and ventral views) and toe or tail clips taken to create a tissue library. Salamanders were released where captured, except those fatally sampled for microscopy and gene expression studies (n=15).

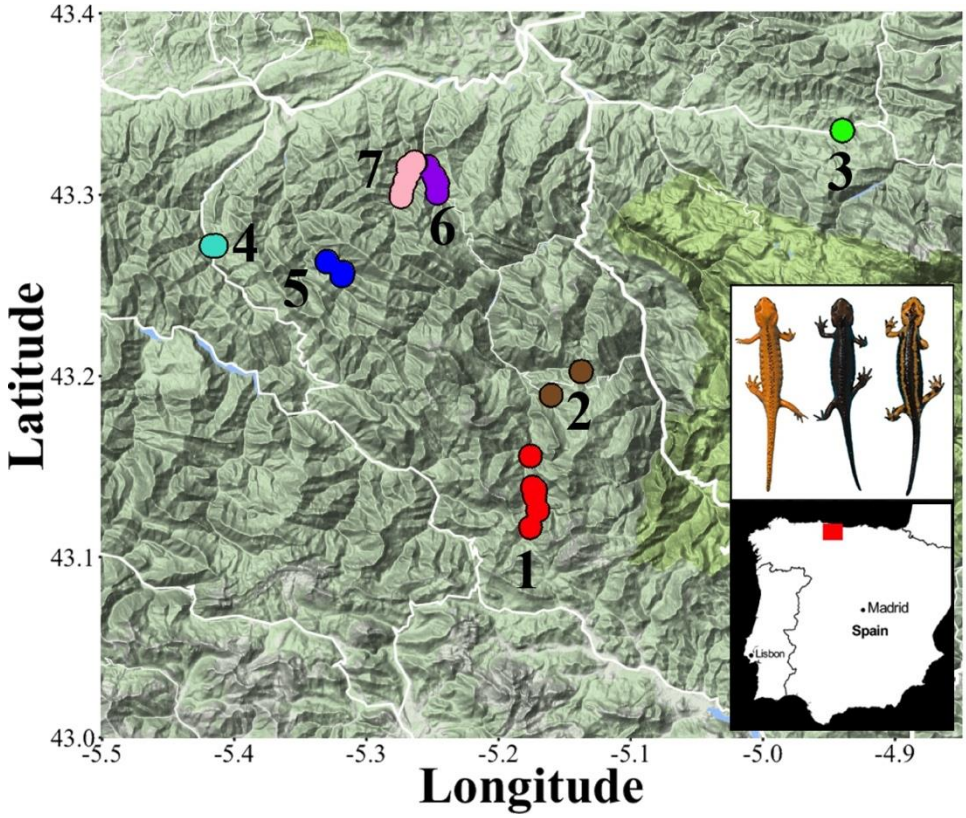


Figure 4.1: Sample site locations in Asturias, Northern Spain: 1 (red) Ventaniella-Soberfoz; 2 (brown) San Juan de Beleño; 3 (green) Avín; 4 (turquoise) La Marea; 5 (blue) Rio del Infierno; 6 (purple) Tendi valley; and 7 (pink) Color valley. Sites 1–3 are monomorphic ('striped-only'); sites 4–7 are polymorphic. Inset: the three representative colour morphs (top) and the location of the sample region (red square) in Spain.

4.3.2. Digital photography

Dorsal images of salamanders were taken in the field using a standardised set-up to control for lighting. Images were taken in RAW format using a DSLR camera, manually controlling for aperture (F8), ISO (800) and exposure (though adjustment of the shutter speed). An X-Rite ColorChecker Passport standard was also included to allow for post-hoc corrections in Adobe Photoshop CS6 in order to standardize images as much as possible: a custom camera profile (generated using X-Rite's DNG ProfileManager) was applied, the white balance of each image was set using the white square of the X-Rite ColorChecker Passport, and the exposure was modified until the RGB (red, green and blue) values of the white ColorChecker square in each image equalled 240:240:240 (±5).

4.3.3. Colour pattern analysis

Dorsal colour patterns were analysed in R using the *Patternize* package, using a modified version of the Galápagos wolf spiders example protocol (R Core Team 2013; Van Belleghem et al. 2017). In order to accurately analyse dorsal colour patterns, only clear images of adult salamanders without obvious body contortions were used. This produced a data set of 277 images, broadly corresponding to 61 striped, 115 xanthic, and 44 hypolutic salamanders from a colour polymorphic population (sample site 7; Fig 4.1) and 57 salamanders from a striped-only monomorphic population (sample site 1; Fig 4.1). The number of females to males per colour group was 1.3, 1.3, 1.6 and 1.0 respectively.

Due to restrictions in the *Patternize* protocol, images first had to be cropped in Photoshop so that only the salamander (minus the legs) remained on a white background. JPEG formatted images were then imported into R, and image registration and colour pattern extraction carried out using the patRegK function. Briefly, all images were transformed into a single coordinate system by designating one as a target, and then colour pattern boundaries were identified using a k-means clustering algorithm that assigns each pixel to a predefined number (k) of RGB clusters; for this analysis the value of k was set to two, as this was the maximum number of discrete colour elements on an individual salamander. The RGB values of these clusters are the averaged RGB values of all the pixels they contain. Clusters were checked visually to make sure that they did not substantially deviate from the original pattern; where noticeably incorrect (due to shading or glare), images were excluded.

The patPCA function was then used to conduct principal component analyses (PCAs) to look at differences between: 1) striped, xanthic and hypolutic salamanders within the polymorphic population; 2) male and female salamanders within the polymorphic population; and, 3) striped salamanders in the polymorphic and monomorphic populations. PC scores were compared *via* one-way ANOVA, and PC1 scores extracted to use as a measure of colour phenotype in morphometric and genetic analyses.

4.3.4. Morphometrics

To determine if there were morphological differences between *S. s. bernardezi* colour morphs, linear measures for six skeletal morphological characters were taken (to the nearest mm) from JPEG images of salamanders using ImageJ (Schneider et al. 2012).

These were the inter-orbital distance, orbital-to-nostril, snout-to-vent length (SVL), tail, and the upper fore- and hindlimb (Fig. 4.2). Measurements were taken from the same images used for *Patternize* analyses, after excluding some for which all measurements could not be obtained. This resulted in a data set of 58 striped, 108 xanthic, and 40 hypolutic salamanders from within the polymorphic population and 56 striped individuals from the monomorphic population; the number of females to males per colour group was 1.2, 1.2, 1.5 and 1.4 respectively. Within the colour polymorphic population, one-way MANOVA analyses were used to look at the association between these morphologies and colour pattern (the explanatory variable): colour patterns were coded as either one of three discrete characters (striped, xanthic and hypolutic) or as a continuous variable using *Patternize* PC1 scores. One-way MANOVA was also used to look at the association between these morphologies and biological sex (within the polymorphic population), and to look for differences between striped salamanders in the polymorphic and monomorphic populations (sample site being the explanatory variable).

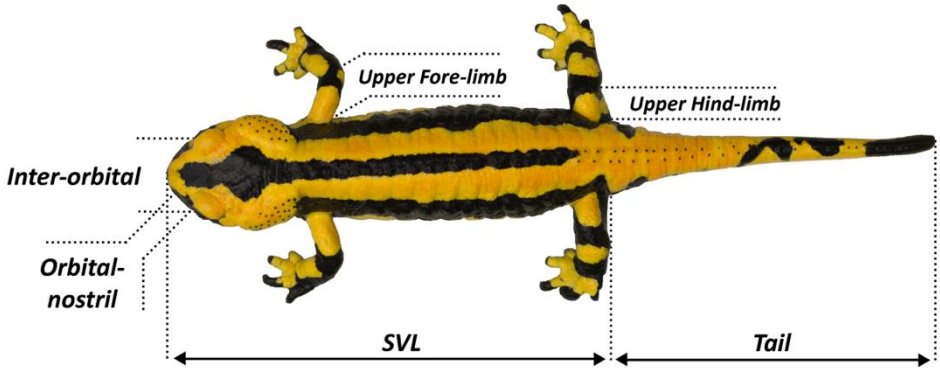


Figure 4.2: Schematic showing the measures taken for morphometric analyses.

4.3.5. Reflectance spectrophotometry

To quantify non-visible aspects of *S. s. bernardezi* colour patterns, three independent reflectance spectrophotometry readings were taken from seven body landmarks on 24 salamanders (Fig. 4.3): 10 hypolutic, 10 xanthic and four striped; the number of females to males per colour group was 1.5, 0.43 and 3 respectively. Readings were taken using an Ocean Optics USB2000 Miniature Fiber Optic Spectrometer with a DHU 2000 Deuterium Tungsten Halogen Light Source, an Ocean Optics ZR100UUV/SR seven fibre probe and

the Ocean Optics SpectraSuit software. A custom probe sheath was made to allow measurements to be taken from a distance of 1 mm at an angle of 90° while excluding background light, and a white Spectralon® diffuse reflectance standard was used to take reference white and dark measurements.

Reflectance spectra were processed using the R package *pavo* (Maia et al. 2013). Spectra were first imported and truncated to between 300nm and 750nm, which covers ultraviolet (UV; <400nm), visual (400-700nm) and near infrared (NIR; >700nm) wavelengths of light. Spectra were explored visually, and the three readings per landmark on a single individual averaged. A noticeable peak between 654nm and 660nm, an artefact created by the lamp, was manually removed. Values corresponding to hue (*pavo* variable H1; traditional colour), saturation (*pavo* variable S8; Chroma/intensity) and brightness (*pavo* variable B2) were extracted—the three primary metrics used with spectral data (Maia et al. 2013)—and compared to colour pattern morph (the explanatory variable) *via* one-way MANOVA for each landmark.

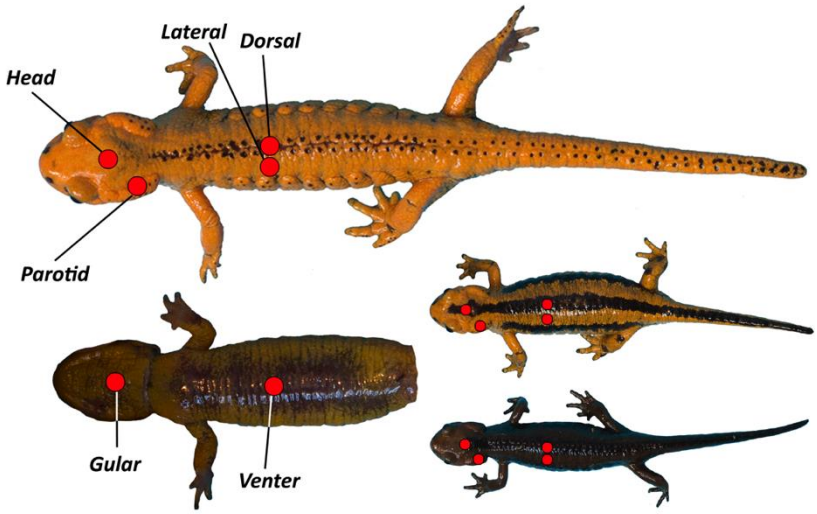


Figure 4.3: Location of spectrophotometry reading landmarks (red points).

To look specifically at spectral differences between skin colours, 17 black (dorsal), 14 yellow (lateral) and nine brown (lateral) individual-averaged spectra were subset. Values for hue, saturation and brightness were extracted across full spectra (300-750), UV (360–400nm), visible (400–700nm) and NIR (700–750nm) wavelengths; the lower end of animal UV visual sensitivity is 360nm (Honkavaara et al. 2002). These values were then compared to skin colour (the explanatory variable) *via* one-way MANOVA.

4.3.6. Transmission electron microscopy (TEM)

Amphibian skin contains up to five kinds of colour producing cells called chromatophores. These are the yellow-red pigment containing xanthophores and erythrophores, the light reflecting iridophores and leucophores, and the dark melanin pigment containing melanophores (Browder 1968; Lyerla and Jameson 1968; Bagnara et al. 2007). However, the distinction between both xanthophores and erythrophores and iridophores and leucophores is vague and not always recognised (see Rudh and Qvarnström 2013). To examine which of these cell types are present in different *Salamandra* colour patterns, two dorsal skin samples were dissected from two hypolutic, two xanthic and two striped salamanders—six individuals and 12 skin landmarks in total (Fig. 4.4); salamanders were euthanized using a 2g/L solution of MS-222 followed by decapitation. This allowed for a comparison of cell structure both within and between individuals.

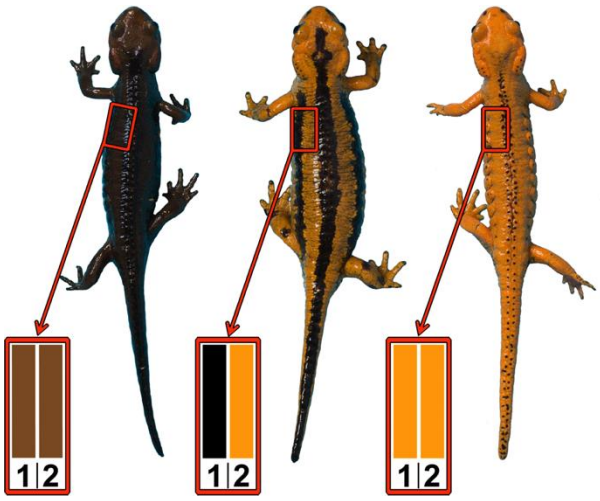


Figure 4.4: Inter- and intra-individual tissue sampling scheme for TEM and gene expression analyses.

Sample processing followed established protocols (e.g. Graham and Orenstein 2007). First, skin samples were fixed for two hours using 2.5% gluteraldehyde in 0.1M sodium cocodylate buffer, before rinsing and storage in pure 0.1M sodium cocodylate buffer at 4°C. Samples were then postfixed in 1% osmium tetroxide, stained in 0.5% aqueous uranyl acetate, dehydrated in graded concentrations of ethanol, and embedded in Araldite epoxy resin. These resin blocks were cut into semi-thin (0.5-1μm) and ultra-thin (60–70nm)

sections using an ultratome in the University of Glasgow's Electron Microscopy Unit. Semi-thin sections were stained with 0.5% toluidine blue for light microscopy and ultra-thin sections with 0.5% aqueous uranyl acetate and lead citrate for TEM. Samples were examined using both a compound light microscope and a FEI Tecnai T20 TEM (University of Glasgow, Kelvin Nanocharacterisation Centre). Extensive TEM imaging was only carried out on one skin sample from a xanthic individual (yellow), one from a hypolutic individual (brown), and both samples from a striped individual (one yellow and one black).

4.3.7. Metabolomics

As *Salamandra* colour patterns are hypothesised to be aposematic, I investigated the relationship between a salamander's colour phenotype and the metabolomic content of its toxic secretions. For this, the metabolomic profiles of *S. s. bernardezi* toxic secretions were examined using gas chromatography–mass spectrometry (GC-MS). Due to the high viscosity of these secretions, a standard volume could not be collected. Instead, secretions were directly expressed into a pre-weighed sterile 1.5ml collection tube by gently squeezing a salamander's parotid gland. Secretions were collected for six striped, six xanthic and six hypolutic salamanders (from sample site 7), and six salamanders of a different subspecies sourced from central Germany (*S. s. terrestris*). The weight of each sample was then determined to the nearest 0.001g, and a ratio of 5mg secretion to 200µl of solvent used during metabolite extractions; the extraction solvent consisted of a 1:3:1 ratio of ice-cold chloroform, methanol and water.

For the extraction, the appropriate volume of solvent was added to each sample before vortexing at 4ºC for 5 minutes. They were then centrifuged at 13,000g for three minutes at 4ºC. Finally the supernatant (~90% of the solvent used) was moved to a fresh sterile tube and stored at -80ºC. GC-MS was conducted by Glasgow Polyomics, and known *Salamandra* alkaloids were identified by comparing mass spectra and gas chromatographic retention times with known data for *Salamandra* alkaloids (Appendix 5 Table A5.1; Mebs and Pogoda 2005).

The intensities (abundance) of identified peaks were investigated statistically in R. First, a PCA was carried out using the *mixOmics* package (Cao et al. 2017). Following this, individual one-way ANOVA comparisons were made for each identified peak (log-transforming data was required for some to meet the assumption of normality, see Appendix 5 Table A5.4) between: 1) striped, xanthic and hypolutic salamanders; 2)

ancestral (striped) and derived (xanthic and hypolutic) salamanders; and 3) between the two derived colour morphs (xanthic and hypolutic).

4.3.8. ddRAD sequencing and de novo assembly

Genomic DNA was extracted from salamander toes clips using the Macherey-Nagel NucleoSpin® Tissue kit (following manufacturer's instructions). This was done for 82 individuals collected from across the seven sample sites (Table 4.1; Fig. 4.1; see Appendix 1 Table A1.3 for collection data). Following this, a double digest Restriction site Associated DNA sequencing (ddRAD-Seq; Peterson et al. 2012) library was prepared as follows (per Recknagel et al. 2015, with modification of Illumina adapters): 1µg of DNA from each individual (and six technical replicates; 88 samples in total) was double-digested using the *PstI-HF*® and *AclI* restriction enzymes (New England Biolabs, Ipswich); modified Illumina adaptors with unique barcodes for each individual were ligated onto this fragmented DNA; samples were multiplexed; and a PippinPrep (Sage Science, Beverly) used to size select fragments around a tight selection of 383 bp (range: 350–416 bp) based on the fragment length distribution identified using a 2200 TapeStation instrument (Agilent Technologies, Santa Clara). Finally, enrichment PCR was performed to amplify the library using forward and reverse RAD primers (see Appendix 2 for detailed protocols). Sequencing was conducted on an Illumina NextSeq™ 500 platform (1.3 runs) at Glasgow Polyomics to generate paired-end reads 75 bp in length; c.4-6M raw reads (paired-end, PE) per individual was the target.

Table 4.1: The number of salamanders ddRAD-Seq genotyped from each sample site.

Site	Name	Striped	Xanthic	Hypolutic
1	Ventaniella-Soberfoz	8	-	-
2	San Juan de Beleño	4	-	-
3	Avín	4	-	-
4	La Marea	4	4	-
5	Rio del Infierno	5	5	-
6	Rio Tendi	6	9	8
7	Rio Color	8	8	9

Raw sequence reads were quality checked using FastQC v.0.11.3 (Andrews 2010). Samples were then de-multiplexed, Illumina adaptors and barcodes removed, and reads truncated to 60 nt using Stacks v.1.44 (Catchen et al. 2011) *process_radtags* pipeline. As sequencing errors on several samples (relating to the enzyme cut-site on the reverse read) resulted in ~50% of the data for ten samples being discarded, the *disable_rad_check* flag was used to recover all forward reads. The cut site on these reads was then checked using the FASTX-Toolkit (fastx_barcode_splitter.pl; Gordon and Hannon 2008). Final processed reads averaged 5.7M per individual (range: 2.2–14.5M; single end).

As there is no reference genome, reads were assembled *de novo* in Stacks v.1.44. The minimum number of identical raw reads required to create a stack was set to six (other settings left on default). Genotype and haplotype corrections in individual samples were then conducted using the Stacks *rxstacks* pipeline: loci for which 25% or more individuals had a confounded match to the catalogue were removed, excess haplotypes were pruned, sequencing errors were removed using the bounded model and SNPs recalled, and catalogue loci with an average log likelihood less than -10.0 were removed (Rochette and Catchen 2017). Following this, the catalogue of loci was rebuilt and each sample mapped against it. Finally, Stacks' *export_sql.pl* function was used to generate a list of all loci containing 1-2 SNPs for use in downstream analyses (a so-called whitelist of loci).

4.3.9. Genotyping errors

Technical replicates for six individuals were sequenced to assess genotyping error during sequencing and SNP calling. For this, samples were filtered through the Stacks *populations* pipeline as a single population, with whitelisted loci retained if they were present in ≥75% of samples, had a minimum individual stack depth of 10, a maximum observed heterozygosity of 0.5, and a minimum minor allele frequency of 0.05. Data for the 4839 retained SNP loci were exported in Variant Call Format (VCF; one SNP per locus) and converted into Phylip format using PGDSpider v2.1.0.3 (Lischer and Excoffier 2012); technical replicates of the same individual shared an average of 4363 loci (range: 4078–4478). Genotype error rates were assessed using custom scripts in R (see Appendix 6), and found to average 1.76% (range: 1.66–2.07%), comparable to those in published protocols (e.g. Recknagel et al. 2015). Finally, RAxML (GTRCAT with 1000 bootstrap replicates; Stamatakis 2014) was used to check that replicates clustered together in an analysis of the full 88 sample data set. Following this, the replicate with the most missing data was removed from further analyses.

4.3.10. Population genomics

In order to determine whether polymorphic sample sites represent an instance of secondary contact between ancestral and derived colour morphs, or sympatric diversification, I conducted population genomic analyses. First, data were re-filtered through Stacks *populations* as above with technical replicates removed. SNP loci (n=4702, one per RAD locus) were exported in GenePop, PLINK, STRUCTURE, VCF and Phylip (including variable sites encoded using IUPAC notation) formats. The average heterozygosity per individual and within sample sites was calculated using VCFtools (Danecek et al. 2011). PCAs were then conducted using the *dudi.pca* function in the *ade4* R package (Dray and Dufour 2007) and hierarchical cluster analysis using the *hclust* function in the *stats* package (both through the *adegenet* package; Jombart 2008; Jombart and Ahmed 2011). From this, two individuals were identified as outliers, removed and the data re-filtered.

Relationships among samples were further investigated using RAxML (GTRCAT with 1000 bootstrap replicates). Shared genetic ancestry was estimated by using STRUCTURE analyses (Pritchard et al. 2000) to test a range of genetic clusters (K) from 2–8 without population or phenotype priors, using an admixture model with correlated alleles and five iterations of 100,000 MCMC reps (after a burn-in of 10,000 generations). The best-fit of K was determined using STRUCTURE HARVESTER (Earl and VonHoldt 2012) to identify the one with the highest ΔK (following Evanno et al. 2005). Next, *Fst* among sample sites, and between colour morphs in sites 6 and 7, were calculated through pairwise differentiation tests in GenoDive (Meirmans and van Tienderen 2004), applying a *post hoc* Bonferroni correction for multiple testing to resulting p-values using the *p.adjust* function in the *stats* R package (R Core Team 2013). Finally, a Mantel test was used to assess isolation by distance between sample sites, using Edwards' genetic distance and Euclidean geographic distance (calculated from sample GPS coordinates).

4.3.11. Gene expression (transcriptomics)

Given the dearth of knowledge available on amphibian colour genetics, I used transcriptomic analyses to identify key genes associated with generating yellow, brown and black skin. For this, total RNA was extracted from ~25mg of fresh dorsal skin tissue using the PureLink™ RNA Mini Kit; this was followed by a lithium chloride precipitation at -20ºC to remove contaminants (see Appendix 2 for detailed protocols). Tissue sampling followed the same pattern as TEM sampling (though using different individuals; Fig. 4.4):

two skin samples were taken from three hypolutic, three xanthic and three striped salamanders—nine individuals and 18 skin landmarks in total. This allowed for a comparison of gene expression between different colours of skin both within and between individuals. Library preparation, using polyA selection (TruSeq stranded mRNA kit), and 75bp PE sequencing on the Illumina NextSeq™ 500 platform (RNA-Seq) was conducted at Glasgow Polyomics.

Raw read pairs averaged 24.85 million per sample (range: 20.17–29.94M), with an average of 89% retained after quality filtering (range: 86–95%) using Trimmomatic v.0.36 (Bolger et al. 2014). Raw reads were aligned to a unannotated *Salamandra* transcriptome assembly (44,165 transcripts; Rodríguez et al. 2017/Chapter 2) using TopHat2 v.2.1.1 (Kim et al. 2013), and had an average alignment rate of 46.54% (range: 44.4–48.9%) when using largely default settings but excluding mixed or discordant reads. Resulting SAM files were converted to BAM format using SAMtools v.1.3.1 (Li et al. 2009), and the hierarchically clustering of contigs for statistical testing carried out using Corset v.1.05 (Davidson and Oshlack 2014).

Differential gene expression analyses were carried out using the R package *DESeq2* (Love et al. 2014). Count data (from Corset) were imported and differential gene expression analysis (DESeq function) conducted; an adjusted p-value cut-off of 0.1 was used to identify significantly differentially expressed (sDE) transcripts. Pairwise comparisons were made for each colour combination (yellow-black, yellow-brown, and brown-black; accounting for skin landmark), between yellow and black skin in striped individuals only, and between the two skin landmarks across all individuals.

The sequence data for sDE transcripts were then extracted from the reference assembly, and genes identified through Blastx and Blastn searches against the NCBI nucleotide and protein databases (NCBI Resource Coordinators 2017). Gene functions and importance to animal colouration were identified through searching available literature and databases, particularly UniProtKB (The UniProt Consortium 2017), AmiGO 2 (Carbon et al. 2009), IMP (Integrative Multi-species Prediction; Wong et al. 2015) and the European Society for Pigment Cell Research (ESPCR) database (Montoliu et al. 2017). Further to this, GO (Gene Ontology; Ashburner et al. 2000; The Gene Ontology Consortium 2015) enrichment analyses were carried out to identify the biological processes, molecular functions and cellular components statistically overrepresented in the sDE genes upregulated in different colours of skin (PANTHER Overrepresentation Test; http://geneontology.org/).

4.3.12. Colour genomics

Data for the 80 ddRAD-Seq genotyped individuals used for population genetic analyses were refiltered as a single population through the Stacks *populations* pipeline (as above, without missing data constraints). Following this, VCFtools was used to remove loci that were out of Hardy-Weinberg equilibrium or present in <75% of samples, generating a data set of 2149 SNP loci. Missing data (13.4%) were imputed using fastPHASE v.1.4.8 (Scheet and Stephens 2006), a requirement for downstream random forest and LFMM analyses, and the output manually converted into PLINK MAP and PED format. PLINK (Purcell et al. 2007) was used to convert SNP loci into binary format. Genotypes were then corrected for population structure through linear regression with PC scores generated by a *dudi.pca* analysis of the data in the R package *adegenet*.

Phenotypes for the 80 genotyped salamanders were scored in two ways: as one of three discrete characters (striped, xanthic and hypolutic), and also using continuous PC1 scores from an analysis of dorsal colour patterns in *Patternize*. One striped salamander from sample site 4 (La Marea) was removed from further analyses as the dorsal image available for it clustered incorrectly during the *Patternize* analysis.

Random forest analyses were conducted in R using the *randomForest* package (Liaw and Wiener 2002). This is a machine learning technique that implements Breiman's random forest algorithm for classification and regression, which allows identification of covarying/interacting loci that differentiate between individuals based on a trait of interest. A full overview of random forest analysis can be found in Boulesteix et al. (2012). Briefly, three independent analyses of 100,000 trees were run using the corrected genotypes and *Patternize* PC1 scores as the phenotype input. The importance for each SNP marker was shown by the permuted importance statistic (%IncMSE), which indicates the ability of that marker to classify an individual by colour pattern when interacting with other loci; a Pearson correlation coefficient of ≥0.95 was used to assess convergence in the importance statistics between runs. The mean importance per locus across the three runs was then calculated and loci ranked by importance.

The top 101 loci—those with an importance of ≥5, roughly corresponding to the ‘upper end of the elbow’ following Laporte et al. (2016)—were subset and a backwards purging process used to determine the cohort that explained the highest amount of variance in the data. First, three random forest analyses (of 10,000 trees) were run for the subset of 101

SNPs and the mean r-squared value (variance explained) across runs calculated. The SNP of lowest importance (smallest %IncMSE) was then removed and the analysis repeated until the data set contained only two SNP loci. The iteration with the highest r-squared indicated the set of co-varying loci that were able to best discriminate between colour pattern morphs. This process was repeated using discrete phenotypes (striped, xanthic and hypolutic) instead of the *Patternize* PC1 scores.

Latent factor mixed models (LFMM; Frichot et al. 2013) analyses were conducted through the R package *LEA* (Frichot and François 2015), using the same corrected genotype and phenotype inputs as the random forest analyses. This package tests associations between loci and environmental (or phenotypic) gradients using an MCMC algorithm for regression analysis, in which the confounding variables are modelled with unobserved (latent) factors. The analysis was run 100 times, which increases the power to detect true associations, assuming one latent factor (K; as this is estimated, an exact value is not essential). Z-scores (the number of standard deviations above or below the population mean a data point is) from these runs were combined and the median value per locus calculated. The genomic inflation factor (λ) was then calculated and used to compute adjusted p-values. Finally, to correct for multiple testing, q-values were estimated for the adjusted p-values using the R package *qvalue*, with a significance cut off of 0.1 used to identify outlier loci.

Loci identified by both random forest and LFMM analyses were then identified, where possible. First, the 60 bp consensus reads for each locus were searched against the NCBI nucleotide (Blastn) and protein (Blastx) databases, a *Salamandra* reference transcriptome (Rodríguez et al. 2017) and the *Ambystoma mexicanum* transcriptome assembly v4.0 (Smith et al. 2005; Keinath et al. 2015); an E-value <0.1 was considered a 'good hit'. As much of the *A. mexicanum* assembly has not been fully annotated, any longer sequences the *Salamandra* short reads accurately matched against were themselves searched against the NCBI Blastn and Blastx databases. Gene function and importance to animal colouration were investigated as described for sDE transcripts (section *4.3.10*.).

Finally, tests were run to look for signals of selection on genomic loci. First, genome scans were carried out in OutFLANK, which identifies *Fst* outliers based on an inferred distribution of neutral *Fst* (Whitlock and Lotterhos 2014), using a left and right trim fraction of 0.05, a minimum per locus heterozygosity of 0.1 and a qvalue threshold of 0.1. This was done across all samples, and for pairwise comparisons between representative colour morphs and ancestral (striped) *vs.* derived (xanthic and hypolutic) salamanders. In a

more permissive and less statistically constrained approach, z-transformed *Fst* (Z*Fst*) was compered to genetic diversity (Δpi) in order to identify loci putatively under selection between colour morphs (following Axelsson et al. 2013; Cagan and Blass 2016). For this, per-site Weir and Cockerham's *Fst* and genetic diversity (nucleotide divergency) were calculated in VCFtools. However, to mitigate population effects, this was carried out using a sub-set of the data containing only individuals from the sympatric sample site 6 + 7 pair, which contained all three representative colour morphs and showed no genetic clustering by colour. Those loci with a Z*Fst* ≥3 standard deviations form the mean were considered putatively under selection. Finally, a STRUCTURE analysis was carried out using just those loci identified as putatively under selection across the three pairwise colour morph comparisons based on Z*Fst*.

4.4. Results

4.4.1. Colour diversity

To quantify colour diversity within *S. s. bernardezi*, I conducted digital image analysis using the R package *Patternize* (Van Belleghem et al. 2017). To compare the widespread (ancestral) black and yellow striped colour pattern with the derived colour patterns, I focused on individuals from two sample sites: site 1 (Fig. 4.1), comprising individuals exclusively of the ancestral striped pattern (typical *S. s. bernardezi*; 797–1312 masl; n=57); and site 7 (Rio Color; typical '*S. s. alfredschmidti'*; 172–414 masl; n=220), which represents a contact zone between all colour morphs. I found no significant difference in colour patterns when comparing striped individuals from site 1 to striped individuals in Rio Color (Appendix 5: Results A5.1.1, Fig. A5.1). Within the Rio Color sample site, the three representative colour morphs (striped, yellow and brown) did not constitute discrete polymorphisms. While *Patternize* PC1 scores—which represented an axis of variation from fully yellow at one end, fully brown at the other, and black-yellow striped in the middle—significantly differed between colour morphs (one-way ANOVA: $F_{(2,217)}= 429.8$; $P<0.005$), a great deal of overlap was seen (Fig 4.5). This was particularly true of 'striped' and 'xanthic' individuals, due to a high degree of variability in the extent of lateral striping. I found no significant differences in colour patterns between males and females (Appendix 5: Results A5.1.1, Fig. A5.1).

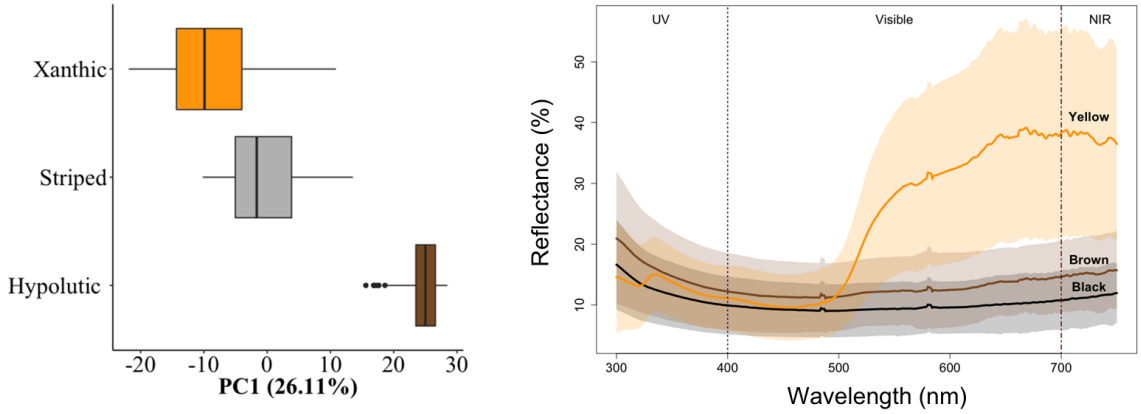


Figure 4.5: Colour phenotype characterisation. **Left:** *Patternize* PC1 scores for salamanders in site 7 (Rio Color) when categorized by pre-assigned colour morph (n = 220). **Right:** Average reflectance spectra of yellow (n = 14), brown (n = 9) and black (n = 17) skin across individuals in different wavelengths of light (shading = standard deviation).

Patternize analyses were also used to elucidate associations between colour phenotype and morphology within the contact zone. When phenotypes were categorically scored as striped (n=58), xanthic (n=108) or hypolutic (n=40), small but significant differences were seen in two of six morphological measures: the orbital-nostril distance (one-way ANOVA: $F_{(2,203)}$=8.29; P<0.005) and snout-to-vent length ($F_{(2,203)}$=3.83; P=0.023; Appendix 5: Results A5.1.2). However, when individual *Patternize* PC1 scores were used these differences were non-significant (one-way MANOVA: Pillai=0.05; $F_{(1,204)}$=1.59; P=0.15), revealing a lack of morphological separation by colour phenotype.

As both *Salamandra* (Przyrembel et al. 1995) and many of their potential predators, like birds (Lind et al. 2013), are UV sensitive, it was important to also quantify non-visible aspects of their colouration. Using the R package *pavo* (Maia et al. 2013), I extracted hue (colour), saturation (Chroma or intensity) and brightness values from reflectance spectrophotometry measurements (300-750nm; Appendix 5: Fig. A5.2, Fig. A5.3). I found significant differences between the three colour morphs at four dorsal body landmarks using one-way MANOVA, but not at either of the two ventral landmarks (Appendix 5: Table A5.2, Fig. A5.3). *Post hoc* ANOVAs indicated that this was predominantly the result of differences in hue values, which corresponds to traditional concepts of colour. Therefore, spectral differences were also examined in a subset of measurements for black, yellow and brown skin (Fig 4.5). Significant spectral differences were found across full

(300-750), UV (360-400nm; truncated to known animal visual limits; Honkavaara et al. 2002), visible (400-700nm) and near-infrared (NIR; 700-750nm) wavelengths (Appendix 5: Table A5.2, Fig. A5.4). However, *post hoc* one-way ANOVA revealed that no single measure from UV spectra differed significantly, although saturation values suggested a trend ($F_{(2,39)}$=3.1; P=0.056)

As the majority of the variation in these data resided in yellow reflectance measurements (Appendix 5: Fig. A5.4), these were removed and analyses re-run to compare just black and brown skin. One way-MANOVA revealed significant differences in both UV and visible wavelengths (Appendix 5: Table A5.2). As before, no individual measurement from UV spectra differed significantly. In the visible spectra, only saturation values differed significantly (one-way ANOVA: $F_{(1,24)}$=5.48; P=0.028).

Through transmission electron microscopy (TEM; Fig. 4.6), I found that yellow skin contained yellow pigment containing epidermal xanthophores (and/or erythrophores) and light-reflecting dermal iridophores (and/or leucophores); dark pigment containing melanophores were also seen in yellow skin as either infrequent single cells or localised clusters in the dermis. By contrast, brown and black skin only contained melanophores, which were found throughout both skin layers.

4.4.2. Associations between colour and toxin

Using GC-MS I identified 18 putative metabolites within an alkaloid region that are shared across most or all salamander samples (Appendix 5: Table A5.3). Through comparison with published data (Habermehl and Spiteller 1967; Mebs and Pogoda 2005) four could be associated with known *Salamandra* alkaloids: cycloneosamandione (Peak02_375), samandarone (Peak07_375) samandaridine (Peak19_417) and samandarine (Peak06_449; Fig. 4.7). A PCA of GC-MS intensities—an indication of abundance—showed that striped, xanthic and hypolutic *S. s. bernardezi* samples clustered together and were more similar to each other than they were to the other subspecies investigated, the yellow-black striped *S. s. terrestris* (Fig 4.7; see Appendix 5 Results *A5.1.3.* for subspecies comparisons).

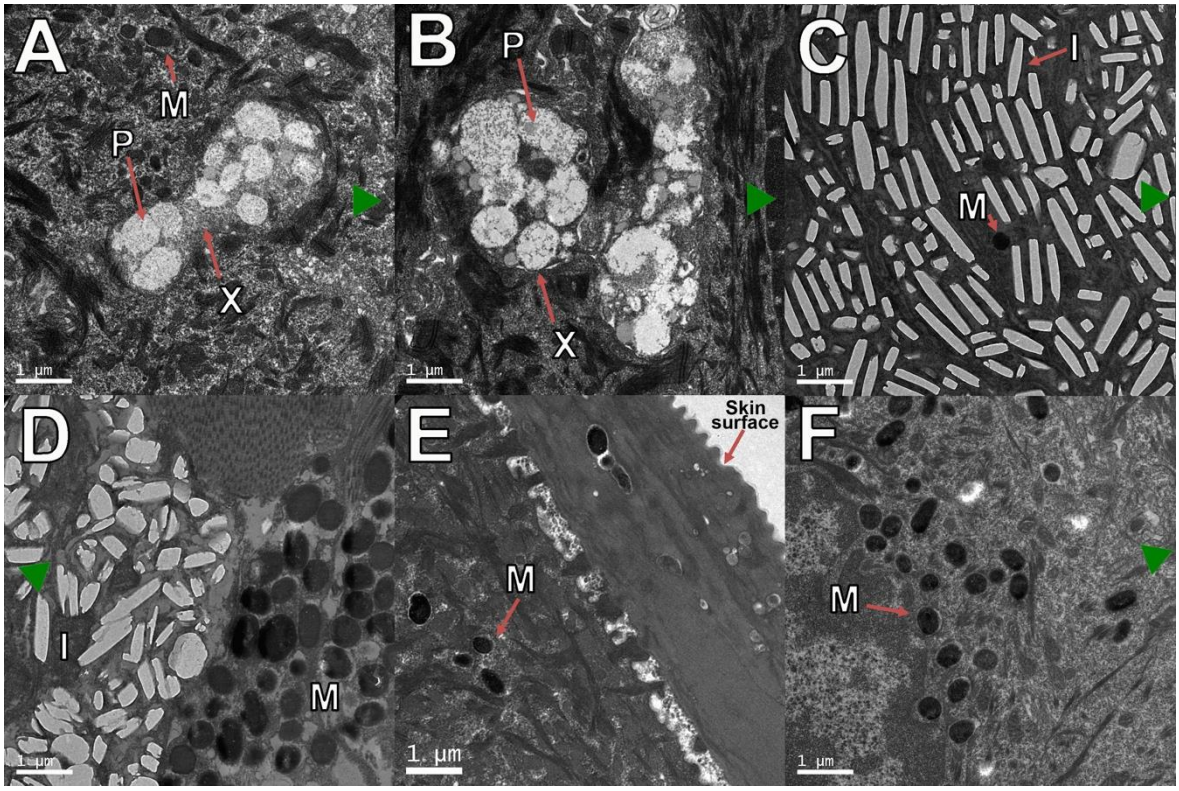


Figure 4.6: TEM images: **A)** epidermal yellow skin from a striped salamander; **B)** epidermal yellow skin from a xanthic salamander; **C)** dermal yellow skin from a striped salamander; **D)** dermal yellow skin from a xanthic salamander; **E**) epidermal brown skin; **F)** dermal black skin. Cellular structures: P= pigment vesicle, M= melanophore, X= xanthophore, I= iridophore. Green triangles point towards the external surface of the skin.

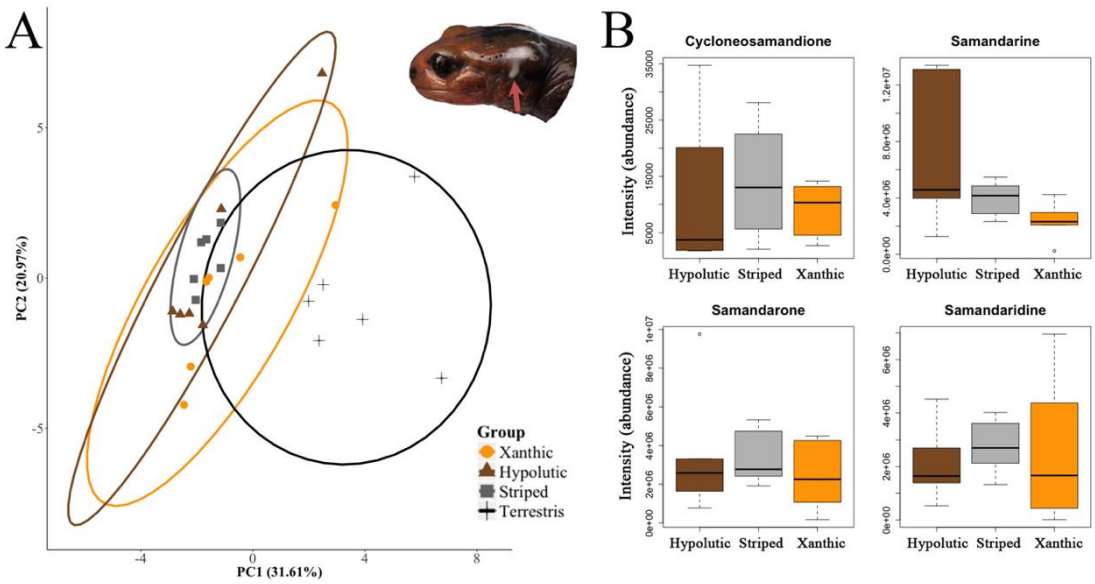


Figure 4.7: **A)** PCA of 18 *Salamandra* toxin metabolite intensities. Inset: *Salamandra* head with toxic secretion on the parotid gland (red arrow). **B)** Intensities for the four identified *Salamandra* alkaloids across representative colour morphs.

Between *S. s. bernardezi* colour morphs, no significant differences were found in the intensities of the 18 identified metabolites (Appendix 5: Table A5.4). However three putative metabolites showed a trend of differing between colours, including samandarine (one-way ANOVA: $F_{(2,15)}$=3.11; P=0.074), Peak04_303 ($F_{(2,15)}$=3.44; P=0.059) and Peak14_486 ($F_{(2,15)}$=2.89; P=0.087). In addition, no significant differences were seen when comparing ancestral (striped) to derived (xanthic and hypolutic) or xanthic to hypolutic salamanders (Appendix 5: Table A5.4). However, a great amount of variability was seen within, as well as between, colour morphs. For example, hypolutic individuals showed the lowest and highest intensities for cycloneosamandione (Peak02_375), as well as presenting two individuals in which it could not be identified at all (Appendix 5: Table A5.3).

4.4.3. Population genomics

In total, 82 salamanders were ddRAD-Seq genotyped at 4702 loci (summary values of diversity and heterozygosity can be found in Appendix 5, Results A5.1.4.). Average *Fst* between sampling localities ranged from 0.018–0.196, all of which were significant except for the sample site 1 *vs.* 2 comparison (Appendix 5: Table A5.5). There was also significant isolation by distance (Mantel test: r=0.795; p<0.005; Appendix 5: Fig. A5.5), with population differentiation increasing with geographic distance.

STRUCTURE analyses identified K = 4 as best fitting the data (ΔK = -264554.28; for full results see Appendix 5: Table A5.6) and grouped samples by sample site and geographic proximity but not colouration (Fig. 4.8). This same pattern was seen in principal component (Appendix 5: Fig. A5.6) and phylogenetic analyses (Fig. 4.8) of the data, which both separated individuals into clusters concordant with geography rather than colouration.

As the above analyses showed that colour polymorphic sites 6 and 7 broadly formed a single genetic cluster, samples from these locations were subset and *Fst* calculated between the three representative colour morphs. However, no significant differences were revealed (xanthic vs hypolutic: *Fst*=0.001, P=1; striped vs hypolutic: *Fst*=0.002, P=0.264; striped vs xanthic: *Fst*=0.003, P=0.123).

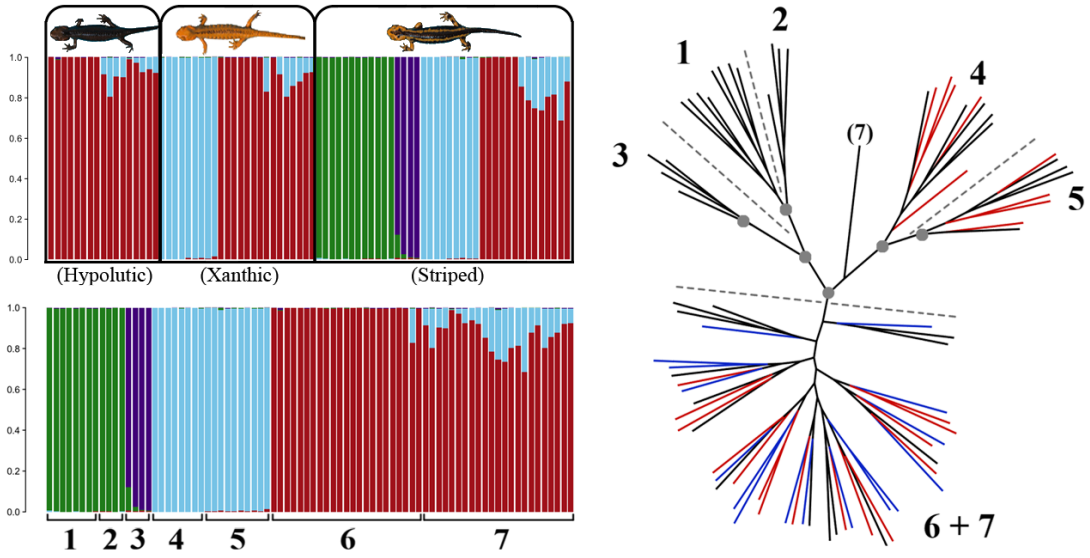


Figure 4.8: Population structure. **Left**: STRUCTURE plots for K = 4, ordered by colour morph (top) and sample site (bottom). **Right**: Unrooted maximum clade credibility tree (RAxML) showing clustering by sample site not colour phenotype: red branches denote xanthic salamanders, blue hypolutic salamanders. Grey circles indicate branches with ≥85% bootstrap support.

4.4.4. Gene expression analyses

In total, 180.8M reads aligned to 35,926 of the 44,165 reference transcripts. Pairwise comparisons using the R package *DESeq2* (Love et al. 2014) identified 218 of these to be significantly differentially expressed (sDE) across skin colours (adjusted p-value < 0.1; see Appendix 5: Fig. A5.7 for MA plots). These transcripts were associated with 167 unique genes (43 were unable to be identified), at least 59 of which are known or suspected to be involved in animal colouration (Fig. 4.9; Appendix 5: Table A5.7–8 ND Fig. A5.8–10).

The comparison between yellow and black skin revealed the fewest sDE genes (42; 28 down regulated and 14 upregulated in yellow; Appendix 5: Fig. A5.8). These could be associated with 35 unique genes, 21 (or 62.9%) of which are known or suspected to be involved in animal colouration (Appendix 5: Table A5.7–8, Fig. A5.8). Notably, the four most down regulated transcripts in yellow skin are related to genes involved in melanin production (*PMEL*, *TRYP1* and *TYR*), and the seven most upregulated transcripts in yellow skin to genes involved in a number of pigment-related roles, including iridophore (*PNP* and *ADA2*), xanthophore (*PAX7* and *SLC2A11*) and leucophore (*SLC2A9*) production (Appendix 5: Fig. A5.8).

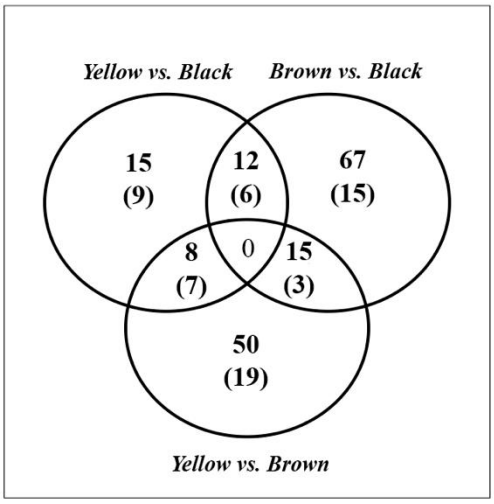


Figure 4.9: Venn diagram showing the number of sDE genes identified across all individuals in pairwise comparisons between the three skin colours. Values in parentheses indicate the number of genes associated with colouration through *post hoc* assessment (for greater detail, see Appendix 5: Table A5.7–8, Fig. A5.8–10).

When comparing yellow to brown skin, more sDE transcripts were identified (95; 43 down regulated and 52 upregulated in yellow). From these, 73 unique genes were identified, 29 (or 39.7%) of which could be associated with animal colouration (Appendix 5: Table A5.7–8, Fig. A5.9). Again, the most up (*DEBF* family and *TFEC*) and down (*MAPK12*, *PMEL*, *MLANA*, *CCM2L* and *PDE1B*) regulated genes were pigmentation related (Appendix 5: Fig. A5.9). Interestingly, all of these are melanin-process related (Appendix 5: Table A5.7–8, Fig. A5.9). Finally, the comparison of brown and black skin resulted in the highest number of sDE transcripts (121; 71 down regulated and 50 upregulated in brown). However, only 24 (or 25.5%) out of the 94 putatively identified genes could be associated with animal colouration (Appendix 5: Fig A5.11).

PANTHER Overrepresentation Tests found a number of gene ontology (GO) terms statistically associated with those genes upregulated in different colours of skin (for full results see Appendix 5: Table A5.9). Broadly, those genes upregulated in yellow skin compared to black were associated with inosine and urate metabolic processes—inosine being important in the production of xanthophores and iridophores (Krauss et al. 2013), and uric acid being important in the production of leucophores (Kimura et al. 2014). In contrast, when compared to yellow, genes upregulated in both brown and black skin were associated with a number of GO terms associated with pigmentation, particularly those relating to melanin biosynthesis and melanosomes. When compared to the zebrafish

reference (*Danio rerio*), gene expression in black skin was also associated with iridophore differentiation. While black skin lacks iridophores (Fig. 4.6), the genes responsible for this—*OCA2* and *TYR*—are known to be involved in both iridophore and melanophore differentiation in teleost fish (Darias et al. 2013; Beirl et al. 2014). Finally, none of the GO terms identified by overrepresentation tests of genes sDE between brown and black skin could be directly associated with animal pigmentation (Appendix 5: Table A5.9).

Comparisons were also made between yellow and black skin in striped individuals only (to reduce the impact of inter-individual variation), and between the two skin landmarks across all individuals (to identify sDE transcripts associated with physiological landmark, not colour). Between yellow and black skin in striped individuals only, 58 sDE transcripts were identified (41 down regulated and 17 upregulated in yellow), which corresponded to 47 unique genes (Appendix 5: Table A5.7). Of these, 19 (or 40.4%) could be associated with animal colouration (Appendix 5: Table A5.8), 15 of which overlapped with the previous yellow-black comparison. Again, PANTHER Overrepresentation tests revealed genes upregulated in black skin to be associated with melanin biosynthesis, melanosomes and iridophore differentiation, and in yellow skin to urate metabolic process (Appendix 5: Table A5.9). Between the two skin landmarks sampled per individual 15 sDE genes were identified (Appendix 5: Fig. A5.12). Only five of these overlapped with previous comparisons (striped-only analyses; *ACAN*, *FAM83A*, *FMOD*, *THBS4*, *VMO1*), and only one of these could be associated with animal colouration (*FAM83A*; Appendix 5: Table A5.7). In addition, sDE transcripts between skin landmarks showed lower levels of log fold expression change (±0.69) as compared to between skin colour comparisons (±1.56–2.21). Combined, these results show that skin landmark did not have a strong effect on pairwise colour comparisons.

4.4.5. Genotype-phenotype association

Random forest (RF) analyses using individual PC1 scores from the analysis of colour pattern variation (*Patternize* PCA; Appendix 5: Fig. A5.13) identified nine loci that together explained 56.8% of the variance in the data (Fig. 4.10). Through Blastn alignment of the corresponding 60 bp RAD-loci to the *Ambystoma mexicanum* transcriptome assembly v4.0 (Smith et al. 2005; Keinath et al. 2015), and subsequent NCBI Blastn of all matched contigs with an E-value <1 (Appendix 5 Table A5.10), one of these nine loci (locus ID 2162) was found to map to the tyrosinase (*TYR*) coding sequence (Appendix 5: Table A5.11). None of the remaining eight loci could be annotated.

When salamander colour phenotypes were scored categorically (striped (n=38), xanthic (n=24) or hypolutic (n=17)) RF results were substantially different. Two subsets of loci were able to differentiate colour morphs equally well: one containing 21 SNPs (61.22% variance explained) and another containing just seven of these 21 (61.24% variance explained; Fig. 4.10). Of these, one was the putative *TYR* locus, which overlapped with the *Patternize* PC score based analysis (Appendix 5: Table A5.11). Through alignment to the *A. mexicanum* assembly, a further six of these 21 loci could be mapped to known genes: *BAZ2A*, *CAMK1*, *LDLRAD4*, *NYNRIN*, *RGR*, and *SURF4* (Appendix 5: Table A5.11). Although none can be directly linked to animal pigmentation, some have potential associations. *BAZ2A* is known to share the enhancers GH12G055965 and GH12G056040 with the colour gene *PEML* (Premelanosome Protein; Rebhan et al. 1997) and *SURF4* has been associated with molecular pathways involved in *MC1R* (Melanocortin 1 receptor) signalling (April et al. 2007). Also, while the probability (Pr) is low, when using the Integrative Multi-species Prediction (IMP) platform, which predicts possible functional roles for genes based on prior published knowledge and data collections, two of these six genes could be associated with colouration. The gene *RGR* is hypothesised to be involved in melanin metabolic (GO:0006582; Pr = 0.052) and biosynthetic (GO: 0042438, Pr = 0.052) processes in zebrafish (*Danio rerio*, gene *rgrb*), and pigmentation (GO:0043473, Pr = 0.012) in mice (*Mus musculus*). In addition, *CAMK1* is potentially involved in guanosine related processes (GO:1901069, Pr = 0.025; GO:1901068, Pr = 0.013) in humans, guanosine being a key constituent of iridophores (Ide and Hama 1972).

SNP-by-SNP association studies in LFMM, when using *Patternize* PC1 scores for each individual, identified eight loci significantly associated with colour phenotype (q-value < 0.05; Fig. 4.10; Appendix 5: Table A5.11). Two of these loci overlapped with the *Patternize* based RF analysis; however, none could be mapped to known genes. When treating colour morphs as categorical, LFMM analyses identified nine significantly correlated SNP loci (q-value < 0.1; Appendix 5: Table A5.11). Of these, two could be mapped to known genes. These were the putative *CAMK1* locus (also identified in the RF analysis for categorical colours) and *prel* (protein preli-like), a gene that is hypothesised to be involved in mouse pigmentation (IMP: gene *Prelid1*: GO:0043473, Pr = 0.012).

Finally, I looked for signs of selection between the three representative colour morphs in the 2149 genomic loci used in genotype-phenotype association analyses. Data lacked statistical power for OutFLANK to detect *Fst* outliers. However, loci putatively under selection between colour morphs were identified by comparing z-transformed *Fst* (Z*Fst*)

and genetic diversity; to mitigate background or population level effects, I focused on just those individuals from the polymorphic sample site pair 6 and 7 (Rio Color and Rio Tendi). Across these 46 salamanders, 142 of the 2149 loci had a Z*Fst* over three standard deviations from the mean, and 64 were over four standard deviations (6.6% and 3% of loci respectively; Fig. 4.11); for a full list of loci putatively under selection between colour morph comparisons see Appendix 5: Table 12–14.

Of those loci found to be putatively under selection, 19 overlapped with loci identified as associated with *Salamandra* colouration through RF and LFMM analyses: 10 in the hypolutic-xanthic comparison, 10 in the hypolutic-striped comparison, and two in the xanthic-striped comparison (overlap was seen in three loci across comparisons; see Appendix 5: Table 12–14). Of these, several are particularly noteworthy. For example, the putative *CAMK1* locus showed increased Z*Fst* and genetic diversity in hypolutic compared to striped and xanthic salamanders. Three important melanin related loci—*TYR*, *SURF4* and *RGR*—also showed signals of selection between hypolutic and striped salamanders. Finally, the *TYR* locus was also found to be putatively under selection between striped and xanthic salamanders, but not between the two derived morphs (xanthic and hypolutic).

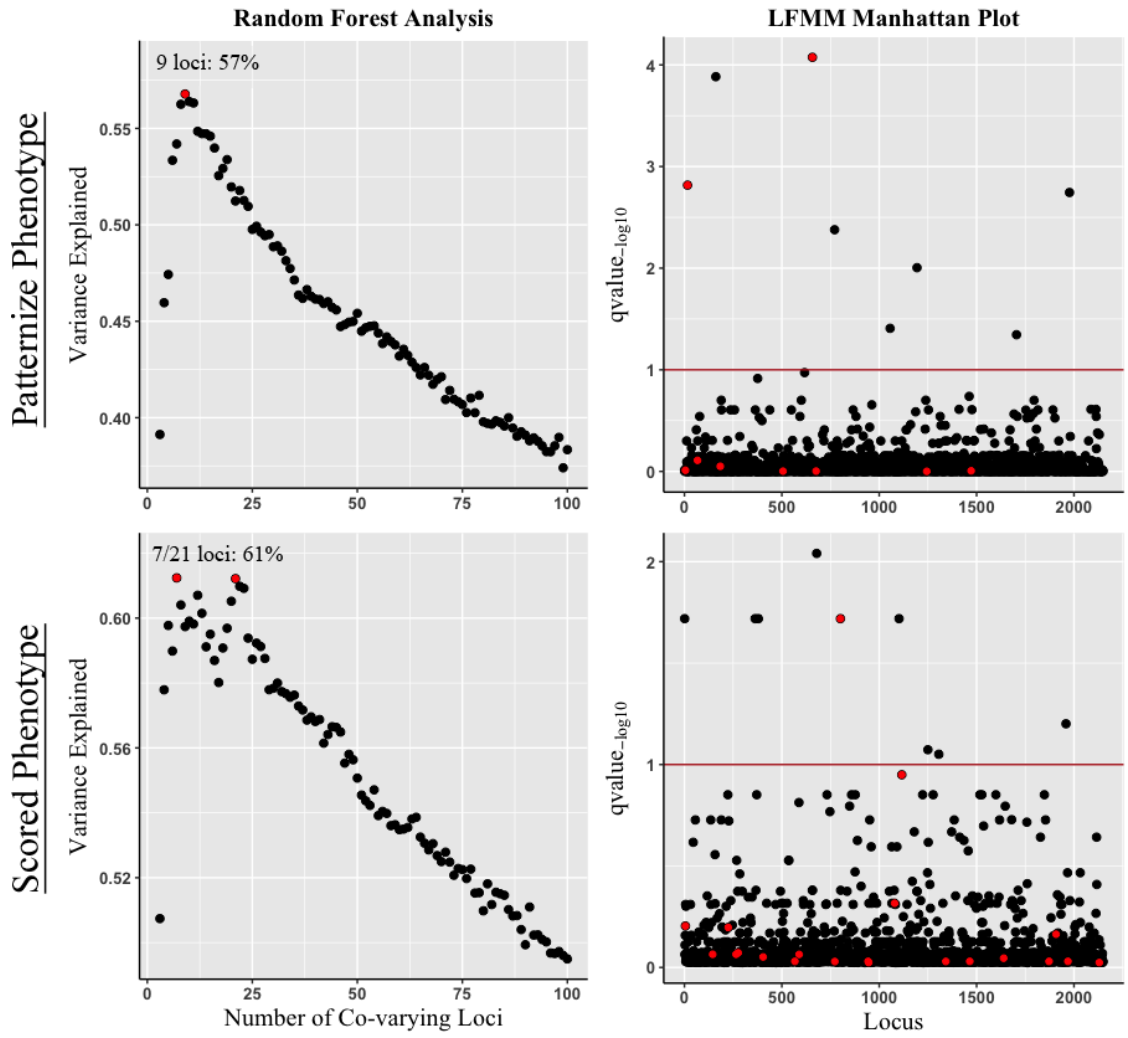


Figure 4.10: Genotype-phenotype association plots. **Left:** The variance explained (as a proportion) by covarying loci during the backwards-purging step of the Random Forest analyses (starting with the top 100 most important loci). Red points indicate the set of loci best able to discriminate between colour phenotypes. **Right:** LFMM Manhattan plots. The red line denotes the significance threshold, with loci above this significantly associated with colour phenotype. Points in red show the sets of loci identified as best able to discriminate between colour phenotypes during Random Forest analysis of the same data.

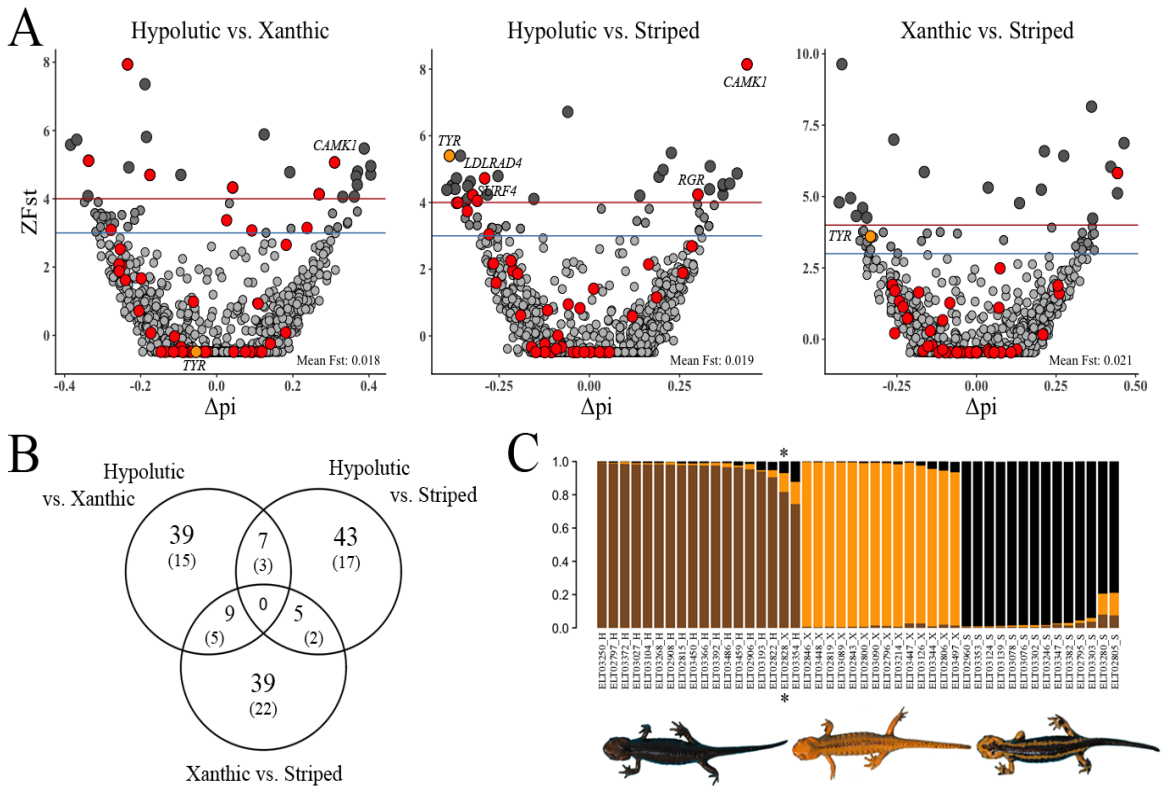


Figure 4.11: Identification of loci showing a signal of selection between representative colour morphs in Rio Color/Rio Tendi (sample sites 6 and 7). **A)** Pairwise comparisons of z-transformed *Fst* and genetic diversity: the blue line delimits a Z*Fst* of three standard deviations from the mean and the red line four standard deviations from the mean. Point shading groups loci by Z*Fst* category: light grey <3; medium grey: 3-4; dark grey: >4. Red points indicate those loci associated with colour phenotype identified across all four RF and LFMM analyses: labelled with a gene ID (where possible) if putatively under selection (the *TYR* loci is highlighted in orange in all cases). **B)** Venn diagram showing the number of loci three (and four) standard deviations from the mean in each comparison. **C)** A STRUCTURE plot (K = 3; see Appendix 5 Table A5.15 for Evanno results table) of the 142 loci across comparisons with a Z*Fst* over three standard deviations from the mean, showing clear clustering by colour morph; the asterisk (*) marks a xanthic individual that grouped within the hypolutic cluster.

4.5. Discussion

In this study, I combined the quantitative analyses of colour phenotype with detailed genetic studies to gain insights into the molecular and evolutionary basis of pigmentation in a highly variable salamander: *S. s. bernardezi*. Through population genomic analyses, I infer a sympatric colour diversification, resulting in a gradient of phenotypic variation from fully yellow to fully brown through the ancestral striped phenotype. However, between representative colour morphs I also find putative signals of selection on 142 genomic loci, some of which are known or suspected to be involved in animal colouration (e.g. *TYR* and *CAMK1*). Genotype-phenotype association studies also identified a total of 43 loci able to discriminate between individuals based on colouration, and transcriptomic analyses identified 196 significantly differentially expressed genes between yellow, brown and black skin, 63 of which are known or suspected to be involved in animal pigmentation. However, I found no correlation between a salamanders colour phenotype and the content of its toxic secretions, calling into question the aposematic function of this colouration.

4.5.1. Evolutionary history of sympatric Salamandra colour morphs

While the yellow and brown coloured populations of *S. s. bernardezi* around the Rio Tendi region of Asturias, Northern Spain, are no longer recognised as a distinct subspecies (Beukema et al. 2016a), it has remained completely unknown how or why these geographically restricted xanthic and hypolutic phenotypes evolved. My data support a sympatric origin, as opposed to vicariance with secondary contact, as *S. s. bernardezi* display small but significant genetic differentiation between geographic populations, yet no genetic clustering or differentiation by colour phenotype. This also indicates no assortative mating by colour, and unlike a recent study on Slovakian yellow-black spotted *S. salamandra* (Balogová and Uhrin 2015), I found no sexual dichromatism. This suggests that *Salamandra* colour patterns are not under sexual selection, as seen in some other polymorphic species like the North American red-backed salamander (*Plethodon cinereus*; Acord et al. 2013). It also suggests that this variation is not indicative of incipient speciation, as has been claimed of geographically sympatric yet genetically differentiated *S. salamandra* in central Germany, which deposit their larvae in either pond or stream habitats (Steinfartz et al. 2007b; Caspers et al. 2015). However, despite this and the gradient of variation seen, putative signals of selection on colour loci were found, showing that polymorphic *S. s. bernardezi* are not maintained simply through genetic drift.

4.5.2. Ecological drivers of colour diversification

One of the most distinctive aspects of this lineage of *Salamandra* is the presence of high numbers of hypolutic (fully brown) individuals (Köhler and Steinfartz 2006). Outside of this geographic region, such colouration is only found as a rare variant in two other *Salamandra* species: the Alpine *S. atra* (Bonato and Steinfartz 2005) and the North African *S. algira tingitana* (Martínez-Solano et al. 2005). Along with black colouration, this falls under the process of melanisation (an increase in the concentration of melanin pigments), with my gene expression and TEM results showing that brown and black skin are molecularly and structurally highly similar. These two skin colours are also nearly indistinguishable in terms of their spectral reflection, indicating that they absorb comparable wavelengths and amounts of solar radiation. This is interesting, as skin melanisation has been hypothesised to be adaptive for thermoregulation at high altitude in *Salamandra* spp. (Vences et al. 2014). However, while salamanders were sampled between c.78–1312 masl, hypolutic individuals were only found at comparatively low altitudes (172–679 masl), suggesting selection for darkening to improve thermoregulation at high altitudes is not a selective driver. I also found no differences in striped colour patterns between my representative low altitude (polymorphic) and high altitude (monomorphic) populations, showing that the proportion of yellow to black does not vary along this altitudinal gradient. These results are similar to those in Alho et al. (2010), which found no evidence for skin melanisation as being adaptive for thermoregulation at high latitudes in the common frog (*Rana temporaria*). It has also been hypothesised that the evolution of hypolutic colouration in *Salamandra* may be related to drier environmental conditions (Beukema et al. 2010). However, as colour divergent *S. s. bernardezi* occupy a broadly sympatric ecological space, with no apparent differences in habitat between neighbouring populations, contemporary environmental pressures are unlikely to be driving this diversification.

The more widely assumed functional basis of *Salamandra* colouration is aposematism, an interspecies warning signal to potential predators used to communicate their toxicity (Beukema et al. 2016b). In those amphibian systems where skin colouration has been found to be a true aposematic signal, such as the poison frog *Oophaga pumilio* (Maan and Cummings 2012), conspicuousness and toxicity have been shown to be tightly correlated. This does not appear to be the case in colour polymorphic *S. s. bernardezi*. Despite the fact that considerable variation was seen in toxin metabolites, which differed between subspecies, this was not correlated to colour phenotype within *S. s. bernardezi*. While a

test such as a mouse LD-50 assay would be required to thoroughly measure toxicity, my data suggest that adaptation for aposematism is not driving colour evolution within this lineage. This is similar to the findings of Vences et al. (2014), who found no correlation between toxicity and colour pattern between *Salamandra* species. However, while they hypothesised that the uniform black colourations of *S. atra* and *S. lanzai* could perhaps serve an aposematic function in certain Alpine habitats, this is unlikely to be a driver of colour diversification in sympatric populations. Additionally, while differences in predation risks can drive variation in aposematic signals between populations (Dreher et al. 2015), they must still be recognisable to potential predators, leading to strong balancing selection on pattern elements within populations (Winters et al. 2017). Such processes would not give rise to the sympatric gradient of variation observed in this study.

While my data would seem to rule out the most commonly hypothesised adaptive functions for amphibian colouration (mate attraction, predator avoidance and physiological regulation), more fine-scale processes may be at work. For example, red-backed salamanders (*Plethodon cinereus*) present two colour morphs: red striped and melanic. These differ in a number of behavioural traits, from microhabitat selection (Anthony et al. 2008; Fisher-Reid et al. 2013; Cosentino et al. 2017) and mating behaviours (Acord et al. 2013), to movement patterns (Grant and Liebgold 2017) and diet selection (Stuczka et al. 2016). They also show differences in basal stress levels (Davis and Milanovich 2010), disease susceptibility and predation rates (Venesky et al. 2015; Grant and Liebgold 2017). While there is currently no evidence for such differences between *S. s. bernardezi* colour morphs, it is clear that the ecological pressures and evolutionary dynamics responsible for generating and maintaining this diversity are more complex than has been previously hypothesised (e.g. aposematism and thermoregulation; Vences et al. 2014).

4.5.3. Colour loci discovery

By utilising two next generation sequencing (NGS) methods as complimentary approaches for gene discovery, I have identified a number of known and candidate colour loci in *S. s. bernardezi*. Through gene expression analyses, I identified 196 significantly differentially expressed (sDE) genes between skin colourations, 63 of which could be associated with animal pigmentation. Of these, 46 are associated with melanophore related pathways, while just 11 are known to be involved in xanthophore and/or iridophore production. While the pathways involved in producing these later two cell types are poorly characterised, they are known to share some developmental pathways with melanophores

(see Darias et al. 2013; Beirl et al. 2014; Woodcock et al. 2017) and my data provides a wealth of candidate loci for further study. For example, the *CAMK1* locus was identified during genotype-phenotype association analyses and found to be putatively under selection between hypolutic and both xanthic and striped salamanders. This is interesting, as I hypothesis *CAMK1* to be involved in iridophore production (see section *4.4.5.*) and these cells have been lost in hypolutic individuals (Fig. 4.6).

Perhaps the most important loci for discriminating between colour phenotypes were melanin related genes in the tyrosinase family (*TYR*, *TYRP1*, *DCT*/*TYRP2*). Of particular note is tyrosinase (*TYR*), an important enzyme involved in converting tyrosine into melanin pigments (Murisier and Beermann 2006). *TYR* had significantly reduced expression levels in yellow skin and showed genomic signals of selection and decreased genetic diversity in individuals with derived (xanthic and hypolutic) compared to ancestral (striped) phenotypes. The identification of *TYR* in both RNA-Seq and RAD-Seq data is remarkable, given that RAD-loci only covered ~0.0005% of the approximately 34Gb *Salamandra* genome (Gregory 2017). Recently, Woodcock et al. (2017) also mapped an albino locus in laboratory-reared axolotl salamanders (*Ambystoma mexicanum*) to a mutant *TYR* allele, hinting at its prominent role in amphibian colouration. My study demonstrates that even with low coverage, genome-scanning methods have the statistical power to detect loci associated with adaptive phenotypes. However, it is interesting to note that the melanin pigment controlling gene *MC1R*, which is a common candidate gene in vertebrate colour genetic studies (see Rosenblum 2006; Herczeg et al. 2010; Poelstra et al. 2013; Roulin and Ducrest 2013; Posso-Terranova and Andres 2017), was not identified within the sequenced transcriptome data. This reflects the findings of Henning et al. (2013), which found no difference in *MC1R* expression between grey and gold Midas cichlids (*Amphilophus citrinellus*). However, my genomic analyses did identify *SURF4*, which is associated with molecular pathways involved in *MC1R* signalling (April et al. 2007).

The combination of RNA-Seq and RAD-Seq data also provides insights into the evolution of brown skin. While yellow and black pigmentation are both components of the ancestral yellow-black striped colour pattern, brown skin is a derived phenotype (see Chapter 3: section *3.5.4.*). However, brown and black melanin pigments do share a common molecular basis, both being eumelanin polymers (Bagnara et al. 1978; Meredith and Sarna 2006; Ito and Wakamatsu 2011). This is interesting, as two genes involved in eumelanin synthesis—*TYR* (Hirobe et al. 2007) and *SURF4* (through its association to *MC1R*; April et al. 2007)—were among those loci identified by genotype-phenotype association analyses

and also found to be putatively under selection. Also, of the 94 sDE genes identified by transcriptomic analyses between brown and black skin, 19 of the 24 genes associated with animal colouration were melanophore related. Of these, 5 were upregulated in brown skin and 14 were down regulated (Appendix 5: Fig. A5.10). A further eight sDE genes (*CIRBP*, *CYR61*, *MMP1*, *MMP3*, *AKR1C1*, *AKR1C3*, *TPBG* and *UPK1B*) are also individually associated with GO terms relating to keratinocytes (Appendix 5: Table A5.8). While the exact mechanisms remain unclear, it is known that keratinocytes interact with melanosomes, either through melanin pigment transfer or uptake (Aspengren et al. 2006). Such results indicate that brown skin may arise as a result of modifications in melanin pathways, which either changes the kinds or ratios of eumelanin pigments produced in melanosomes. This would be consistent with recent studies on human skin pigmentation (e.g. Ebanks et al. 2011; Crawford et al. 2017), and could potentially explain the phenotypic similarity, yet visual difference, seen between black and brown skin.

4.6. Conclusion

The data presented here contribute substantially to our understanding of the molecular basis of amphibian colouration. They also highlight a number of candidate colour genes warranting further study and provide a valuable resource for comparative colour genetic studies. In a uniquely colour variable lineage of *S. salamandra*, I have identified 196 significantly differentially expressed genes across skin colour comparisons, 63 of which are known or suspected to be involved in animal colouration. For the remaining 133 genes, this study likely represents their first association to vertebrate pigmentation. Further to this, genotype-phenotype association analyses identified 43 genomic loci statistically associated with dorsal colour pattern in *S. s. bernardezi*. These include the melanin pigment related gene *TYR* and *CAMK1*, which due to its association with guanosine related biological processes may be involved in the production of iridophores. Importantly, I also found signals of selection on 143 genomic loci, indicating that selective forces are involved in driving colour diversification within this lineage. This is the first genomic inference of this kind in caudate amphibians that I am aware of. However, the lack of a reference genome prevented the identification of 35 of the 43 RAD-loci identified across association analyses: whether these represent unidentified coding sequences, regulatory regions or other genetic markers will require the development of more robust genomic resources.

Interestingly, while my results help to elucidate the molecular basis of amphibian colour patterns, they call into question the way in which adaptive functionality is attributed to them. I found no association between a salamander's colouration and the metabolomic content of its toxic secretions. Also, through the inferred symparty of divergent colour morphs, I find no association between dorsal colouration in *S. s. bernardezi* and environmental gradients like altitude. As such, I find no support for the long-standing assumptions that aposematism and adaptation for thermoregulation underpin colour evolution within *Salamandra*. However, while the ecological pressures behind this localised chromatic diversification remain unclear, finer scale physiological or behavioural differences between colour morphs may be involved and warrant further investigation. Overall, this study highlights the need to integrate phenotypic, ecological and evolutionary analyses in testing hypothesis about supposedly adaptive traits and provides a valuable resource for future studies on vertebrate pigmentation genetics.

Chapter 5: General Discussion

In this thesis, all chapters had distinct yet complimentary aims. In **Chapter 2**, I aimed to resolve the shallow phylogenetic relationships between currently recognised *Salamandra* species using multiple phylogenomic data sets (RNA-Seq, RAD-Seq and full mitochondrial genomes) and approaches (maximum likelihood and Bayesian based concatenated and species tree analyses). Through this, we also aimed to validate the use of such approaches for resolving interspecies relationships within systematically challenging taxa like *Salamandra*, and determine if these different molecular datasets provided complimentary phylogenetic signals. In **Chapter 3**, having shown the appropriateness and power of our RAD-Seq phylogenomic approach, I aimed to further resolve species and subspecies relationships within the genus using a more taxonomically comprehensive dataset than that used in **Chapter 2**. Following this, having constructed a robust phylogenetic hypothesis, I aimed to identify cases of parallel/convergent evolution in reproductive and colour phenotypes through ancestral state reconstruction analyses. Finally, in **Chapter 4**, I leveraged a uniquely colour diverse lineage of *Salamandra salamandra* from northern Spain (*S. s. bernardezi*) to identify genetic associations with colour using a suite of approaches. Specifically, I aimed to resolve the evolutionary origins of sympatric colour morphs, assess the hypothesised ecological drivers of colour diversification within this lineage (aposematism and thermoregulation), and identify the underlying genetic basis of these colour patterns.

5.1. Key Findings

In **Chapter 2** we demonstrate the potential of phylogenomic data sets to elucidate shallow relationships among systematically challenging, closely related taxa. Using 3070 nuclear protein-coding genes from RNA-Seq, 7440 SNP-loci obtained by RAD-Seq, full mitochondrial genomes, and a range of phylogenetic analyses, we accurately reconstructed the controversial phylogeny of the genus *Salamandra*. However, we also note that traditional phylogenetic support metrics, like bootstrap values and Bayesian posterior probabilities, often tend to be overinflated in phylogenomic analyses, resulting in high confidence for potentially spurious relationships. Maximum likelihood (ML) and Bayesian analyses of concatenated RNA-Seq and RAD-Seq data returned fully congruent topologies, despite their differing biological properties. These analyses found the North African *S.*

algira to be sister to the European *S. salamandra*, with these two species being the sister group to a clade containing *S. corsica*, *S. atra* and *S. lanzai* (the later two forming sister taxa). Finally, the Near Eastern *S. infraimmaculata* was found to be sister to all other *Salamandra* species. However, the topological position of *S. corsica* was found to be unstable, with species tree analyses of RNA-Seq and RAD-Seq data placing it as sister to a clade containing *S. algira, S. salamandra, S. atra* and *S. lanzai*, and mitochondrial genomes supporting its placement as sister to *S. atra*. This uncertainty regarding the phylogenetic position of *S. corsica* was also found in **Chapter 3**.

In **Chapter 3**, I conducted one of the most comprehensive phylogenetic analyses of the genus *Salamandra* to date and identified cases of parallel/convergent phenotype evolution. This included extensive geographic and subspecies sampling and the analysis of 294,300 nt of sequence data. ML and Bayesian analyses of concatenated RAD-Seq data are largely congruent with the inter-species relationships found in **Chapter 2**, however, the placement of *S. corsica* is again unclear, either being placed as sister to the *S. atra* + *S. lanzai* clade or the *S. algira* + *S. salamandra* clade. Notably, I also find that the parameters used to filter RAD-Seq data, particularly the setting of an upper bound on the number of permissible variable sites per locus, can have a large impact on the topologies of downstream phylogenetic reconstructions.

Further to this, within the North African *S. algira*, I confirm the monophyly of three out of four currently recognised subspecies, but also identify previously undescribed diversity in the Moroccan Mid-Atlas Mountains. This makes the nominative subspecies (*S. algira algira*) paraphyletic, and perhaps warrants the description of a fifth *S. algira* subspecies. In contrast, within the European *S. salamandra*, I only find that eight of the 14 currently recognised subspecies meet a criterion of monophyly (*S. s. bernardezi*, *S. s. fastuosa*, *S. s. almanzoris*, *S. s. crespoi*, *S. s. morenica*, *S. s. longirostris*, *S. s. bejarae* and *S. s. salamandra*). The rest (*S. s. beschkovi*, *S. s. gallaica*, *S. s. gigliolii*, *S. s. hispanica*, *S. s. terrestris* and *S. s. werneri*) likely representing cases of over-splitting based on regional phenotypic variation in unreliable characters, like colour patterns

In **Chapter 3** I also tested the history of convergent phenotype evolution. Through ancestral state reconstruction analyses, I find that pueriparity (giving birth to fully metamorphosed juveniles) has independently arisen in at least four lineages, melanism in at least five, and a yellow-black striped phenotype in at least two, all from a common yellow-black spotted larviparous (larvae depositing) ancestor. Importantly, I also found a

striking correlation between pueriparity and melanism, traits that display almost identical evolutionary histories in our ASR analyses.

Finally, in **Chapter 4,** I have significantly contributed to our understanding of the molecular basis of amphibian colouration and assessed its ecological function. For this, I utilised a contact zone within a rare, highly colour-variable lineage of *S. salamandra* from northern Spain (*S. s. bernardezi*) to identify genetic associations with colour. First, I quantitatively measured colour phenotypes using digital image analyses. From this, I identified a gradient of phenotypic variation in colour variable *S. s. bernardezi*, ranging from fully brown (hypolutic) at one extreme to fully yellow (xanthic) at the other, through the ancestral yellow-black striped phenotype; however, individuals broadly fell into one of these three categories. I then used spectrophotometry and transmission electron microscopy (TEM) to phenotypically characterise skin colours. I found that yellow skin was the most distinctive, showing a unique reflectance profile and cellular structure containing epidermal xanthophores and dermal iridophores. In contrast, both brown and black skin only contained dark pigment containing melanophores, and were almost indistinguishable in terms of spectral reflectance. However, they were distinguishable in terms of gene expression differences.

Across pairwise comparisons of yellow, brown and black skin, I identified 196 significantly differentially expressed (sDE) genes, 63 of which are known or suspected to be involved in animal pigmentation. When comparing yellow and black skin, 62.9% of the sDE genes were related to animal colouration, with those genes upregulated in yellow skin broadly relating to iridophores and xanthophores, and those downregulated in yellow skin relating to melanophores. Between yellow and brown skin, 39.7% of sDE genes were colour related, with the most up- and downregulated genes in yellow skin all being melanophore related. Finally, between brown and black skin I found the most sDE genes (94), indicating that despite their phenotypic similarity, they are generated through distinct and separate molecular pathways. However, only 25.5% of these 94 genes are known to be colour related.

Interestingly, the ecological drivers of colour diversification in *S. s. bernardezi* remain unclear. The sympatry of different colour morphs suggests environmental processes are not associated, and a lack of assortative mating indicates that sexual selection is not involved. Importantly, I also found no association between toxin metabolite intensities and colour phenotypes, which calls into question the supposed aposematic function of *Salamandra*

colour patters. However, between sympatric striped, xanthic and hypolutic colour morphs I did find putative signals of selection on 142 genomic loci, some of which are known or suspected to be involved in animal colouration (e.g. *TYR* and *CAMK1*). This indicates that there are selective pressures on colour phenotype in this lineage. Genotype-phenotype association studies also identified a total of 43 loci able to discriminate between individuals based on colouration, 19 of which show putative signals of selection and one of which (*TYR*) was also identified during gene expression analyses. As such, the data presented here comprise one of the most comprehensive studies of amphibian colour genetics to date, provide a valuable resource for comparative colour genetic studies, and display the power of NGS techniques for studying the molecular basis of ecologically adaptive traits in wild non-model organisms.

5.2. Study limitations

5.2.1. The lack of a reference genome

Perhaps the most limiting factor in this study is the size of the *Salamandra* genome. At approximately ~34Gbp in size (Gregory 2017), generating a reference genome assembly remains computationally and financially prohibitive. In addition, as caudate amphibians in general have large genome sizes (Mohlhenrich and Mueller 2016), there is no closely related reference genome available. As a result of this I was only able to associate eight of the 43 genomic loci identified across genotype-phenotype association analyses in **Chapter 4** with a known gene. Further, these eight genes could only be identified by initially aligning reads to the *Ambystoma mexicanum* reference transcriptome assembly v4.0 (Smith et al. 2005; Keinath et al. 2015) and subsequent NCBI searches of mapped *A. mexicanum* contigs. Whether the remaining 35 loci represent unidentified protein coding sequences, regulatory regions, or other kinds of non-coding DNA is unclear. That said, it is remarkable to note that even at such low coverage, with RAD-loci in **Chapter 4** only corresponding to ~0.0005% of the *Salamandra* genome, there was still enough power to identify important colour loci, like the melanin pigment related gene *TYR*. Therefore, this study provides a robust example of the utility of reduced representation genome sequencing for studying the genetic basis of adaptive phenotypes in wild non-model organism, even those with large genome sizes.

A lack of reference genome was also limiting during RAD-Seq data filtering for phylogenetic analyses. Aligning reads to a reference genome helps to identify paralogues and sequencing errors, and typically retains more loci than *de novo* based pipelines (Shafer et al. 2017). It also eliminates the need to apply an 'upper bound' on the number of variable sites per locus, as it removes ambiguity over the 'realness' of highly variable RAD-loci (although the impact of these loci on phylogenetic reconstructions is unclear). For small genome sized organisms (<1000Mbp) it is even possible to sequence multiple entire genomes for phylogenomic analyses (Geneva et al. 2015). However, in the absence of a reference genome assembly, sequence data must be assembled *de novo*. While *de novo* assembly allows for genomic studies in wild non-model organisms, the filtering parameters chosen during the assembly and SNP calling stages can have an enormous effect on downstream analyses (Shafer et al. 2017). This was clearly seen during phylogenetic analyses in **Chapter 3**, where interspecies topologies were highly dependent on the amount of missing data and the number of variable sites allowed per locus. Unfortunately, there are no 'standard procedures' for filtering *de novo* assembled RAD-Seq data, and recent studies in this area have typically focused on either the impact of missing data (e.g. Huang and Knowles 2014; Leaché et al. 2015b; Darwell et al. 2016; Díaz-Arce et al. 2016; Eaton et al. 2017) or simply obtaining the greatest number of loci (e.g. Pante et al. 2014; Puritz et al. 2014). As such, my results highlight an often overlooked, yet highly important, stage of RAD-Seq data filtering in organisms that lack pre-existing genomic resources: the maximum number of variable sites allowed per locus. Therefore, improving data processing and analytical tools for *de novo* assembled phylogenomic data sets is of high priority.

Finally, a reference genome would also have helped me to better elucidate the inter-species relationships within *Salamandra.* While I believe my phylogenomic analyses to be robust given general topological congruence between multiple, independent phylogenomic data sets, some ambiguity remains, particularly in the position of *S. corsica*. In **Chapter 2**, concatenated *vs.* species tree analyses of RNA-Seq and RAD-Seq data conflicted in their placement of *S. corsica*. This is important, as it is known that concatenated *vs.* species tree approaches tend to disagree only at weakly supported nodes with short internal branch lengths (Lambert et al. 2015), and that without gene tree discordance the two approaches should give near identical reconstructions (Edwards 2009). The position of *S. corsica* within the genus was also the only major conflict found between these analyses and trees inferred from whole mitochondrial genomes (**Chapter 2**), and the only species level topological conflict identified through different analyses of concatenated RAD-Seq data in

Chapter 3. This suggests either a rapid divergence event, leading to incomplete lineage sorting, or a complex history of introgression, which cannot be resolved simply through the addition of more sequence data (as noted in Philippe et al. 2011). This is especially true as coalescent-based species tree approaches, which are generally more accurate for reconstructing shallow divergences, become increasingly computationally prohibitive with the number of loci used (reviewed in Lambert et al. 2015).

One potential way to more robustly resolve the inter-species relationships within the genus may be through the identification of loci and/or genomic regions showing signs of ongoing gene flow and historic introgression. Through this, it would be possible to assess the phylogenetic signal in these regions compared to the rest of the genome, and determine if they are responsible for the topological conflicts seen between analyses. It is possible to quantify genome-wide levels of ongoing gene flow and historical introgression, for example, through the use of F_3, F_4, and Patterson's D (or ABBA-BABA) statistics (Eaton and Ree 2013; Rheindt et al. 2014; Hou et al. 2015; Skoglund et al. 2015; Barbato et al. 2017). However, these measures are most effective when used on whole genome sequence (WGS) data; the D-statistic in particular has been shown to become unreliable when using smaller genomic regions (Martin et al. 2015). When combined with reference informed WGS data and a sliding window approach, such statistics can also be used to identify specific regions of the genome and loci showing signals of introgression (Geneva et al. 2015; Yoshida et al. 2016; Medugorac et al. 2017; Teng et al. 2017). Unfortunately, this is not currently feasible with *Salamandra,* as it lacks a reference genome.

As a result of this work, it is clear that the generation of a *Salamandra* reference genome would be highly valuable. While currently considered unfeasible due to its large size (~34 Gbp), it is important to note that reference assemblies have been created for the ~22 Gbp genome of the loblolly pine (*Pinus taeda*; Zimin et al. 2014) and the ~20 Gbp genome of the Iberian ribbed newt (*Pleurodeles waltl*; Elewa et al. 2017). Though still computationally challenging, these examples show that it is possible to sequence very large genomes, something worth exploring for future research on *Salamandra*.

5.2.2. Geographic sampling for phylogenetic analyses

The results of my phylogenetic analyses in **Chapter 3** suggest the need for taxonomic revisions within the genus *Salamandra*. This is especially true for *S. salamandra* subspecies, where many appear to have been described based on non-contiguous

distributions and unreliable phenotypic characters. For example, while phylogenomic analyses recovered both *Salamandra salamandra morenica* and *S. s. crespoi* as monophyletic sister taxa, samples largely clustered by geographic proximity. As these subspecies are also known to overlap in both colour pattern phenotypes and morphology (Sparreboom and Arntzen 2014; Velo-Antón and Buckley 2015; Speybroeck et al. 2016), it is possible that greater sampling may show genetic differences to simply represent geographic distance, not taxonomic distinctiveness. Further to this, phylogenetic analyses indicate that some currently recognised subspecies designations in *S. algira* and *S. salamandra* do not meet a criterion of monophyly (following Nixon and Wheeler 1990). However, some caution must be taken due to the fact that there were noticeable gaps in both taxonomic and geographic sampling. This includes the absence of some subspecies, like *S. atra aurorae, S. a. prenjensis*, *S. infraimmaculata orientalis* and *S. i. semenovi*, and geographic breaks, like a lack of sampling in France. In addition, several lineages were represented by only a few samples, for example *S. lanzai* and *S. salamandra bejarae*. While I do not believe this to be of concern for interspecies relationships, given the strong support for species monophyly and broad topological congruence across analyses (**Chapter 2** and **3**), it is limiting for our ability to delimit subspecies relationships. Therefore, suggesting taxonomic revisions lies outside the scope of this thesis and should be the focus of future study.

5.2.3. Amphibian colour loci discovery

One of the primary research aims of this thesis was to identify loci associated with *S. s. bernardezi* colouration. The motivation for this is also one of the most limiting aspects of the study: a lack of data on the genetic basis of amphibian colouration (Hoffman and Blouin 2000; Rudh and Qvarnström 2013). Currently, most of our knowledge on the genetics of animal colouration comes from studies on just a few mammalian and avian model organisms (McLean et al. 2017), although an increasing number of studies are looking into the genetic basis of colouration in fish (Lamason et al. 2005; Braasch et al. 2007; Salzburger et al. 2007; Diepeveen and Salzburger 2011; Henning et al. 2013; Singh and Nüsslein-Volhard 2015; Zhu et al. 2016). As mammals and birds only present melanophores, this means that our understanding of the molecular pathways involved in melanin-pigment related traits are well characterised (Hoekstra 2006; Braasch et al. 2007; Poelstra et al. 2015). However, poikilothermic vertebrates (amphibians, fish and nonavian reptiles) also present carotenoid and/or pteridine containing xanthophores/erythrophores,

and structurally based light reflecting iridophores/leucophores (Bagnara et al. 1968; Hoffman and Blouin 2000; Rudh and Qvarnström 2013).

While carotenoid metabolic pathways are well studied, our understanding of their role in animal pigmentation comes mostly from birds (von Lintig 2010; Walsh et al. 2012; Lopes et al. 2016; Mundy et al. 2016). In contrast, the underlying genetics of pteridine based pigmentation and purine based structural colouration in vertebrates remains remarkably understudied (McLean et al. 2017). This means that there is an acute lack of information on the colour genetics of terrestrial poikilothermic vertebrates (Hoffman and Blouin 2000; Olsson et al. 2013; Rudh and Qvarnström 2013); although see McLean et al. (2017) for a recent study on gene expression between different colours of skin in the Australian tawny dragon, *Ctenophorus decresii*. Therefore, while NGS technques allow us to study the molecular evolution of adaptive traits like colouration in wild non-model organisms, it often remains challenging to identify the colour pattern related functional role of identified loci. For example, of the 196 significantly differentially expressed (sDE) genes identified through gene expression analyses of different *Salamandra* skin colours in **Chapter 4**, 133 can not currently be associated to a known animal colouration function, with this study likely representing their first association to vertebrate colour production. Identifying the specific colour related role of these genes will require detailed and time consuming research on their exact molecular pathways and interactions.

5.3. Future directions

Animal colour patterns provide a strong system in which to study the molecular basis of adaptation. Generally, it is difficult to understand adaptation in an environmental context, as phenotypes, environments and behaviours are complex and multifaceted, with the interactions between them often poorly understood. However, animal colouration is known to be ecologically adaptive, there are a range of hypothesis explaining its evolution, and it is conspicuously affected by natural selection, all of which make it an ideal trait for studying genotype-phenotype associations (Caro 2005; Hoekstra 2006; Rudh and Qvarnström 2013). For example, colouration in guppies (*Poecilia reticulata*) is known to be driven by the competing pressures of mate attraction and predation risk (Endler 1980; Houde 1997; Gordon et al. 2012), light and dark background matching (crypsis) in *Peromyscus* mice occupying different habitats is driven by visual predation pressures (Vignieri et al. 2010), and colour patterning in South American Corydoradinae catfishes

(Alexandrou et al. 2011) and *Heliconius* butterflies (Kronforst and Papa 2015) represent cases of both both Batesian and Müllerian mimicry. In addition, a great deal of evolutionary parallelism and convergence is seen in animal colour patterns, with convergent colour phenotypes inherently suggesting adaptation even in the absence of obvious ecological explanations (Losos 2011). This offers the opportunity to test hypothesis on the predictability of evolution and, through comparative analyses across different taxonomic scales, allows us to identify molecular constraints on evolution by natural selection. For example, if genomic analyses show that parallel phenotypes in closely related species have a high degree of molecular heterogeneity this suggests little evolutionary constraint, while conserved molecular functions across highly divergent taxa suggest a great deal of constraint (Kronforst et al. 2012).

Unfortunately, the genetic and phenotypic basis of colouration has only been investigated for a few systems, and while this short list includes examples of mammals, fish, birds, reptiles and invertebrates (see Kronforst et al. 2012), studies on amphibian taxa are lacking. Therefore, this thesis presents one of the most in-depth assessments of amphibian colour genetics to date, providing a valuable resource for comparative studies across distant animal lineages. However, while this study has uncovered many novel genetic associations to colouration in *Salamandra*, the functional role of the majority of the genes identified in **Chapter 4** remain unknown, and population genomic and toxin metabolomic analyses appear to eliminate the currently hypothesised ecological and environmental explanations for *Salamandra* colour pattern evolution: aposematism and thermoregulation. This means that there is still considerable work to be done in order to elucidate the molecular and adaptive basis of *Salamandra* colour pattern evolution.

5.3.1. Amphibian colour loci discovery

This study provides a significant contribution to our understanding of amphibian colour genetics. However, the functional role many of the candidate colour genes identified in **Chapter 4** play in vertebrate pigmentation remains unclear. Therefore, a logical next step in this work is to determine how these genes are involved in producing *S. s. bernardezi* colour patterns. Many studies on vertebrate pigmentation have confirmed gene functions through the use of gene knock-out lines in model organisms like mice (*Mus musculus*), where the activation of a specific gene is disrupted to prevent its expression (Hall et al. 2009). In non-model organisms, the development of gene editing techniques, like CRISPR/Cas9, is opening new opportunities for gene manipulation (Chen et al. 2016;

Elewa et al. 2017). However, such genome editing techniques are costly and time consuming to engineer. Alternatively, greater understanding of a genes functional role can be obtained through more detailed phenotypic characterisation. For example, in their study of *C. decresii* lizard colouration, McLean et al. (2017) used GC-MS metabolomic analyses to identify specific carotenoid and pteridine pigments in different colours of skin. This allowed them to better understand the biochemical pathays involved in generating these pigmentation patterns, thereby allowing them to more accuratly interprite the results of transcriptomic analyses.

While a previous study has identified a number of pteridine and flavin pigments in both yellow and black *Salamandra* skin (*S. atra aurorae*; Pederzoli and Trevisan 1990), this has never been combined with genetic data. As such, a detailed examination of the biological pigments underlying *S. s. bernardezi* colour patterns would greatly aid our ability to understand the functioanl role of those loci identified through transcriptomic and genomic analyses in **Chapter 4**. This would be particularly valuable for determining the moleular basis of brown and black *Salamandra* skin, which TEM and reflectance spectrophotomerty revealed to be nearly indestinghishable. However, they are visually different, and a pairwise comparison of brown and black skin displayed the highest number of sDE transcripts of any colour comparison (see **Chapter 4**). Extracting and analysing the pigments present in these skin colours would help us to determine if pigment density, melanin polymer or the presence of different pigments like pteridines or flavins (Pederzoli and Trevisan 1990) is driving phenotypic differences between brown and black skin. This, in turn, could help us assign new vertebrate pigmentation related roles to the sDE genes found between these skin colours.

Transcriptomic analyses of *Salamandra* colouration may also benefit from a more candidate gene based approach to colour loci discovery. In this thesis, I looked at genes significantly differentially expressed between skin colours. However, over 350 pigmentation related genes have been identified in animal models (Crawford et al. 2017), with up to 100 in zebrafish alone (Higdon et al. 2013; Singh and Nüsslein-Volhard 2015), which is notable given the similar chromatophore structures shared between fish and amphibians (Bagnara et al. 1968; Hawkes 1974). The recent study by McLean et al. (2017) also identified 489 sDE genes between *C. decresii* skin colours, most of which do not currently have a known association to animal pigmentation. It is possible that some of these genes are present within the non-sDE *Salamandra* transcripts, with expression profiles still be informative even in the absence of statistical significance. However, it is

also notable that some known colour genes were not identified in my transcriptomic data. For example, differential expression of endothelin 3 (*Edn3*) is known to be important in colour patterning in a diverse range of vertebrates, from domestic cats (Kaelin et al. 2012) to *Ambystoma* salamanders (Woodcock et al. 2017). However, none of the 35,926 transcripts that aligned to our reference transcriptome, whether sDE or not, could be associated with *Edn3*. While this may indicate that *Edn3* is not playing an important role in *Salamandra* colouration, it may also represent the need to further develop the reference transcriptome assembly, which is supported by the fact that 51.1–55.6% of the RNA-Seq reads per sample failed to map to it. Therefore, a valuable next step in this work would be to robustly annotate the reference transcriptome assembly, compile a list of all known vertebrate colour genes, identify which of these genes are present in *Salamandra* gene expression data and determine whether they show expression differences between skin colours/pattern morphs.

5.3.2. Adaptive colouration in Salamandra

Within *Salamandra*, colour patterns have been typically assumed to be adaptive for either aposematism (yellow-black patterns) or thermoregulation (melanism; Vences et al. 2014; Beukema et al. 2016b). However, my results from **Chapter 4** call into question both of these hypotheses, showing the need to more thoroughly assess the adaptive functionality of these colourations. First, in terms of aposematism, I found no correlation between the intensities (abundance) of toxin metabolites and colour phenotype, similar to a between species analyses by Vences et al. (2014). Although, it must be noted that sampling was limited in our analyses and toxic *Salamandra* secretions were both challenging to collect and showed a great deal of inter-individual variability. A logical first step would be increasing the sample size to obtain greater statistical power, or to conduct non-lethal mouse based toxicity trials, as has been done for poison frogs (e.g. Darst and Cummings 2006; Darst et al. 2006; Maan and Cummings 2012). However, this lack of correlation between colour and toxicity is not wholly unsurprising.

The supposed aposematic role of *Salamandra* colouration is based mainly on a correlation between their bright contrasting colour patterns and toxic secretions (Vences et al. 2014; Beukema et al. 2016b). However, the most robust test of this to date comes from a study looking at 'predation' on representative plasticine models (Velo-antón and Cordero-rivera 2011). Interestingly, this study found no evidence of bird attacks, which is notable, given that visually based predation by birds is considered a strong selective force in the

maintenance of colour polymorphisms (e.g. *Plethodon cinereus*; Fitzpatrick et al. 2009) and aposematic patterns (e.g. *Oophaga pumilio*; Dreher et al. 2015) in other amphibian species. In fact, predation on adult *Salamandra* in general is thought to be rare, with only a few documented cases by rats, snakes and possibly mink (Velo-antón and Cordero-rivera 2011; Böhme et al. 2013; Pezaro et al. 2017). It is also unclear whether *Salamandra* actively display their colour patterns when threatened. For example, some aposematic species, like fire-bellied toads (*Bombina* spp.; Bajger 1980) and rough skinned newt (*Taricha granulosa*; Johnson and Brodie 1975), will remain motionless and display their bright colour patterns when disturbed by a predator. Behavioural responses to predation risks are also know to differ between colour phenotypes in some polymorphic amphibians. For example, the North American red-backed salamander (*Plethodon cinereus*) presents both red-striped and lead phase (i.e. melanic) individuals. When disturbed by a predator, it has been found that red-striped individuals are more likely to remain immobile and display their colouration, while melanic individuals are more likely to run away (Venesky and Anthony 2007). Therefore, conducting similar behavioural studies on the response of *S. s. berhardezi* colour morphs to predation would be valuable in helping to evaluate the potential inter-species signalling of their colour patterns.

The assumptions of aposematism is also based largely on the human visual system, which is not always biologically relevant (Endler 1990; Stevens et al. 2007). For example, it has been shown that apparently conspicuous markings, which differ significantly from the background substrate, can offer a degree of camouflage to an animal by breaking up the outline of its body and making it hard for a predator to identify (Stevens et al. 2006). Recent studies have also combined spectrophotometry data with known animal photoreceptor sensitivities to model the conspicuousness of poison frogs to potential predators (Maan and Cummings 2012). Conducting similar studies with *S. s. bernardezi* would help to further elucidate the potential predator avoidance mechanisms underlying *Salamandra* colour patterns. However, it is also important to note that *Salamandra* spp. are predominantly nocturnal (Pezaro et al. 2017), which reduces the effectiveness of visual signalling.

A further problem with the assumption of aposematism in *Salamandra* is that this thesis represents the second study to find no correlation between colour phenotype and toxin metabolites in the genus (after Vences et al. 2014). Therefore, even if some *Salamandra* colour patterns do act as a true aposematic signal, this is clearly not the only selective pressure maintaining toxicity within the genus. Vences et al. (2014) speculate that the

melanic patterns of *S. atra* and *S. lanzai* may be highly conspicuous in certain Alpine habitats, although this this remains untested. If true, this would also deviate significantly from our current understanding of aposematic signals, which are thought to operate strictly through bright contrasting colour patterns (Prudic et al. 2007; Summers et al. 2015). An alternative explanation for the maintenance of toxicity in *Salamandra* could come from the fact that toxic steroidal alkaloid secretions in amphibians are known to have a distinctive smell and unpalatable or bitter taste (Myers and Daly 1976; Schulte et al. 2017; pers. obvs.). Therefore, it is possible that toxicity in melanic/hypolutic salamanders is being maintained through olfactory or gustatory based predator avoidance, and is not strongly linked to visually based aposematic signalling. Testing this would be reasonably straightforward. For example, studies looking at aposematism in poison frogs have shown that young birds can learn to associate unpalatability with colouration (Darst and Cummings 2006; Dreher et al. 2015). A similar study looking at olfactory or gustatory based predator avoidance could easily be conducted with a '*Salamandra* predator' like rats, by offering visually identical food items that differ in smell or taste depending on 'palatability'.

The other regularly hypothesised adaptive function of *Salamandra* colouration is thermoregulation at high altitudes (Vences et al. 2014). This is largely due to other studies showing that melanisation provides a thermal advantage in ectotherms (Clusella-Trullas et al. 2007), including both nonavian reptiles (Clusella-Trullas et al. 2008) and amphibians (e.g. *Rana temporaria*; Vences et al. 2002). While *S. s. bernardezi* do not display fully melanic morphs, like the Alpine species *S. atra* and *S. lanzai* (Bonato and Steinfartz 2005), within the rio Tendi/rio Color region of Asturias (northern Spain) they do display hypolutic (or brown) colouration (**Chapter 4**; Köhler and Steinfartz 2006; Beukema et al. 2016a). This is a variant of melanism, with both brown and black pigments being polymers of eumelanin (Bagnara et al. 1978; Meredith and Sarna 2006; Ito and Wakamatsu 2011). However, the broadly sympatric ecological space occupied by xanthic (yellow), striped and hypolutic *S. s. bernardezi* would seem to rule out large-scale environmental processes (like altitude) as the evolutionary driver of divergent colour patterns in this lineage. Although it is possible that more fine scale physiological and environmental processes are at work.

Given the lack of support found for aposematism and thermoregulation as drivers of colour diversification in *S. s. bernardezi*, future research must focus on identifying more subtle biological differences between colour morphs that may explain their evolutionary persistence. While this will take long, time consuming work, lessons can be learned from

studies on what may be the most well studied colour polymorphic salamander: *Plethodon cinereus*. In this species, red-striped and melanic salamanders differ in a number of ways, including their choice of microhabitat (Anthony et al. 2008; Fisher-Reid et al. 2013; Cosentino et al. 2017), mating behaviours (Acord et al. 2013), dispersal and movement patterns (Grant and Liebgold 2017) and diets (Stuczka et al. 2016). They also show differences in basal stress levels (Davis and Milanovich 2010), disease susceptibility and predation rates (Venesky et al. 2015; Grant and Liebgold 2017). Currently, there are no studies looking at such fine scale behavioural and ecological differences between *S. s. bernardezi* colour morphs, but they are all feasible. For instance, radio-telemetry has been used to study dispersal, movement/activity patterns, microhabitat selection and other behaviours in North American *Ambystoma* salamanders (Madison 1997, 1998; Madison and Farrand 1998; Montieth and Paton 2006; Rittenhouse and Semlitsch 2006; Veysey et al. 2009). However, while such studies may help identify subtle environmental, ecological or behavioural niche partitioning among *S. s. bernardezi* colour morphs, thereby helping to explain their evolutionary maintenance, it does not explain why this chromatic diversity is so geographically restricted (Köhler and Steinfartz 2006; Beukema et al. 2016a).

An interesting hypothesis that may help explain this regional variation in *Salamandra* colour patterning comes from Beukema et al. (2010). In this study, the authors hypothesised that the tendency of some populations of the North African *Salamandra algira tingitana* to be both hypolutic and pueriparous (giving birth to fully formed terrestrial juveniles) may be related to low levels of precipitation and a lack of available standing water. This is notable, as our ancestral state reconstruction analyses in **Chapter 3** showed a striking correlation in the evolutionary histories of both melanism/hypolutism and pueriparity. This potentially indicates that similar selective pressures can led to the evolution of these two complex traits; although it should be noted that *S. s. fastuosa* displays pueriparity and not melanism (García-París et al. 2003), and while *S. s. gallaica* does display both phenotypes, these occur in different populations (Velo-Antón et al. 2012; Velo-Antón and Buckley 2015). With these caveats in mind, it is interesting to consider that melanism in *Salamandra*, in its broadest sense, could represent a vertebrate example of the melanisation-desiccation hypothesis. Originally conceived through observations and experimental evolution studies in insect systems, like *Drosophila* spp. and mosquitoes, this hypothesis suggests that increased melanisation may be an adaptation for desiccation resistance (Rajpurohit et al. 2008, 2013; Rajpurohit and Nedved 2013; Ramniwas et al. 2013; Farnesi et al. 2017).This is thought to be the result of hydrophobic melanin pigments thickening or decreasing the permeability of the cuticle. While the mechanism is likely to

be different in *Salamandra*, this hypothesis would be easy to test by measuring rates of evaporative water loss (Tracy et al. 2008; Young et al. 2015) and statistically comparing this to *S. s. bernardezi* colour phenotype.

Above experiments could also be integrated with genomic analyses. In **Chapter 4**, I identified 43 genomic loci able to discriminate between our three representative colour morphs, and 142 genomic loci showing signals of selection. Using demographic modelling programmes, like fastsimcoal2 (Excoffier et al. 2013), it may be possible estimate divergence dates between colour morphs based on these sets these loci; a similar approach was recently used to estimate divergence times between discriminatory marine and freshwater stickleback alleles (Nelson and Cresko 2017; pre-print). Following this, it may then be possible to associate divergence estimates between colour phenotype associated alleles with past climatic events in northern Spain (such as cooling or changes in precipitation), thereby helping to elucidate potential climatic drivers of colour diversification in *Salamandra*.

5.3.3. Parallel and convergent colour pattern evolution

A particularly exciting outcome of this thesis is the groundwork it has laid for studying the parallel and convergent evolution of ecologically adaptive colour phenotypes. While the distinction is often vague and highly debated, parallel evolution typically refers to the independent emergence of similar phenotypes within closely related lineages, while convergent evolution is applied to the same process in more distantly related taxa (reviewed in Elmer and Meyer 2011). Further, with high-throughput sequencing techniques allowing for genomic studies on wild non-model organisms, parallel evolution has increasingly come to refer to both phenotypic and molecular parallelism (e.g. Bernatchez et al. 2016). Parallel systems are interesting as they provide natural evolutionary replicates in which to test the flexibility of evolutionary patterns and processes. For example, have the same or different molecular and phenotypic routes been taken? Or have different evolutionary trajectories converged on the same contemporary phenotype? Studies seeking to answer such questions have greatly benefited from the use of ecological genomics, which allows the relative contributions of stochastic genetic processes and natural selection, as well as standing genetic variation and *de novo* mutations, to be assessed (Elmer and Meyer 2011).

In **Chapter 3,** I provide a robust phylogenetic hypothesis for the genus *Salamandra* and identify the repeated evolution of two colour pattern phenotypes. One of these is the evolution of a striped phenotype in both *S. s. terrestris* and the *S. s. bernardezi* + *S. s. fastuosa* clade. Notably, *S. s. terrestris* has contemporary spotted and striped populations, and *S. s. bernardezi* striped and patternless phenotypes. Detailed transcriptomic and genomic comparisons of these two subspecies may reveal important loci involved in amphibian stripe and spot formation. The other parallel colour phenotype is melanism, which was found to have independently evolved in five lineages. This includes four taxa displaying fully melanic phenotypes (*S. atra atra*, *S. lanzai*, *S. salamandra gallaica*, and *S. algira tingitana*) and three displaying hypolutic (brown) colouration (*S. atra atra*, *S. algira tingitana* and *S. s. bernardezi*). Notably, all of these taxa, with the exception of *S. lanzai*, have closely related populations or subspecies displaying yellow-black patterned phenotypes, and two (*S. atra atra* and *S. algira tingitana*) display both melanic and hypolutic colouration. When this colour pattern diversity is combined with the quantitative colour phenotyping and candidate colour loci identified in **Chapter 4**, it establishes *Salamandra* as an ideal system in which to study parallelism in animal colouration at the phenotypic, molecular and environmental level.

More broadly, the work presented in this thesis also helps to fill a known knowledge gap in amphibian colour genetics (Hoffman and Blouin 2000; Rudh and Qvarnström 2013) and displays the utility of both transcriptomic and genomic NGS methods in identifying colour loci in large genome sized non-model organisms. Through this, my research will help facilitate studies on the convergence of colour patterns across vertebrates. For example, stripes and melanism are common colour pattern elements found throughout the animal kingdom. In addition, many toxic amphibian species have converged on a yellow-black spotted phenotype (e.g. *S. salamandra, A. maculatum* and *Ranitomeya vanzolinii*), but it is unclear if they have taken parallel or convergent molecular routes to these similar phenotypes. By expanding the work presented in this thesis to include other vertebrate taxa displaying similar colour phenotypes, we will be able to assess the repeatability and predictability of, as well as the constraints on, vertebrate colour pattern evolution by natural selection.

5.4. Conclusion

In conclusion, the results contained within this thesis present a robust phylogenetic hypothesis for the 'true' salamanders in the genus *Salamandra* and substantially contribute to our understanding of amphibian colour genetics. While some doubt remains regarding the exact topological positions of some taxa, our data shows that comparative phylogenomic analyses provides an effective route for assessing the evolutionary relationships between closely related species. However, our analyses also display the limitations of such methods, especially in the face of suspected introgression, which highlights the need to develop new analytical tools. In particular, a lack of reference genome and a lack of clarity regarding the impact of certain RAD-Seq data filtering parameters, specifically the placing of an upper bound on the number of variable sites per locus, were limiting factors in this study and are a priority for future work. However, through a combination of phylogenomics and ancestral state reconstruction analyses, I was able to confirm the independent, repeated evolution of two colour phenotypes: stripes and melanism. When combined with the candidate colour loci identified, and insights gained from the quantitative characterisation of colour phenotypes, this establishes *Salamandra* as an ideal system for studying the parallel evolution of adaptive colouration in animals. This will greatly contribute to our understanding of the repeatability and predictability of evolution by natural selection. However, this thesis also calls into question the means by which adaptive significance is attributed to animal colouration. While long assumed to be aposematic in function, *S. s. bernardezi* colour patterns do not correlate to toxin metabolite intensities. The sympatric nature of divergent colour morphs and an apparent lack of assortative mating by colour phenotype also suggests that broad environmental pressures like adaptation for thermoregulation and sexual selection are not important factors in the evolution of colour polymorphisms within this lineage. However, as signals of selection on genomic loci were found between individuals displaying different dorsal colour patterns, this suggests that more fine scale physiological or behavioural processes may be responsible for generating and maintaining this diversity, for example, desiccation resistance. While the ecological drivers of *S. s. bernardezi* colour diversification remain unclear and warrant further investigation, the results presented in this thesis lay the groundwork for a number of future evolutionary studies. Also, by shedding light on the genetic basis of amphibian colouration, which to date remains remarkably poorly understood, this thesis provides a rich resources for the comparative study of vertebrate pigmentation genetics.

Appendix 1: ddRAD-Seq sample metadata

Supplementary Table A1.1: Information for ddRAD-Seq genotyped samples included in comparative phylogenomic analyses (**Chapter 2**). Asterisks (*) = approximate GPS coordinates based on sample site information.

Sample No.	Species	Subspecies	Locality	Latitude	Longitude		Collector Names
ELT00415	*Salamandra corsica*	-	E Col de Badell, Corsica, France	41.809904	9.247292		Vences
ELT04698	*Salamandra salamandra*	*salamandra*	Nasenbach, Bavaria, Germany	48.189167	12.455833		Steinfartz
ELT04713	*Salamandra corsica*	-	Corsica, France	-	-		Liebetrau
ELT04734	*Salamandra salamandra*	*bernardezi*	Oviedo, Spain	43.361499	-5.849688	*	Liebetrau
ELT05454	*Salamandra lanzai*	-	Valle Po, Piedmont, Cuneo, Italy	44.942894	7.006494	*	Köpernik
ELT05493	*Salamandra infraimaculata*	*infraimaculata*	Ain Ghadran, Syria	-	-	*	Bogaerts
ELT05498	*Salamandra atra*	*pasubiensis*	Terra typica - Monte Pasubio, Italy	45.793528	11.176011	*	Steinfartz, Bonato, Romanazzi
ELT05541	*Salamandra algira*	*algira (?)*	Morocco	-	-	*	Vences, Joger
ELT05559	*Salamandra algira*	*algira*	Algeria	36.77785	4.966938	*	Merabat
ELT05575	*Lyciasalamandra billae*	-	Kale Tepe, Turkey	36.753315	30.553276	*	Veith
ELT05576	*Lyciasalamandra flavimembris*	-	Cicekli KöyÜ, Turkey	36.894451	28.280963	*	Veith
ELT05596	*Salamandra atra*	*atra*	Switzerland	46.875687	8.337644	*	Werner
ELT05609	*Salamandra infraimaculata*	*infraimaculata*	Tel Dan , Israel	33.249332	35.650891	*	Steinfartz

Supplementary Table A1.2: Information for ddRAD-Seq genotyped samples included in phylogenomic analyses of *Salamandra* species and subspecies (**Chapter 3**). Sample number superscripts: a = sample removed due to low coverage or misidentification; b = sample removed due to an excess of individuals from this locality; c = originally identified as *S. s. fastuosa*, but genetically clustered with *S. s. gallaica*; d = originally identified as *S. s. terrestris*, but genetically clustered with *S. s. salamandra*; e = originally identified as *S. s. salamandra*, but genetically clustered with *S. s. terrestris*; f = originally identified as *S. s. bejarae*, but geographically and genetically clustered with *S. s. gallaica*. Asterisks (*) = approximate GPS coordinates based on sample site information.

Sample No.	Species	Subspecies	Locality	Latitude	Longitude	Collector Names
ELT00415	*Salamandra corsica*	-	E Col de Badell, Corsica, France	41.809904	9.247292	Vences
ELT02678	*Salamandra corsica*	-	Corsica, France	-	-	Vences
ELT02690	*Salamandra salamandra*	*bernardezi*	Parque Natural de Ponga, Asturias, Spain	43.115833	-5.1765	Burgon
ELT02726	*Salamandra salamandra*	*bernardezi*	Parque Natural de Ponga, Asturias, Spain	43.125333	-5.170833	Burgon
ELT02806	*Salamandra salamandra*	*alfredschmidti*	Tendi Valley, Asturias, Spain	43.314333	-5.254167	Burgon
ELT02828	*Salamandra salamandra*	*alfredschmidti*	Tendi Valley, Asturias, Spain	43.305167	-5.246667	Burgon
ELT03076	*Salamandra salamandra*	*alfredschmidti*	Color Valley, Asturias, Spain	43.297833	-5.274	Burgon
ELT03078	*Salamandra salamandra*	*alfredschmidti*	Color Valley, Asturias, Spain	43.297833	-5.274	Burgon
ELT03089	*Salamandra salamandra*	*alfredschmidti*	Color Valley, Asturias, Spain	43.313333	-5.266167	Burgon
ELT03090	*Salamandra salamandra*	*alfredschmidti*	Color Valley, Asturias, Spain	43.313333	-5.266167	Burgon
ELT03104	*Salamandra salamandra*	*alfredschmidti*	Color Valley, Asturias, Spain	43.3005	-5.2735	Burgon
ELT03126	*Salamandra salamandra*	*alfredschmidti*	Color Valley, Asturias, Spain	43.297833	-5.274	Burgon
ELT03250	*Salamandra salamandra*	*alfredschmidti*	Tendi Valley, Asturias, Spain	43.305333	-5.246833	Burgon
ELT03268	*Salamandra salamandra*	*alfredschmidti*	Tendi Valley, Asturias, Spain	43.314333	-5.254	Burgon
ELT03366	*Salamandra salamandra*	*alfredschmidti*	Color Valley, Asturias, Spain	43.310833	-5.268	Burgon
ELT03459	*Salamandra salamandra*	*alfredschmidti*	Color Valley, Asturias, Spain	43.314	-5.268833	Burgon
ELT04698	*Salamandra salamandra*	*salamandra*	Nasenbach, Bavaria, Germany	48.189167	12.455833	Steinfartz
ELT04699	*Salamandra salamandra*	*salamandra*	Nasenbach, Bavaria, Germany	48.189167	12.455833	Steinfartz
ELT04700	*Salamandra algira*	*tingitana*	South Ceuta, Morroco	35.877894	-5.355353 *	Liebetrau
ELT04701	*Salamandra algira*	*tingitana*	South Ceuta, Morroco	35.877894	-5.355353 *	Liebetrau
ELT04702[a]	*Salamandra infraimaculata*	*infraimaculata*	Fevcipasa, Turkey	40.234148	28.252066 *	Liebetrau
ELT04703	*Salamandra salamandra*	*bernardezi*	Somiedo, Spain	43.110393	-6.257644 *	Liebetrau
ELT04704[a]	*Salamandra salamandra*	*bernardezi*	Somiedo, Spain	43.110393	-6.257644 *	Liebetrau
ELT04706	*Salamandra salamandra*	*gallaica*	Coimbra, Portugal	40.20342	-8.410251 *	Liebetrau
ELT04707	*Salamandra salamandra*	*gallaica*	Serra Estrela, Portugal	40.323121	-7.596032 *	Liebetrau

ELT04708	*Salamandra salamandra*	*gallaica*	Serra Estrela, Portugal	40.323121	-7.596032	*	Liebetrau
ELT04709	*Salamandra salamandra*	*bejarae*	Terra typica, Spain	40.324828	-5.97068	*	Liebetrau
ELT04710	*Salamandra salamandra*	*bejarae*	Terra typica, Spain	40.324828	-5.97068	*	Liebetrau
ELT04711	*Salamandra infraimaculata*	*infraimaculata*	Bater, Lebanon	33.601408	35.618063	*	Liebetrau
ELT04712	*Salamandra infraimaculata*	*infraimaculata*	Bater, Lebanon	33.601408	35.618063	*	Liebetrau
ELT04713	*Salamandra corsica*	-	Corsica, France	-	-	*	Liebetrau
ELT04714	*Salamandra salamandra*	*hispanica*	Sierra de Montseny, Spain	41.808909	2.382783	*	Liebetrau
ELT04715	*Salamandra salamandra*	*hispanica*	Sierra de Montseny, Spain	41.808909	2.382783	*	Liebetrau
ELT04717	*Salamandra salamandra*	*bernardezi*	Val de Los Infieles, Spain	43.312644	-5.059957	*	Liebetrau
ELT04718	*Salamandra salamandra*	*bernardezi*	Val de Los Infieles, Spain	43.312644	-5.059957	*	Liebetrau
ELT04719	*Salamandra salamandra*	*fastuosa*	Gorbeja, Basque Country, Spain	43.034961	-2.77988	*	Liebetrau
ELT04720	*Salamandra salamandra*	*fastuosa*	Gorbeja, Basque Country, Spain	43.034961	-2.77988	*	Liebetrau
ELT04721	*Salamandra salamandra*	*morenica*	Mittlere Sierra de Morena, north of Cordoba, Spain	38.417017	-4.766255	*	Liebetrau
ELT04722	*Salamandra salamandra*	*gallaica*	Sintra, Portugal, Spain	38.802875	-9.381652	*	Liebetrau
ELT04723	*Salamandra salamandra*	*crespoi*	Sierra de Monchique, Portugal	37.311126	-8.594759	*	Liebetrau
ELT04724	*Salamandra salamandra*	*crespoi*	Sierra de Monchique, Portugal	37.311126	-8.594759	*	Liebetrau
ELT04725	*Salamandra salamandra*	*salamandra*	Lago de Lugano, Tessin, Switzerland	45.989944	8.970312	*	Liebetrau
ELT04726	*Salamandra salamandra*	*salamandra*	Lago de Lugano, Tessin, Switzerland	45.989944	8.970312	*	Liebetrau
ELT04727	*Salamandra salamandra*	*terrestris*	Solling, Germany	51.819498	9.565885	*	Liebetrau
ELT04728	*Salamandra salamandra*	*salamandra*	Makarska, Croatia	43.311225	17.05175	*	Liebetrau
ELT04732	*Salamandra salamandra*	*beschkovi*	Pirin, Bulgaria	41.711527	23.401611	*	Liebetrau
ELT04733	*Salamandra salamandra*	*beschkovi*	Pirin, Bulgaria	41.711527	23.401611	*	Liebetrau
ELT04734	*Salamandra salamandra*	*bernardezi*	Oviedo, Spain	43.361499	-5.849688	*	Liebetrau
ELT04735	*Salamandra salamandra*	*bernardezi*	Oviedo, Spain	43.361499	-5.849688	*	Liebetrau
ELT04737[a]	*Salamandra salamandra*	*bejarae*	Central Spain, Spain	40.324828	-5.97068	*	Liebetrau
ELT04738	*Salamandra salamandra*	*terrestris*	Hanstedt south of Hamburg, Germany	53.258058	10.014472	*	Liebetrau
ELT04739	*Salamandra salamandra*	*terrestris*	Vannes,, France	47.657079	-2.757744	*	Liebetrau
ELT04740	*Salamandra salamandra*	*almanzoris*	Terra typica, Spain	40.253235	-5.27703	*	Liebetrau
ELT04741	*Salamandra salamandra*	*almanzoris*	Terra typica, Spain	40.253235	-5.27703	*	Liebetrau
ELT04742[c]	*Salamandra salamandra*	*gallaica*	south of lake Ebro, Spain	42.900064	-4.052433	*	Liebetrau
ELT04743[c]	*Salamandra salamandra*	*gallaica*	south of lake Ebro, Spain	42.900064	-4.052433	*	Liebetrau
ELT04744	*Salamandra salamandra*	*salamandra*	Lake Garda, Italy	45.655142	10.668407	*	Liebetrau
ELT04745	*Salamandra salamandra*	*salamandra*	Lake Garda, Italy	45.655142	10.668407	*	Liebetrau
ELT04746	*Salamandra salamandra*	*salamandra*	Stuttgart, Germany	48.778735	9.182707	*	Liebetrau
ELT04748[d]	*Salamandra salamandra*	*salamandra*	Carpates, Ukraine	48.510588	23.745884	*	Liebetrau
ELT04749	*Salamandra salamandra*	*salamandra*	Carpates, Ukraine	48.510588	23.745884	*	Liebetrau
ELT04750	*Salamandra salamandra*	*salamandra*	High Tatra, near Martin, Slovakia	49.167092	20.131615	*	Liebetrau

ELT04751	*Salamandra salamandra*	*salamandra*	High Tatra, near Martin, Slovakia	49.167092	20.131615	*	Liebetrau
ELT04752	*Salamandra salamandra*	*salamandra*	Loiblpass, Slovenia	46.433429	14.266635	*	Liebetrau
ELT04753	*Salamandra salamandra*	*salamandra*	Loiblpass, Slovenia	46.433429	14.266635	*	Liebetrau
ELT04754	*Salamandra salamandra*	*gigliolii*	Serra San Bruno, Italy	38.562652	16.322762	*	Liebetrau
ELT04755	*Salamandra salamandra*	*gigliolii*	Serra San Bruno, Italy	38.562652	16.322762	*	Liebetrau
ELT04756	*Salamandra salamandra*	*werneri*	Sparta, Greece	37.07614	22.441157	*	Liebetrau
ELT04757	*Salamandra salamandra*	*werneri*	Sparta, Greece	37.07614	22.441157	*	Liebetrau
ELT04758	*Salamandra salamandra*	*gallaica*	Serra de Grandola, Portugal	38.099579	-8.628056	*	Liebetrau
ELT04759	*Salamandra salamandra*	*gallaica*	Serra de Grandola, Portugal	38.099579	-8.628056	*	Liebetrau
ELT04760[b]	*Salamandra salamandra*	*terrestris*	Kottenforst stream (KOGC), Germany	50.659309	7.061018	*	Steinfartz
ELT04761[b]	*Salamandra salamandra*	*terrestris*	Kottenforst stream (KOGC), Germany	50.659309	7.061018	*	Steinfartz
ELT04762[b]	*Salamandra salamandra*	*terrestris*	Kottenforst stream (KOGC), Germany	50.659309	7.061018	*	Steinfartz
ELT04763[b]	*Salamandra salamandra*	*terrestris*	Kottenforst pond (KOIA), Germany	50.659309	7.061018	*	Steinfartz
ELT04764	*Salamandra salamandra*	*terrestris*	Kottenforst pond (KOIA), Germany	50.659309	7.061018	*	Steinfartz
ELT04765[b]	*Salamandra salamandra*	*terrestris*	Kottenforst pond (KOIA), Germany	50.659309	7.061018	*	Steinfartz
ELT04766	*Salamandra salamandra*	*terrestris*	Eifel stream (HeA), Germany	50.444809	6.679526	*	Steinfartz
ELT04767	*Salamandra salamandra*	*terrestris*	Eifel stream (HeA), Germany	50.444809	6.679526	*	Steinfartz
ELT04768	*Salamandra salamandra*	*terrestris*	Eifel stream (HeA), Germany	50.444809	6.679526	*	Steinfartz
ELT04769	*Salamandra salamandra*	*terrestris*	Eifel stream (Alt B), Germany	50.444809	6.679526	*	Steinfartz
ELT04770	*Salamandra salamandra*	*terrestris*	Eifel stream (Alt B), Germany	50.444809	6.679526	*	Steinfartz
ELT04771	*Salamandra salamandra*	*terrestris*	Eifel stream (Alt B), Germany	50.444809	6.679526	*	Steinfartz
ELT04772[b]	*Salamandra salamandra*	*terrestris*	Kottenforst pond (KOM), Germany	50.659309	7.061018	*	Steinfartz
ELT04773[b]	*Salamandra salamandra*	*terrestris*	Kottenforst pond (KOM), Germany	50.659309	7.061018	*	Steinfartz
ELT04774	*Salamandra salamandra*	*terrestris*	Kottenforst pond (KOM), Germany	50.659309	7.061018	*	Steinfartz
ELT04775	*Salamandra salamandra*	*terrestris*	Kottenforst stream (KOGb), Germany	50.659309	7.061018	*	Steinfartz
ELT04776	*Salamandra salamandra*	*terrestris*	Kottenforst stream (KOGb), Germany	50.659309	7.061018	*	Steinfartz
ELT04777[a]	*Salamandra salamandra*	*terrestris*	Kottenforst stream (KOGb), Germany	50.659309	7.061018	*	Steinfartz
ELT04778[b]	*Salamandra salamandra*	*terrestris*	Kottenforst stream (KOGd), Germany	50.659309	7.061018	*	Steinfartz
ELT04779	*Salamandra salamandra*	*terrestris*	Kottenforst stream (KOGd), Germany	50.659309	7.061018	*	Steinfartz
ELT04780[b]	*Salamandra salamandra*	*terrestris*	Kottenforst stream (KOGd), Germany	50.659309	7.061018	*	Steinfartz

ELT04781[b]	*Salamandra salamandra*	*terrestris*	Kottenforst pond (KOIC), Germany	50.659309	7.061018	*	Steinfartz
ELT04782[b]	*Salamandra salamandra*	*terrestris*	Kottenforst pond (KOIC), Germany	50.659309	7.061018	*	Steinfartz
ELT04783[b]	*Salamandra salamandra*	*terrestris*	Kottenforst pond (KOIC), Germany	50.659309	7.061018	*	Steinfartz
ELT04784	*Salamandra salamandra*	*terrestris*	Kottenforst pond (KORa), Germany	50.659309	7.061018	*	Steinfartz
ELT04785[b]	*Salamandra salamandra*	*terrestris*	Kottenforst pond (KORa), Germany	50.659309	7.061018	*	Steinfartz
ELT04786[b]	*Salamandra salamandra*	*terrestris*	Kottenforst pond (KORa), Germany	50.659309	7.061018	*	Steinfartz
ELT04787[b]	*Salamandra salamandra*	*terrestris*	Kottenforst pond (KOE), Germany	50.659309	7.061018	*	Steinfartz
ELT04788[b]	*Salamandra salamandra*	*terrestris*	Kottenforst pond (KOE), Germany	50.659309	7.061018	*	Steinfartz
ELT04789[b]	*Salamandra salamandra*	*terrestris*	Kottenforst pond (KOE), Germany	50.659309	7.061018	*	Steinfartz
ELT05441	*Salamandra salamandra*	*salamandra*	Buk, Hungary	47.385398	16.769027	*	Köpernik
ELT05442	*Salamandra salamandra*	*bernardezi*	Calabrez, Asturias, Spain	43.434897	-5.137931	*	Köpernik
ELT05443	*Salamandra salamandra*	*terrestris*	Heiden	47.444025	9.531792	*	Köpernik
ELT05444	*Salamandra salamandra*	*gigliolii*	Valdurasca, Liguria, Italy	44.146678	9.818026	*	Köpernik
ELT05445	*Salamandra salamandra*	*gallaica*	Serra da Arrábida, Portugal	38.500034	-9	*	Köpernik
ELT05446	*Salamandra salamandra*	*gigliolii*	Mount Creto, Italy	44.476236	9.009533	*	Köpernik
ELT05447	*Salamandra salamandra*	*werneri*	Pelion, Central Greece, Greece	39.4383	23.047532	*	Köpernik
ELT05448	*Salamandra salamandra*	*bernardezi*	Lake Ercina, Picos de Europa, Asturias, Spain	43.267699	-4.980999	*	Köpernik
ELT05449	*Salamandra salamandra*	*terrestris*	Pezzo, Italy	46.288134	10.521954	*	Köpernik
ELT05450	*Salamandra salamandra*	*longirostris*	Garzzalema, Spain	36.77359	-5.379402	*	Köpernik
ELT05451	*Salamandra salamandra*	*morenica*	Spain	-	-		Köpernik
ELT05452	*Salamandra salamandra*	*gallaica*	North Serra da Arrábida, Portugal	38.500034	-9	*	Köpernik
ELT05453[a]	*Salamandra salamandra*	*gallaica*	Monte do Faro (Pt), Faro, Portugal	37.036557	-7.924131	*	Köpernik
ELT05454	*Salamandra lanzai*	-	Valle Po, Piedmont, Cuneo, Italy	44.942894	7.006494	*	Köpernik
ELT05455	*Salamandra salamandra*	*fastuosa*	Bagneres de Bigorre, Midi-Pyrénées, France	43.06056	0.14389	*	Köpernik
ELT05456[d]	*Salamandra salamandra*	*salamandra*	East Thuringia, Germany	50.859665	11.521467	*	Köpernik
ELT05458[a]	*Salamandra salamandra*	*salamandra*	Lake Garda (Rocco Mountain), Italy	45.655142	10.668407	*	Köpernik
ELT05459	*Salamandra salamandra*	*fastuosa*	Markina-Xemein, Basque Country, Spain	43.268078	-2.496905	*	Köpernik
ELT05460	*Salamandra salamandra*	*gigliolii*	La Spezia, Liguria, Italy	44.110794	9.832547	*	Wawrzyniak
ELT05461	*Salamandra salamandra*	*gigliolii*	Carrara, Tuscany, Italy	44.078129	10.099207	*	Wawrzyniak
ELT05462	*Salamandra salamandra*	*fastuosa*	Bilbao, Biscay, Spain	43.261442	-2.935043	*	Wawrzyniak

ELT05464	*Salamandra infraimaculata*	*infraimaculata*	Harbiye, Turkey	36.133477	36.147716	*	Bogaerts
ELT05467[a]	*Salamandra infraimaculata*	*semenovi*	Kurdestan, Iran	38.830254	45.912813	*	Nick Poyarkov
ELT05469	*Salamandra infraimaculata*	*infraimaculata*	Ilic, Erzincan, Turkey	39.455681	38.561089	*	Bogaerts
ELT05470	*Salamandra infraimaculata*	*infraimaculata*	Kahramanmaras, Turkey	37.576713	36.936263	*	Bogaerts
ELT05474	*Salamandra infraimaculata*	*infraimaculata*	Kemaliye, Turkey	39.260935	38.496853	*	Bogaerts
ELT05476	*Salamandra infraimaculata*	*infraimaculata*	Aslantepe, Turkey	38.343394	38.344006	*	Bogaerts
ELT05478	*Salamandra infraimaculata*	*infraimaculata*	Eskiköy, Turkey	38.148102	38.02772	*	Bogaerts
ELT05483	*Salamandra infraimaculata*	*infraimaculata*	Fevzipasa, Turkey	37.093686	36.650833	*	Bogaerts
ELT05484	*Salamandra infraimaculata*	*infraimaculata*	Fevzipasa, Turkey	37.093686	36.650833	*	Bogaerts
ELT05485	*Salamandra infraimaculata*	*infraimaculata*	Lebanon	-	-		Bogaerts
ELT05486	*Salamandra infraimaculata*	*infraimaculata*	Lebanon	-	-		Bogaerts
ELT05487	*Salamandra infraimaculata*	*infraimaculata*	Lebanon	-	-		Bogaerts
ELT05493	*Salamandra infraimaculata*	*infraimaculata*	Ain Ghadran, Syria	-	-		Bogaerts
ELT05494	*Salamandra infraimaculata*	*infraimaculata*	Roman well source, Syria	-	-		Bogaerts
ELT05495	*Salamandra infraimaculata*	*infraimaculata*	Roman well source, Syria	-	-		Bogaerts
ELT05496	*Salamandra infraimaculata*	*infraimaculata*	Roman well source, Syria	-	-		Bogaerts
ELT05497	*Salamandra atra*	*pasubiensis*	Terra typica - Monte Pasubio, Italy	45.793528	11.176011	*	Steinfartz, Bonato, Romanazzi
ELT05498	*Salamandra atra*	*pasubiensis*	Terra typica - Monte Pasubio, Italy	45.793528	11.176011	*	Steinfartz, Bonato, Romanazzi
ELT05499	*Salamandra atra*	*pasubiensis*	Terra typica - Monte Pasubio, Italy	45.793528	11.176011	*	Steinfartz, Bonato, Romanazzi
ELT05500[e]	*Salamandra salamandra*	*terrestris*	Bielefeld, Teutoburger forest, Germany	52.018087	8.534875	*	Steinfartz, Krause
ELT05501[e]	*Salamandra salamandra*	*terrestris*	Bielefeld, Teutoburger forest, Germany	52.018087	8.534875	*	Steinfartz, Krause
ELT05503	*Salamandra salamandra*	*salamandra*	Poland	50.554869	16.431172	*	Pabijan
ELT05504	*Salamandra salamandra*	*salamandra*	Poland	50.554869	16.431172	*	Pabijan
ELT05505[f]	*Salamandra salamandra*	*gallaica*	Eume, Galicia, Spain	43.401672	-7.996936	*	Steinfartz, Vieteis
ELT05506[f]	*Salamandra salamandra*	*gallaica*	Eume, Galicia, Spain	43.401672	-7.996936	*	Steinfartz, Vieteis
ELT05507[f]	*Salamandra salamandra*	*gallaica*	Eume, Galicia, Spain	43.401672	-7.996936	*	Steinfartz, Vieteis
ELT05508[e]	*Salamandra salamandra*	*terrestris*	Detmold, Germany	51.938249	8.879773	*	Steinfartz, Krause
ELT05509[e]	*Salamandra salamandra*	*terrestris*	Detmold, Germany	51.938249	8.879773	*	Steinfartz, Krause
ELT05510[e]	*Salamandra salamandra*	*terrestris*	Kolvenbach-Eifel, Germany	50.525265	6.743456	*	Steinfartz
ELT05511[e]	*Salamandra salamandra*	*terrestris*	Kolvenbach-Eifel, Germany	50.525265	6.743456	*	Steinfartz
ELT05512	*Salamandra salamandra*	*salamandra*	Bayerischer Wald, Germany	48.998937	12.670165	*	Steinfartz
ELT05513	*Salamandra salamandra*	*salamandra*	Bayerischer Wald, Germany	48.998937	12.670165	*	Steinfartz
ELT05514[a]	*Salamandra salamandra*	*salamandra*	Gufflham (Altötting), Germany	48.232798	12.67185	*	Miller
ELT05515	*Salamandra salamandra*	*salamandra*	Spessart, Germany	49.917437	9.422391	*	Steinfartz
ELT05516	*Salamandra salamandra*	*salamandra*	Spessart, Germany	49.917437	9.422391	*	Steinfartz
ELT05517[e]	*Salamandra salamandra*	*terrestris*	Münster, Germany	51.961851	7.625465	*	Steinfartz

ELT05518[e]	*Salamandra salamandra*	*terrestris*	Münster, Germany	51.961851	7.625465	*	Steinfartz
ELT05519[e]	*Salamandra salamandra*	*terrestris*	Felderbachtal, Germany	51.339694	7.187215	*	Steinfartz
ELT05520[e]	*Salamandra salamandra*	*terrestris*	Felderbachtal, Germany	51.339694	7.187215	*	Steinfartz
ELT05521[f]	*Salamandra salamandra*	*gallaica*	Caurel-moutnains, Galicia, Spain	42.595601	-7.180929	*	Steinfartz, Vieites
ELT05522[f]	*Salamandra salamandra*	*gallaica*	Caurel-moutnains, Galicia, Spain	42.595601	-7.180929	*	Steinfartz, Vieites
ELT05523[a,f]	*Salamandra salamandra*	*gallaica*	Caurel-moutnains, Galicia, Spain	42.595601	-7.180929	*	Steinfartz, Vieites
ELT05524	*Salamandra salamandra*	*salamandra*	Ilz, Bavaria, Germany	48.696699	13.422388	*	Steinfartz
ELT05525	*Salamandra salamandra*	*salamandra*	Ilz, Bavaria, Germany	48.696699	13.422388	*	Steinfartz
ELT05526[e]	*Salamandra salamandra*	*terrestris*	Solling-Schießhaus, Germany	51.819498	9.565885	*	Steinfartz, Seidel
ELT05527[a,e]	*Salamandra salamandra*	*terrestris*	Solling-Schießhaus, Germany	51.819498	9.565885	*	Steinfartz, Seidel
ELT05528[e]	*Salamandra salamandra*	*terrestris*	Ahaus, Germany	52.100306	6.967512	*	Steinfartz
ELT05529	*Salamandra salamandra*	*terrestris*	Ahaus, Germany	52.100306	6.967512	*	Steinfartz
ELT05534	*Salamandra salamandra*	*salamandra*	Burgk (Saale), Germany	50.573591	11.686282	*	Steinfartz
ELT05535	*Salamandra salamandra*	*salamandra*	Burgk (Saale), Germany	50.573591	11.686282	*	Steinfartz
ELT05536	*Salamandra salamandra*	*salamandra*	Odenwald, Germany	49.590688	9.007306	*	Steinfartz
ELT05537	*Salamandra salamandra*	*salamandra*	Odenwald, Germany	49.590688	9.007306	*	Steinfartz
ELT05538	*Salamandra salamandra*	*salamandra*	Odenwald, Germany	49.590688	9.007306	*	Steinfartz
ELT05539[e]	*Salamandra salamandra*	*terrestris*	Westerwald, Germany	50.455234	7.482275	*	Nolte
ELT05540[e]	*Salamandra salamandra*	*terrestris*	Westerwald, Germany	50.455234	7.482275	*	Nolte
ELT05541	*Salamandra algira*	*algira* (MA)	Morocco	-	-		Vences, Joger
ELT05542	*Salamandra algira*	*algira* (MA)	Morocco	-	-		Vences, Joger
ELT05545	*Salamandra algira*	*algira*	Algeria	36.77785	4.966938	*	Merabat
ELT05546	*Salamandra algira*	*algira*	Algeria	36.77785	4.966938	*	Merabat
ELT05547	*Salamandra algira*	*algira*	Algeria	36.77785	4.966938	*	Merabat
ELT05548	*Salamandra algira*	*algira*	Algeria	36.77785	4.966938	*	Merabat
ELT05556[a]	*Salamandra algira*	*algira*	Algeria	36.77785	4.966938	*	Merabat
ELT05558	*Salamandra algira*	*algira*	Algeria	36.77785	4.966938	*	Merabat
ELT05559	*Salamandra algira*	*algira*	Algeria	36.77785	4.966938	*	Merabat
ELT05560	*Salamandra algira*	*algira*	Algeria	36.77785	4.966938	*	Merabat
ELT05561[a]	*Salamandra algira*	*algira*	Algeria	36.77785	4.966938	*	Merabat
ELT05562	*Salamandra algira*	*algira*	Algeria	36.77785	4.966938	*	Merabat
ELT05564	*Salamandra algira*	*algira*	Algeria	36.77785	4.966938	*	Merabat
ELT05567	*Salamandra algira*	*algira*	Algeria	36.77785	4.966938	*	Merabat
ELT05575	*Lyciasalamandra billae*	-	Kale Tepe, Turkey	36.753315	30.553276	*	Veith
ELT05576	*Lyciasalamandra flavimembris*	-	Cicekli KöyÜ, Turkey	36.894451	28.280963	*	Veith
ELT05578[a]	*Salamandra algira*	?	Morocco	-	-		Hauswaldt
ELT05579[a]	*Salamandra algira*	?	Morocco	-	-		Hauswaldt
ELT05580[a]	*Salamandra algira*	?	Morocco	-	-		Hauswaldt
ELT05582[a]	*Salamandra algira*	?	Morocco	-	-		Hauswaldt

ELT05583[a]	*Salamandra algira*	?	Morocco	-	-	Hauswaldt
ELT05585[a]	*Salamandra salamandra*	*longirostris*	Juanar (Marbella), Spain	36.580646	-4.885182 *	Vences, Donaire
ELT05586[a]	*Salamandra salamandra*	*longirostris*	Juanar (Marbella), Spain	36.580646	-4.885182 *	Vences, Donaire
ELT05587	*Salamandra salamandra*	*fastuosa*	Respomuso, Aragon, Spain	42.813159	-0.289534 *	Vences
ELT05588	*Salamandra salamandra*	*terrestris*	Ellhausen, Lohmar, Germany	50.85363	7.24862 *	Steinfartz
ELT05589	*Salamandra salamandra*	*terrestris*	Ellhausen, Lohmar, Germany	50.85363	7.24862 *	Steinfartz
ELT05590[b]	*Salamandra salamandra*	*terrestris*	Kottenforst (KO J), Germany	36.77785	4.966938 *	Steinfartz
ELT05591[b]	*Salamandra salamandra*	*terrestris*	Kottenforst (KO J), Germany	36.77785	4.966938 *	Steinfartz
ELT05592	*Salamandra salamandra*	*terrestris*	Königsdorf (near Cologne), Germany	50.94337	6.746704 *	Steinfartz
ELT05593	*Salamandra salamandra*	*terrestris*	Königsdorf (near Cologne), Germany	50.94337	6.746704 *	Steinfartz
ELT05594[a]	*Salamandra salamandra*	*terrestris*	Königsdorf (near Cologne), Germany	50.94337	6.746704 *	Steinfartz
ELT05595	*Salamandra salamandra*	*terrestris*	Königsdorf (near Cologne), Germany	50.94337	6.746704 *	Steinfartz
ELT05596	*Salamandra atra*	*atra*	Study site PhD Philine Werner,	46.875687	8.337644 *	Werner
ELT05597	*Salamandra atra*	*atra*	Study site PhD Philine Werner,	46.875687	8.337644 *	Werner
ELT05598	*Salamandra salamandra*	*terrestris*	Hohnstedter Holz (Wolfsburg), Germany	52.384388	10.699453 *	Steinfartz
ELT05599	*Salamandra salamandra*	*terrestris*	Hohnstedter Holz (Wolfsburg), Germany	52.384388	10.699453 *	Steinfartz
ELT05600	*Salamandra salamandra*	*terrestris*	Reitlingstal (Elm), Germany	52.209694	10.760311 *	Steinfartz
ELT05601	*Salamandra salamandra*	*terrestris*	Reitlingstal (Elm), Germany	52.209694	10.760311 *	Steinfartz
ELT05602	*Salamandra infraimaculata*	*infraimaculata*	Secher (Carmel mountain, Israel), Israel	32.764611	35.048568 *	Steinfartz
ELT05603	*Salamandra infraimaculata*	*infraimaculata*	Karreman (Carmel mountain, Israel), Israel	32.645435	34.991533 *	Steinfartz
ELT05604	*Salamandra infraimaculata*	*infraimaculata*	El Kamon (Lower Galilee, Israel), Israel	32.912779	35.3607 *	Steinfartz
ELT05605	*Salamandra infraimaculata*	*infraimaculata*	Harashim (Lower Galilee, Israel), Israel	32.957361	35.328756 *	Steinfartz
ELT05606	*Salamandra infraimaculata*	*infraimaculata*	Tel Dan, Israel	33.249332	35.650891 *	Steinfartz
ELT05607	*Salamandra infraimaculata*	*infraimaculata*	Tel Dan, Israel	33.249332	35.650891 *	Steinfartz
ELT05608	*Salamandra infraimaculata*	*infraimaculata*	Tel Dan, Israel	33.249332	35.650891 *	Steinfartz
ELT05609	*Salamandra infraimaculata*	*infraimaculata*	Tel Dan, Israel	33.249332	35.650891 *	Steinfartz
ELT06851	*Salamandra salamandra*	*morenica*	Miraflores, Riopar, Acbacete, Spain	38.494184	-2.483468 *	Donaire
ELT06852	*Salamandra salamandra*	*morenica*	Fuenfria, Albacete, Spain	38.56032	-2.402326 *	Donaire
ELT06853	*Salamandra salamandra*	*morenica*	Paymogo, Spain	37.740694	-7.343562 *	Donaire
ELT06854	*Salamandra salamandra*	*morenica*	Las Chinas, Aracena, Spain	37.932156	-6.724486 *	Donaire
ELT06855	*Salamandra salamandra*	*morenica*	Fuenteheridos, Spain	37.89767	-6.65952	Donaire

ELT06856	*Salamandra salamandra*	*morenica*	Valdearcos, Spain	37.94871	6.6808	Donaire
ELT06857	*Salamandra salamandra*	*morenica*	Sta Ana Real, Spain	37.8707	6.70956	Donaire
ELT06858	*Salamandra salamandra*	*morenica*	Fuenteheridos, Spain	37.89767	6.65952	Donaire
ELT06859	*Salamandra salamandra*	*longirostris*	M.P. de Jerez, Spain	36.54445	-5.66163	Donaire
ELT06861	*Salamandra salamandra*	*longirostris*	Spain	36.53741	-5.64304	Donaire
ELT06862	*Salamandra salamandra*	*longirostris*	Picacho, Laguna del Ingenl, Spain	36.52134	-5.6484	Donaire
ELT06863	*Salamandra salamandra*	*longirostris*	Srra Montecoche, Spain	36.27773	-5.57833	Donaire
ELT06864	*Salamandra salamandra*	*longirostris*	Srra de Montecoche, Spain	36.27773	-5.57833	Donaire
ELT06865	*Salamandra salamandra*	*longirostris*	Spain	36.51426	-5.66124	Donaire
ELT06866	*Salamandra salamandra*	*longirostris*	Picacho, Spain	36.52134	-5.6484	Donaire
ELT06867[a]	*Salamandra algira*	*algira* (MA)	9km south of Taza, Morocco	34.14792	-4.01038	Donaire
ELT06868	*Salamandra algira*	*algira* (MA)	Tazekka NP, Morocco	34.02011	-4.16876	Donaire
ELT06869	*Salamandra algira*	*algira* (MA)	Tazekka NP, Morocco	34.0053	-4.18821	Donaire
ELT06870	*Salamandra algira*	*algira* (MA)	Near Tazekka NP, Morocco	34.01545	-4.10036	Donaire
ELT06871	*Salamandra algira*	*splendens*	Morocco	34.99214	-4.82329	Donaire
ELT06872[a]	*Salamandra algira*	*tingitana*	Ain Remmel, Morocco	35.77172	-5.58754	Donaire
ELT06873	*Salamandra algira*	*tingitana*	Beni arus sub., Morocco	35.37035	-5.55413	Donaire
ELT06874	*Salamandra algira*	*tingitana*	Zemmij, Morocco	35.69422	-5.61039	Donaire
ELT06876	*Salamandra algira*	*tingitana*	Yebel Sougna, Morocco	35.12967	-5.36517	Donaire
ELT06877	*Salamandra algira*	*tingitana*	Yebel Sougna, Morocco	35.12967	-5.36517	Donaire
ELT06878	*Salamandra algira*	*splendens*	Sur Targuist, Morocco	34.75613	-4.2575	Donaire
ELT06880	*Salamandra algira*	*tingitana*	Bouhachem, Morocco	35.17441	-5.341705 *	Donaire
ELT06881	*Salamandra algira*	*tingitana*	Morocco	35.29043	-5.22994	Donaire
ELT06882	*Salamandra algira*	*splendens*	Xauen, Morocco	35.181541	-5.259572 *	Donaire
ELT06883	*Salamandra algira*	*splendens*	Xauen, Morocco	35.181541	-5.259572 *	Donaire
ELT06884	*Salamandra algira*	*splendens*	Morocco	35.20083	-5.2525	Donaire
ELT06885[a]	*Salamandra algira*	*tingitana*	Morocco	35.32806	-5.53361	Donaire
ELT06886	*Salamandra algira*	*splendens*	Ketama, Morocco	34.96064	-4.68036	Donaire
ELT06887	*Salamandra algira*	*splendens*	Ketama, Morocco	34.93607	-4.62913	Donaire
ELT06888	*Salamandra algira*	*tingitana*	Morocco	35.26391	-5.48813	Donaire
ELT06889	*Salamandra algira*	*tingitana*	Masmuda, Morocco	35.27453	-5.49386	Donaire
ELT06890	*Salamandra algira*	*tingitana*	Tangier population, Morocco	35.69422	-5.61039	Donaire
ELT06891	*Salamandra algira*	*tingitana*	Tangier population, Morocco	35.69422	-5.61039	Donaire
ELT06895	*Salamandra algira*	*splendens*	Chefchaoun 2, Morocco	35.165724	-5.269865 *	Donaire
ELT06896	*Salamandra algira*	*spelea*	Berkane beni snassen, Morocco	34.813436	-2.384466 *	Donaire
ELT06897	*Salamandra algira*	*spelea*	Berkane beni snassen 2, Morocco	34.813436	-2.384466 *	Donaire

ELT06898	*Salamandra algira*	*spelea*	Berkane beni snassen 2, Morocco	34.813436	-2.384466 *	Donaire
ELT06899	*Salamandra algira*	*splendens*	Chefchaoun 10, Morocco	35.165724	-5.269865	Donaire
ELT06900[a]	*Salamandra salamandra*	*salamandra*	Fagaras Mountains, Fagaras, Romania	45.652683	24.603267	Recknagel
ELT08282	*Salamandra algira*	*tingitana*	molay abdslam 3, Morocco	35.318644	-5.508324	Donaire
ELT08283	*Salamandra salamandra*	*morenica*	Jaen, Srra Cazorla, Spain	38.33453	-2.62129	Donaire
ELT08284	*Salamandra algira*	*tingitana*	West of Ain Lahcen, Morocco	35.55157	-5.56406	Donaire
ELT08285	*Salamandra algira*	*splendens*	Cudia Sbaa, Morocco	35.02526	-5.01965	Donaire
ELT08286	*Salamandra algira*	*tingitana*	Molay Abdslam, Morocco	35.318644	-5.508324 *	Donaire
ELT08287	*Salamandra algira*	*tingitana*	Molay Abdslam, Morocco	35.318644	-5.508324 *	Donaire
ELT08288	*Salamandra salamandra*	*longirostris*	M.P. de Jerez, Spain	36.54445	-5.66163	Donaire

Supplementary Table A1.3: Information for ddRAD-Seq genotyped samples included in population genetic and colour phenotype-genotype association analyses (**Chapter 4**).

Sample No.	Colour morph	Sex	Locality	Latitude	Longitude	Altitude (masl)	Collector Names
ELT02473	Striped	Female	Peloño	43.20158	-5.13812	999	Elmer, Vieites, Fernández
ELT02474	Striped	Female	Peloño	43.20158	-5.13812	999	Elmer, Vieites, Fernández
ELT02690	Striped	Female	rio Ponga Valley (south of Sobrefoz)	43.115833	-5.1765	1294	Burgon, Elmer, Williams, Ramos
ELT02716	Striped	Male	rio Ponga Valley (south of Sobrefoz)	43.138167	-5.175	945	Burgon, Elmer, Williams, Ramos
ELT02726	Striped	Female	rio Ponga Valley (south of Sobrefoz)	43.125333	-5.170833	994	Burgon, Elmer, Williams, Ramos
ELT02726	Striped	Female	rio Ponga Valley (south of Sobrefoz)	43.125333	-5.170833	994	Burgon, Elmer, Williams, Ramos
ELT02742	Striped	Female	rio Ponga Valley (south of Sobrefoz)	43.1335	-5.1735	993	Burgon, Elmer, Williams, Ramos
ELT02745	Striped	Female	rio Ponga Valley (south of Sobrefoz)	43.1365	-5.1725	944	Burgon, Elmer, Williams, Ramos
ELT02770	Striped	Female	rio Tendi Valley	43.155167	-5.176333	876	Burgon, Elmer, Williams, Ramos
ELT02783	Striped	Male	rio Tendi Valley	43.132667	-5.173	998	Burgon, Elmer, Williams, Ramos
ELT02795	Striped	Male	rio Tendi Valley	43.314333	-5.254167	259	Burgon, Elmer, Williams, Ramos
ELT02796	Yellow	Male	rio Tendi Valley	43.314333	-5.254167	259	Burgon, Elmer, Williams, Ramos
ELT02797	Brown	Male	rio Tendi Valley	43.314333	-5.254167	259	Burgon, Elmer, Williams, Ramos
ELT02800	Yellow	Female	rio Tendi Valley	43.314333	-5.254167	259	Burgon, Elmer, Williams, Ramos
ELT02805	Striped	Male	rio Tendi Valley	43.314333	-5.254167	259	Burgon, Elmer, Williams, Ramos
ELT02806	Yellow	Male	rio Tendi Valley	43.314333	-5.254167	259	Burgon, Elmer, Williams, Ramos
ELT02815	Brown	Male	rio Tendi Valley	43.308667	-5.248833	284	Burgon, Elmer, Williams, Ramos
ELT02819	Yellow	Male	rio Tendi Valley	43.308667	-5.248833	284	Burgon, Elmer, Williams, Ramos
ELT02822	Brown	Male	rio Tendi Valley	43.305167	-5.246667	283	Burgon, Elmer, Williams, Ramos
ELT02828	Yellow	Male	rio Tendi Valley	43.305167	-5.246667	283	Burgon, Elmer, Williams, Ramos
ELT02843	Yellow	Female	Rio de la Marea Valley	43.305167	-5.246667	283	Burgon, Elmer, Williams, Ramos
ELT02846	Yellow	Male	Rio de la Marea Valley	43.305167	-5.246667	283	Burgon, Elmer, Williams, Ramos
ELT02855	Yellow	Male	Rio de la Marea Valley	43.270833	-5.4135	464	Burgon, Elmer, Williams, Ramos
ELT02856	Striped	Female	Rio de la Marea Valley	43.270833	-5.4135	464	Burgon, Elmer, Williams, Ramos
ELT02861	Yellow	Male	Rio de la Marea Valley	43.270833	-5.4135	464	Burgon, Elmer, Williams, Ramos
ELT02862	Yellow	Male	Rio de la Marea Valley	43.270833	-5.4135	464	Burgon, Elmer, Williams, Ramos
ELT02863	Yellow	-	Rio de la Marea Valley	43.270833	-5.4135	464	Burgon, Elmer, Williams, Ramos
ELT02864	Yellow	-	Rio de la Marea Valley	43.270833	-5.4135	464	Burgon, Elmer, Williams, Ramos
ELT02866	Yellow	Female	rio Tendi Valley	43.270833	-5.417	78	Burgon, Elmer, Williams, Ramos
ELT02868	Striped	-	rio Tendi Valley	43.270833	-5.417	78	Burgon, Elmer, Williams, Ramos

ELT02906	Brown	Female	rio Color Valley	43.299833	-5.246667	310	Burgon, Elmer, Williams, Ramos
ELT02908	Brown	Female	rio Color Valley	43.299833	-5.246667	310	Burgon, Elmer, Williams, Ramos
ELT02960	Brown	Male	rio Color Valley	43.305833	-5.271	310	Burgon, Williams, Ramos
ELT03027	Brown	Male	rio Color Valley	43.317	-5.263833	302	Burgon, Williams, Ramos
ELT03076	Striped	Female	rio Color Valley	43.297833	-5.274	414	Burgon, Williams, Ramos
ELT03078	Striped	Male	rio Color Valley	43.297833	-5.274	414	Burgon, Williams, Ramos
ELT03089	Yellow	Male	rio Color Valley	43.313333	-5.266167	233	Burgon, Williams, Ramos
ELT03090	Yellow	Female	rio Color Valley	43.313333	-5.266167	233	Burgon, Williams, Ramos
ELT03104	Brown	Male	rio Color Valley	43.3005	-5.2735	322	Burgon, Williams, Ramos
ELT03124	Striped	Male	rio Color Valley	43.297833	-5.274	414	Burgon, Williams, Ramos
ELT03126	Yellow	Male	rio del Infierno Valley	43.297833	-5.274	414	Burgon, Williams, Ramos
ELT03139	Striped	Female	rio del Infierno Valley	43.3005	-5.2735	322	Burgon, Williams, Ramos
ELT03165	Yellow	Female	rio del Infierno Valley	43.2545	-5.3185	552	Burgon, Williams, Ramos
ELT03166	Yellow	Male	rio del Infierno Valley	43.2545	-5.3185	552	Burgon, Williams, Ramos
ELT03167	Striped	Male	rio del Infierno Valley	43.2545	-5.3185	552	Burgon, Williams, Ramos
ELT03170	Striped	Female	rio del Infierno Valley	43.262167	-5.329833	450	Burgon, Williams, Ramos
ELT03177	Yellow	Male	rio del Infierno Valley	43.2545	-5.3185	552	Burgon, Williams, Ramos
ELT03178	Striped	Female	rio del Infierno Valley	43.2545	-5.3185	552	Burgon, Williams, Ramos
ELT03181	Striped	Female	rio del Infierno Valley	43.262167	-5.329833	450	Burgon, Williams, Ramos
ELT03185	Yellow	Male	rio del Infierno Valley	43.262167	-5.329833	450	Burgon, Williams, Ramos
ELT03188	Striped	Female	rio Tendi Valley	43.262167	-5.329833	450	Burgon, Williams, Ramos
ELT03190	Yellow	Female	rio Tendi Valley	43.2565	-5.317667	528	Burgon, Williams, Ramos
ELT03193	Brown	Female	rio Tendi Valley	43.314167	-5.254	225	Burgon, Williams, Ramos
ELT03214	Yellow	Male	rio Tendi Valley	43.305333	-5.246833	277	Burgon, Williams, Ramos
ELT03246	Striped	Female	rio Tendi Valley	43.305333	-5.246833	277	Burgon, Williams, Ramos
ELT03250	Brown	Female	rio Tendi Valley	43.305333	-5.246833	277	Burgon, Williams, Ramos
ELT03268	Brown	Female	rio Tendi Valley	43.314333	-5.254	253	Burgon, Williams, Ramos
ELT03280	Striped	Female	rio Tendi Valley	43.305333	-5.246833	277	Burgon, Williams, Ramos
ELT03302	Yellow	Female	rio Color Valley	43.305333	-5.246833	277	Burgon, Williams, Ramos
ELT03303	Yellow	Female	rio Color Valley	43.305333	-5.246833	277	Burgon, Williams, Ramos
ELT03344	Yellow	Male	rio Color Valley	43.306	-5.271	273	Burgon, Williams, Ramos
ELT03347	Striped	Female	rio Color Valley	43.306	-5.271	273	Burgon, Williams, Ramos
ELT03353	Striped	Female	rio Color Valley	43.306	-5.271	273	Burgon, Williams, Ramos
ELT03354	Brown	Female	rio Color Valley	43.306	-5.271	273	Burgon, Williams, Ramos

ELT03366	Brown	Male	rio Color Valley	43.310833	-5.268	283	Burgon, Williams, Ramos
ELT03371	Yellow	Male	rio Color Valley	43.310833	-5.268	283	Burgon, Williams, Ramos
ELT03372	Brown	Male	rio Color Valley	43.310833	-5.268	283	Burgon, Williams, Ramos
ELT03382	Striped	Male	rio Color Valley	43.306	-5.271	273	Burgon, Williams, Ramos
ELT03392	Brown	Male	rio Color Valley	43.3165	-5.2635	317	Burgon, Williams, Ramos
ELT03447	Yellow	Male	rio Color Valley	43.310833	-5.268	283	Burgon, Williams, Ramos
ELT03448	Yellow	Female	rio Color Valley	43.314	-5.268833	343	Burgon, Williams, Ramos
ELT03450	Brown	Male	rio Color Valley	43.314	-5.268833	343	Burgon, Williams, Ramos
ELT03459	Brown	Male	rio Color Valley	43.314	-5.268833	343	Burgon, Williams, Ramos
ELT03486	Brown	Female	rio Color Valley	43.310833	-5.268	283	Burgon, Williams, Ramos
ELT03497	Yellow	Female	rio Color Valley	43.310833	-5.268	283	Burgon, Williams, Ramos
ELT03498	Yellow	Female	rio Color Valley	43.310833	-5.268	283	Burgon, Williams, Ramos
ELT05627	Striped	Male	Near Avín	43.334167	-4.941167	-	Elmer, Vieites, Fernández
ELT05630	Striped	Male	Near Avín	43.334167	-4.941167	-	Elmer, Vieites, Fernández
ELT05632	Striped	Female	Near Avín	43.334167	-4.941167	-	Elmer, Vieites, Fernández
ELT05633	Striped	Female	Near Avín	43.334167	-4.941167	-	Elmer, Vieites, Fernández
ELT05639	Striped	Female	Belleno	43.18871	-5.16089	679	Elmer, Vieites, Fernández
ELT05640	Striped	Male	Belleno	43.18871	-5.16089	679	Elmer, Vieites, Fernández

Appendix 2: Molecular protocols

A2.1 Genomic DNA extraction Protocol

(Macherey-Nagel NucleoSpin® Tissue kit)

1. Disintegrate tissue in 180µl Buffer T1
2. Add 25µl Proteinase K
   - Vortex and spin down
3. Incubate at 56°C overnight in shaking incubator
4. Turn on hot plate (70°C) and put elution buffer (EB) in the oven
5. Vortex samples; add 200µl Buffer B3
   - Vortex and incubate at 70°C for 10 min
6. Add 210µl absolute (100%) ethanol to each sample and vortex
7. Move samples to fresh spin column and centrifuge for 1min (11krpm)
   - Discard flow through
8. Wash 1:
   - Add 500µl Buffer BW and centrifuge for 1min (11krpm)
   - Discard flow through
9. Wash 2:
   - Add 600µl Buffer B5 and centrifuge for 1min (11krpm)
   - Discard flow through
10. Spin dry 1 min (11krpm)
11. Elute DNA
    - Move spin column to labelled elution tube
    - Add 50µl pre-warmed Buffer EB
    - Stand for 2min
    - Centrifuge for 1min (11krpm)
    - Repeat
12. Store at 4°C
13. Run sample on agarose gel
    - 120ml 2% gel with ethidium bromide
      - 5µl Ladder to either side of samples
      - Samples: 3µl Orange G + 5µl DNA elution
      - Run at 100V for ~30min

A2.2 Illumina ddRAD-Seq protocol

Modified from Peterson et al. 2012 *PLoS ONE* (see also Recknagel et al. 2015 *Mol. Ecol. Res*). This protocol is a modified version of one written by A. Jacobs and H. Recknagel.

1) DNA extraction

1. Extract genomic DNA from tissue using the NucleoSpin® Tissue Kit (Macherey-Nagel). *See above*.
2. Quantify DNA using a Qubit® 2.0 Fluorometer (Invitrogen): c_{DNA} values should be ≥24ng/µl.

2) Digestion with double restriction enzyme digestion

1. Digest 1µg of DNA with two restriction enzymes:
   - 5µl of 10x New England Biolabs CutSmart® Buffer
   - 1µl of *PstI-HF*® (20kU/ml)
   - 1µl of *AclI* (5kU/ml)
   - 1µg of DNA
   - fill up with ddH_2O to end volume of 50µl

   → incubate at 37ºC for 3 hours

2. Let the reaction cool down to room temperature (if you proceed next day, store reaction at 4ºC) and proceed with cleaning with NucleoSpin® Gel and PCR Clean-up (Macherey-Nagel), elute to 22 µl and measure concentrations with the Qubit® 2.0 Fluorometer.

3) Adapter ligation (*see table A2.1 for the barcodes used in each library*)

1. Prepare stocks of adapter single strand oligos (100µ*M* stock in 1X Elution Buffer EB: 10 m*M* Tris-Cl, pH 8.5). For the working solution combine adapters at 10µ*M* stock in 1X AB (10X AB: 500 m*M* NaCl, 100 m*M* Tris-Cl, pH 7.5-8.0; e.g. 1µl P1_forward + 1µl P1_reverse + 8µl 1xAB). Put combined adapters in a thermocycler and incubate at 97.5ºC for 2:30 min, then cool 3ºC/minute down to 21ºC. Store at 4ºC.
2. Assign barcode combination to each individual and note it down (IMPORTANT).

3. Prepare in PCR single tubes:
    - 0.5µl T4 ligase (2,000 U/µl)
    - 4µl 10x T4 ligation buffer
    - 0.5µl Barcoded P1 Adapter (10µ*M*)
    - 0.5µl Barcoded P2 Adapter (10µ*M*)
    - up to 20µl DNA sample
    - ddH_2O up to 40.0µl total volume

Note: Use the DNA sample with the smallest concentration as basis (multiply the concentration by 20µl, your total volume) and then use the same amount of DNA for all samples (e.g. if smallest concentration is 30ng/µl, then your sample contains 600ng in 20µl. Adjust the volume of all other samples to contain the same amount of DNA, e.g. if concentration is 50 ng/µl then use 12 µl).

Ligate for 30 min at 25ºC, followed by heat kill @ 65ºC for 10 min, then slowly cool down to room temperature in thermocycler (2ºC per 90 sec).

4) Size selection

1. Samples should be multiplexed before the size selection.
2. Clean multiplexed samples using NucleoSpin® Gel and PCR Clean-up kit (Macherey-Nagel). Elute in 32µl (16µl + 16µl) for Pippin Prep Size selection.
3. Check fragment size distribution on a 2200 TapeStation instrument (Genomic DNA Tape; Agilent Technologies)
4. Size select on a gel. For efficiently sequencing on an Illumina sequencer, fragments should be between 150–400 nucleotides. Increasing or decreasing the range can optimize the amount of total loci.
    - Run sample on 2% agarose dye-free cassette (CEF-2010) with Sage Science PippinPrep and marker E.
    - Elute in 40µl PippinPrep buffer
    - Quantify 1µl of size selected DNA using Qubit® 2.0 Fluorometer (high sensitivity DNA assay)

5) RAD tag enrichment

1. Perform 5-8 PCR amplifications using Phusion® High Fidelity DNA Polymerase MasterMix kit.

- In thin-walled PCR tube, combine:
  10µl Phusion High Fidelity MasterMix
  1.0µl of RAD F primer (10µ*M*): CCACTACGCCTCCGCTTTCC
  1.0µl of RAD R primer (10µ*M*): CCATCTCATCCCTGCGTGTCT
  2.5–10ng of RAD library template (max. 8 µl)
  Fill up to 20 µl with double-distilled water (ddH_2O).
- Perform 10 cycles of amplification in thermal cycler: 30 sec 98ºC, 9X [10 sec 98ºC, 30 sec 65ºC, 30 sec 72ºC], 5 min 72ºC, hold 4ºC.
- Run 5.0 µl PCR product in 1X Orange Loading Dye out on 1.0% agarose gel next to ¼ of the volume you used as DNA template and 2.0µl GeneRuler 100 bp DNA marker. The PCR sample should be much brighter then the template.
- Combine all reactions and clean sample using NucleoSpin Gel and PCR Clean up kit (Macherey-Nagel) and elute in 20µl

2. Run a gel for cleaning from contaminants
   - Load entire sample in 1X Orange Loading Dye on a 1.25% agarose, 1X TBE gel (with SYBRsafe) and run for 45min at 100V with 2.0µl 100 bp DNA Ladder for size reference
   - Take a picture of the gel (use filter for bluelight). Extract DNA using Qiagen MinElute Gel Extraction Kit following manufacturer's instructions. Only 400mg of Gel slices and maximally 5µg can be processed per column
   - Elute each column probe in 2x10µl EB and combine all probes

3. Quantify the DNA using Qubit® 2.0 Fluorometer. Concentrations will range from 1–20ng/µl. In addition, run the sample on the Bioanalyzer/TapeStation compared to control (size selected DNA, if there is any left!) and your library standard (the first library of your project as reference). Compare the range of fragments on the gel picture and the concentration of your Qubit® 2.0 Fluorometer results with the Bioanalyzer results (including range and mean of fragment sizes and total quantity). Assess that the size matches what you cut from the pippin (accommodating any increase after enrichment)

6) Sequencing (by Polyomics)

1. Send library for sequencing on Illumina platform. Retain 10µl back in the lab as reference for future library sizes

A2.3 Total RNA extraction protocol

1) Tissue homogenisation:

1. Add lysis buffer to Fast-Prep Matrix A tubes (1000µl per 50mg tissue), and label all necessary tubes
2. Add tissue and Fast-Prep: speed 5, 25sec, x4 (leaving a short brake between runs)
   - Ensure tissue is fully homogenised, if not run again
3. Let the tubes settle (~1min), then carefully move supernatant to clean labelled 1.5ml microcentrifuge tubes
   - Avoid the foam at the top of the tube and the deposits at the bottom
4. Centrifuge: 2600 rpm, 5min, room temperature.
5. Transfer supernatant to new 1.5ml tubes, avoiding the pellet of debris
   - Max 700µl, use more than one tube if necessary

2) Extraction (Ambion® PureLink™ RNA Mini Kit):

1. Add 1 volume 70% EtOH (max 700µl): vortex
2. Add 700µl of sample to new labelled spin column (in collection tube)
3. Centrifuge: 13000 rpm, 30–60sec, room temperature
   - Discard flow through
4. Repeat the above step until your entire sample has passed through the column
5. Add 350µl Wash buffer 1, centrifuge: 13,000 rpm, 30-60sec, room temp.
   - Discard flow through
6. Apply DNAse treatment (80µl)- incubate, room temp. 15min
7. Add 350µl Wash buffer 1, centrifuge: 13,000 rpm, 30-60sec, room temp
   - Discard flow through
   - Move spin column to new collection tube
8. Add 500µl Wash buffer 2, centrifuge: 13,000 rpm, 1min, room temp
   - Discard flow through
9. Repeat Wash buffer 2
10. Wash with 750µl 80% EtOH, centrifuge: 13,000 rpm, 1min, room temp.
    - Discard flow through
11. Repeat 70% alcohol (EtOH) wash
12. Move spin column to new 1.5ml tube (with lid cut off)
13. Spin dry: 13,000 rpm, 5min, room temp
14. Place spin columns into labelled recovery tubes

15. Elute RNA:
    - Add 30µl RNase free water
    - Incubate for 1min at room temperature
    - Centrifuge: 13,000 rpm, 2min, room temperature
    - Repeat elution step
16. Store at -70°C

3) Lithium Chloride clean up (4M):

1. Add 1 volume 4M LiCl, spin down, and store overnight at -20 °C
2. Centrifuge: max speed, 30min, 4°C
    - Chill 70% EtOH (on ice or freezer)
3. Remove supernatant: discard
4. Wash pellet in chilled 70% EtOH (180µl)
5. Centrifuge: max speed, 2min, 4°C
6. Remove supernatant (set pipette to 200µl)
7. Dry pellet: 5–10min MAX! (do not over dry)
8. Re-suspend in 10–20µl RNase free water
9. Store at -70°C

A2.4 ddRAD-Seq barcode table

Table A2.1: Combinatorial barcodes for each sample used in ddRAD-Seq libraries, with in house library name (ID) and associated data chapters.

Library	ID	Chapters	ELT	P1		P2	
1	EZT	3/4	ELT02690	**sP1_1**	CTCC	**P2_2**	CCT
1	EZT	3/4	ELT03089	**sP1_2**	TGCA	**P2_2**	CCT
1	EZT	3/4	ELT03104	**sP1_3**	ACTA	**P2_2**	CCT
2	Geno	4	ELT03448	**sP1_1**	CTCC	**P2_1**	TAG
2	Geno	4	ELT03190_2	**sP1_2**	TGCA	**P2_1**	TAG
2	Geno	4	ELT02856	**sP1_3**	ACTA	**P2_1**	TAG
2	Geno	4	ELT02866_1	**sP1_12**	TGCGA	**P2_1**	TAG
2	Geno	4	ELT03498	**sP1_13**	CGCTT	**P2_1**	TAG
2	Geno	4	ELT03177	**sP1_14**	TCACC	**P2_1**	TAG
2	Geno	4	ELT03372	**sP1_34**	GGTTGT	**P2_1**	TAG
2	Geno	4	ELT02796	**sP1_35**	CCAGCT	**P2_1**	TAG
2	Geno	4	ELT02738	**sP1_36**	TTCAGA	**P2_1**	TAG
2	Geno	4	ELT03188	**sP1_57**	CTTGCTT	**P2_1**	TAG
2	Geno	4	ELT05639	**sP1_58**	ATGAAAC	**P2_1**	TAG
2	Geno	4	ELT03371_2	**sP1_60**	GAATTCA	**P2_1**	TAG
2	Geno	4	ELT03392	**sP1_78**	ACGACTAC	**P2_1**	TAG
2	Geno	4	ELT02783	**sP1_79**	TAGCATGC	**P2_1**	TAG
2	Geno	4	ELT05633	**sP1_83**	TAGGCCAT	**P2_1**	TAG
2	Geno	4	ELT05627	**sP1_1**	CTCC	**P2_4**	GAGC
2	Geno	4	ELT02819_1	**sP1_2**	TGCA	**P2_4**	GAGC
2	Geno	4	ELT02861_1	**sP1_3**	ACTA	**P2_4**	GAGC
2	Geno	4	ELT02960	**sP1_12**	TGCGA	**P2_4**	GAGC
2	Geno	4	ELT03497	**sP1_13**	CGCTT	**P2_4**	GAGC
2	Geno	4	ELT03303	**sP1_14**	TCACC	**P2_4**	GAGC
2	Geno	4	ELT03353	**sP1_34**	GGTTGT	**P2_4**	GAGC
2	Geno	4	ELT03371_1	**sP1_35**	CCAGCT	**P2_4**	GAGC
2	Geno	4	ELT03178	**sP1_36**	TTCAGA	**P2_4**	GAGC
2	Geno	4	ELT02855	**sP1_57**	CTTGCTT	**P2_4**	GAGC
2	Geno	4	ELT03167	**sP1_58**	ATGAAAC	**P2_4**	GAGC
2	Geno	4	ELT02862	**sP1_60**	GAATTCA	**P2_4**	GAGC
2	Geno	4	ELT03302	**sP1_78**	ACGACTAC	**P2_4**	GAGC
2	Geno	4	ELT02868	**sP1_79**	TAGCATGC	**P2_4**	GAGC
2	Geno	4	ELT03166	**sP1_83**	TAGGCCAT	**P2_4**	GAGC
2	Geno	4	ELT03447	**sP1_1**	CTCC	P2_5	CTAA
2	Geno	4	ELT03185	**sP1_2**	TGCA	P2_5	CTAA
2	Geno	4	ELT02908	**sP1_3**	ACTA	P2_5	CTAA
2	Geno	4	ELT02864	**sP1_12**	TGCGA	P2_5	CTAA
2	Geno	4	ELT05640_2	**sP1_13**	CGCTT	P2_5	CTAA
2	Geno	4	ELT03124	**sP1_14**	TCACC	P2_5	CTAA
2	Geno	4	ELT03486	**sP1_34**	GGTTGT	P2_5	CTAA
2	Geno	4	ELT02843	**sP1_35**	CCAGCT	P2_5	CTAA

2	Geno	4	ELT02822	**sP1_36**	TTCAGA	P2_5	CTAA
2	Geno	4	ELT02805	**sP1_57**	CTTGCTT	P2_5	CTAA
2	Geno	4	ELT03190_1	**sP1_58**	ATGAAAC	P2_5	CTAA
2	Geno	4	ELT02866_2	**sP1_60**	GAATTCA	P2_5	CTAA
2	Geno	4	ELT02797	**sP1_78**	ACGACTAC	P2_5	CTAA
2	Geno	4	ELT02473	**sP1_79**	TAGCATGC	P2_5	CTAA
2	Geno	4	ELT02474	**sP1_1**	CTCC	**P2_2**	CCT
2	Geno	4	ELT02863	**sP1_2**	TGCA	**P2_2**	CCT
2	Geno	4	ELT03280	**sP1_3**	ACTA	**P2_2**	CCT
2	Geno	4	ELT03181	**sP1_12**	TGCGA	**P2_2**	CCT
2	Geno	4	ELT02861_2	**sP1_13**	CGCTT	**P2_2**	CCT
2	Geno	4	ELT02819_2	**sP1_14**	TCACC	**P2_2**	CCT
2	Geno	4	ELT05640_1	**sP1_34**	GGTTGT	**P2_2**	CCT
2	Geno	4	ELT03450	**sP1_35**	CCAGCT	**P2_2**	CCT
2	Geno	4	ELT02800	**sP1_36**	TTCAGA	**P2_2**	CCT
2	Geno	4	ELT03214	**sP1_57**	CTTGCTT	**P2_2**	CCT
2	Geno	4	ELT02815	**sP1_58**	ATGAAAC	**P2_2**	CCT
2	Geno	4	ELT03347	**sP1_60**	GAATTCA	**P2_2**	CCT
2	Geno	4	ELT05630	**sP1_78**	ACGACTAC	**P2_2**	CCT
2	Geno	4	ELT03193	**sP1_79**	TAGCATGC	**P2_2**	CCT
2	Geno	4	ELT05632	**sP1_83**	TAGGCCAT	**P2_2**	CCT
2	Geno	4	ELT03139	**sP1_1**	CTCC	**P2_3**	ATCG
2	Geno	4	ELT03027	**sP1_2**	TGCA	**P2_3**	ATCG
2	Geno	4	ELT02846	**sP1_3**	ACTA	**P2_3**	ATCG
2	Geno	4	ELT03344	**sP1_12**	TGCGA	**P2_3**	ATCG
2	Geno	4	ELT02906	**sP1_13**	CGCTT	**P2_3**	ATCG
2	Geno	4	ELT02770	**sP1_14**	TCACC	**P2_3**	ATCG
2	Geno	4	ELT02716	**sP1_34**	GGTTGT	**P2_3**	ATCG
2	Geno	4	ELT02795	**sP1_35**	CCAGCT	**P2_3**	ATCG
2	Geno	4	ELT03170	**sP1_36**	TTCAGA	**P2_3**	ATCG
2	Geno	4	ELT03354	**sP1_57**	CTTGCTT	**P2_3**	ATCG
2	Geno	4	ELT03246	**sP1_58**	ATGAAAC	**P2_3**	ATCG
2	Geno	4	ELT02745	**sP1_60**	GAATTCA	**P2_3**	ATCG
2	Geno	4	ELT03165	**sP1_78**	ACGACTAC	**P2_3**	ATCG
2	Geno	4	ELT03382	**sP1_79**	TAGCATGC	**P2_3**	ATCG
2	Geno	4	ELT02742	**sP1_83**	TAGGCCAT	**P2_3**	ATCG
3	Phy	3	ELT04756	**sP1_1**	CTCC	**P2_1**	TAG
3	Phy	3	ELT05450	**sP1_2**	TGCA	**P2_1**	TAG
3	Phy	3	ELT05587_2	**sP1_3**	ACTA	**P2_1**	TAG
3	Phy	3	ELT05539	**sP1_12**	TGCGA	**P2_1**	TAG
3	Phy	3	ELT05587	**sP1_13**	CGCTT	**P2_1**	TAG
3	Phy	2/3	ELT05609	**sP1_14**	TCACC	**P2_1**	TAG
3	Phy	3	ELT04738	**sP1_34**	GGTTGT	**P2_1**	TAG
3	Phy	3/4	ELT02828	**sP1_35**	CCAGCT	**P2_1**	TAG
3	Phy	3	ELT04749	**sP1_36**	TTCAGA	**P2_1**	TAG
3	Phy	3	ELT05496	**sP1_57**	CTTGCTT	**P2_1**	TAG
3	Phy	3	ELT05505	**sP1_58**	ATGAAAC	**P2_1**	TAG

3	Phy	3/4	ELT02726	**sP1_60**	GAATTCA	**P2_1**	TAG
3	Phy	3	ELT04732	**sP1_78**	ACGACTAC	**P2_1**	TAG
3	Phy	3/4	ELT02806	**sP1_79**	TAGCATGC	**P2_1**	TAG
3	Phy	3	ELT05458	**sP1_83**	TAGGCCAT	**P2_1**	TAG
3	Phy	3	ELT05469	**sP1_1**	CTCC	**P2_2**	CCT
3	Phy	3	ELT05456	**sP1_2**	TGCA	**P2_2**	CCT
3	Phy	3	ELT04723	**sP1_3**	ACTA	**P2_2**	CCT
3	Phy	3	ELT05582	**sP1_12**	TGCGA	**P2_2**	CCT
3	Phy	3	ELT05467	**sP1_13**	CGCTT	**P2_2**	CCT
3	Phy	3	ELT04739	**sP1_14**	TCACC	**P2_2**	CCT
3	Phy	3	ELT04715	**sP1_34**	GGTTGT	**P2_2**	CCT
3	Phy	3	ELT04750	**sP1_35**	CCAGCT	**P2_2**	CCT
3	Phy	3	ELT04700	**sP1_36**	TTCAGA	**P2_2**	CCT
3	Phy	2/3	ELT05559	**sP1_57**	CTTGCTT	**P2_2**	CCT
3	Phy	3	ELT04787	**sP1_58**	ATGAAAC	**P2_2**	CCT
3	Phy	3	ELT05497	**sP1_60**	GAATTCA	**P2_2**	CCT
3	Phy	3	ELT04706	**sP1_78**	ACGACTAC	**P2_2**	CCT
3	Phy	3	ELT04737	**sP1_79**	TAGCATGC	**P2_2**	CCT
3	Phy	3/4	ELT03078	**sP1_1**	CTCC	**P2_3**	ATCG
3	Phy	2/3	ELT05576_2	**sP1_2**	TGCA	**P2_3**	ATCG
3	Phy	2/3	ELT05541	**sP1_3**	ACTA	**P2_3**	ATCG
3	Phy	3	ELT04721	**sP1_12**	TGCGA	**P2_3**	ATCG
3	Phy	3	ELT04746	**sP1_13**	CGCTT	**P2_3**	ATCG
3	Phy	3	ELT02678	**sP1_14**	TCACC	**P2_3**	ATCG
3	Phy	3	ELT05534_2	**sP1_34**	GGTTGT	**P2_3**	ATCG
3	Phy	3	ELT04708	**sP1_35**	CCAGCT	**P2_3**	ATCG
3	Phy	3/4	ELT03268	**sP1_36**	TTCAGA	**P2_3**	ATCG
3	Phy	3	ELT04755	**sP1_57**	CTTGCTT	**P2_3**	ATCG
3	Phy	3	ELT05460	**sP1_58**	ATGAAAC	**P2_3**	ATCG
3	Phy	3	ELT05586	**sP1_60**	GAATTCA	**P2_3**	ATCG
3	Phy	3	ELT05585	**sP1_78**	ACGACTAC	**P2_3**	ATCG
3	Phy	2/3	ELT04698	**sP1_79**	TAGCATGC	**P2_3**	ATCG
3	Phy	3	ELT04701	**sP1_83**	TAGGCCAT	**P2_3**	ATCG
3	Phy	2/3	ELT04713	**sP1_1**	CTCC	**P2_4**	GAGC
3	Phy	3	ELT05534	**sP1_2**	TGCA	**P2_4**	GAGC
3	Phy	3	ELT05455	**sP1_3**	ACTA	**P2_4**	GAGC
3	Phy	3	ELT05447	**sP1_12**	TGCGA	**P2_4**	GAGC
3	Phy	3	ELT05441	**sP1_13**	CGCTT	**P2_4**	GAGC
3	Phy	3/4	ELT03076	**sP1_14**	TCACC	**P2_4**	GAGC
3	Phy	3	ELT05578	**sP1_34**	GGTTGT	**P2_4**	GAGC
3	Phy	3	ELT04759	**sP1_35**	CCAGCT	**P2_4**	GAGC
3	Phy	3	ELT05521	**sP1_36**	TTCAGA	**P2_4**	GAGC
3	Phy	2/3	ELT05575	**sP1_57**	CTTGCTT	**P2_4**	GAGC
3	Phy	3	ELT04733	**sP1_58**	ATGAAAC	**P2_4**	GAGC
3	Phy	3	ELT04741	**sP1_60**	GAATTCA	**P2_4**	GAGC
3	Phy	3	ELT05576	**sP1_78**	ACGACTAC	**P2_4**	GAGC
3	Phy	3	ELT05469_2	**sP1_79**	TAGCATGC	**P2_4**	GAGC

3	Phy	2/3	ELT05493	**sP1_1**	CTCC	**P2_5**	CTAA
3	Phy	3	ELT04714	**sP1_2**	TGCA	**P2_5**	CTAA
3	Phy	3	ELT04728	**sP1_3**	ACTA	**P2_5**	CTAA
3	Phy	3/4	ELT03366	**sP1_12**	TGCGA	**P2_5**	CTAA
3	Phy	3	ELT05578_2	**sP1_13**	CGCTT	**P2_5**	CTAA
3	Phy	3	ELT04720	**sP1_14**	TCACC	**P2_5**	CTAA
3	Phy	2/3	ELT04734	**sP1_34**	GGTTGT	**P2_5**	CTAA
3	Phy	2/3	ELT05596	**sP1_35**	CCAGCT	**P2_5**	CTAA
3	Phy	3	ELT04757	**sP1_36**	TTCAGA	**P2_5**	CTAA
3	Phy	3	ELT04724	**sP1_57**	CTTGCTT	**P2_5**	CTAA
3	Phy	3	ELT04754	**sP1_58**	ATGAAAC	**P2_5**	CTAA
3	Phy	2/3	ELT00415	**sP1_60**	GAATTCA	**P2_5**	CTAA
3	Phy	3	ELT04718	**sP1_78**	ACGACTAC	**P2_5**	CTAA
3	Phy	2/3	ELT05498	**sP1_79**	TAGCATGC	**P2_5**	CTAA
3	Phy	3	ELT05451	**sP1_83**	TAGGCCAT	**P2_5**	CTAA
3	Phy	3/4	ELT03090	**sP1_1**	CTCC	**P2_6**	TTGC
3	Phy	3	ELT04704	**sP1_2**	TGCA	**P2_6**	TTGC
3	Phy	3	ELT04740	**sP1_3**	ACTA	**P2_6**	TTGC
3	Phy	3	ELT05522	**sP1_12**	TGCGA	**P2_6**	TTGC
3	Phy	3	ELT05484	**sP1_13**	CGCTT	**P2_6**	TTGC
3	Phy	3/4	ELT03126	**sP1_14**	TCACC	**P2_6**	TTGC
3	Phy	3	ELT04722	**sP1_34**	GGTTGT	**P2_6**	TTGC
3	Phy	3	ELT05499	**sP1_35**	CCAGCT	**P2_6**	TTGC
3	Phy	2/3	ELT05454	**sP1_36**	TTCAGA	**P2_6**	TTGC
3	Phy	3/4	ELT03250	**sP1_57**	CTTGCTT	**P2_6**	TTGC
3	Phy	3	ELT04743	**sP1_58**	ATGAAAC	**P2_6**	TTGC
3	Phy	3	ELT05597	**sP1_60**	GAATTCA	**P2_6**	TTGC
3	Phy	3	ELT05461	**sP1_78**	ACGACTAC	**P2_6**	TTGC
3	Phy	3/4	ELT03459	**sP1_79**	TAGCATGC	**P2_6**	TTGC
4	Seb1	3	ELT05470	**sP1_1**	CTCC	**P2_1**	TAG
4	Seb1	3	ELT05548	**sP1_2**	TGCA	**P2_1**	TAG
4	Seb1	3	ELT05444	**sP1_3**	ACTA	**P2_1**	TAG
4	Seb1	3	ELT04773	**sP1_12**	TGCGA	**P2_1**	TAG
4	Seb1	3	ELT04784	**sP1_13**	CGCTT	**P2_1**	TAG
4	Seb1	3	ELT05580	**sP1_14**	TCACC	**P2_1**	TAG
4	Seb1	3	ELT05519_2	**sP1_34**	GGTTGT	**P2_1**	TAG
4	Seb1	3	ELT05510_1	**sP1_35**	CCAGCT	**P2_1**	TAG
4	Seb1	3	ELT05535	**sP1_36**	TTCAGA	**P2_1**	TAG
4	Seb1	3	ELT04748	**sP1_57**	CTTGCTT	**P2_1**	TAG
4	Seb1	3	ELT05567_2	**sP1_58**	ATGAAAC	**P2_1**	TAG
4	Seb1	3	ELT04786	**sP1_60**	GAATTCA	**P2_1**	TAG
4	Seb1	3	ELT04772	**sP1_78**	ACGACTAC	**P2_1**	TAG
4	Seb1	3	ELT05524	**sP1_79**	TAGCATGC	**P2_1**	TAG
4	Seb1	3	ELT04771	**sP1_83**	TAGGCCAT	**P2_1**	TAG
4	Seb1	3	ELT05588	**sP1_1**	CTCC	**P2_4**	GAGC
4	Seb1	3	ELT04788	**sP1_2**	TGCA	**P2_4**	GAGC
4	Seb1	3	ELT04753	**sP1_3**	ACTA	**P2_4**	GAGC

4	Seb1	3	ELT05474	**sP1_13**	CGCTT	**P2_4**	GAGC
4	Seb1	3	ELT05560	**sP1_1**	CTCC	**P2_2**	CCT
4	Seb1	3	ELT05483	**sP1_2**	TGCA	**P2_2**	CCT
4	Seb1	3	ELT05589	**sP1_3**	ACTA	**P2_2**	CCT
4	Seb1	3	ELT04726	**sP1_12**	TGCGA	**P2_2**	CCT
4	Seb1	3	ELT04711	**sP1_13**	CGCTT	**P2_2**	CCT
4	Seb1	3	ELT04742	**sP1_14**	TCACC	**P2_2**	CCT
4	Seb1	3	ELT05606	**sP1_34**	GGTTGT	**P2_2**	CCT
4	Seb1	3	ELT05442	**sP1_35**	CCAGCT	**P2_2**	CCT
4	Seb1	3	ELT04699	**sP1_36**	TTCAGA	**P2_2**	CCT
4	Seb1	3	ELT04758	**sP1_57**	CTTGCTT	**P2_2**	CCT
4	Seb1	3	ELT04779	**sP1_58**	ATGAAAC	**P2_2**	CCT
4	Seb1	3	ELT05600	**sP1_60**	GAATTCA	**P2_2**	CCT
4	Seb1	3	ELT05562	**sP1_78**	ACGACTAC	**P2_2**	CCT
4	Seb1	3	ELT05567_1	**sP1_79**	TAGCATGC	**P2_2**	CCT
4	Seb1	3	ELT05558	**sP1_83**	TAGGCCAT	**P2_2**	CCT
4	Seb1	3	ELT05540	**sP1_1**	CTCC	**P2_3**	ATCG
4	Seb1	3	ELT05445	**sP1_2**	TGCA	**P2_3**	ATCG
4	Seb1	3	ELT05486	**sP1_3**	ACTA	**P2_3**	ATCG
4	Seb1	3	ELT04783	**sP1_12**	TGCGA	**P2_3**	ATCG
4	Seb1	3	ELT05478	**sP1_13**	CGCTT	**P2_3**	ATCG
4	Seb1	3	ELT05510_2	**sP1_14**	TCACC	**P2_3**	ATCG
4	Seb1	3	ELT05515	**sP1_34**	GGTTGT	**P2_3**	ATCG
4	Seb1	3	ELT05603	**sP1_35**	CCAGCT	**P2_3**	ATCG
4	Seb1	3	ELT04710	**sP1_36**	TTCAGA	**P2_3**	ATCG
4	Seb1	3	ELT04766_2	**sP1_57**	CTTGCTT	**P2_3**	ATCG
4	Seb1	3	ELT05519_1	**sP1_58**	ATGAAAC	**P2_3**	ATCG
4	Seb1	3	ELT05452	**sP1_60**	GAATTCA	**P2_3**	ATCG
4	Seb1	3	ELT04707	**sP1_78**	ACGACTAC	**P2_3**	ATCG
4	Seb1	3	ELT04744	**sP1_79**	TAGCATGC	**P2_3**	ATCG
4	Seb1	3	ELT04766_1	**sP1_83**	TAGGCCAT	**P2_3**	ATCG
5	Seb2	3	ELT05453	**sP1_1**	CTCC	**P2_1**	TAG
5	Seb2	3	ELT05583	**sP1_2**	TGCA	**P2_1**	TAG
5	Seb2	3	ELT05528	**sP1_3**	ACTA	**P2_1**	TAG
5	Seb2	3	ELT05564	**sP1_12**	TGCGA	**P2_1**	TAG
5	Seb2	3	ELT05511	**sP1_13**	CGCTT	**P2_1**	TAG
5	Seb2	3	ELT05605	**sP1_14**	TCACC	**P2_1**	TAG
5	Seb2	3	ELT05547	**sP1_34**	GGTTGT	**P2_1**	TAG
5	Seb2	3	ELT05591	**sP1_35**	CCAGCT	**P2_1**	TAG
5	Seb2	3	ELT05494	**sP1_36**	TTCAGA	**P2_1**	TAG
5	Seb2	3	ELT05476	**sP1_57**	CTTGCTT	**P2_1**	TAG
5	Seb2	3	ELT05556	**sP1_58**	ATGAAAC	**P2_1**	TAG
5	Seb2	3	ELT04751	**sP1_60**	GAATTCA	**P2_1**	TAG
5	Seb2	3	ELT05504_2	**sP1_78**	ACGACTAC	**P2_1**	TAG
5	Seb2	3	ELT05545	**sP1_79**	TAGCATGC	**P2_1**	TAG
5	Seb2	3	ELT04727	**sP1_83**	TAGGCCAT	**P2_1**	TAG
5	Seb2	3	ELT05459	**sP1_1**	CTCC	**P2_4**	GAGC

5	Seb2	3	ELT05517	**sP1_2**	TGCA	**P2_4**	GAGC
5	Seb2	3	ELT04764	**sP1_3**	ACTA	**P2_4**	GAGC
5	Seb2	3	ELT05546	**sP1_12**	TGCGA	**P2_4**	GAGC
5	Seb2	3	ELT05601	**sP1_13**	CGCTT	**P2_4**	GAGC
5	Seb2	3	ELT05529_1	**sP1_1**	CTCC	**P2_2**	CCT
5	Seb2	3	ELT05590_1	**sP1_2**	TGCA	**P2_2**	CCT
5	Seb2	3	ELT04762	**sP1_3**	ACTA	**P2_2**	CCT
5	Seb2	3	ELT04789	**sP1_12**	TGCGA	**P2_2**	CCT
5	Seb2	3	ELT04775_1	**sP1_13**	CGCTT	**P2_2**	CCT
5	Seb2	3	ELT05462	**sP1_14**	TCACC	**P2_2**	CCT
5	Seb2	3	ELT05598	**sP1_34**	GGTTGT	**P2_2**	CCT
5	Seb2	3	ELT04763	**sP1_35**	CCAGCT	**P2_2**	CCT
5	Seb2	3	ELT05449	**sP1_36**	TTCAGA	**P2_2**	CCT
5	Seb2	3	ELT04752	**sP1_57**	CTTGCTT	**P2_2**	CCT
5	Seb2	3	ELT05485	**sP1_58**	ATGAAAC	**P2_2**	CCT
5	Seb2	3	ELT05500	**sP1_60**	GAATTCA	**P2_2**	CCT
5	Seb2	3	ELT04775_2	**sP1_78**	ACGACTAC	**P2_2**	CCT
5	Seb2	3	ELT05590_2	**sP1_79**	TAGCATGC	**P2_2**	CCT
5	Seb2	3	ELT05607	**sP1_83**	TAGGCCAT	**P2_2**	CCT
5	Seb2	3	ELT05508	**sP1_1**	CTCC	**P2_3**	ATCG
5	Seb2	3	ELT05504_1	**sP1_2**	TGCA	**P2_3**	ATCG
5	Seb2	3	ELT05513	**sP1_3**	ACTA	**P2_3**	ATCG
5	Seb2	3	ELT04725	**sP1_12**	TGCGA	**P2_3**	ATCG
5	Seb2	3	ELT04780	**sP1_13**	CGCTT	**P2_3**	ATCG
5	Seb2	3	ELT04760	**sP1_14**	TCACC	**P2_3**	ATCG
5	Seb2	3	ELT05501	**sP1_34**	GGTTGT	**P2_3**	ATCG
5	Seb2	3	ELT05526	**sP1_35**	CCAGCT	**P2_3**	ATCG
5	Seb2	3	ELT05448	**sP1_36**	TTCAGA	**P2_3**	ATCG
5	Seb2	3	ELT05529_2	**sP1_57**	CTTGCTT	**P2_3**	ATCG
5	Seb2	3	ELT05507	**sP1_58**	ATGAAAC	**P2_3**	ATCG
5	Seb2	3	ELT05518	**sP1_60**	GAATTCA	**P2_3**	ATCG
5	Seb2	3	ELT05495	**sP1_78**	ACGACTAC	**P2_3**	ATCG
5	Seb2	3	ELT05509	**sP1_79**	TAGCATGC	**P2_3**	ATCG
5	Seb2	3	ELT05593	**sP1_83**	TAGGCCAT	**P2_3**	ATCG
6	Seb3	3	ELT04745	**sP1_1**	CTCC	**P2_1**	TAG
6	Seb3	3	ELT04703	**sP1_2**	TGCA	**P2_1**	TAG
6	Seb3	3	ELT05595	**sP1_3**	ACTA	**P2_1**	TAG
6	Seb3	3	ELT05594	**sP1_12**	TGCGA	**P2_1**	TAG
6	Seb3	3	ELT05523	**sP1_13**	CGCTT	**P2_1**	TAG
6	Seb3	3	ELT05516	**sP1_14**	TCACC	**P2_1**	TAG
6	Seb3	3	ELT05561	**sP1_34**	GGTTGT	**P2_1**	TAG
6	Seb3	3	ELT05536_2	**sP1_35**	CCAGCT	**P2_1**	TAG
6	Seb3	3	ELT05536_1	**sP1_36**	TTCAGA	**P2_1**	TAG
6	Seb3	3	ELT05514	**sP1_57**	CTTGCTT	**P2_1**	TAG
6	Seb3	3	ELT04777	**sP1_58**	ATGAAAC	**P2_1**	TAG
6	Seb3	3	ELT05446	**sP1_60**	GAATTCA	**P2_1**	TAG
6	Seb3	3	ELT04769_1	**sP1_78**	ACGACTAC	**P2_1**	TAG

6	Seb3	3	ELT04702	**sP1_79**	TAGCATGC	**P2_1**	TAG
6	Seb3	3	ELT05527	**sP1_83**	TAGGCCAT	**P2_1**	TAG
6	Seb3	3	ELT04776	**sP1_1**	CTCC	**P2_4**	GAGC
6	Seb3	3	ELT05608	**sP1_2**	TGCA	**P2_4**	GAGC
6	Seb3	3	ELT05525	**sP1_3**	ACTA	**P2_4**	GAGC
6	Seb3	3	ELT05512	**sP1_12**	TGCGA	**P2_4**	GAGC
6	Seb3	3	ELT05599	**sP1_13**	CGCTT	**P2_4**	GAGC
6	Seb3	3	ELT04717	**sP1_14**	TCACC	**P2_4**	GAGC
6	Seb3	3	ELT04712	**sP1_1**	CTCC	**P2_2**	CCT
6	Seb3	3	ELT04774	**sP1_2**	TGCA	**P2_2**	CCT
6	Seb3	3	ELT04709	**sP1_3**	ACTA	**P2_2**	CCT
6	Seb3	3	ELT04735	**sP1_12**	TGCGA	**P2_2**	CCT
6	Seb3	3	ELT05506	**sP1_13**	CGCTT	**P2_2**	CCT
6	Seb3	3	ELT05487	**sP1_14**	TCACC	**P2_2**	CCT
6	Seb3	3	ELT05464	**sP1_34**	GGTTGT	**P2_2**	CCT
6	Seb3	3	ELT04768	**sP1_35**	CCAGCT	**P2_2**	CCT
6	Seb3	3	ELT04761	**sP1_36**	TTCAGA	**P2_2**	CCT
6	Seb3	3	ELT04767	**sP1_57**	CTTGCTT	**P2_2**	CCT
6	Seb3	3	ELT05520_2	**sP1_58**	ATGAAAC	**P2_2**	CCT
6	Seb3	3	ELT05538_2	**sP1_60**	GAATTCA	**P2_2**	CCT
6	Seb3	3	ELT05443	**sP1_78**	ACGACTAC	**P2_2**	CCT
6	Seb3	3	ELT04770	**sP1_79**	TAGCATGC	**P2_2**	CCT
6	Seb3	3	ELT04765	**sP1_83**	TAGGCCAT	**P2_2**	CCT
6	Seb3	3	ELT05592	**sP1_1**	CTCC	**P2_3**	ATCG
6	Seb3	3	ELT05579	**sP1_2**	TGCA	**P2_3**	ATCG
6	Seb3	3	ELT04781	**sP1_3**	ACTA	**P2_3**	ATCG
6	Seb3	3	ELT04719	**sP1_12**	TGCGA	**P2_3**	ATCG
6	Seb3	3	ELT04782	**sP1_13**	CGCTT	**P2_3**	ATCG
6	Seb3	3	ELT05537	**sP1_14**	TCACC	**P2_3**	ATCG
6	Seb3	3	ELT05604	**sP1_34**	GGTTGT	**P2_3**	ATCG
6	Seb3	3	ELT04769_2	**sP1_35**	CCAGCT	**P2_3**	ATCG
6	Seb3	3	ELT05538_1	**sP1_36**	TTCAGA	**P2_3**	ATCG
6	Seb3	3	ELT04778	**sP1_57**	CTTGCTT	**P2_3**	ATCG
6	Seb3	3	ELT04785	**sP1_58**	ATGAAAC	**P2_3**	ATCG
6	Seb3	3	ELT05520_1	**sP1_60**	GAATTCA	**P2_3**	ATCG
6	Seb3	3	ELT05602	**sP1_78**	ACGACTAC	**P2_3**	ATCG
6	Seb3	3	ELT05503	**sP1_79**	TAGCATGC	**P2_3**	ATCG
6	Seb3	3	ELT05542	**sP1_83**	TAGGCCAT	**P2_3**	ATCG
7	Seb4	3	ELT08283	**sP1_1**	CTCC	**P2_1**	TAG
7	Seb4	3	ELT06877	**sP1_2**	TGCA	**P2_1**	TAG
7	Seb4	3	ELT06852	**sP1_3**	ACTA	**P2_1**	TAG
7	Seb4	3	ELT06865	**sP1_12**	TGCGA	**P2_1**	TAG
7	Seb4	3	ELT06871	**sP1_13**	CGCTT	**P2_1**	TAG
7	Seb4	3	ELT06858	**sP1_14**	TCACC	**P2_1**	TAG
7	Seb4	3	ELT08284	**sP1_34**	GGTTGT	**P2_1**	TAG
7	Seb4	3	ELT06856	**sP1_35**	CCAGCT	**P2_1**	TAG
7	Seb4	3	ELT06898	**sP1_36**	TTCAGA	**P2_1**	TAG

7	Seb4	3	ELT06869	**sP1_57**	CTTGCTT	**P2_1**	TAG
7	Seb4	3	ELT06885	**sP1_58**	ATGAAAC	**P2_1**	TAG
7	Seb4	3	ELT06883	**sP1_60**	GAATTCA	**P2_1**	TAG
7	Seb4	3	ELT06882	**sP1_78**	ACGACTAC	**P2_1**	TAG
7	Seb4	3	ELT06888	**sP1_79**	TAGCATGC	**P2_1**	TAG
7	Seb4	3	ELT06900	**sP1_83**	TAGGCCAT	**P2_1**	TAG
7	Seb4	3	ELT06866	**sP1_1**	CTCC	**P2_4**	GAGC
7	Seb4	3	ELT06889	**sP1_2**	TGCA	**P2_4**	GAGC
7	Seb4	3	ELT06891	**sP1_3**	ACTA	**P2_4**	GAGC
7	Seb4	3	ELT06896	**sP1_12**	TGCGA	**P2_4**	GAGC
7	Seb4	3	ELT08282	**sP1_13**	CGCTT	**P2_4**	GAGC
7	Seb4	3	ELT08288	**sP1_14**	TCACC	**P2_4**	GAGC
7	Seb4	3	ELT06854	**sP1_1**	CTCC	**P2_2**	CCT
7	Seb4	3	ELT06864	**sP1_2**	TGCA	**P2_2**	CCT
7	Seb4	3	ELT06884	**sP1_3**	ACTA	**P2_2**	CCT
7	Seb4	3	ELT06887	**sP1_12**	TGCGA	**P2_2**	CCT
7	Seb4	3	ELT06895	**sP1_13**	CGCTT	**P2_2**	CCT
7	Seb4	3	ELT06861	**sP1_14**	TCACC	**P2_2**	CCT
7	Seb4	3	ELT06873	**sP1_34**	GGTTGT	**P2_2**	CCT
7	Seb4	3	ELT06880	**sP1_35**	CCAGCT	**P2_2**	CCT
7	Seb4	3	ELT08285	**sP1_36**	TTCAGA	**P2_2**	CCT
7	Seb4	3	ELT08287	**sP1_57**	CTTGCTT	**P2_2**	CCT
7	Seb4	3	ELT06862	**sP1_58**	ATGAAAC	**P2_2**	CCT
7	Seb4	3	ELT06855	**sP1_60**	GAATTCA	**P2_2**	CCT
7	Seb4	3	ELT06878	**sP1_78**	ACGACTAC	**P2_2**	CCT
7	Seb4	3	ELT06890	**sP1_79**	TAGCATGC	**P2_2**	CCT
7	Seb4	3	ELT06872	**sP1_83**	TAGGCCAT	**P2_2**	CCT
7	Seb4	3	ELT06853	**sP1_1**	CTCC	**P2_3**	ATCG
7	Seb4	3	ELT06874	**sP1_2**	TGCA	**P2_3**	ATCG
7	Seb4	3	ELT06863	**sP1_3**	ACTA	**P2_3**	ATCG
7	Seb4	3	ELT06886	**sP1_12**	TGCGA	**P2_3**	ATCG
7	Seb4	3	ELT06868	**sP1_13**	CGCTT	**P2_3**	ATCG
7	Seb4	3	ELT06851	**sP1_14**	TCACC	**P2_3**	ATCG
7	Seb4	3	ELT06881	**sP1_34**	GGTTGT	**P2_3**	ATCG
7	Seb4	3	ELT06876	**sP1_35**	CCAGCT	**P2_3**	ATCG
7	Seb4	3	ELT08286	**sP1_36**	TTCAGA	**P2_3**	ATCG
7	Seb4	3	ELT06897	**sP1_57**	CTTGCTT	**P2_3**	ATCG
7	Seb4	3	ELT06859	**sP1_58**	ATGAAAC	**P2_3**	ATCG
7	Seb4	3	ELT06870	**sP1_60**	GAATTCA	**P2_3**	ATCG
7	Seb4	3	ELT06899	**sP1_78**	ACGACTAC	**P2_3**	ATCG
7	Seb4	3	ELT06857	**sP1_79**	TAGCATGC	**P2_3**	ATCG
7	Seb4	3	ELT06867	**sP1_83**	TAGGCCAT	**P2_3**	ATCG

Appendix 3: supplementary material for Chapter 2

A3.1: Supplementary methods

RNA extraction protocol

Tissues (100mg) from each salamander were homogenized in 2ml tubes with steal beads in 1ml trizol using the tissue lyser Precellys (speed 6000, 2x 55 sec in between 10 sec break). The samples were centrifuged for 10min at 12,000g at 4°C. The solution was transferred to a new tube and incubated at room temperature (RT) for 5 min, after which 200µl chloroform was added, samples were vortexed for 15 sec, incubated for 3min at RT and centrifuged for 15min at 12,000g at 4°C. The upper phase was transferred to a new tube and 500µl isopropanol was added. The solution was incubated for 10min at RT and centrifuged for 15min at 12,000g at 4°C. The supernatant was removed,1ml 75% EtOH was added to the pellet (using DEPC water) and the sample was vortexed until the pellet was resuspended. The samples were centrifuged for 5min at 7.500g at 4°C and the supernatant was removed. The pellet was dried and dissolved in DEPC water. The RNA was precipitated by adding 200µl 5M LiCl solution and incubated for 1.5h at -20°C. After incubation, the solution was centrifuged for 20min at 14.000g at 4°C. The supernatant was removed, the pellet washed with 500µl 75% EtOH and then centrifuged for 10min at 14.000g at 4°C. The pellet was again dried and dissolved in DEPC water.

A3.2: Supplementary Tables

Supplementary Table A3.1: Gene ontology assessment for gene trees (from loci obtained from the RNAseq data) supporting the alternative clades containing *S. atra* and *S. corsica* (441 trees), and *S. atra* and *S. lanzai* (430 trees).

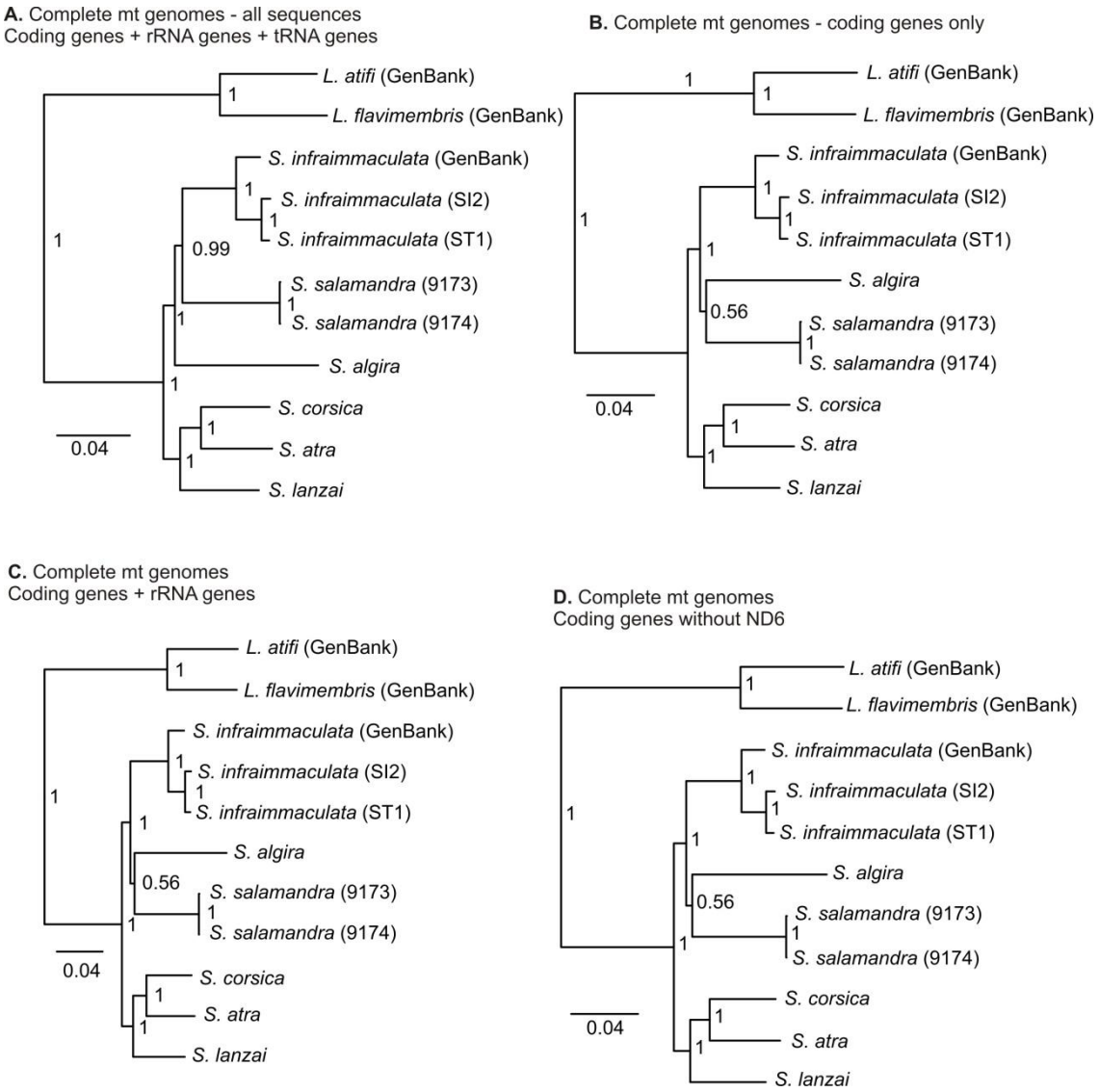
Gene product properties	***atra-corsica***	***atra-lanzai***
Molecular function	319	322
transcription factor activity, transcription factor binding	10	7
transcription factor activity, sequence-specific DNA binding	16	14
catalytic activity	153	173
signal transducer activity	14	11
receptor activity	9	6
structural molecule activity	14	14
transporter activity	26	21
binding	242	234
electron carrier activity	2	8
antioxidant activity	3	3
metallochaperone activity	1	0
chemorepellent activity	1	1
molecular function regulator	25	21
Cellular component	393	404
extracellular region	89	71
basement membrane	4	0
cell	361	369
membrane	175	180
cell junction	19	17
extracellular matrix	9	5
membrane-enclosed lumen	110	105
macromolecular complex	109	104
mitochondrial nucleoid	2	2
organelle	325	323
extracellular matrix component	0	2
extracellular region part	84	65
organelle part	228	220
membrane part	122	134
synapse part	5	4
cell part	361	369
synapse	7	11
polymeric cytoskeletal fiber	7	7
Biological process	369	373
reproduction	11	7
cartilage condensation	0	1
cell killing	0	2
immune system process	22	25
cell adhesion	0	14
behavior	5	5
metabolic process	230	254
cellular process	321	318
antioxidant activity	0	3
reproductive process	11	7
biological adhesion	11	0
multicellular organismal process	79	73
developmental process	82	71

growth	0	9
locomotion	11	16
tissue regeneration	2	0
single-organism process	275	275
single organism signaling	66	47
rhythmic process	4	4
response to stimulus	118	106
localization	99	100
multi-organism process	22	22
biological regulation	201	178
cellular component organization or biogenesis	118	103
cellular oxidant detoxification	3	0
presynaptic process involved in chemical synaptic transmission	1	1

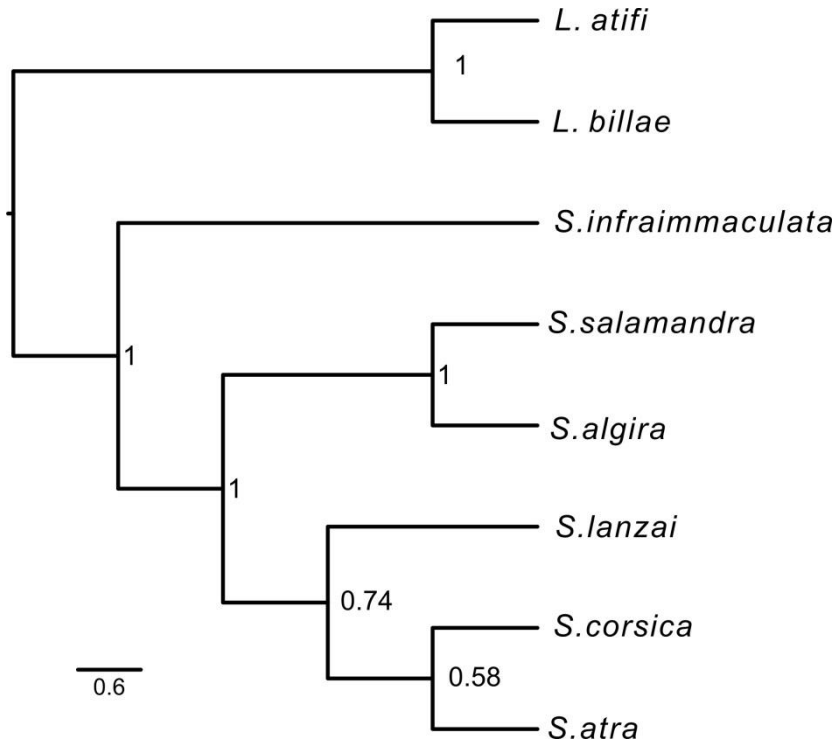
Supplementary Table A3.2: Nuclear genes from the RNAseq data set functionally connected to mitochondrial genes that alternatively support the *atra-corsica* or *atra-lanzai* clade (gene ID = UniProt entry)

	Support for *atra-corsica* clade	**Support for *atra-lanzai* clade**
Genes tightly connected to the mitochondrial respiratory chain	P02721, A3KMZ6, Q8N8Q8, Q02379	Q08DG6, B0VYY3, P00169, P23935, Q02372, Q02369, Q5RBS1, P17694
Genes weakly connected to mitochondrial functions	Q99NB1, P13216, B8JMH0, Q5RBC8, Q3T131, Q9NUL7, P33316, P42126, Q5RC31, Q14197, Q28I39, Q3ZBW7, Q6PBC3, Q10713, A9JTX2, P11181, Q9Z2Z7, Q32N55, Q9BYD6, P0C2C4, Q9BYD1, Q2TBS2, Q9BYC8, Q9Y399, Q9Y3D3, P82930, Q6GLA2, Q9WUM5, Q8N3R3, Q2KIR8, Q5M7K0, Q5IS35, Q69BJ6	Q5RF40, Q641Y1, Q6AY04, Q9H3J6, Q0V9D9, Q6DF46, Q99807, O75208, O46521, Q8NBI2, A5PKR8, P55931, Q922E6, P11183, P00390, P36551, O57478, Q8K3A0, O00142, Q5XGI1, Q5HZE0, Q5ZKP2, Q9N285, Q5RDE7, Q5R655, A1L2L5, P0C089, Q3MHI7, Q3ZBX6, Q99N92, Q6DJI4, P82927, Q28GD1, A6QNM2, P82924, Q924D0, Q5XHA0, Q90XD2, Q6PI48

A3.3. Supplementary Figures



Supplementary Figure A3.1: Majority-Rule consensus trees obtained by partitioned Bayesian Inference from different subsets of genes from complete or almost complete mitochondrial genomes. Numbers after species names refer to laboratory sample numbers.



Supplementary Figure A3.2: Species tree obtained from ML gene trees of each of the 3072 orthologous loci from the RNA-Seq analysis, summarized with ASTRAL II, and including the outgroup (*Lyciasalamandra*). Branch support was estimated by computing the local posterior probability.

Appendix 4: supplementary material for Chapter 3

A4.1. Supplementary Tables

Supplementary Table A4.1: Character coding for ancestral state reconstruction analyses (for detailed phenotype descriptions see Chapter 3 section *3.3.7.*). The N, C and S after *S. s. gallaica* indicate the clades approximate geographic position in western Iberia (northern, central and southern respectively).

Species	Subspecies	Parity	Melanism	Colour pattern
Salamandra algira	*algira*	Larviparity	No	Spotted
	algira (Mid-Atlas)	Larviparity	No	Spotted
	tingitana	Both	Yes	Spotted
	spelaea	Larviparity	No	Spotted
	splendens	Larviparity	No	Spotted
Salamandra atra	*atra*	Pueriparity	Yes	No pattern
	pasubiensis	Pueriparity	No	Spotted
Salamandra corsica	-	Larviparity	No	Spotted
Salamandra infraimmaculata	*infraimmaculata*	Larviparity	No	Spotted
Salamandra lanzai	-	Pueriparity	Yes	No pattern
Salamandra salamandra	*almanzoris*	Larviparity	No	Spotted
	bejarae	Larviparity	No	Spotted
	bernardezi	Both	Yes	Striped
	beschkovi	Larviparity	No	Spotted
	crespoi	Larviparity	No	Spotted
	fastuosa	Both	No	Striped
	gallaica	Both	Yes	Spotted
	gallaica	Larviparity	No	Spotted
	gallaica	Larviparity	No	Spotted
	gigliolii	Larviparity	No	Spotted
	hispanica	Larviparity	No	Spotted
	longirostris	Larviparity	No	Spotted
	morenica	Larviparity	No	Spotted
	salamandra	Larviparity	No	Spotted
	terrestris	Larviparity	No	Striped
	werneri	Larviparity	No	Spotted
Lyciasalamandra spp.	-	Pueriparity	No	-

Supplementary Table A4.2: P-distances within and between the six currently recognised *Salamandra* species and the outgroup *Lyciasalamandra* (calculated in MEGA 7).

Species	Samples	Within Group p-distance (%)	Between Group p-distance (%)							
			-	1	2	3	4	5	6	7
S. salamandra	151	1.6	**1**	-						
S. algira	44	1.5	**2**	2.1	-					
S. corsica	3	0.4	**3**	2.6	2.7	-				
S. lanzai	1	-	**4**	2.3	2.4	2.4	-			
S. atra	5	0.8	**5**	2.4	2.5	2.3	2.1	-		
S. infraimmaculata	25	1.6	**6**	2.9	2.9	2.8	2.7	2.7	-	
Lyciasalamandra	2	0.7	**7**	4.0	4.0	3.5	3.5	3.5	3.5	-

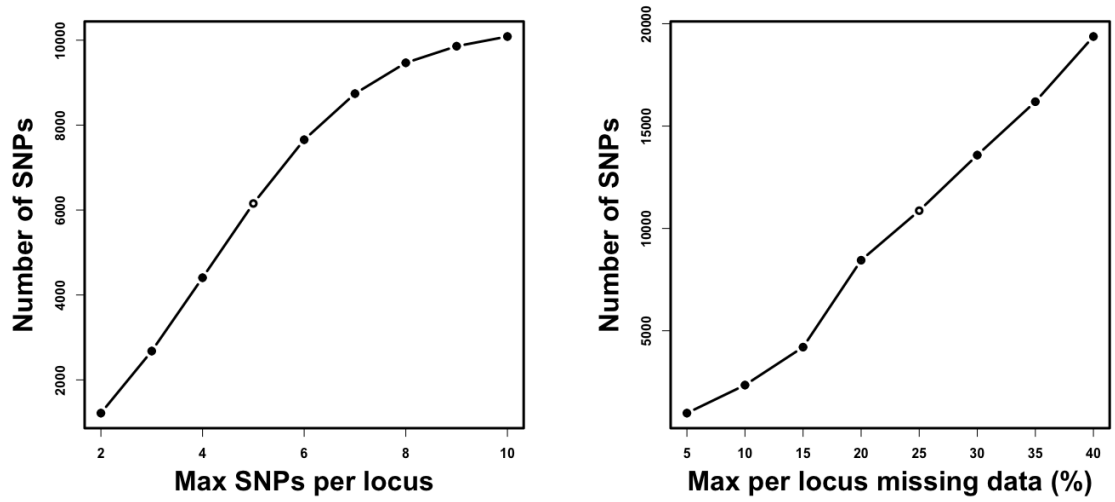
Supplementary Table A4.3: P-distances within and between the five *S. algira* clades identified during phylogenetic analyses (calculated in MEGA 7).

Subspecies	Samples	Within Group p-distance (%)	Between Group p-distance (%)					
			-	1	2	3	4	5
S. a. spelaea	3	0.8	**1**	-				
S. a. algira (Algeria)	10	1	**2**	1.4	-			
S. a. splendens	10	1.2	**3**	1.5	1.4	-		
S. a. tingitana	16	1.3	**4**	1.7	1.6	1.5	-	
S. a. algira (Mid-Atlas)	5	1.5	**5**	1.6	1.7	1.6	1.8	-

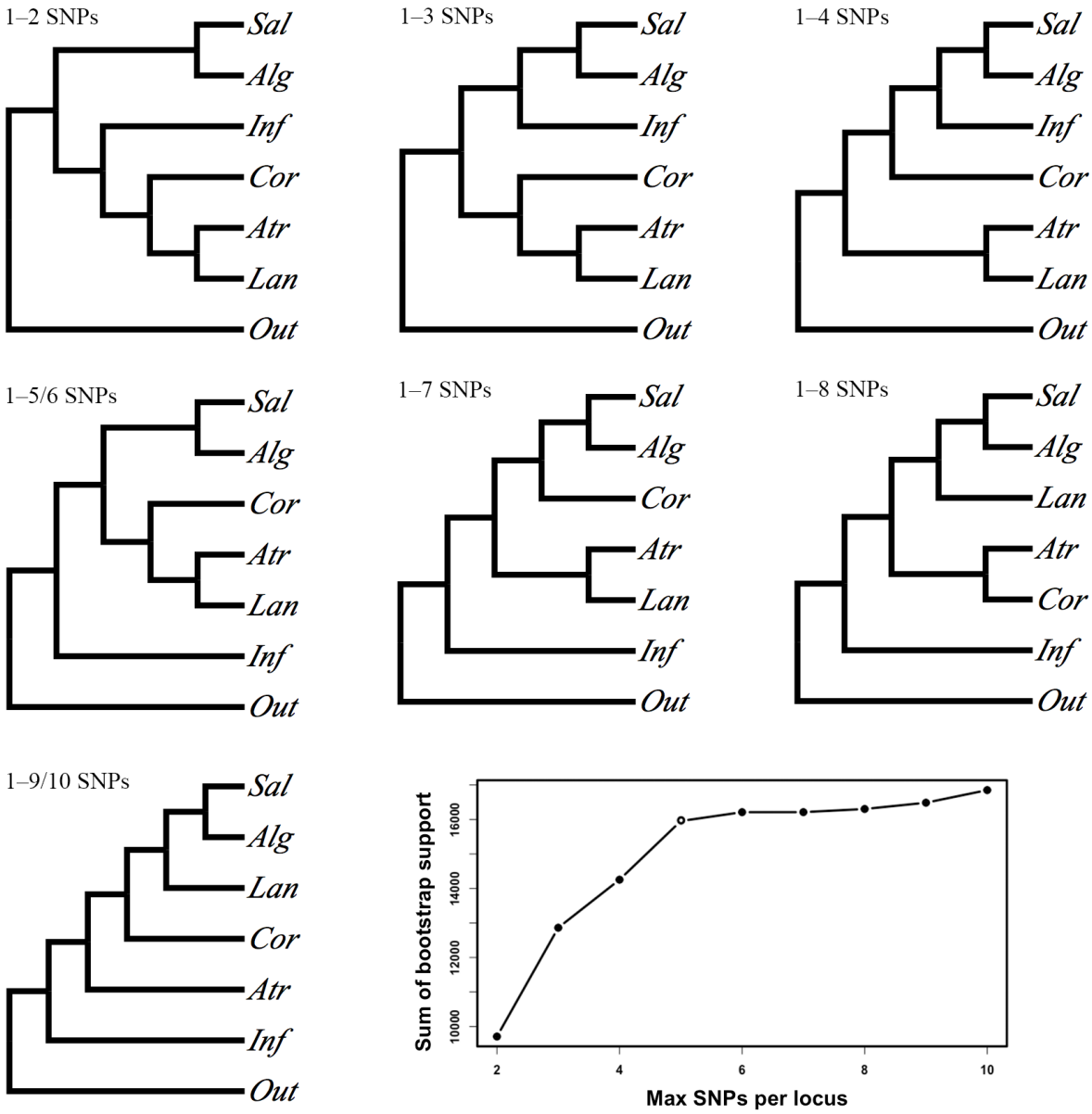
Supplementary Table A4.4: P-distances within and between *S. salamandra* subspecies (calculated in MEGA 7).

Subspecies	**Samples**	**Within Group p-distance (%)**	**Between Group p-distance (%)**														
			-	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**
S. s. bernardezi	9	1.8	**1**	-														
S. s. alfredschmidti	12	0.8	**2**	1.4	-													
S. s. salamandra	43	1.3	**3**	2.0	1.7	-												
S. s. gallaica	15	1.3	**4**	1.9	1.6	1.4	-											
S. s. bejarae	2	0.4	**5**	1.8	1.4	1.3	1.1	-										
S. s. hispanica	2	1.1	**6**	1.9	1.6	1.3	1.3	1.2	-									
S. s. fastuosa	6	1.5	**7**	1.9	1.5	1.9	1.8	1.7	1.8	-								
S. s. morenica	11	1.2	**8**	2.1	1.8	1.7	1.5	1.3	1.5	2.0	-							
S. s. crespoi	2	2.4	**9**	2.5	2.2	2.2	2.0	1.9	2.0	2.4	1.8	-						
S. s. terrestris	41	1.2	**10**	2.0	1.7	1.7	1.4	1.3	1.3	1.9	1.6	2.2	-					
S. s. beschkovi	2	0.9	**11**	1.9	1.6	1.6	1.3	1.2	1.2	1.8	1.5	2.0	1.3	-				
S. s. almanzoris	2	0.2	**12**	1.9	1.6	1.6	1.4	1.1	1.5	1.8	1.5	2.0	1.6	1.5	-			
S. s. gigliolii	6	1.6	**13**	2.0	1.7	1.7	2.0	1.9	2.0	1.6	2.1	2.5	2.0	2.0	1.9	-		
S. s. werneri	3	0.8	**14**	1.9	1.5	1.5	1.2	1.1	1.1	1.7	1.5	2.0	1.2	1.0	1.4	1.9	-	
S. s. longirostris	9	0.4	**15**	2.0	1.7	1.7	1.5	1.3	1.5	1.9	1.5	1.9	1.6	1.5	1.4	2.1	1.4	-

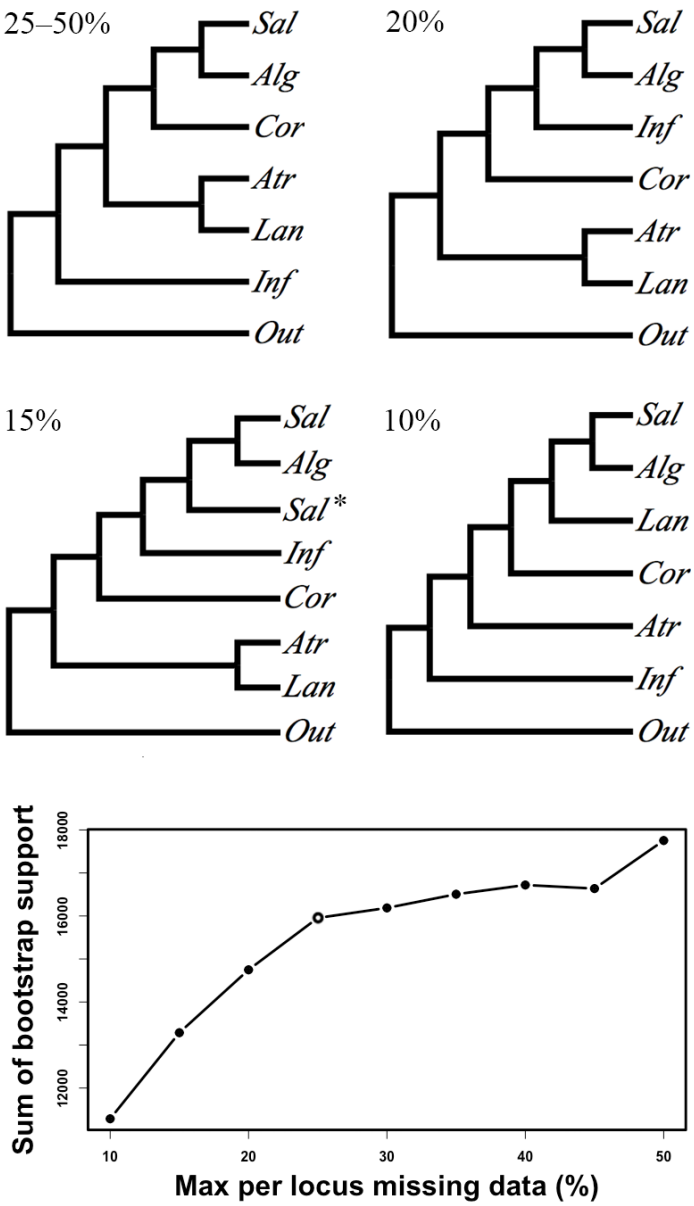
A4.2. Supplementary Figures



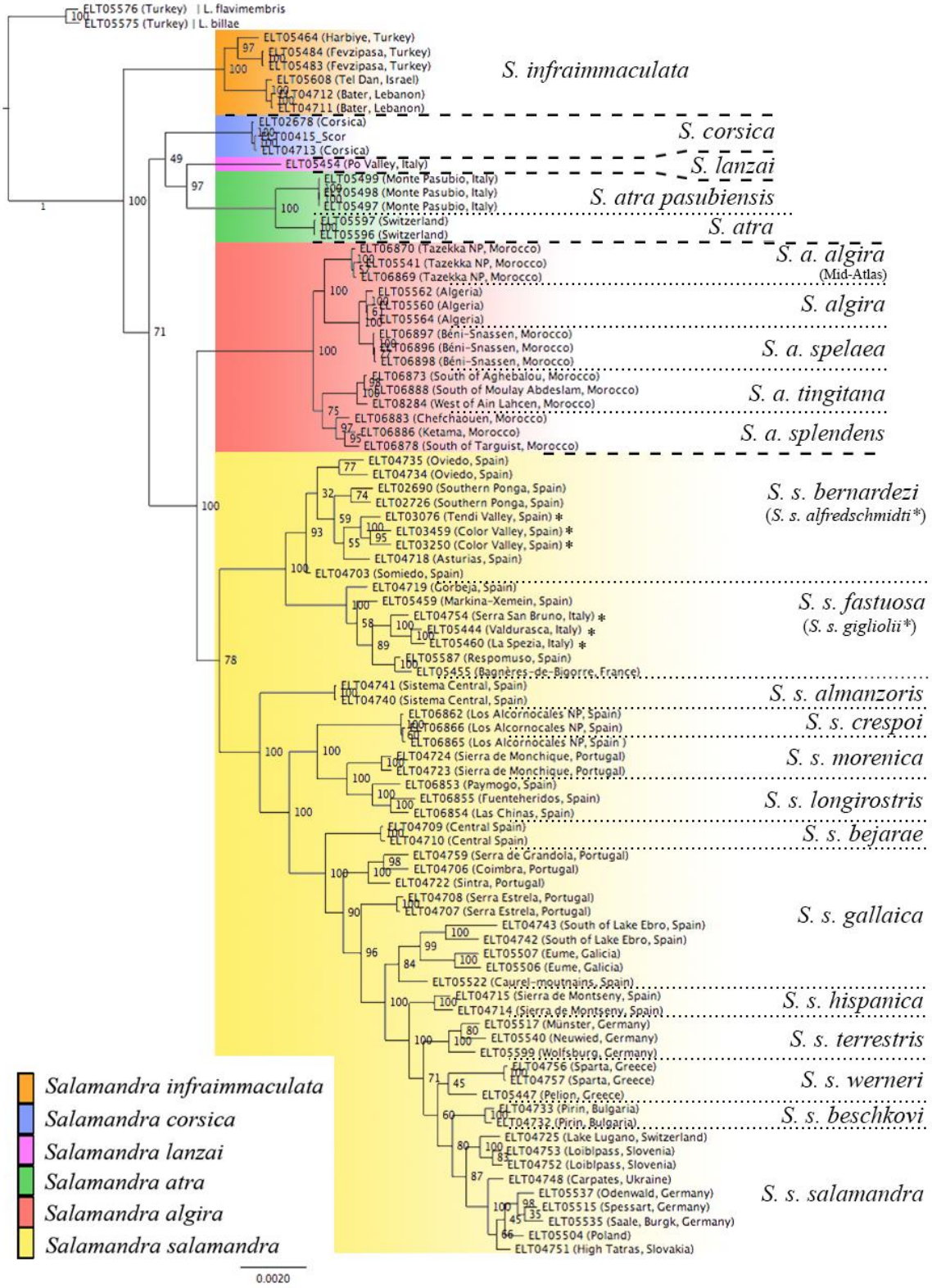
Supplementary Figure A4.1: SNP calling optimisation trials. **Left**: the total number of SNPs retained, allowing for different numbers of per locus variable sites (SNPs). **Right**: the total number of SNPs retained allowing for different levels of per locus missing data. White/hollow points show the values selected for later analyses.



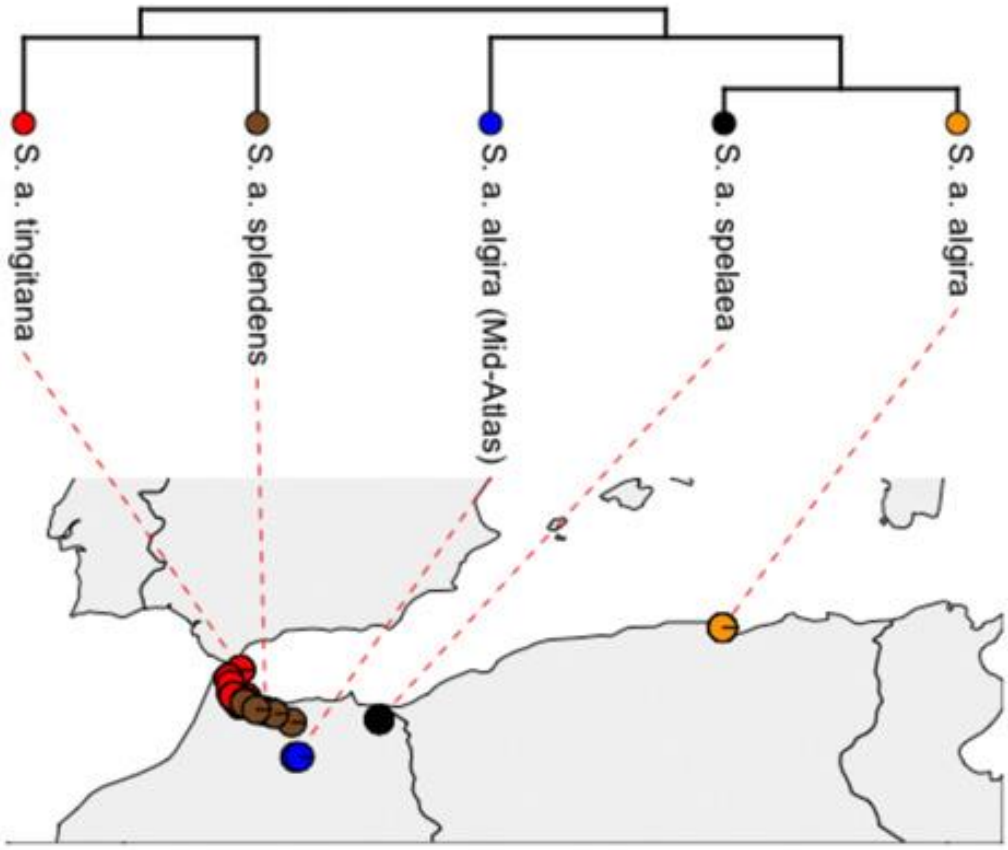
Supplementary Figure A4.2: Alternative topologies between the six currently recognised *Salamandra* species returned by RAxML analysis of SNP data when allowing different numbers of maximum SNPs per locus (tips within species have been collapsed; branch lengths are arbitrary and do not reflect evolutionary rates). The maximum per locus missing data in each instance was set to 25%. Tip labels: *Sal = Salamandra salamandra*; *Alg = Salamandra algira*; *Cor = Salamandra corsica*; *Atr = Salamandra atra*; *Lan= Salamandra lanzai*; *Inf = Salamandra infraimmaculata*; *Out = Lyciasalamandra*. The plot shows the sum of bootstrap support for each analysis (the white/hollow point shows the value selected for later analyses).



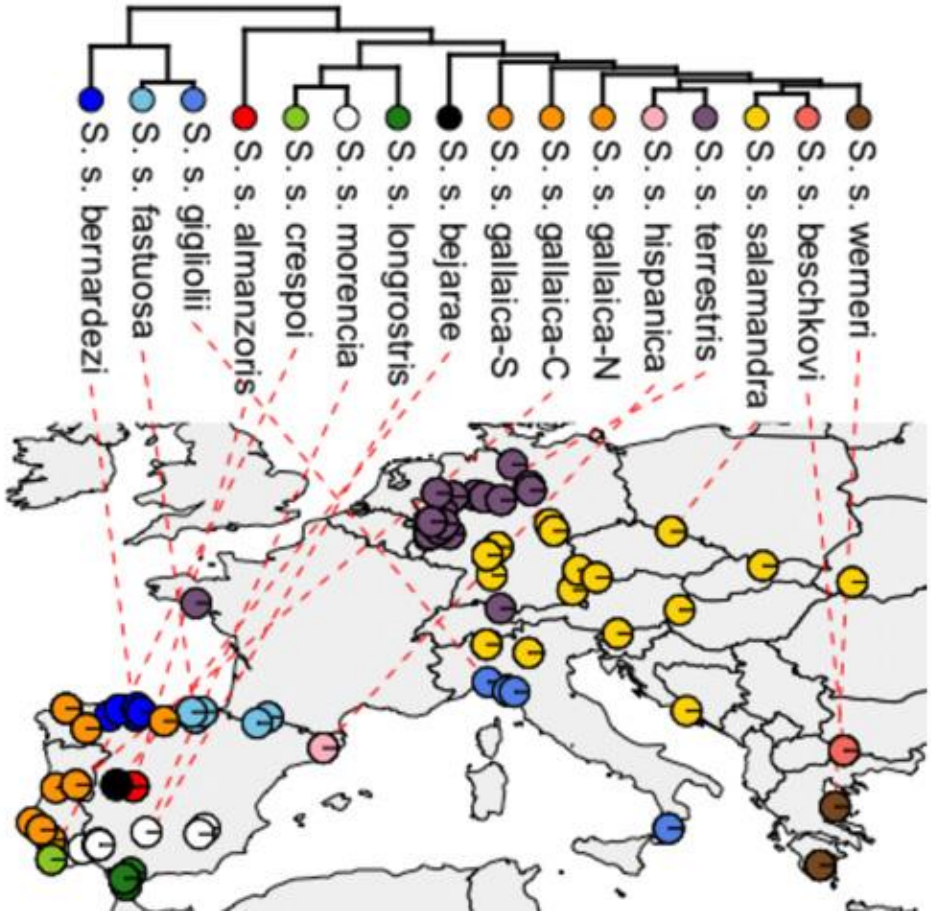
Supplementary Figure A4.3: Alternative topologies between the six currently recognised *Salamandra* species returned by RAxML analysis of SNP data when allowing different levels of maximum per locus missing data (tips within species have been collapsed; branch lengths are arbitrary and do not reflect evolutionary rates). The maximum SNPs per locus in each instance was set to five. Taxa codes: *Sal*/*Sal** = *Salamandra salamandra*; *Alg* = *Salamandra algira*; *Cor* = *Salamandra corsica*; *Atr* = *Salamandra atra*; *Lan*= *Salamandra lanzai*; *Inf* = *Salamandra infraimmaculata*; *Out* = *Lyciasalamandra*. The plot shows the sum of bootstrap support for each analysis (the white/hollow point shows the value selected for later analyses).



Supplementary Figure A4.4: RAxML phylogenetic tree of 89 individuals based on GTRCAT analysis of 4905 RAD-loci (294,300 nt). Coloured blocks highlight currently recognised species. Node values show bootstrap support. Dashed lines separate species; dotted lines separate subspecies.



Supplementary Figure A4.5: Intraspecific relationships and geographic sampling distributions of *S. algira* subspecies (phylogeny simplified from Fig. 3.5; sample localities are for the 44 individuals included in PCA analyses; Fig. 3.5).



Supplementary Figure A4.6: Intraspecific relationships and geographic sampling distributions of *S. salamandra* subspecies (phylogeny simplified from Fig. 3.5; sample localities are for the 151 individuals for this species group included in PCA analyses; Fig. 3.6). The N, C and S after *S. s. gallaica* indicate the clades approximate geographic position in western Iberia (northern, central and southern respectively)

Appendix 5: supplementary material for Chapter 4

A5.1. Supplementary results

A5.1.1. Colour pattern diversity: sample site and sex

To assess whether the striped phenotype within the polymorphic population is comparable to the more widely distributed (ancestral) phenotype, colour pattern features were extracted from 57 salamanders from a monomorphic population (population 1; Fig. 4.1) and 61 striped salamanders from the polymorphic population; the female to male ratio (F:M) in each group was 1.3:1 and 1:1 respectively. A *Patternize* PCA found the two groups to completely overlap (Sup. Fig. A5.1), with no significant difference seen in either PC1 (One-way ANOVA: $F_{(1,116)}$=; 2.35; P= 0.128) or PC2 scores ($F_{(1,116)}$=0.1; P= 0.755).

Patternize PCA also revealed overlapping clusters between male (n=95) and female (n=125) salamanders in the polymorphic population (Sup. Fig. A5.1). The first axis (PC1) corresponded to both the extent and brightness of lateral stripes and the second axis (PC2) largely corresponded to the presence and width of a black dorsal stripe. As PC2 also appeared to contain elements of shading, an artefact of lighting during imaging, only PC1 was used to interpret colour phenotype in subsequent analyses. PC1 scores were compared *via* one-way ANOVA; as the data were found to be binomial, they were first split into two groups, one corresponding to xanthic and striped salamanders and the other hypolutic salamanders. No significant difference was seen between male and female colour patterns in either the xanthic-striped ($F_{(1,175)}$= 2.36; P= 0.127) or hypolutic ($F_{(1,35)}$= 2.65; P= 0.112) groupings.

A5.1.2. Morphological variation

Six morphological measures were taken from 206 salamanders polymorphic salamanders from sample site 7 (F:M=1.26:1): the Inter-orbital, Orbital-Nostril, Snout-to-vent length (SVL), Tail, Upper-Forelimb and Upper-Hindlimb. When using the PC1 scores from the *Patternize* analysis as a measure of colour phenotype, no significant association was seen between morphology and colour pattern (one-way MANOVA: Pillai=0.05; $F_{(1,204)}$=1.59; P=0.15). However, when ternary ranks were used as the phenotype input—striped (n=58), xanthic (n=108) or hypolutic (n=40)—small but significant differences were seen (one-

way MANOVA: Pillai=0.15; $F_{(2,203)}$=2.69; P= 0.002; Partial eta^2=0.08). One-way ANOVA revealed that the O-N ($F_{(2,203)}$=8.29; P<0.005) and SVL ($F_{(2,203)}$=3.83; P=0.023) differed significantly by colour morph.

Significant differences in morphology were also seen between male and female salamanders (one-way MANOVA: Pillai=0.23; $F_{(1,204)}$=9.97; P<0.005; Partial eta^2=0.23). *Post hoc* one-way ANOVA revealed that females had a significantly a larger I-O ($F_{(1,204)}$= 16.84; P<0.005) and SVL ($F_{(1,204)}$= 6.74; P=0.01), while males had larger UH ($F_{(1,204)}$= 5.9; P=0.02).

A5.1.3. Associations between colour and toxin: a subspecies comparison

Comparisons were made between the metabolomic contents of toxic secretions from *S. s. bernardezi* and *S. s. terrestris*. One-way ANOVA revealed that 11 of the putative metabolites identified by GC-MS differed significantly between the two subspecies (Sup. Tab. A5.9), including the alkaloids, samandarone ($F_{(1,20)}$=132.5; P<0.005), samandaridine ($F_{(1,21)}$= 22.26; P<0.005) and cycloneosamandione ($F_{(1,22)}$=32.29; P<0.005). Similar results were seen when comparing just striped *S. s. bernardezi* to *S. s. terrestris* (Sup. Tab. A5.9), with one additional feature (Peak 17_490; $F_{(1,10)}$= 6.07; P=0.033).

A5.1.4. Population genomics

For population genomic analyses, 82 salamanders were ddRAD-Seq genotyped. Of the 268,491 catalogue loci, 4702 were polymorphic, passed filtering criteria and were retained in the final data set. Observed heterozygosity per individual averaged 0.2 (range: 0.16–0.23), which was lower than the expected value (0.28). By sample site, the average observed heterozygosities were: 1) 19.8%; 2) 20.3%; 3) 17.9%; 4) 19.2%; 5) 19.9%; 6) 20.8%; 7) 21.1%. While low, significant *Fst* differences between these populations were seen (Sup. Tab. A5.2) and they also presented strong isolation by distance (Mantel test: r=0.795; p<0.005; Sup. Fig. A5.6).

A principal component analyses of the data broadly grouped samples by putative population, not colour phenotype, with overlap seen between populations 1–2, 4–5, and 6–7, similar to RAxML analyses (Fig 4.7). However, the first two axes (PCs 1 and 2) only explained 12.09% of the variance, showing that this clustering was weak.

A5.2. Supplementary Tables

Supplementary Table A5.1: Mass matrix table for known *Salamandra* alkaloids (based on Mebs and Pogoda 2005).

Name	Formula	M+	-NH	-OH	[M+#TMS]+	[M+#TMS]+ -CH3
Samandarine	C19H31NO2	305	1	1	449	434
Samandarone	C19H29NO2	303	1	0	375	360
Samandaridine	C21H31NO3	345	1	0	417	402
Samandinine	C24H39NO3	389	1	0	461	446
Cycloneosamandione	C19H29NO2	303	0	1	375	360
Samandenone	C22H33NO2	343	1	0	415	400
Samanine	C19H33NO	291	1	1	435	420

Supplementary Table A5.2: One-way MANOVA and *post hoc* one-way ANOVA outputs for spectrophotometry data analyses.

Analysis	Spctral range (nm)	MANOVA	Partial eta^2	Hue (aov)	Saturation (aov)	Brightness (aov)
Head	300-750	Pillai=0.93; F(2,21)= 5.73; P<0.005***	0.46	F(2,21)=47.73; P<0.005***	F(2,21)=7.96; P=0.003**	F(2,21)=0.25; P=0.782
Parotid	300-750	Pillai=0.53; F(2,21)=2.42; P=0.043*	0.27	F(2,21)=6.02; P=0.009**	F(2,21)=3.23; P=0.06.	F(2,21)=1.69; P=0.208
Dorsal	300-750	Pillai=0.54; F(2,21)=2.49; P=0.039*	0.27	F(2,21)=8.06; P=0.003**	F(2,21)=0.76; P=0.481	F(2,21)=0.74; P=0.489
Lateral	300-750	Pillai=1.02; F(2,21)=6.89; P<0.005***	0.51	F(2,21)=21.66; P<0.005***	F(2,21)=4.70; P=0.02*	F(2,21)=9.83; P<0.005***
Gular	300-750	Pillai=0.48; F(2,21)=2.1; P=0.075	0.24	F(2,21)=3.29; P=0.057	F(2,21)=3.24; P=0.06	F(2,21)=0.4; P=0.68
Venter	300-750	Pillai=0. 33; F(2,21)=1.29; P=0.282	0.16	F(2,37)=0.83; P=0.45	F(2,37)=3.48; P=0.05*	F(2,37)=0.54; P=0.59
Skin Colour	300-750	Pillai=0.9; F(2,37)=9.75; P<0.005***	0.45	F(2,37)=36.22; P<0.005***	F(2,37)=13.56; P<0.005***	F(2,37)=30.14; P<0.005***
Skin Colour (UV)	360-400	Pillai=0.3; F(2,39)=2.28; P=0.045*	0.15	F(2,39)=0.73; P=0.5	F(2,39)=3.1; P=0.056	F(2,39)=0.65; P=0.53
Skin Colour (Visible)	400-700	Pillai=0.96; F(2,39)=11.64; P<0.005***	0.48	F(2,39)=2.05; P=0.142	F(2,39)=150.24; P<0.005***	F(2,39)=14.5; P<0.005***
Skin Colour (NIR)	700-750	Pillai=0.8; F(2,39)=8.53; P<0.005***	0.4	F(2,39)=41.19; P<0.005***	F(2,39)=7.76; P=0.002**	F(2,39)=25.61; P<0.005***
Black vs. Brown	300-750	Pillai=0.18; F(1,24)=0.6; P=0.62	0.08	F(1,24)=0.001; P=0.97	F(1,24)=0.12; P=0.74	F(1,24)=1.93; P=0.18
Black vs. Brown (UV)	360-400	Pillai=0.38; F(1,24)=4.58; P=0.012*	0.38	F(1,24)=0.52; P=0.48	F(1,24)=3.02; P=0.095	F(1,24)=1.21; P=0.28
Black vs. Brown (Visible)	400-700	Pillai=0.41; F(1,24)=5.13; P=0.008**	0.41	F(1,24)=1.45; P=0.24	F(1,24)=5.48; P=0.028*	F(1,24)=1.9; P=0.18
Black vs. Brown (NIR)	700-750	Pillai=0.19; F(1,24)=1.68; P=0.2	0.19	F(1,24)=1.07; P=0.31	F(1,24)=1.3; P=0.27	F(1,24)=3.02; P=0.095

Supplementary Table A5.3: GC-MS intensity values for the 18 putatively identified metabolites. Known alkaloids: Cycloneosamandione = Peak02_375; Samandarone = Peak07_375; Samandaridine = Peak19_417; and Samandarine = Peak06_449.

ELT no.	Morph	Subspecies	Peak 01_363	Peak 02_375	Peak 03_377	Peak 04_303	Peak 05_449	Peak 06_449	Peak 07_375	Peak 08_349	Peak 09_419	Peak 10_530	Peak 12_568	Peak 13_506	Peak 14_486	Peak 15_535	Peak 16_463	Peak 17_490	Peak 18_535	Peak 19_417
05622	Yellow	*bernardezi*	1103623	2736	159639	201969	24755	242803	162113	20385	20041	597658	*NA*	*NA*	40409	*NA*	37217	*NA*	*NA*	12120
05648	Yellow	*bernardezi*	3567291	14181	4458512	4290232	278195	2143805	1758918	2548514	18731693	589734	19391	11552	275302	108753	1024662	141938	*NA*	6960121
05618	Yellow	*bernardezi*	5804022	8432	818480	1926784	203656	2474448	4260276	1045565	15404586	999890	17587	10279	372994	40990	343792	42908	6767427	4379110
05655	Yellow	*bernardezi*	2025222	13214	469752	672211	127040	2086602	1065219	55706	559751	861682	1775	NA	94890	1098	39069	2721	*NA*	441298
05613	Yellow	*bernardezi*	5616679	12306	1308417	1609419	198990	4234035	2749899	447203	5437810	1351919	33609	6515	404358	24798	523064	19066	4607857	1644931
05656	Yellow	*bernardezi*	4891202	4554	254953	631786	714812	2969379	4490475	368126	9784873	1096422	5148	4743	298879	37050	230530	70072	4257527	1689037
05641	Brown	*bernardezi*	5451665	*NA*	310278	323278	629350	3973043	2126206	148442	3823116	457461	7473	4309	364957	5080	157476	25717	2040150	1646049
05625	Brown	*bernardezi*	7189651	1988	1673776	1375718	224192	13396982	3307158	416701	10688393	967759	11721	5190	627537	27037	377201	72459	6405722	4522946
05653	Brown	*bernardezi*	2188501	1775	18488	28999	134797	1267494	1635391	39349	26314	724064	*NA*	1184	212530	35515	47221	50145	3765602	2698300
05617	Brown	*bernardezi*	3920156	*NA*	116299	127438	226716	4600645	3032273	56516	3493032	751527	24619	5097	284937	21848	87361	17678	3885074	1392454
05651	Brown	*bernardezi*	5720429	34737	291210	132628	*NA*	4540367	770627	121008	2657188	519319	2097	*NA*	384193	5630	298338	2123	2909754	531724
05612	Brown	*bernardezi*	6480767	5514	674968	532009	*NA*	13106628	9767592	690340	20860872	342530	83313	30135	589653	175218	1463669	174316	*NA*	*NA*
05631	Striped	*bernardezi*	5824947	28089	189752	402656	462688	2325664	2414119	327378	8269160	590179	3261	*NA*	402062	27862	379784	13346	5282411	1329894
05630	Striped	*bernardezi*	4543946	5685	204096	175437	44128	4855687	2867645	132878	5565363	642843	4294	*NA*	408451	17754	148265	28587	4430260	2126444
05637	Striped	*bernardezi*	6646512	7351	705649	624864	38481	3877517	1912342	668261	14284412	508644	7687	5320	371995	26403	412343	48470	6038484	2725624
05634	Striped	*bernardezi*	6930441	2068	66657	238394	632912	2888901	4739667	227648	10277701	652614	8339	2492	471957	31860	314504	21766	7319324	2672709
05627	Striped	*bernardezi*	5222426	18769	308523	555532	752302	4441831	5332309	192032	6083399	986942	10285	2850	352379	50655	340849	68865	8045991	3619175
05639	Striped	*bernardezi*	6508826	22521	494418	448532	1181763	5464871	2687545	305298	7094910	1219295	3740	*NA*	382424	27714	346472	82917	4341422	4022649
02668	Striped	*terrestris*	2086225	1170765	2694224	5412853	*NA*	4627499	14750316	1019986	25100261	556145	52742	27181	138316	31271	241052	69845	4611736	6156966
02671	Striped	*terrestris*	57670	4495064	4144186	9343032	*NA*	3418265	22439992	1708771	40724250	760297	1749667	491981	7612	29657	228869	77887	5451126	6726405
02675	Striped	*terrestris*	146548	6574573	3387469	11349848	*NA*	748406	8704008	1943624	17534999	123179	165463	88350	34553	41132	715640	54180	7159822	5427344
02669	Striped	*terrestris*	590336	540494	2781673	5055674	*NA*	7163080	15603466	692590	29252999	937424	29411	22284	23800	21433	177945	81951	6835458	7231850
02673	Striped	*terrestris*	403154	1070580	3554170	8718849	73551	2025390	2133707	2527236	19549150	179245	120954	49891	28479	25806	198156	72535	3893071	4382032
02672	Striped	*terrestris*	779357	15496196	4273561	10671412	*NA*	2998904	21407871	1777659	32982363	1136755	413268	116214	61262	143453	908355	175105	*NA*	7066820

Supplementary Table A5.4: One-way ANOVA outputs for metabolomic data analyses.

Comparison	Peak	Alkaloid?	Log transformed	F-value	Degrees of freedom	P-value
S. s. bernardezi vs. S. s. terrestris	01_363	-	No	32.56	1,22	<0.005
	02_375	Cycloneosamandione	Yes	132.50	1,20	<0.005
	03_377	-	Yes	19.32	1,22	<0.005
	04_303	-	No	109.50	1,22	<0.005
	05_449	-	No	0.75	1,15	0.401
	06_449	Samandarine	Yes	0.13	1,22	0.719
	07_375	Samandarone	No	32.29	1,22	<0.005
	08_349	-	Yes	13.16	1,22	0.002
	09_419	-	No	36.74	1,22	<0.005
	10_530	-	No	1.09	1,22	0.307
	12_568	-	Yes	27.95	1,20	<0.005
	13_506	-	Yes	31.69	1,16	<0.005
	14_486	-	No	24.93	1,22	<0.005
	15_535	-	Yes	0.83	1,21	0.372
	16_463	-	Yes	0.51	1,22	0.481
	17_490	-	Yes	3.52	1,21	0.075
	18_535	-	No	0.46	1,17	0.506
	19_417	Samandaridine	No	22.26	1,21	<0.005
S. s. bernardezi (striped only) vs. S. s. terrestris	01_363	-	No	118.60	1,10	<0.005
	02_375	Cycloneosamandione	Yes	69.88	1,10	<0.005
	03_377	-	Yes	55.53	1,10	<0.005
	04_303	-	No	54.93	1,10	<0.005
	05_449	-	No	0.88	1,5	0.392
	06_449	Samandarine	Yes	0.75	1,10	0.408
	07_375	Samandarone	No	11.48	1,10	0.007
	08_349	-	Yes	32.76	1,10	<0.005
	09_419	-	No	25.04	1,10	<0.005
	10_530	-	No	0.57	1,10	0.469
	12_568	-	Yes	28.74	1,10	<0.005
	13_506	-	Yes	19.69	1,7	0.003
	14_486	-	No	185.70	1,10	<0.005
	15_535	-	Yes	0.77	1,10	0.4
	16_463	-	Yes	0.04	1,10	0.849
	17_490	-	Yes	6.07	1,10	0.033
	18_535	-	No	0.13	1,9	0.728
	19_417	Samandaridine	No	32.48	1,10	<0.005
S. s. bernardezi (striped, xanthic & hypolutic)	01_363	-	No	2.57	2,15	0.11
	02_375	Cycloneosamandione	Yes	0.33	2,13	0.727
	03_377	-	Yes	1.29	2,15	0.305
	04_303	-	No	3.44	2,15	0.059
	05_449	-	No	1.05	2,13	0.379
	06_449	Samandarine	Yes	3.11	2,15	0.074
	07_375	Samandarone	No	0.37	2,15	0.697
	08_349	-	Yes	1.32	2,15	0.296

	09_419	-	No	0.11	2,15	0.895
	10_530	-	No	1.74	2,15	0.209
	12_568	-	Yes	0.91	2,13	0.429
	13_506	-	Yes	0.89	2,9	0.444
	14_486	-	No	2.89	2,15	0.087
	15_535	-	Yes	0.18	2,14	0.836
	16_463	-	Yes	0.28	2,15	0.761
	17_490	-	Yes	0.06	2,14	0.947
	18_535	-	No	2.60	2,11	0.119
	19_417	Samandaridine	No	0.14	2,14	0.874
S. s. bernardezi (ancestral vs. derived)	01_363	-	No	2.98	1,16	0.104
	02_375	Cycloneosamandione	Yes	0.63	1,14	0.441
	03_377	-	Yes	0.53	1,16	0.475
	04_303	-	No	0.17	1,16	0.686
	05_449	-	No	2.20	1,14	0.16
	06_449	Samandarine	Yes	0.26	1,16	0.62
	07_375	Samandarone	No	0.97	1,16	0.34
	08_349	-	Yes	0.19	1,16	0.67
	09_419	-	No	0.09	1,16	0.765
	10_530	-	No	0	1,16	0.973
	12_568	-	Yes	1.77	1,14	0.205
	13_506	-	Yes	1.35	1,10	0.272
	14_486	-	No	0.91	1,16	0.354
	15_535	-	Yes	0.23	1,15	0.636
	16_463	-	Yes	0.57	1,16	0.462
	17_490	-	Yes	0.12	1,15	0.737
	18_535	-	No	3.46	1,12	0.088
	19_417	Samandaridine	No	0.18	1,15	0.677
S. s. bernardezi (xanthic vs. hypolutic)	01_363	-	No	1.48	1,10	0.252
	02_375	Cycloneosamandione	Yes	0.46	1,8	0.515
	03_377	-	Yes	1.58	1,10	0.238
	04_303	-	No	3.13	1,10	0.107
	05_449	-	No	0.42	1,8	0.536
	06_449	Samandarine	Yes	3.68	1,10	0.084
	07_375	Samandarone	No	0.58	1,10	0.465
	08_349	-	Yes	1.54	1,10	0.243
	09_419	-	No	0.10	1,10	0.76
	10_530	-	No	3.57	1,10	0.088
	12_568	-	Yes	0.10	1,8	0.76
	13_506	-	Yes	0.41	1,7	0.544
	14_486	-	No	3.22	1,10	0.103
	15_535	-	Yes	0	1,9	0.988
	16_463	-	Yes	0.02	1,10	0.905
	17_490	-	Yes	0	1,9	0.991
	18_535	-	No	1.55	1,6	0.259
	19_417	Samandaridine	No	0.07	1,9	0.794

Supplementary Table A5.5: Values of *Fst* for all pairs of sample sites calculated from 4702 SNP loci across 80 individuals (GenoDive pairwise differentiation test). Light cells contain *Fst* statistics; shaded cells contain Bonferroni corrected adjusted p-values.

Sample site	1	2	3	4	5	6	7	
1	--	0.168	0.142	0.18	0.168	0.131	0.115	Fst
2	0.546	--	0.046	0.196	0.176	0.155	0.14	
3	0.042	0.042	--	0.183	0.164	0.152	0.135	
4	0.042	0.042	>0.001	--	0.068	0.164	0.131	
5	0.021	0.021	>0.001	>0.001	--	0.14	0.103	
6	>0.001	>0.001	>0.001	>0.001	>0.001	--	0.018	
7	>0.001	>0.001	>0.001	>0.001	>0.001	>0.001	--	
	p-value							

Supplementary Table A5.6: Likelihood values obtained for different levels of genetic groups (K) calculated by STRUCTURE analyses; the best fitting value (K = 4) is shown in bold. (Evanno table generated by STRUCTURE HARVESTER.) Each K was evaluated using 5 replicates.

K	**Mean LnP(K)**	**Stdev LnP(K)**	**Ln'(K)**	**|Ln''(K)|**	**Delta K**
1	-288667.26	18.937872	—	—	—
2	-275222.7	8.621485	13444.56	4512.86	523.44348
3	-266291	13.848285	8931.7	7194.98	519.557472
4	**-264554.28**	**96.450775**	**1736.72**	**148067.88**	**1535.165269**
5	-410885.44	201221.0042	-146331.16	24758.52	0.123041
6	-532458.08	175545.6096	-121572.64	53320.38	0.303741
7	-707351.1	427320.0841	-174893.02	64945.9	0.151984
8	-817298.22	256150.8757	-109947.12	—	—

Supplementary Table A5.7: A list of putative genes identified as differentially expressed in various comparisons (Y = Yellow skin; Bl = Black skin; Br = Brown skin; Striped = yellow *vs.* black comparison in striped individuals only; L1/L2 = Landmark 1/2) that have known/suspected colour associations (M =Melanophore related; I = Iridophore related; X = Xanthophore related; P = Pigmentation related; L = Leucophore related; * = putative colour association).

Putative Gene	Putative Protein (UniProtKB-human)	Colour Related?	Comparisons				
ABCG4	ATP-binding cassette sub-family G member 4	-	-	-	Y-Br	-	-
ACAN	Aggrecan core protein	-	-	-	-	Striped	L1-L2
ACER2	Alkaline ceramidase 2	-	-	-	Y-Br	-	-
ACOT8	Acyl-coenzyme A thioesterase 8	M*	Y-Bl	-	-	-	-
ACY1	Aminoacylase-1	M*	Y-Bl	Br-Bl	-	-	-
ADA2	Adenosine deaminase 2	I	Y-Bl	-	Y-Br	-	-
ADGRG1	Adhesion G-protein coupled receptor G1	-	-	Br-Bl	-	-	-
ADI1	1,2-dihydroxy-3-keto-5-methylthiopentene dioxygenase	-	-	Br-Bl	Y-Br	-	-
ADPRHL2	Poly(ADP-ribose) glycohydrolase ARH3	-	-	-	Y-Br	-	-
AGR2	Anterior gradient protein 2 homolog	M/X	-	-	Y-Br	Striped	-
AKR1C1	Aldo-keto reductase family 1 member C1	Possibly	-	Br-Bl	-	-	-
AKR1C2	Aldo-keto reductase family 1 member C2	M*	-	Br-Bl	-	-	-
AKR1C3	Aldo-keto reductase family 1 member C3	M*	-	Br-Bl	-	Striped	-
ANKH	Progressive ankylosis protein homolog	-	-	-	-	Striped	-
ANTXR2	Anthrax toxin receptor 2	M*	-	Br-Bl	-	-	-
AP1M1	AP-1 complex subunit mu-1	M	-	Br-Bl	-	-	-
AP3B2	AP-3 complex subunit beta-2	Possibly	-	Br-Bl	-	-	-
APEX1	DNA-(apurinic or apyrimidinic site) lyase	-	-	Br-Bl	Y-Br	-	-
ATP5S	ATP synthase subunit s, mitochondrial	-	-	Br-Bl	Y-Br	-	-
ATP6AP1	V-type proton ATPase subunit S1	M	-	-	Y-Br	-	-
AWAT1	Acyl-CoA wax alcohol acyltransferase 1	-	Y-Bl	-	-	-	-
B3GNT2	N-acetyllactosaminide beta-1,3-N-acetylglucosaminyltransferase 2	M*	-	-	-	Striped	-
BARX2	Homeobox protein BarH-like 2	P*	-	-	-	-	L1-L2
BCAS1	Breast carcinoma-amplified sequence 1	-	-	-	Y-Br	-	-
BCAT1	Branched-chain-amino-acid aminotransferase, cytosolic	-	-	-	-	Striped	-
C15orf39	Uncharacterized protein C15orf39	-	-	Br-Bl	-	-	-
c1galt1	Glycoprotein-N-acetylgalactosamine 3-beta-galactosyltransferase 1-A/B	-	-	Br-Bl	-	-	-

C21orf33	ES1 protein homolog, mitochondrial	-	-	Br-Bl	-	-	-
C3orf58	Deleted in autism protein 1	-	-	-	Y-Br	-	-
CALM	Calmodulin	M	-	Br-Bl	-	-	-
CAPSL	Calcyphosin-like protein	-	-	-	-	Striped	-
CCL19	C-C motif chemokine 19	-	-	-	-	-	L1-L2
CCM2L	Cerebral cavernous malformations 2 protein-like	M/X*	-	-	Y-Br	-	-
CD164L2	CD164 sialomucin-like 2 protein	-	-	Br-Bl	-	-	-
CECR5	Cat eye syndrome critical region protein 5	-	-	-	-	Striped	-
CIART	Circadian-associated transcriptional repressor	-	-	Br-Bl	-	-	-
CILP	Cartilage intermediate layer protein 1	-	-	Br-Bl	-	-	-
CIRBP	Cold-inducible RNA-binding protein	Possibly	-	Br-Bl	-	-	-
CLU	Clusterin	-	-	-	-	Striped	-
COL1A1	Collagen alpha-1(I) chain	-	-	-	-	-	L1-L2
COL2A1	Collagen alpha-1(II) chain	-	-	-	-	Striped	-
COMT	Catechol O-methyltransferase	M*	-	Br-Bl	-	-	-
CPE	Carboxypeptidase E	-	-	Br-Bl	-	-	-
CRYAB	Alpha-crystallin B chain	P*	-	-	Y-Br	-	-
CYB561	Cytochrome b561	-	Y-Bl	Br-Bl	-	Striped	-
CYP1A5/4	Cytochrome P450 1A5/4	M*	-	Br-Bl	-	-	-
Cyp2g1	Cytochrome P450 2G1	-	-	Br-Bl	Y-Br	-	-
CYP7A1	Cholesterol 7-alpha-monooxygenase	-	-	-	-	Striped	-
CYR61	Protein CYR61	Possibly	-	Br-Bl	-	-	-
DCT/TYRP2	L-dopachrome tautomerase	M	Y-Bl	-	Y-Br	-	-
DEFB family	Beta-defensin 1 (?)	M*	Y-Bl	-	Y-Br	Striped	-
DHCR7	7-dehydrocholesterol reductase	-	-	Br-Bl	-	-	-
DIXDC1	Dixin	M*	Y-Bl	-	-	Striped	-
DNAAF5	Dynein assembly factor 5, axonemal	-	-	Br-Bl	-	-	-
DST	Dystonin	M	-	-	Y-Br	-	-
DYRK2	Dual specificity tyrosine-phosphorylation-regulated kinase 2	M*	-	Br-Bl	-	-	-
ELMOD1	ELMO domain-containing prot. 1	M*	Y-Bl	Br-Bl	-	Striped	-
ELOVL7	Elongation of very long chain fatty acids protein 7	M*	-	-	Y-Br	Striped	-
ENTPD8	Ectonucleoside triphosphate diphosphohydrolase 8	M*	-	-	Y-Br	-	-
EPB41L1	Band 4.1-like protein 1	-	Y-Bl	-	-	-	-
ERN1	Serine/threonine-protein kinase/endoribonuclease IRE1	-	-	Br-Bl	-	-	-
FAM204A	Protein FAM204A	-	-	Br-Bl	-	-	-
FAM234A	Protein FAM234A	-	-	Br-Bl	-	-	-
FAM83A	Protein FAM83A	M*	-	-	-	Striped	L1-L2
FAR1	Fatty acyl-CoA reductase 1	P*	Y-Bl	Br-Bl	-	-	-
FCGBP	IgGFc-binding protein	-	-	Br-Bl	-	-	-
FCN2	Ficolin-2	-	-	-	-	Striped	-

FMOD	Fibromodulin	-	-	-	-	Striped	L1-L2
FMOGS-OX1	Flavin-containing monooxygenase FMO GS-OX1	-	-	-	-	Striped	-
FNIP2	Folliculin-interacting protein 2	-	Y-Bl	-	-	-	-
FOXA3	Hepatocyte nuclear factor 3-gamma	P*	-	Br-Bl	-	-	-
Ftl1	Ferritin light chain 1	-	-	Br-Bl	-	-	-
FUT2	Galactoside 2-alpha-L-fucosyltransferase 2	-	-	Br-Bl	-	-	-
GALC	Galactocerebrosidase	P*	-	Br-Bl	-	-	-
GCNT7	Beta-1,3-galactosyl-O-glycosyl-glycoprotein beta-1,6-N-acetylglucosaminyltransferase 7	-	-	Br-Bl	-	Striped	-
GDPD4	Glycerophosphodiester phosphodiesterase GDPD4	-	-	-	-	Striped	-
Ggta1l1	N-acetyllactosaminide alpha-1,3-galactosyltransferase-like 1	-	-	Br-Bl	-	-	-
GPAT3	Glycerol-3-phosphate acyltransferase 3	-	-	Br-Bl	-	-	-
GSTT1	Glutathione S-transferase theta-1	M*	-	-	Y-Br	-	-
GTF2I	General transcription factor II-I	-	Y-Bl	-	-	-	-
HADHB	Trifunctional enzyme subunit beta, mitochondrial	-	-	-	Y-Br	-	-
HMGCS2	Hydroxymethylglutaryl-CoA synthase, mitochondrial	-	-	Br-Bl	-	-	-
HNMT	Histamine N-methyltransferase	-	-	Br-Bl	-	-	-
HS3ST3B1	Heparan sulfate glucosamine 3-O-sulfotransferase 3B1	P*	-	Br-Bl	Y-Br	-	-
HSD17B11	Estradiol 17-beta-dehydrogenase 11	-	-	-	-	-	L1-L2
HSD17B14	17-beta-hydroxysteroid dehydrogenase 14	-	Y-Bl	Br-Bl	-	-	-
IAH1	Isoamyl acetate-hydrolyzing esterase 1 homolog	-	Y-Bl	-	-	-	-
IFT172	Intraflagellar transport protein 172 homolog	P*	-	-	Y-Br	-	-
IGLV5-48	Protein IGLV5-48	-	-	Br-Bl	Y-Br	-	-
IGSF8	Immunoglobulin superfamily member 8	-	-	Br-Bl	-	-	-
JCHAIN	Immunoglobulin J chain	-	-	-	Y-Br	-	-
KCNMA1	Calcium-activated potassium channel subunit alpha-1	-	-	-	-	-	L1-L2
KDSR	3-ketodihydrosphingosine reductase	P*	-	Br-Bl	-	-	-
KIAA0319L	Dyslexia-associated protein KIAA0319-like protein	-	-	-	-	Striped	-
KLHDC7A	Kelch domain-containing protein 7A	-	-	-	-	Striped	-
KPNA4	Importin subunit alpha-3	-	-	Br-Bl	-	-	-
LACE1	Lactation elevated protein 1	-	-	-	Y-Br	-	-
LMX1B	LIM homeobox transcription factor 1-beta	M	-	-	Y-Br	-	-
LRP5	Low-density lipoprotein receptor-related protein 5	M*	Y-Bl	Br-Bl	-	-	-
LTC4S	Leukotriene C4 synthase	-	-	Br-Bl	-	-	-
MAL	Myelin and lymphocyte protein	-	-	Br-Bl	-	Striped	-
MAPK12	Mitogen-activated protein kinase 12	M*	-	Br-Bl	Y-Br	-	-
MAPRE2	Microtubule-associated protein RP/EB family member 2	M*	Y-Bl	-	-	-	-

MATN1	Cartilage matrix protein	-	-	-	-	Striped	-
MBLAC1	Metallo-beta-lactamase domain-containing protein 1	-	-	-	Y-Br	-	-
MGC75872	Uncharacterized protein	-	-	-	-	Striped	-
MICAL2	[F-actin]-methionine sulfoxide oxidase MICAL2	P*	-	Br-Bl	-	-	-
MLANA	Melanoma antigen recognized by T-cells 1	M	Y-Bl	-	Y-Br	Striped	-
MMP1	Interstitial collagenase	Possibly	-	Br-Bl	Y-Br	-	-
MMP3	Stromelysin-1	Possibly	-	Br-Bl	-	-	-
MROH1	Maestro heat-like repeat-containing protein family member 1	-	Y-Bl	Br-Bl	-	Striped	-
MTHFR	Methylenetetrahydrofolate reductase	Possibly	-	Br-Bl	-	-	-
MXRA7	Matrix-remodeling-associated protein 7	-	-	Br-Bl	-	-	-
MYH7	Myosin-7	-	-	Br-Bl	-	-	-
MYO1E	Unconventional myosin-Ie	Possibly	-	-	Y-Br	-	-
NACC2	Nucleus accumbens-associated protein 2	-	-	-	-	Striped	-
NDUFAF5	Arginine-hydroxylase NDUFAF5, mitochondrial	-	-	-	Y-Br	-	-
NIPAL4	Magnesium transporter NIPA4	M*	-	Br-Bl	-	-	-
NLRP3	NACHT, LRR and PYD domains-containing protein 3	-	-	Br-Bl	-	-	-
NME1	Nucleoside diphosphate kinase A	Possibly	-	-	-	Striped	-
NR4A2	Nuclear receptor subfamily 4 group A member 2	M	-	-	Y-Br	-	-
NUAK2	NUAK family SNF1-like kinase 2	P*	-	-	Y-Br	-	-
NUDT7	Peroxisomal coenzyme A diphosphatase NUDT7	-	-	Br-Bl	-	-	-
NXPE1 or *2*	NXPE family member 1/2	-	-	Br-Bl	Y-Br	-	-
OCA2	P protein	M/I	Y-Bl	-	-	Striped	-
OPLAH	5-oxoprolinase	-	-	-	Y-Br	-	-
P2RX1	P2X purinoceptor 1	Possibly	-	Br-Bl	-	-	-
PAFAH2	Platelet-activating factor acetylhydrolase 2, cytoplasmic	-	-	-	Y-Br	-	-
PAX7	Paired box protein Pax-7	X	Y-Bl	-	-	Striped	-
PCSK1	Neuroendocrine convertase 1	M	Y-Bl	Br-Bl	-	Striped	-
PDE1B	Calcium/calmodulin-dependent 3',5'-cyclic nucleotide phosphodiesterase 1B	P	-	-	Y-Br	-	-
PIGY	Phosphatidylinositol N-acetylglucosaminyltransferase subunit Y	-	-	-	Y-Br	-	-
PLIN2	*perilipin-2-like*	P*	Y-Bl	-	-	-	-
PMEL	Melanocyte protein PMEL	M	Y-Bl	-	Y-Br	Striped	-
PNP	Purine nucleoside phosphorylase	I	Y-Bl	-	Y-Br	Striped	-
POLD4	DNA polymerase delta subunit 4	-	-	-	Y-Br	-	-
PPP1R1B	Protein phosphatase 1 regulatory subunit 1B	A	-	-	Y-Br	-	-
PREX1	Phosphatidylinositol 3,4,5-trisphosphate-dependent Rac exchanger 1 protein	-	-	-	Y-Br	-	-
PSME1	Proteasome activator complex subunit 1	-	-	-	Y-Br	-	-

PTPN3	Tyrosine-protein phosphatase non-receptor type 3	M	-	-	Y-Br	-	-
PTX4	Pentraxin-4	-	-	-	-	-	L1-L2
RASEF	Ras and EF-hand domain-containing protein	Possibly	-	Br-Bl	-	-	-
RCAN1	Calcipressin-1	Possibly	Y-Bl	-	Y-Br	-	-
RD3L	Protein RD3-like	-	-	Br-Bl	Y-Br	-	-
RETSAT	All-trans-retinol 13,14-reductase	-	Y-Bl	Br-Bl	-	-	-
RGS16	Regulator of G-protein signaling 16	Possibly	-	-	Y-Br	-	-
RNF40	E3 ubiquitin-protein ligase BRE1B	-	-	-	Y-Br	-	-
ROGDI	Protein rogdi homolog	M*	-	-	Y-Br	-	-
RPUSD1	RNA pseudouridylate synthase domain-containing protein 1	-	-	Br-Bl	Y-Br	-	-
RUNDC3B	RUN domain-containing protein 3B	M*	-	Br-Bl	-	-	-
SCNN1B	Amiloride-sensitive sodium channel subunit beta	-	-	Br-Bl	-	-	-
SEC14L1	SEC14-like protein 1	-	-	-	Y-Br	-	-
SEL1L3	Protein sel-1 homolog 3	M*	-	-	Y-Br	-	-
SEMA4F	Semaphorin-4F	Possibly	-	-	Y-Br	-	-
SHC4	SHC-transforming protein 4	M	-	-	Y-Br	-	-
SLC2A11	SLC2A11	X/M	Y-Bl	-	-	-	-
SLC2A9	Solute carrier family 2, facilitated glucose transporter member 9	L/X*	Y-Bl	-	-	Striped	-
SLC35F6	Solute carrier family 35 member F6	Possibly	-	Br-Bl	-	-	-
SLC46A2	Thymic stromal cotransporter homolog	Possibly	-	Br-Bl	-	-	-
SLC7A14	Probable cationic amino acid transporter	Possibly	-	-	-	Striped	-
SMOC1	SPARC-related modular calcium-binding protein 1	-	-	Br-Bl	-	-	-
SMPDL3B	Acid sphingomyelinase-like phosphodiesterase 3b	-	-	Br-Bl	-	-	-
SNAI1	Zinc finger protein SNAI1	Possibly	-	-	Y-Br	-	-
SPDEF	SAM pointed domain-containing Ets transcription factor	-	Y-Bl	Br-Bl	-	Striped	-
SRD5A3	Polyprenol reductase	M*	-	Br-Bl	-	-	-
SUSD3	Sushi domain-containing protein 3	-	-	-	Y-Br	Striped	-
SYNRG	Synergin gamma	-	-	Br-Bl	-	-	-
TAP1	Antigen peptide transporter 1	-	-	Br-Bl	-	-	-
TBX3	T-box transcription factor TBX3	Possibly	-	-	-	-	L1-L2
TF	Serotransferrin	-	-	-	Y-Br	-	-
TFEC	Transcription factor EC (up or down regulates transcription)	M	-	-	Y-Br	Striped	-
TGFBI	Transforming growth factor-beta-induced protein ig-h3	M*	-	-	-	-	L1-L2
THAP9	DNA transposase THAP9	-	-	-	Y-Br	-	-
THBS4	Thrombospondin-4	-	-	-	-	Striped	L1-L2
TLCD1	Calfacilitin	-	Y-Bl	-	-	-	-
TMEM98	Transmembrane protein 98	Possibly	-	Br-Bl	Y-Br	-	-
TNC	Tenascin C	M*	-	-	-	-	L1-L2
TNFAIP2	Tumor necrosis factor alpha-induced protein 2	-	-	-	Y-Br	-	-

TNK1	Non-receptor tyrosine-protein kinase TNK1	-	-	-	Y-Br	-	-
TOR2A	Prosalusin	-	-	-	Y-Br	-	-
TPBG	Trophoblast glycoprotein	Possibly	-	Br-Bl	-	-	-
TRAPPC3L	Trafficking protein particle complex subunit 3-like protein	Possibly	-	Br-Bl	-	-	-
TRH	Pro-thyrotropin-releasing hormone	-	-	-	-	Striped	-
TRIM25	E3 ubiquitin/ISG15 ligase TRIM25	Possibly	-	-	Y-Br	-	-
TRIM32	E3 ubiquitin-protein ligase TRIM32	M	-	Br-Bl	Y-Br	-	-
TRIM39	E3 ubiquitin-protein ligase TRIM39	Possibly	-	Br-Bl	-	-	-
TRPV5	Transient receptor potential cation channel subfamily V member 5	-	-	Br-Bl	-	-	-
TSPAN1	Tetraspanin-1	M*	-	-	-	Striped	-
TTC19	Tetratricopeptide repeat protein 19, mitochondrial	-	-	-	Y-Br	-	-
TTC39A	Tetratricopeptide repeat protein 39A	-	Y-Bl	Br-Bl	-	-	-
TVP23B	Golgi apparatus membrane protein TVP23 homolog B	-	-	Br-Bl	Y-Br	-	-
TYR	Tyrosinase	M/I	Y-Bl	-	Y-Br	Striped	-
TYRP1	5,6-dihydroxyindole-2-carboxylic acid oxidase	M	Y-Bl	-	-	Striped	-
UMAD1	UBAP1-MVB12-associated (UMA)-domain containing protein 1	-	-	Br-Bl	-	-	-
UPK1B	Uroplakin-1b	Possibly	-	Br-Bl	-	-	-
UQCC2	Ubiquinol-cytochrome-c reductase complex assembly factor 2	-	-	-	Y-Br	-	-
VEGFA	Vascular endothelial growth factor A	M*	Y-Bl	Br-Bl	-	-	-
VMO1	Vitelline membrane outer layer protein 1	-	-	-	-	Striped	L1-L2
WBSCR27	Williams-Beuren syndrome chromosomal region 27 prot.	-	-	Br-Bl	-	-	-
WDCP	WD repeat and coiled-coil-containing protein	-	-	Br-Bl	Y-Br	-	-
ZDHHC2	Palmitoyltransferase ZDHHC2	-	-	Br-Bl	-	-	-

Supplementary Table A5.8: Known or suspected colour-related genes from Sup. Fig. A5.3 with identified function (M=Melanophore related; I=I ridophore related; X= Xanthophore related; P= Pigmentation related; L= Leucophore related; *= putative association). IMP= Integrative Multi-species Prediction platform; ESPCR: European Society for Pigment Cell Research.

Putative Gene	Colour Relation	IMP predicted function	GO Direct annotation (AmiGO 2)	Literature search
ACOT8	M*	H: Melanin metabolic process GO:0006582 (Pr=0.012)		-
ACY1	M*	ZF: Melanin metabolic process GO:0006582 (Pr=0.014)	-	-
ADA2	I	-	-	Purine metabolism (Pei et al. 2016)
AGR2	M/X	-	Tyrosine biosynthetic process (GO:0006571); Riboflavin biosynthetic process (GO:0009231)	-
AKR1C1 or *3*	Possibly	-	Retinoid metabolic process (GO:0001523); keratinocyte differentiation (GO:0030216)	-
AKR1C2	M*	-	-	Expression related to melanocytes and kerationcytes (Marin and Lin 2008)
AKR1C3	M*	H: Melanin metabolic process GO:0006582 (Pr=0.014)	Retinoid metabolic process (GO:0001523); keratinocyte differentiation (GO:0030216)	-
ANTXR2	M*	ZF: Melanin metabolic process GO:0006582 (Pr=0.012)	Melanin metabolic process (GO:0006582)	-
AP1M1	M	-	Melanosome organization (GO:0032438)	Melanosome organization (Alves et al. 2016)
AP3B2	Possibly	-	-	*Ap3b1* is in the ESPCR database (melanosome construction)
ATP6AP1	M	-	*ATP6AP1B*: melanosome organization (GO:0032438)	Melanosome/melanophore related (Amsterdam et al. 1999; Nuckels et al. 2009); *ATP6V1B2* & *ATP6V1C1* in (Poelstra et al. 2015); *ATP7A* and *ATP7B* are in the ESPCR database
B3GNT2	M*	H: Melanin metabolic process GO:0006582 (Pr=0.012)	-	-
BARX2	P*	ZF: cellular pigmentation GO:0033059 (Pr=0.012)	-	-
CALM	M	-	-	Melanophore related (Poelstra et al. 2015)

CCM2L	M/X*	ZF: Melanocyte apoptotic process GO:1902362 (Pr=0.024); xanthophore differentiation GO:0050936 (Pr=0.020)	-	-
CIRBP	Possibly	H: Keratinocyte differentiation GO:0030216 (Pr=0.011); negative regulation of keratinocyte differentiation GO:0045617 (Pr=0.011)	-	-
COMT	z	ZF: melanin metabolic process GO:0006582 (Pr=0.028); regulation of melanocyte differentiation GO:0045634 (Pr=0.017); pigment granule organization GO:0048753 (Pr=0.013)	-	-
CRYAB	P*	ZF: Pigmentation GO:0043473 (Pr=0.010)	-	-
CYP1A5/ *CYP1A4*	M*	ZF: Melanocyte migration GO:0097324 (Pr=0.033); pigmentation GO:0043473 (Pr=0.160)	-	-
CYR61	Possibly	H: Positive regulation of keratinocyte migration GO:0051549 (Pr=0.027); regulation of keratinocyte migration GO:0051547 (Pr=0.026)	-	-
DCT/*TYRP2*	M	-	Extensive: melanosome related.	Used in melanin biosynthesis in early melanoblasts and melanocyte stem cells (Curran et al. 2010; Ng'oma et al. 2014; Poelstra et al. 2015); In the ESPCR database: dilution of eumelanin colour (DOPAchrome tautomerase, melanosomal enzyme)
DEFB family	M*	ZF: melanin metabolic process GO:0006582 (Pr=0.017)	-	Could modulate melanocortin receptor signalling (Candille et al. 2007)
DIXDC1	M*	H: Melanin metabolic process GO:0006582 (Pr=0.011); ZF: pigmentation GO:0043473 (Pr=0.011); melanin metabolic process GO:0006582 (Pr=0.011)	-	-
DST	M	-	-	Melanophore related (Poelstra et al. 2015); In the ESPCR database: Knockout causes pale skin mutant.
DYRK2	M*	ZF: Melanin metabolic process GO:0006582 (Pr=0.014)	-	Melanin related
ELMOD1	M*	H: Melanin metabolic process GO:0006582 (Pr=0.012)	-	-
ELOVL7	M*	H: Melanin metabolic process GO:0006582 (Pr=0.011)	-	*ELOVL4* in (Poelstra et al. 2015)

ENTPD8	M*	ZF: Melanin metabolic process GO:0006582 (Pr=0.012)	-	-
FAM83A	M*	H: Melanin metabolic process GO:0006582 (Pr=0.011)	-	-
FAR1	P*	ZF: Pigmentation GO:0043473 (Pr=0.042); pigment cell differentiation GO:0050931 (Pr=0.017); developmental pigmentation GO:0048066 (Pr=0.016); cellular pigmentation GO:0033059 (Pr=0.013)	-	-
FOXA3	P*	ZF: Pigmentation GO:0043473 (Pr=0.539); pigment cell differentiation GO:0050931 (Pr=0.365); developmental pigmentation GO:0048066 (Pr=0.264); cellular pigmentation GO:0033059 (Pr=0.046)	-	*Foxd3* and *Foxn1* are in the ESPCR database: developmental functions
GALC	P*	ZF: Pigment cell differentiation GO:0050931 (Pr=0.015); developmental pigmentation GO:0048066 (Pr=0.010)	-	-
GSTT1	M*	ZF: Melanin metabolic process GO:0006582 (Pr=0.026)	-	-
HS3ST3B1	P*	ZF: Pigmentation GO:0043473 (Pr=0.021)	-	*HS2ST1* in (Poelstra et al. 2015); similar enzyme
IFT172	P*	ZF: Cellular pigmentation GO:0033059 (Pr=0.089); establishment of melanosome localization GO:0032401 (Pr=0.015)	-	-
KDSR	P*	ZF: Pigmentation GO:0043473 (Pr=0.014)	-	-
LMX1B	M	ZF: Pigment cell differentiation GO:0050931 (Pr=0.173)	-	Melanophore related (Burbach et al. 2000); *LMX1A* is in the ESPCR database: Partial or complete white belt and/or belly spot
LRP5	M*	H: Melanin metabolic process GO:0006582 (Pr=0.016)	-	*Ndp* is in the ESPCR database, and *LRP5* is associated (hyperpigmentation)
MAPK12	M*	-		Melanocyte related (Watahiki et al. 2004)
MAPRE2	M*	H: Melanin metabolic process GO:0006582 (Pr=0.091)	-	-
MICAL2	P*	ZF: Pigmentation GO:0043473 (Pr=0.051); developmental pigmentation GO:0048066 (Pr=0.024)	-	-
MLANA	M	-	Melanosome (GO:0042470)	Melanosome biogenesis (Poelstra et al. 2015)

MMP1	Possibly	H: Positive regulation of keratinocyte migration GO:0051549 (Pr=0.041); regulation of keratinocyte migration GO:0051547 (Pr=0.079)	-	-
MMP3	Possibly	H: Regulation of keratinocyte migration GO:0051547 (Pr=0.489); positive regulation of keratinocyte migration GO:0051549 (Pr=0.096); keratinocyte migration GO:0051546 (Pr=0.071)	-	-
MTHFR	Possibly	-	Flavin adenine dinucleotide binding (GO:0050660): flavins are xanthophore pigment related.	-
MYO1E	Possibly	-	-	No, but *Myo5a* and *Myo7a* are in the ESPCR database (melanosome transport)
NIPAL4	M*	ZF: Melanosome transport GO:0032402 (Pr=0.011); pigmentation GO:0043473 (Pr=0.014); cellular pigmentation GO:0033059 (Pr=0.014)	-	-
NME1	Possibly	H: Regulation of keratinocyte differentiation GO:0045616 (Pr=0.079); keratinocyte differentiation GO:0030216 (Pr=0.014); positive regulation of keratinocyte differentiation GO:0045618 (Pr=0.010)	-	-
NR4A2	M	-	-	Melanophore related (Poelstra et al. 2015)
NUAK2	P*	ZF: Pigmentation GO:0043473 (Pr=0.011)	-	-
OCA2	M/I	(Not run)	Extensive: mainly melanosome related, but also iridophore differentiation (GO:0050935)	Melanin precursor (Ng'oma et al. 2014; Poelstra et al. 2015); In the ESPCR database: Melanosome biogenesis and size.
P2RX1	Possibly	-	-	May be involved in melanin aggregation (Kumazawa and Fujii 1984)
PAX7	X	-	Regulation of xanthophore differentiation (GO:0050938); developmental pigmentation (GO:0048066)	A precursor to melanophores, xanthophores, and iridophores. In cichlids and zebrafish, higher *Pax3/7* dosage correlates with development of fewer, larger melanophores. Orange blotched cichlids are also associated to higher *pax7* expression Xanthophores (Roberts et al. 2007; Nelms and Labosky 2010; Maan and Sefc 2013); In the ESPCR database: Retinal Pigment Epithelium related

PCSK1	M	-	-	Involved in melanophore-stimulating hormone synthesis, e.g. alpha MSH (Chrétien and Mbikay 2016)
PDE1B	P	-	-	Pigment concentration/aggregation (Oshima et al. 1998; Reilein et al. 1998; Milograna et al. 2016)
PLIN2	P*	ZF: pigmentation GO:0043473 (Pr=0.011)	-	-
PMEL	M	-	Developmental pigmentation (GO:0048066); melanosome (GO:0042470); melanosome organization(GO:0032438)	Melanosomes biogenesis (Poelstra et al. 2015); In the ESPCR database: Retinal Pigment Epithelium related Silvering with postnatal melanocyte loss in eumelanic animals (varying with strain background) in *Pmel*. Mild effect on visible pigmentation, substantial reduction in eumelanin content in hair and spherical melanosomes in *Pmel*
PNP	I	-	Multiple purine related GO annotations.	Early expressing enzyme in iridophore differentiation (Curran et al. 2010)
PPP1R1B	A	-	-	Involved in the cAMP/PKA pathway, which is involved in the erythrophore pigment movement (Mantione et al. 2012)
PTPN3	M	ZF: Melanin metabolic process GO:0006582 (Pr=0.010)	-	Melanin related
RASEF	Possibly	-	-	Unknown function, but found in two papers looking at colour related gene expression in fish (Henning et al. 2013; Ng'oma et al. 2014)
RCAN1	Possibly	H: Regulation of keratinocyte migration GO:0051547 (Pr=0.020)	-	-
RGS16	Possibly	H: Regulation of keratinocyte migration GO:0051547 (Pr=0.021)	-	Melanin related (Poelstra et al. 2015)
ROGDI	M*	ZF: Establishment of melanosome localization GO:0032401 (Pr=0.047); melanosome transport GO:0032402 (Pr=0.027); cellular pigmentation GO:0033059 (Pr=0.019); pigment granule localization GO:0051875 (Pr=0.015); pigment cell differentiation GO:0050931 (Pr=0.010)	-	-
RUNDC3B	M*	H: Melanin metabolic process GO:0006582 (Pr=0.010)	-	-

SEL1L3	M*	H: Melanin metabolic process GO:0006582 (Pr=0.011)	-	-
SEMA4F	Possibly	-	-	*SEMA3C*/*SEMA4C*/*SEMA6A* are in (Poelstra et al. 2015); *SEMA3C* (causes some skin hypopigmentation and related to ectopic pigment in internal organs) and *SEMA4A* (Abnormal Retinal Pigment Epithelium, postnatal depigmentation of eye), both are in the ESPCR database
SHC4	M	-	-	Involved in the melanogenesis pathway (Shin and Lee 2013)
SLC2A11	X/M	ZF: Melanin metabolic process GO:0006582 (Pr=0.023); melanin biosynthetic process GO:0042438 (Pr=0.019)	-	Involved in xanthophore differentiation (Kimura et al. 2014)
SLC2A9	L/X*		Urate metabolic process (GO:0046415)	Uric acid related, the primary purine in leucophores (Kimura et al. 2014)
SLC35F6	Possibly	-	-	*SCL* genes are in the ESPCR database (melanosome related)
SLC46A2	Possibly	-	-	*SCL* genes are in the ESPCR database (melanosome related)
SLC7A14	Possibly	-	-	*SLC7A5* and *SLC7A11* is (Poelstra et al. 2015); SCL genes are in the ESPCR database (melanosome related)
SNAI1	Possibly	H: Regulation of keratinocyte migration GO:0051547 (Pr=0.042)	-	SNAI2 is in (Poelstra et al. 2015) and the ESPCR database: developmentally related. Knockout causes spotting, head blaze, and pale hair and skin
SRD5A3	M*	H: Melanin metabolic process GO:0006582 (Pr=0.010)	-	-
TBX3	Possibly	-	-	*TBX10*/*2*/*15* is in (Poelstra et al. 2015); *TBX10*/*15* are in the ESPCR database
TFEC	M	-	-	*MIFT* (Melanogenesis Associated Transcription Factor) parlog; Collaborates with *MITF* in target gene activation by similarity (Acton 2013; Spencer 2015)
TGFBI	M*	ZF: Pigmentation GO:0043473 (Pr=0.028)	-	-
TMEM98	Possibly	-	-	*TMEM51* is in (Poelstra et al. 2015)

TNC	M*	ZF: developmental pigmentation GO:0048066 (Pr=0.028)	-	-
TPBG	Possibly	H: Positive regulation of keratinocyte migration GO:0051549 (Pr=0.013)	-	-
TRAPPC3L	Possibly	-	-	*TRAPPC6A* is in the ESPCR database (knockout causes pale patches)
TRIM25	Possibly	-	-	*TRIM2* is in (Poelstra et al. 2015); see (Sardiello et al. 2008) for duscussion on *TRIM* family
TRIM32	M	-	Melanosome transport (GO:0032402)	*TRIM2* is in (Poelstra et al. 2015); see (Sardiello et al. 2008) for duscussion on *TRIM* family
TRIM39	Possibly	-	-	*TRIM2* is in (Poelstra et al. 2015); see (Sardiello et al. 2008) for duscussion on *TRIM* family
TSPAN1	M*	H: Melanin metabolic process GO:0006582 (Pr=0.011)	-	-
TYR	M/I	(Not run)	Extensive: mainly melanosome related, but also iridophore differentiation (GO:0050935)	Melanin production (Ng'oma et al. 2014; Poelstra et al. 2015); In the ESPCR database: Melanogenic enzyme; no pigment in null mice (multiple allelic variants)
TYRP1	M	-	Extensive: melanosome related.	Melanin synthesis/melanosome organisation (Ng'oma et al. 2014; Alves et al. 2016); In the ESPCR database: melanosomal enzyme/stabilizing factor. Involved in the production of eumelanin.
UPK1B	Possibly	H: Keratinocyte differentiation GO:0030216 (Pr=0.014)	-	-
VEGFA	M*	ZF: Melanocyte migration GO:0097324 (Pr=0.028)	-	-

Supplementary Table A5.9: PANTHER Overrepresentation Test results for lists of sDE genes upregulated in different colours of skin. Associated reference indicates the reference list used to statistically determine over- or under-representation on ontology categories: H = *Homo sapiens*; A = *Anolis carolinensis*; D = *Danio rerio*; M = *Mus musculus*; X = *Xenopus tropicalis*.

Comparison	Annotation Data Set	Skin colour	Associated GO annotation	Lower Hierarchy GO annotations	Associated reference
Yellow *vs*. Black	GO biological process complete	Yellow	inosine metabolic process (GO:0046102)	na	*H*
			urate metabolic process (GO:0046415)	na	*A*
		Black	developmental pigmentation (GO:0048066)	na	*H/M/D/X*
				pigmentation (GO:0043473)	*H/M/D*
			iridophore differentiation (GO:0050935)	na	*D*
			melanin biosynthetic process from tyrosine (GO:0006583)	melanin biosynthetic process (GO:0042438)	*H/M/D/X*
				na	*H/M*
				melanin metabolic process (GO:0006582)	*H/M/D/X*
				organic hydroxy compound biosynthetic process (GO:1901617)	*H/M/D/X*
				organic hydroxy compound metabolic process (GO:1901615)	*H/M/D*
				phenol-containing compound biosynthetic process (GO:0046189)	*H/M/D/A/X*
				phenol-containing compound metabolic process (GO:0018958)	*H/M/D/X*
				pigment biosynthetic process (GO:0046148)	*H/M/D*
				pigment metabolic process (GO:0042440)	*H/M/D*
				secondary metabolic process (GO:0019748)	*H/M/D*
				secondary metabolite biosynthetic process (GO:0044550)	*H/M/D*

Yellow vs. Black (cont.)	GO molecular function complete	Yellow	-	-	-
		Black	oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen, another compound as one donor, and incorporation of one atom of oxygen (GO:0016716)	na	*H/M/A*
			oxidoreductase activity (GO:0016491)	na	*X*
	GO cellular component complete	Yellow	-	-	-
		Black	melanosome membrane (GO:0033162)	cytoplasmic vesicle (GO:0031410)	*M/D*
				intracellular vesicle (GO:0097708)	*M/D*
				chitosome (GO:0045009)	*H/M/D*
				melanosome (GO:0042470)	*H/M/D/X*
				na	*H/M/D*
				pigment granule (GO:0048770)	*H/M/D/X*
				pigment granule membrane (GO:0090741)	*H/M*
Yellow vs. Brown	GO biological process complete	Yellow	mitochondrial respiratory chain complex III assembly (GO:0034551)	na	*D/X*
				respiratory chain complex III assembly (GO:0017062)	*D/X*
				mitochondrial respiratory chain complex III biogenesis (GO:0097033)	*D*
				mitochondrial respiratory chain complex assembly (GO:0033108)	*D*
		Brown	melanin biosynthetic process from tyrosine (GO:0006583)	na	*H*
				melanin biosynthetic process (GO:0042438)	*H/M*
				melanin metabolic process (GO:0006582)	*H/M*
				secondary metabolite biosynthetic process (GO:0044550)	*H/M*
				phenol-containing compound metabolic process (GO:0018958)	*H/M*
	GO molecular function complete	Yellow	-	-	-
		Brown	-	-	-

Yellow vs. Brown (cont.)	GO cellular component complete	Yellow	-	-	-
		Brown	melanosome (GO:0042470)	na	*H/M/A*
				pigment granule (GO:0048770)	*H/M/A*
				cytoplasm (GO:0005737)	*M/A*
Brown vs. Black	GO biological process complete	Brown	single-organism metabolic process (GO:0044710)	na	*H*
		Black	cellular response to jasmonic acid stimulus (GO:0071395)	na	*H*
				response to jasmonic acid (GO:0009753)	*H*
			doxorubicin metabolic process (GO:0044598)	na	*H*
				polyketide metabolic process (GO:0030638)	*H*
			daunorubicin metabolic process (GO:0044597)	na	*H*
				aminoglycoside antibiotic metabolic process (GO:0030647)	*H*
			hormone metabolic process (GO:0042445)	na	*H*
				regulation of hormone levels (GO:0010817)	*H*
			steroid metabolic process (GO:0008202)	na	*H*
	GO molecular function complete	Brown	-	-	-
		Black	ketosteroid monooxygenase activity (GO:0047086)	na	*H*
			phenanthrene 9,10-monooxygenase activity (GO:0018636)	na	*H*
			trans-1,2-dihydrobenzene-1,2-diol dehydrogenase activity (GO:0047115)	na	*H*
			alditol:NADP+ 1-oxidoreductase activity (GO:0004032)	na	*H*
				alcohol dehydrogenase (NADP+) activity (GO:0008106)	*H*
			oxidoreductase activity, acting on the CH-OH group of donors, NAD or NADP as acceptor (GO:0016616)	na	*H*
			extracellular matrix binding (GO:0050840)	na	*M*
	GO cellular component complete	Brown	-	-	-
		Black	cytoplasmic part (GO:0044444)	-	*M*

Striped Individuals only: Yellow *vs.* Black	GO molecular function complete	Black	oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen, another compound as one donor, and incorporation of one atom of oxygen (GO:0016716)	-	*H/M/A*
		Yellow	extracellular matrix structural constituent (GO:0005201)	-	*M*
	GO biological process complete	Black	pigmentation (GO:0043473)	-	*A*
		Black	organic hydroxy compound biosynthetic process (GO:1901617)	-	*D*
			melanin biosynthetic process (GO:0042438)	pigment biosynthetic process (GO:0046148)	*H/M*
				pigment metabolic process (GO:0042440)	*H/M*
				secondary metabolic process (GO:0019748)	*H/M/A*
				-	*H/M/A/X*
				melanin metabolic process (GO:0006582)	*H/M/A/X*
				organic hydroxy compound biosynthetic process (GO:1901617)	*H/M/A/X*
				secondary metabolite biosynthetic process (GO:0044550)	*H/M/A/X*
				phenol-containing compound biosynthetic process (GO:0046189)	*H/M/D/A/X*
				organic hydroxy compound metabolic process (GO:1901615)	*X*
				phenol-containing compound metabolic process (GO:0018958)	*X*
			developmental pigmentation (GO:0048066)	-	*H*
		Yellow	collagen fibril organization (GO:0030199)	-	*H*
			cartilage development (GO:0051216)	na	*H/M*
			cartilage development (GO:0051216)	connective tissue development (GO:0061448)	*M*
			urate metabolic process (GO:0046415)	-	*A*
	GO cellular component complete	Black	melanosome membrane (GO:0033162)	chitosome (GO:0045009)	*H/A*
				cytoplasm (GO:0005737)	*M*
				cytoplasmic part (GO:0044444)	*M*

Striped Individuals only: Yellow vs. Black (cont.)	*GO cellular component complete (cont.)*	*Black (cont.)*	*melanosome membrane (GO:0033162)*	cytoplasmic vesicle (GO:0031410)	*M*
				intracellular vesicle (GO:0097708)	*M*
				melanosome (GO:0042470)	*H/M/A*
				-	*H/A*
				pigment granule (GO:0048770)	*H/M/A*
				pigment granule membrane (GO:0090741)	*H/A*
		Yellow	basement membrane (GO:0005604)	-	*H/M*
			basement membrane (GO:0005604)	proteinaceous extracellular matrix (GO:0005578)	*H/M/A/X*
			basement membrane (GO:0005604)	extracellular matrix (GO:0031012)	*H/M/A/X*
			basement membrane (GO:0005604)	extracellular region (GO:0005576)	*X*

Supplementary Table A5.10: List of Sal-Site aligned contigs (E-value < 0.1) for each locus identified by Random Forest and LFMM analyses, with putative IDs based on subsequent NCBI Blastn and Blastx searches.

Salamandra Locus ID	Sal-Site assembly V4.0 (blastn) contigs	Score (bits)	E-value	NCBI Blastx	NCBI Blastn
1731	C2441847	36	0.51	*prel*	-
	C1617008	36	0.51	-	-
	C0848330	36	0.51	-	-
	C0825510	36	0.51	-	-
	C0440014	36	0.51	-	-
2162	C0337977	38	0.13	-	-
	C2580440	36	0.51	-	*TYR*
	C2568590	36	0.51	-	*TYR*
	C2311585	36	0.51	-	*TYR*
	C1977621	36	0.51	-	*TYR*
	C1425995	36	0.51	-	*TYR*
	C0973835	36	0.51	-	-
	C0842872	36	0.51	-	*TYR*
	C0702554	36	0.51	-	*TYR*
	C0558032	36	0.51	-	-
	C0339823	36	0.51	-	-
	C0011074	36	0.51	-	*TYR*
	C0010387	36	0.51	-	*TYR*
2594	C1049484	40	0.033	-	-
	C1010361	38	0.13	-	-
	C0260081	38	0.13	-	-
4074	C0581618	40	0.033	-	-
	C0463091	36	0.51	-	-
	C0216261	36	0.51	-	-
7391	C0670500	36	0.51	-	-
	C0323270	36	0.51	-	-
8055	C2316492	38	0.13	-	-
	C0607966	38	0.13	-	-
	C0534871	38	0.13	-	-
	C0048910	38	0.13	-	-
	C0961942	36	0.51	-	-
9006	C1296724	36	0.51	-	-
10562	C1154183	38	0.13	-	-
	C1190132	36	0.51	-	-
11822	C0562604	44	0.002	-	-
	C1634650	38	0.13	-	-
12135	C1543346	38	0.13	-	-
	C1482148	38	0.13	-	-
	C0315098	38	0.13	-	-

15595	C2106582	38	0.13	-	-
	C1594431	38	0.13	-	-
	C1362112	38	0.13	-	-
	C0166862	38	0.13	-	-
	C1054905	36	0.51	-	-
	C0448541	36	0.51	-	-
	C0320974	36	0.51	-	-
16124	C2388088	36	0.51	-	-
	C1857851	36	0.51	-	-
	C0908141	36	0.51	-	-
	C0871863	36	0.51	-	-
	C0176608	36	0.51	-	-
16739	C2433923	36	0.51	-	-
	C2267950	36	0.51	-	-
	C1011339	36	0.51	-	-
	C0188756	36	0.51	-	*LDLRAD4*
19416	C1919670	42	0.008	-	-
	C1648278	38	0.13	-	-
	C2320111	36	0.51	-	-
21423	C0698022	38	0.13	-	-
	C0899319	36	0.51	-	-
22038	C2186795	36	0.51	-	-
23809	C0351058	36	0.51	-	-
24580	C0512343	42	0.008	-	-
	C1676158	38	0.13	-	-
	C0462920	38	0.13	-	-
24694	C0800137	36	0.51	-	-
27346	C2023392	36	0.51	-	-
27356	C2346428	40	0.033	-	-
	C1209840	38	0.13	-	-
28130	C1719848	36	0.51	*CAMK1*	-
	C0792708	36	0.51	-	-
	C0767549	36	0.51	*A4R28_09265* (weak)	-
	C0106941	36	0.51	-	-
32624	C1883732	38	0.13	-	-
32743	C0589497	36	0.51	-	-
35922	C1985035	40	0.033	-	-
	C0050379	38	0.13	-	-
	C2476219	36	0.51	-	-
	C1273142	36	0.51	-	-
	C0548768	36	0.51	-	-
36542	C1140091	38	0.13	-	-
	C0435285	38	0.13	-	-
	C0418185	36	0.51	-	-
	C0330467	36	0.51	*BAZ2A*	*BAZ2A*
	C0294476	36	0.51	-	-

37375	C2186292	36	0.51		
	C0210156	36	0.51		
37821	C0395557	40	0.033	*RGR*	-
40024	C2191566	36	0.51	-	-
	C0690983	36	0.51	-	-
41242	C0800716	40	0.033	-	-
	C2522041	36	0.51	-	-
	C1436294	36	0.51	-	-
	C0372305	36	0.51	-	-
41341	C2259173	36	0.51	-	-
	C0497205	36	0.51	-	-
42961	C2241015	38	0.13	-	-
	C1708121	38	0.13	-	-
	C0212876	38	0.13	-	-
43841	C2160922	38	0.13	-	-
	C2277369	36	0.51	-	-
	C0890620	36	0.51	-	-
47160	C0445727	38	0.13	-	-
	C0328235	36	0.51	-	-
47493	C2469665	36	0.51	-	-
	C1857949	36	0.51	-	-
52344	C1491194	36	0.51	-	-
	C0859623	36	0.51	-	-
	C0465191	36	0.51	-	-
54397	-	-	-	-	-
58978	C2100133	42	0.008	*SURF4*	-
	C0999918	38	0.13	-	-
	C0632768	38	0.13	-	-
	C0258488	38	0.13	-	-
59960	-	-	-	-	-
61782	-	-	-	-	-
62160	C0339929	38	0.13	*NYNRIN*	-
	C1676666	36	0.51	*NYNRIN*	-
	C1030385	36	0.51	-	-
	C0763884	36	0.51	*NYNRIN*	-
	C0522727	36	0.51	*NYNRIN*	-
	C0402205	36	0.51	-	-
	C0336896	36	0.51	-	-
	C0087584	36	0.51	-	-
	C0060139	36	0.51	-	-
62524	C2566365	36	0.51	-	-
	C1607486	36	0.51	-	-
84159	C2026845	38	0.13	-	-
	C0321127	36	0.51	-	-

Supplementary Table A5.11: List of loci identified by Random Forest and LFMM analyses, with putative gene IDs. Asterisks (*) mark those loci found in both the 7 and 21 co-varying loci sets found during the discrete-phenotype Random Forest analysis. (Alignment to a reference *Salamandra* transcriptome and the NCBI nucleotide and protein databases returned no hits so have been excluded from the table).

Analysis	Phenotype Input	Locus ID	In other Datasets?	Sal-Site assembly V4.0 (Flowed by NCBI blast)
Random Forest	*Patternize PC1*	2162	RF-Discrete	*TYR*
Random Forest	*Patternize PC1*	2594	LFMM-Patternize	-
Random Forest	*Patternize PC1*	4074	-	-
Random Forest	*Patternize PC1*	9006	-	-
Random Forest	*Patternize PC1*	19416	-	-
Random Forest	*Patternize PC1*	23809	LFMM-Patternize	-
Random Forest	*Patternize PC1*	24580	-	-
Random Forest	*Patternize PC1*	41242	-	-
Random Forest	*Patternize PC1*	47493	-	-
LFMM	*Patternize PC1*	2594	RF-Patternize	-
LFMM	*Patternize PC1*	8055	-	-
LFMM	*Patternize PC1*	23809	RF-Patternize	-
LFMM	*Patternize PC1*	27346	-	-
LFMM	*Patternize PC1*	35922	-	-
LFMM	*Patternize PC1*	40024	-	-
LFMM	*Patternize PC1*	54397	-	-
LFMM	*Patternize PC1*	62524	-	-
Random Forest	Categorical	2162	RF-Patternize	*TYR*
Random Forest	Categorical	7391	-	-
Random Forest	Categorical	10562	-	-
Random Forest	Categorical	11822	-	-
Random Forest	Categorical	12135	-	-
Random Forest	Categorical	21423	-	-
Random Forest	Categorical	22038	-	-
Random Forest	Categorical	27356	-	-
Random Forest	Categorical	32624	-	-
Random Forest	Categorical	32743	-	-
Random Forest	Categorical	36542	-	*BAZ2A*
Random Forest	Categorical	47160	-	-
Random Forest	Categorical	52344	-	-
Random Forest	Categorical	58978	-	*SURF4*
Random Forest	Categorical	16739*	-	*LDLRAD4*
Random Forest	Categorical	28130*	LFMM- Discrete	*CAMK1*
Random Forest	Categorical	37821*	-	*RGR*
Random Forest	Categorical	43841*	-	-
Random Forest	Categorical	59960*	-	-
Random Forest	Categorical	62160*	-	*NYNRIN*

Random Forest	Categorical	84159*	-	-
LFMM	Categorical	1731	-	*prel*
LFMM	Categorical	15595	-	-
LFMM	Categorical	16124	-	-
LFMM	Categorical	24694	-	-
LFMM	Categorical	28130	RF- Discrete	*CAMK1*
LFMM	Categorical	37375	-	-
LFMM	Categorical	41341	-	-
LFMM	Categorical	42961	-	-
LFMM	Categorical	61782	-	-

Supplementary Table A5.12: Genomic loci showing a Z*Fst* ≥3 standard deviations from the mean between hypolutic (brown) and xanthic (yellow) salamanders in the rio Color/rio Tendi contact zone. Also shown are those loci found in random forest (RF) and LFMM genotype-phenotype association analyses (with putative gene ID). Loci are ordered by Z*Fst*. (Pi = genetic diversity.)

Locus ID	*Fst*	pi (Hypolutic)	pi (Xanthic)	Δpi	Z*Fst*	Genotype-Phenotype Association	Gene
8055	0.320	0.271	0.505	-0.234	7.936	LFMM: Patternize	-
49998	0.297	0.315	0.503	-0.189	7.354	-	-
55792	0.242	0.492	0.368	0.124	5.882	-	-
56929	0.239	0.323	0.508	-0.185	5.809	-	-
63029	0.236	0.000	0.368	-0.368	5.727	-	-
44551	0.230	0.059	0.443	-0.384	5.581	-	-
4116	0.226	0.471	0.083	0.387	5.471	-	-
9006	0.212	0.000	0.337	-0.337	5.114	RF: Patternize	-
28130	0.211	0.508	0.198	0.310	5.065	RF/LFMM: Categorical	*CAMK1*
45537	0.206	0.405	0.000	0.405	4.956	-	-
43187	0.205	0.287	0.519	-0.231	4.919	-	-
20713	0.200	0.370	0.000	0.370	4.794	-	-
7342	0.200	0.515	0.323	0.192	4.777	-	-
25236	0.197	0.405	0.000	0.405	4.709	-	-
10697	0.197	0.395	0.489	-0.094	4.698	-	-
27346	0.197	0.349	0.525	-0.176	4.697	LFMM: Patternize	-
25876	0.195	0.443	0.077	0.366	4.653	-	-
51208	0.186	0.369	0.000	0.369	4.407	-	-
2594	0.183	0.476	0.434	0.042	4.329	RF/LFMM: Patternize	-
40024	0.175	0.508	0.239	0.269	4.134	LFMM: Patternize	-
38596	0.173	0.077	0.416	-0.339	4.080	-	-
47058	0.172	0.452	0.091	0.361	4.062	-	-
49730	0.172	0.331	0.000	0.331	4.055	-	-
66003	0.167	0.000	0.312	-0.312	3.904	-	-
42589	0.166	0.083	0.431	-0.348	3.904	-	-
9164	0.165	0.476	0.443	0.033	3.865	-	-
22261	0.157	0.323	0.524	-0.201	3.660	-	-
26593	0.156	0.369	0.000	0.369	3.626	-	-
32809	0.154	0.299	0.515	-0.215	3.588	-	-
15079	0.154	0.114	0.434	-0.320	3.579	-	-
8673	0.154	0.000	0.304	-0.304	3.571	-	-
46372	0.154	0.304	0.000	0.304	3.571	-	-
36327	0.152	0.258	0.508	-0.249	3.526	-	-
31773	0.151	0.129	0.443	-0.314	3.495	-	-
3676	0.151	0.228	0.505	-0.277	3.487	-	-

56077	0.151	0.000	0.290	-0.290	3.484	-	-
56257	0.150	0.186	0.485	-0.299	3.476	-	-
4760	0.147	0.299	0.516	-0.216	3.382	-	-
54397	0.146	0.480	0.455	0.026	3.370	LFMM: Patternize	-
46403	0.145	0.000	0.304	-0.304	3.347	-	-
13699	0.144	0.344	0.000	0.344	3.310	-	-
64994	0.144	0.344	0.000	0.344	3.310	-	-
47466	0.143	0.000	0.268	-0.268	3.288	-	-
46344	0.143	0.212	0.492	-0.280	3.277	-	-
54344	0.142	0.443	0.100	0.343	3.271	-	-
32134	0.139	0.287	0.521	-0.234	3.180	-	-
57248	0.139	0.138	0.443	-0.306	3.173	-	-
45237	0.138	0.389	0.071	0.317	3.166	-	-
35922	0.138	0.508	0.271	0.237	3.145	LFMM: Patternize	-
14438	0.137	0.520	0.325	0.195	3.124	-	-
61782	0.136	0.198	0.476	-0.278	3.091	LFMM: Categorical	-
9226	0.135	0.294	0.000	0.294	3.084	-	-
62524	0.135	0.508	0.416	0.092	3.072	LFMM: Patternize	-
40716	0.134	0.077	0.395	-0.318	3.038	-	-
27297	0.132	0.518	0.389	0.129	3.003	-	-

Supplementary Table A5.13: Genomic loci showing a Z*Fst* ≥3 standard deviations from the mean between hypolutic (brown) and striped salamanders in the rio Color/rio Tendi contact zone. Also shown are those loci found in random forest (RF) and LFMM genotype-phenotype association analyses (with putative gene ID). Loci are ordered by Z*Fst*. (Pi = genetic diversity).

Locus ID	*Fst*	pi (Hypolutic)	pi (Striped)	Δpi	Z*Fst*	Genotype-Phenotype Association	Gene
28130	0.331	0.508	0.071	0.437	8.141	RF/LFMM: Categorical	*CAMK1*
10697	0.276	0.395	0.455	-0.060	6.715	-	-
63206	0.229	0.517	0.290	0.227	5.487	-	-
2162	0.226	0.083	0.471	-0.387	5.401	RF: Patternize/ Categorical	*TYR*
57248	0.226	0.138	0.495	-0.357	5.400	-	-
7040	0.214	0.508	0.173	0.335	5.086	-	-
57073	0.210	0.516	0.312	0.204	4.974	-	-
26560	0.206	0.409	0.000	0.409	4.868	-	-
39926	0.203	0.271	0.524	-0.253	4.793	-	-
5429	0.202	0.516	0.323	0.193	4.772	-	-
16739	0.200	0.228	0.518	-0.290	4.724	RF: Categorical	*LDLRAD4*
32462	0.200	0.000	0.368	-0.368	4.724	-	-
36099	0.196	0.159	0.489	-0.330	4.624	-	-
51379	0.194	0.389	0.000	0.389	4.563	-	-
22572	0.193	0.370	0.000	0.370	4.540	-	-
30557	0.192	0.091	0.471	-0.380	4.509	-	-
19196	0.191	0.371	0.000	0.371	4.500	-	-
45788	0.191	0.000	0.337	-0.337	4.489	-	-
35159	0.188	0.071	0.443	-0.372	4.407	-	-
38732	0.187	0.492	0.159	0.333	4.396	-	-
38469	0.187	0.000	0.395	-0.395	4.379	-	-
37821	0.181	0.499	0.198	0.301	4.237	RF: Categorical	*RGR*
44956	0.181	0.226	0.508	-0.282	4.236	-	-
5856	0.181	0.370	0.000	0.370	4.232	-	-
12135	0.181	0.000	0.323	-0.323	4.220	RF: Categorical	-
15681	0.180	0.508	0.368	0.140	4.204	-	-
38866	0.177	0.000	0.337	-0.337	4.114	-	-
29135	0.176	0.369	0.523	-0.154	4.106	-	-
58978	0.174	0.166	0.476	-0.310	4.052	RF: Categorical	*SURF4*
30137	0.173	0.000	0.344	-0.344	4.013	-	-
47493	0.172	0.100	0.464	-0.364	3.997	RF: Patternize	-
37375	0.172	0.000	0.368	-0.368	3.986	LFMM: Categorical	-
58325	0.170	0.498	0.198	0.300	3.945	-	-
64049	0.169	0.417	0.071	0.346	3.918	-	-

9164	0.165	0.476	0.443	0.033	3.812	-	-
51455	0.165	0.515	0.349	0.166	3.806	-	-
16124	0.162	0.071	0.409	-0.338	3.742	LFMM: Catagorical	-
45161	0.160	0.495	0.189	0.305	3.687	-	-
39263	0.155	0.121	0.431	-0.310	3.554	-	-
79226	0.155	0.466	0.148	0.318	3.548	-	-
37037	0.154	0.000	0.344	-0.344	3.528	-	-
51458	0.151	0.523	0.391	0.132	3.444	-	-
42541	0.149	0.521	0.366	0.155	3.391	-	-
57821	0.146	0.083	0.416	-0.332	3.327	-	-
11240	0.146	0.059	0.369	-0.310	3.315	-	-
23700	0.145	0.331	0.000	0.331	3.296	-	-
38696	0.145	0.331	0.000	0.331	3.296	-	-
32134	0.145	0.287	0.518	-0.231	3.292	-	-
41880	0.143	0.000	0.271	-0.271	3.244	-	-
47466	0.143	0.000	0.268	-0.268	3.242	-	-
12592	0.142	0.331	0.522	-0.191	3.216	-	-
6232	0.142	0.254	0.507	-0.253	3.214	-	-
8314	0.141	0.304	0.000	0.304	3.179	-	-
45132	0.140	0.464	0.159	0.304	3.157	-	-
61782	0.136	0.198	0.476	-0.278	3.047	LFMM: Categorical	-

Supplementary Table A5.14: Genomic loci showing a Z*Fst* ≥3 standard deviations from the mean between xanthic (yellow) and striped salamanders in the rio Color/rio Tendi contact zone. Also shown are those loci found in random forest (RF) and LFMM genotype-phenotype association analyses (with putative gene ID). Loci are ordered by Z*Fst*. (Pi = genetic diversity).

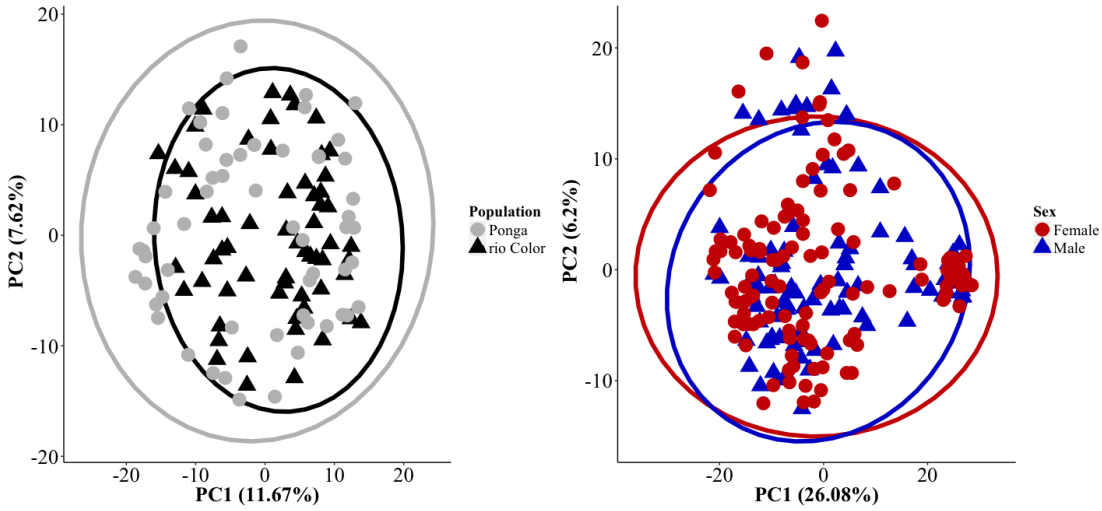
Locus ID	*Fst*	pi (Xanthic)	pi (Striped)	Δpi	Z*Fst*	Genotype-Phenotype Association	Gene
29135	0.452	0.100	0.523	-0.423	9.639	-	-
45132	0.385	0.521	0.159	0.362	8.152	-	-
53576	0.334	0.247	0.506	-0.260	6.994	-	-
30975	0.328	0.464	0.000	0.464	6.873	-	-
8772	0.315	0.503	0.290	0.213	6.589	-	-
38131	0.308	0.521	0.247	0.274	6.425	-	-
41217	0.291	0.505	0.083	0.422	6.041	-	-
14438	0.283	0.325	0.489	-0.164	5.861	-	-
22038	0.281	0.443	0.000	0.443	5.825	RF: Categorical	-
45813	0.258	0.452	0.416	0.037	5.313	-	-
56929	0.255	0.508	0.304	0.204	5.243	-	-
44551	0.250	0.443	0.000	0.443	5.116	-	-
59167	0.242	0.083	0.479	-0.396	4.948	-	-
25236	0.235	0.000	0.431	-0.431	4.797	-	-
49998	0.234	0.503	0.368	0.135	4.770	-	-
23292	0.226	0.148	0.505	-0.358	4.601	-	-
25876	0.214	0.077	0.455	-0.378	4.319	-	-
51003	0.211	0.148	0.492	-0.345	4.264	-	-
37008	0.210	0.442	0.077	0.365	4.228	-	-
34794	0.197	0.464	0.100	0.364	3.945	-	-
60959	0.196	0.526	0.290	0.236	3.913	-	-
56982	0.191	0.077	0.431	-0.354	3.805	-	-
7382	0.189	0.100	0.458	-0.358	3.770	-	-
37627	0.189	0.391	0.505	-0.114	3.755	-	-
4953	0.188	0.228	0.521	-0.293	3.747	-	-
46344	0.187	0.492	0.159	0.333	3.728	-	-
3604	0.187	0.368	0.523	-0.155	3.711	-	-
14984	0.186	0.425	0.485	-0.060	3.705	-	-
53871	0.186	0.370	0.000	0.370	3.698	-	-
15079	0.184	0.434	0.071	0.363	3.653	-	-
2162	0.182	0.138	0.471	-0.333	3.604	RF: Patternize/ Categorical	*TYR*
15626	0.181	0.000	0.323	-0.323	3.578	-	-
51884	0.177	0.000	0.337	-0.337	3.487	-	-
25542	0.176	0.495	0.409	0.085	3.475	-	-
35159	0.175	0.077	0.443	-0.366	3.453	-	-

28039	0.175	0.344	0.517	-0.173	3.444	-	-
3318	0.174	0.503	0.173	0.330	3.438	-	-
78834	0.173	0.000	0.344	-0.344	3.414	-	-
4223	0.173	0.409	0.071	0.338	3.404	-	-
49412	0.172	0.349	0.000	0.349	3.378	-	-
11939	0.171	0.476	0.148	0.328	3.362	-	-
19991	0.168	0.431	0.077	0.354	3.288	-	-
11872	0.167	0.442	0.100	0.342	3.265	-	-
62853	0.167	0.312	0.000	0.312	3.261	-	-
4116	0.166	0.083	0.431	-0.348	3.261	-	-
30936	0.166	0.239	0.508	-0.269	3.250	-	-
12669	0.166	0.331	0.000	0.331	3.244	-	-
32071	0.161	0.000	0.337	-0.337	3.147	-	-
22680	0.161	0.416	0.083	0.332	3.137	-	-
55474	0.161	0.083	0.416	-0.332	3.137	-	-
27140	0.160	0.370	0.000	0.370	3.122	-	-
38866	0.156	0.000	0.337	-0.337	3.031	-	-
29604	0.155	0.271	0.518	-0.247	3.008	-	-

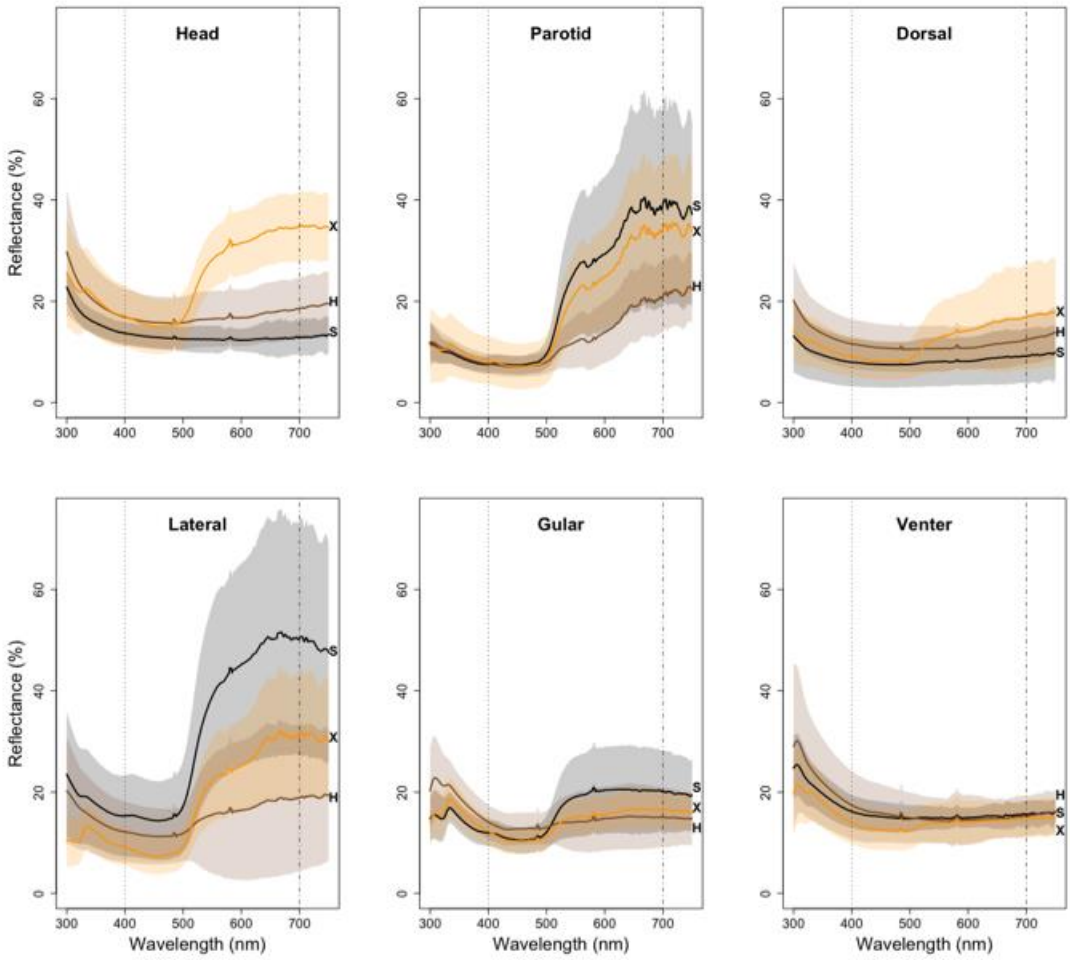
Supplementary Table A5.15: Likelihood values obtained for different levels of genetic groups (K), calculated by STRUCTURE analyses of the 142 loci across colour morph comparisons with a Z*Fst* over three standard deviations from the mean; the best fitting value (K = 3) is shown in bold. Each value of K was evaluated over 5 replicates. (Evanno table generated by STRUCTURE HARVESTER.)

K	Mean LnP(K)	Stdev LnP(K)	Ln'(K)	|Ln''(K)|	Delta K
1	-5330.04	0.45607	—	—	—
2	-5171.46	4.655427	158.58	0.08	0.017184
3	**-5012.8**	**0.547723**	**158.66**	**240.8**	**439.638639**
4	-5094.94	72.463943	-82.14	85.2	1.175757
5	-5091.88	47.859137	3.06	28.26	0.590483
6	-5117.08	18.072963	-25.2	—	—

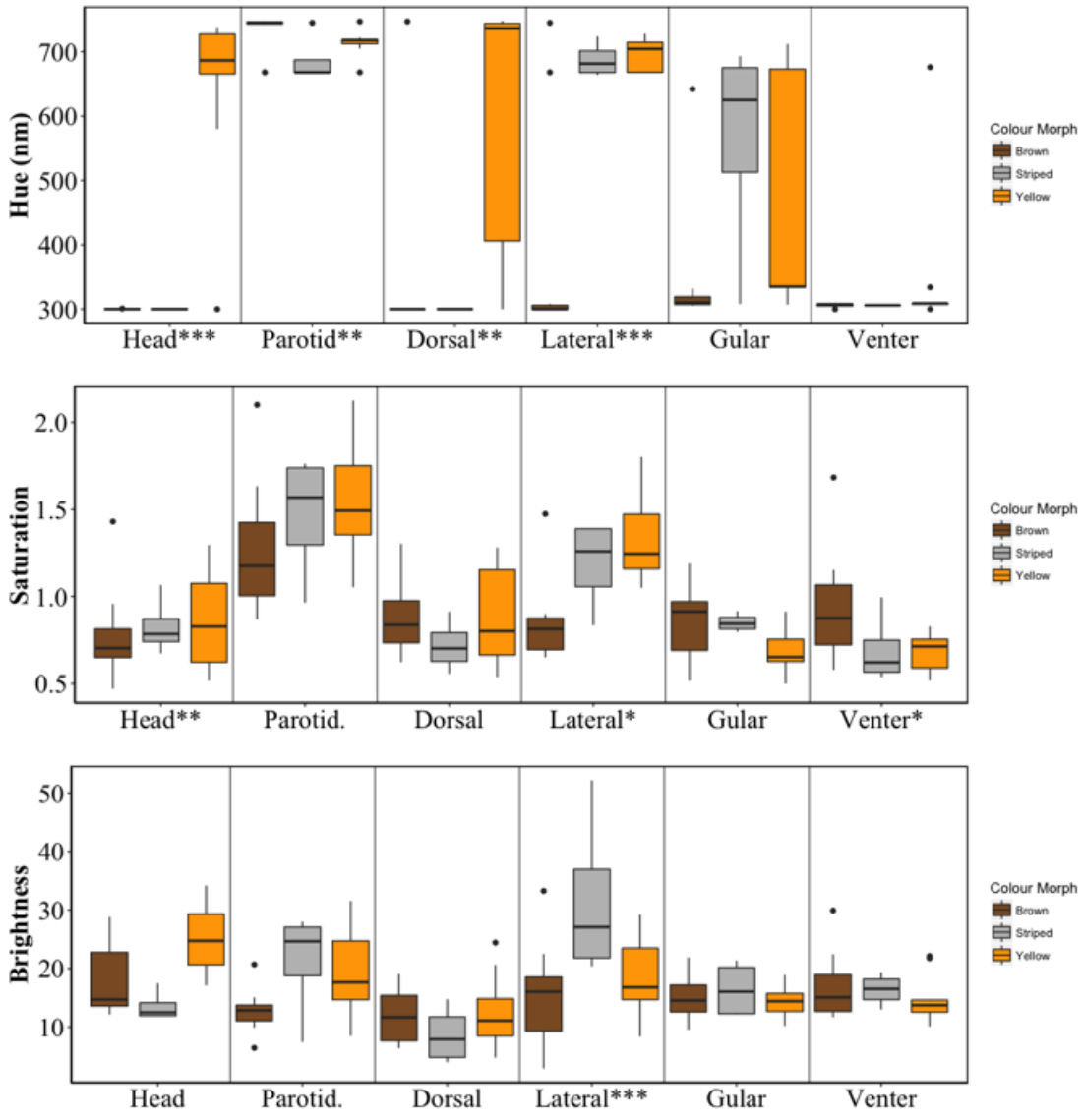
A5.3. Supplementary Figures (A5.2–5 and 8-13)



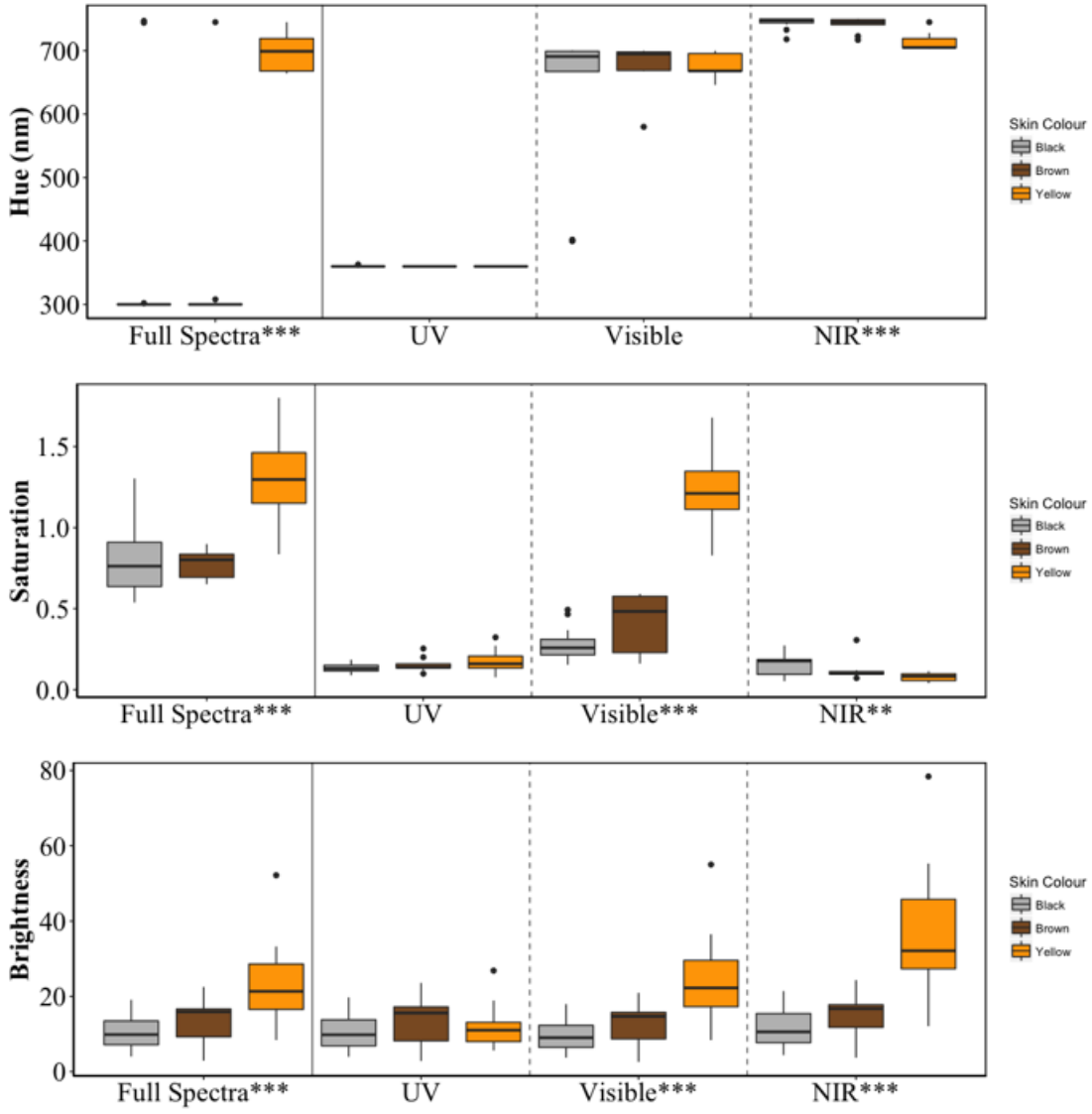
Supplementary Figure A5.1: *Patternize* PCA plots: **left**) comparing the colour patterns of striped salamanders from sample site 7 to salamanders from sample site 1; **right**) *Patternize* PCA plot comparing colour patterns between male and female salamanders in the colour variable contact region (site 7).



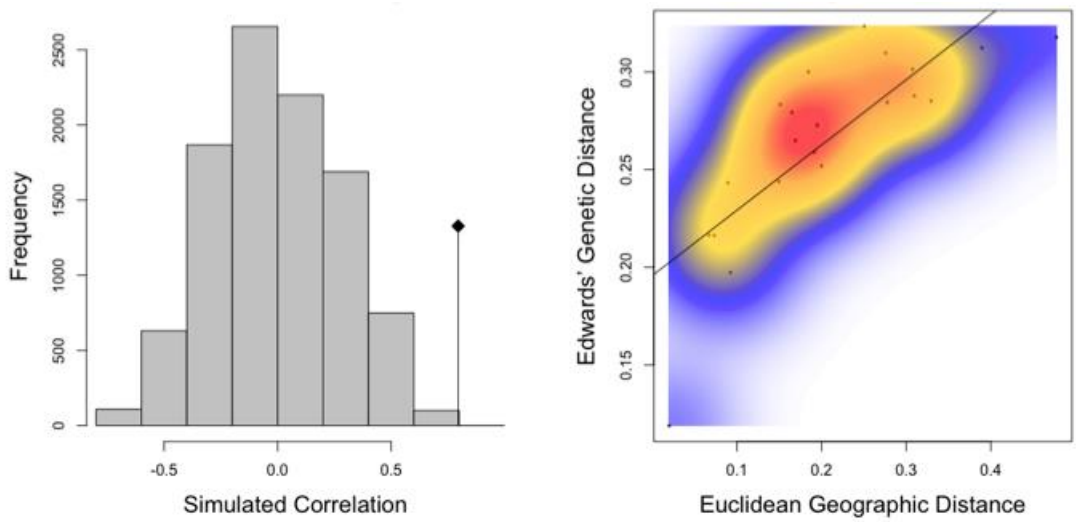
Supplementary Figure A5.2: Averaged reflectance spectra for each representative colour morph at each body landmark (with standard deviation shown by shading): X= xanthic; H=hypoleutic; S= black-yellow striped.



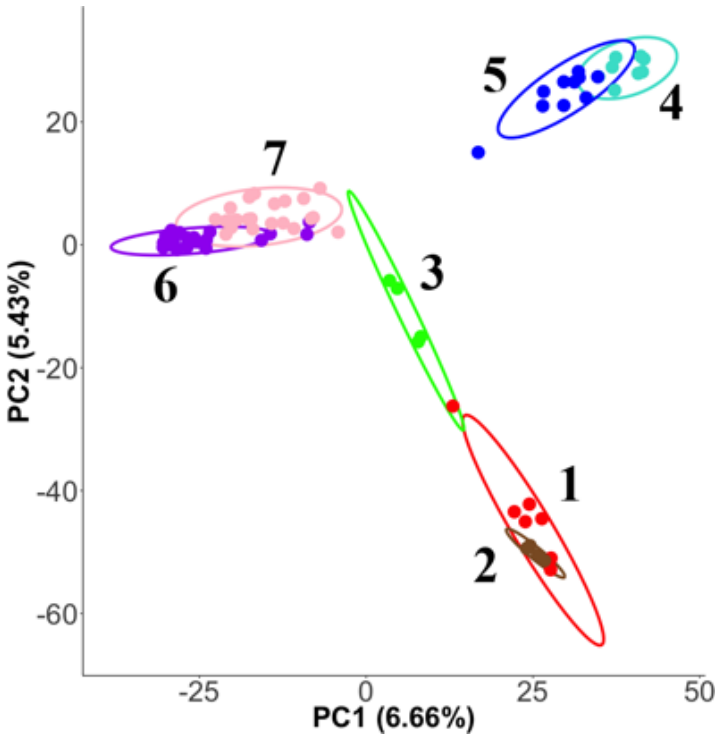
Supplementary Figure A5.3: Hue, saturation and brightness values extracted from reflectance spectra for different body landmarks across representative colour morphs.



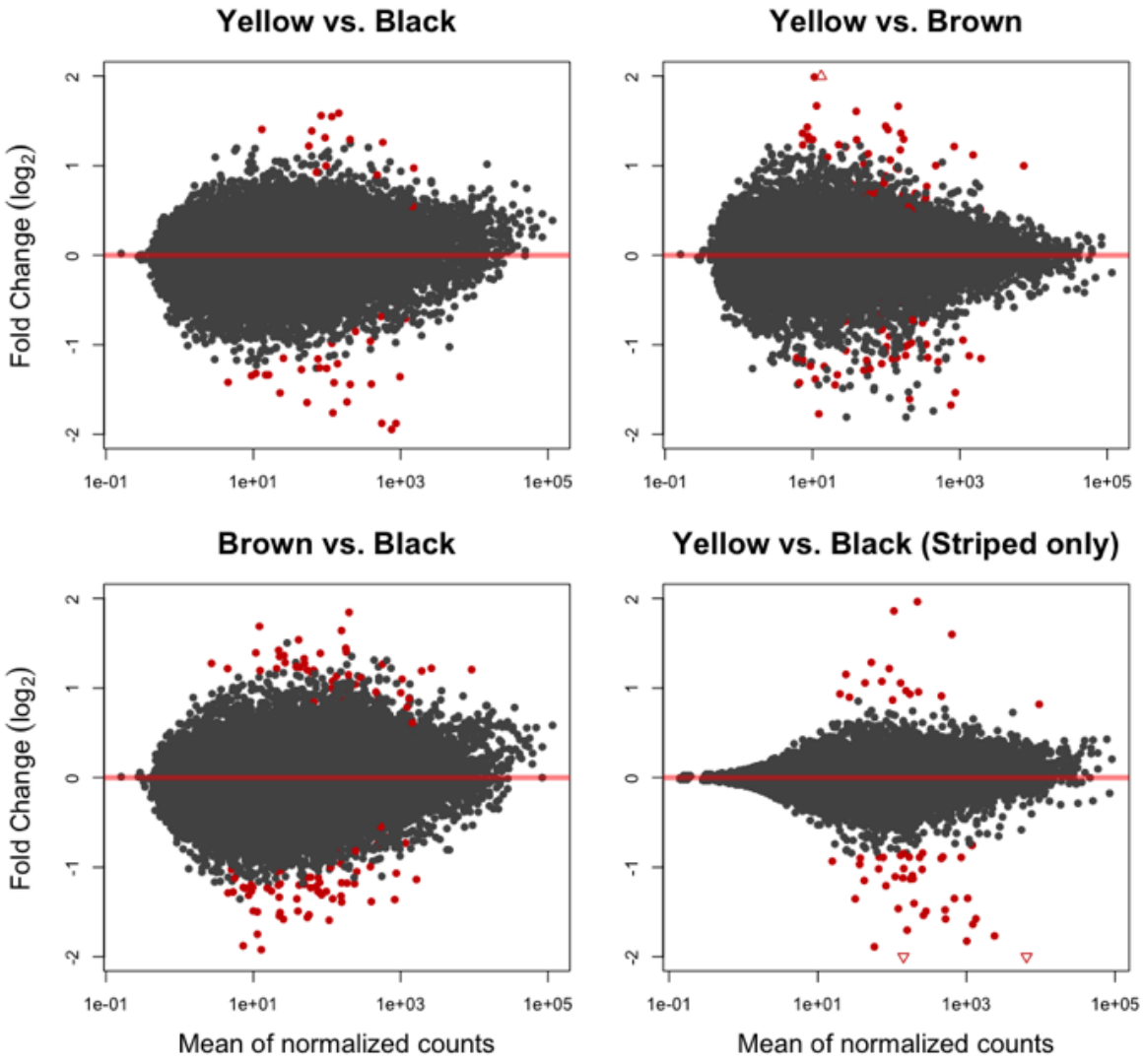
Supplementary Figure A5.4: Hue, saturation and brightness values extracted from yellow, brown and black exemplar spectra in different wavelengths of light.



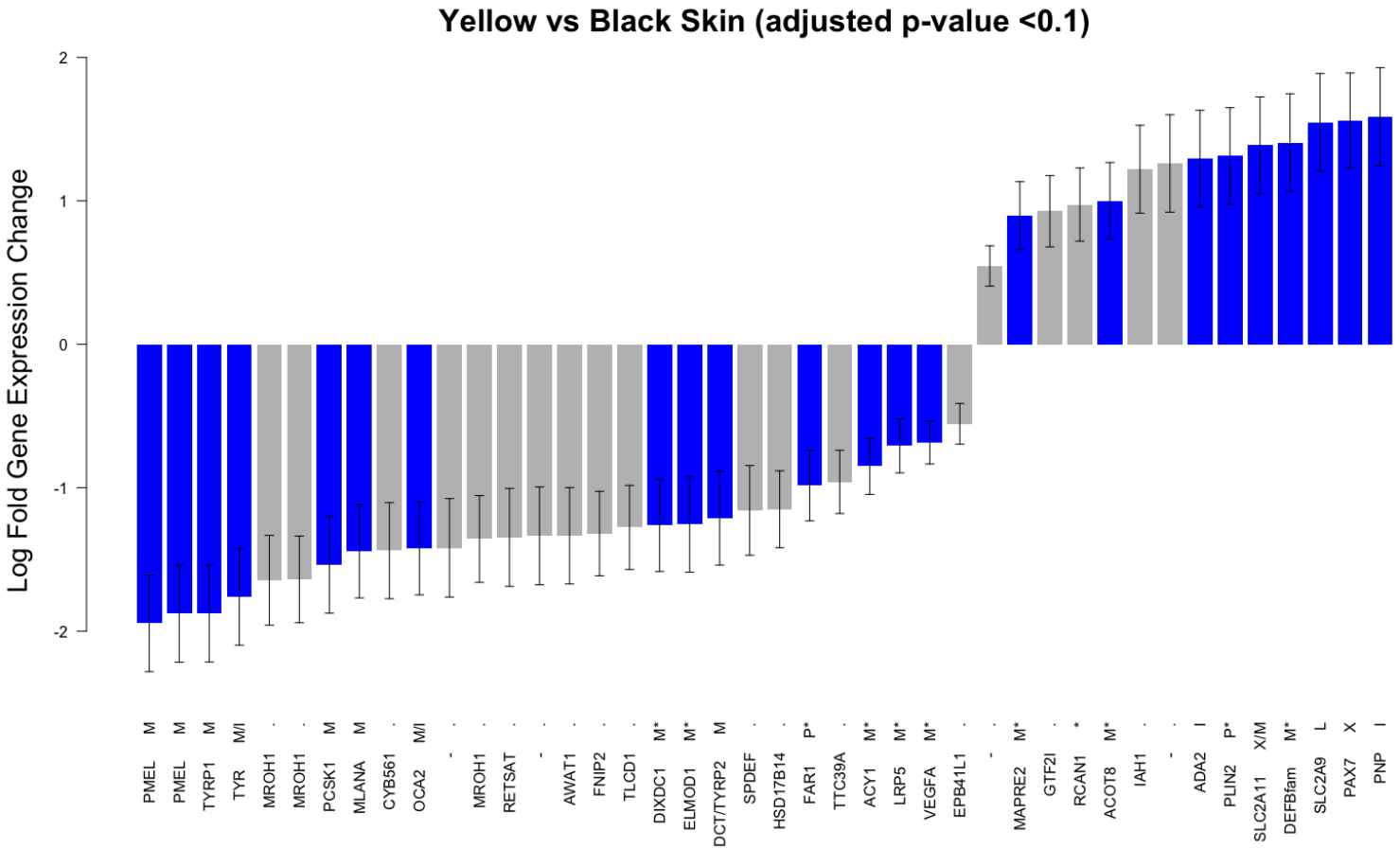
Supplementary Figure A5.5: **Left**: Mantel test histogram showing the actual correlation between the genetic and geographic matrixes (point) and the distribution of correlation values obtained through data permutation (i.e. when no geographic structure is assumed). **Right**: isolation by distance plot. Both show significant isolation by distance (Mantel test: r=0.795; p<0.005) between sample sites 1–7.



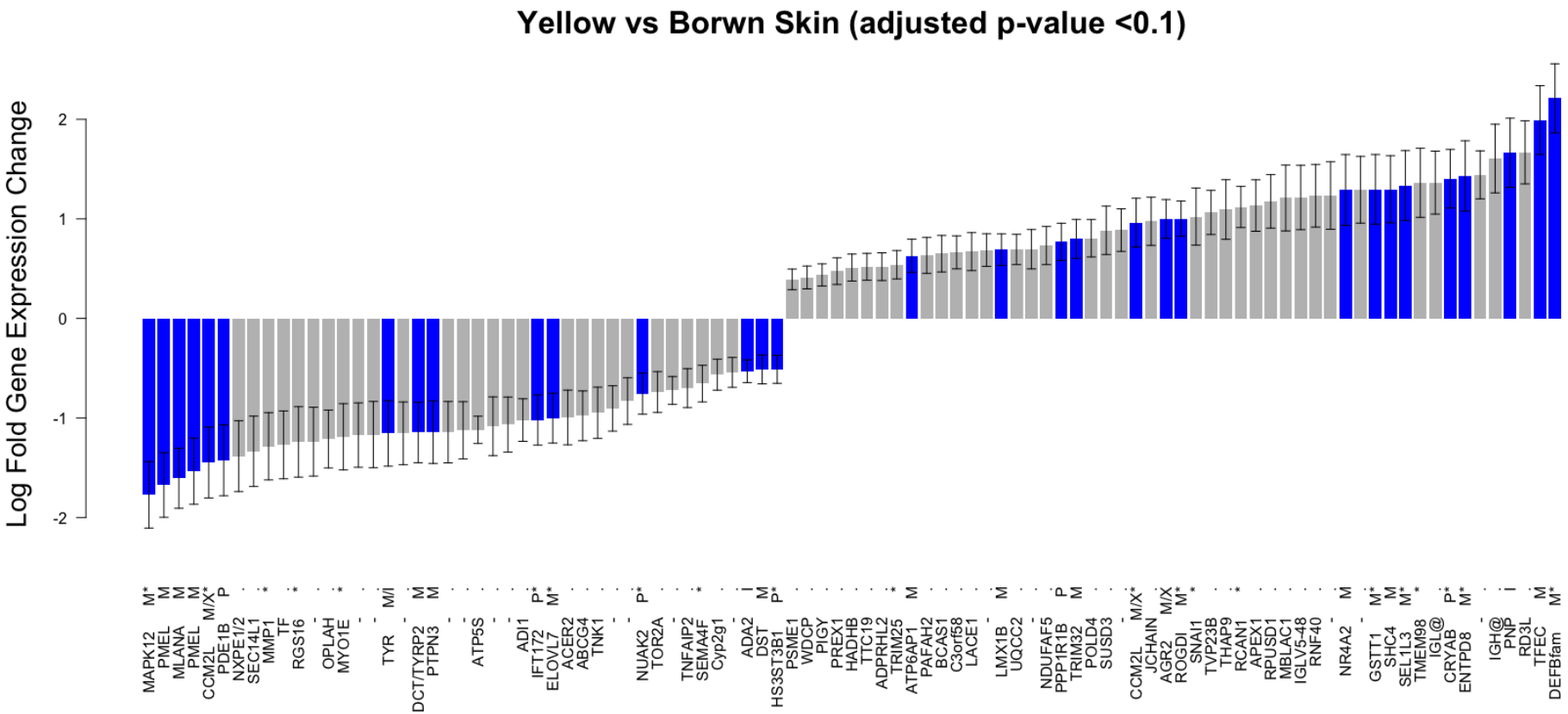
Supplementary Figure A5.6: Principal component analysis comparing variation within and between sample sites 1–7.



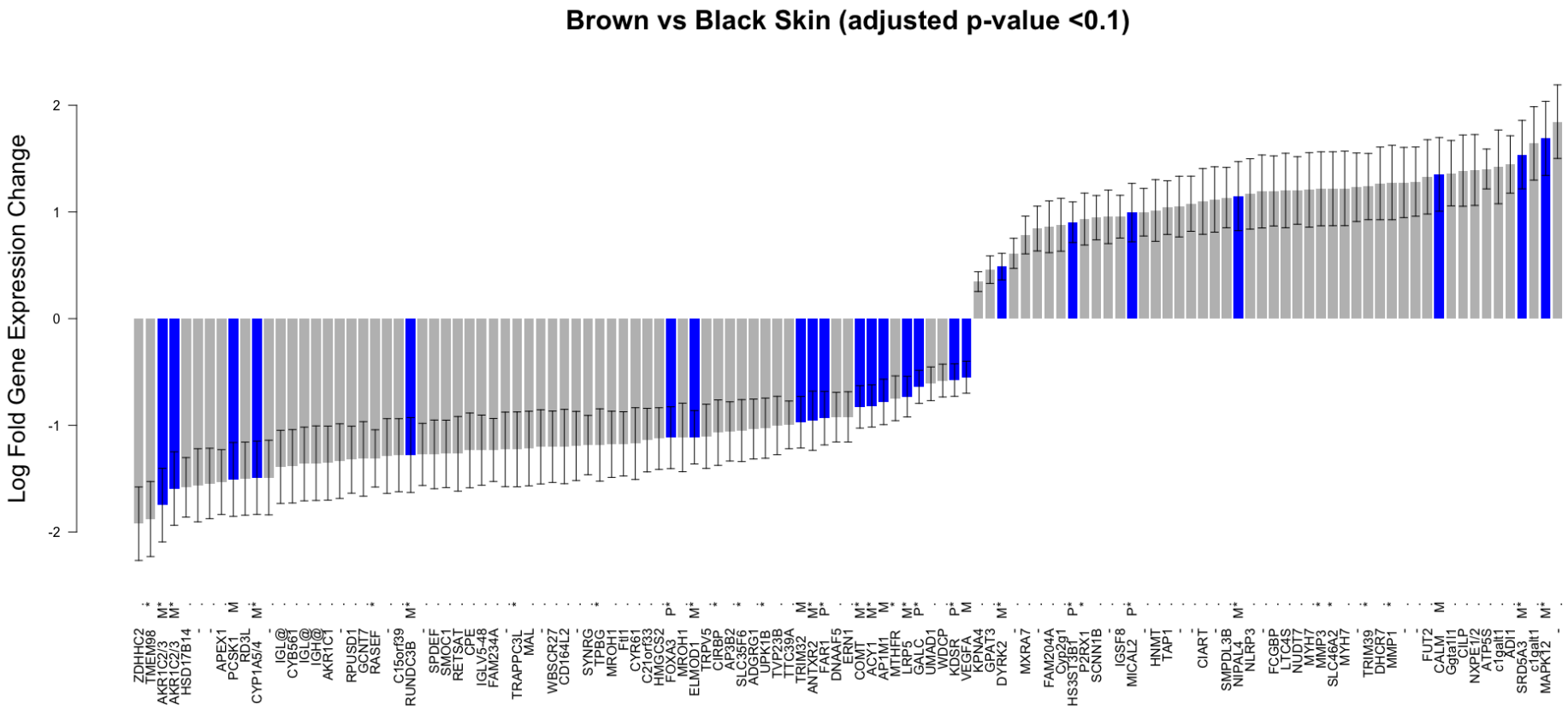
Supplementary Figure A5.7: MA plots of differential gene expression between pairwise comparisons of skin colours. Transcripts highlighted in red are significantly differentially expressed (p-adjust <0.1).



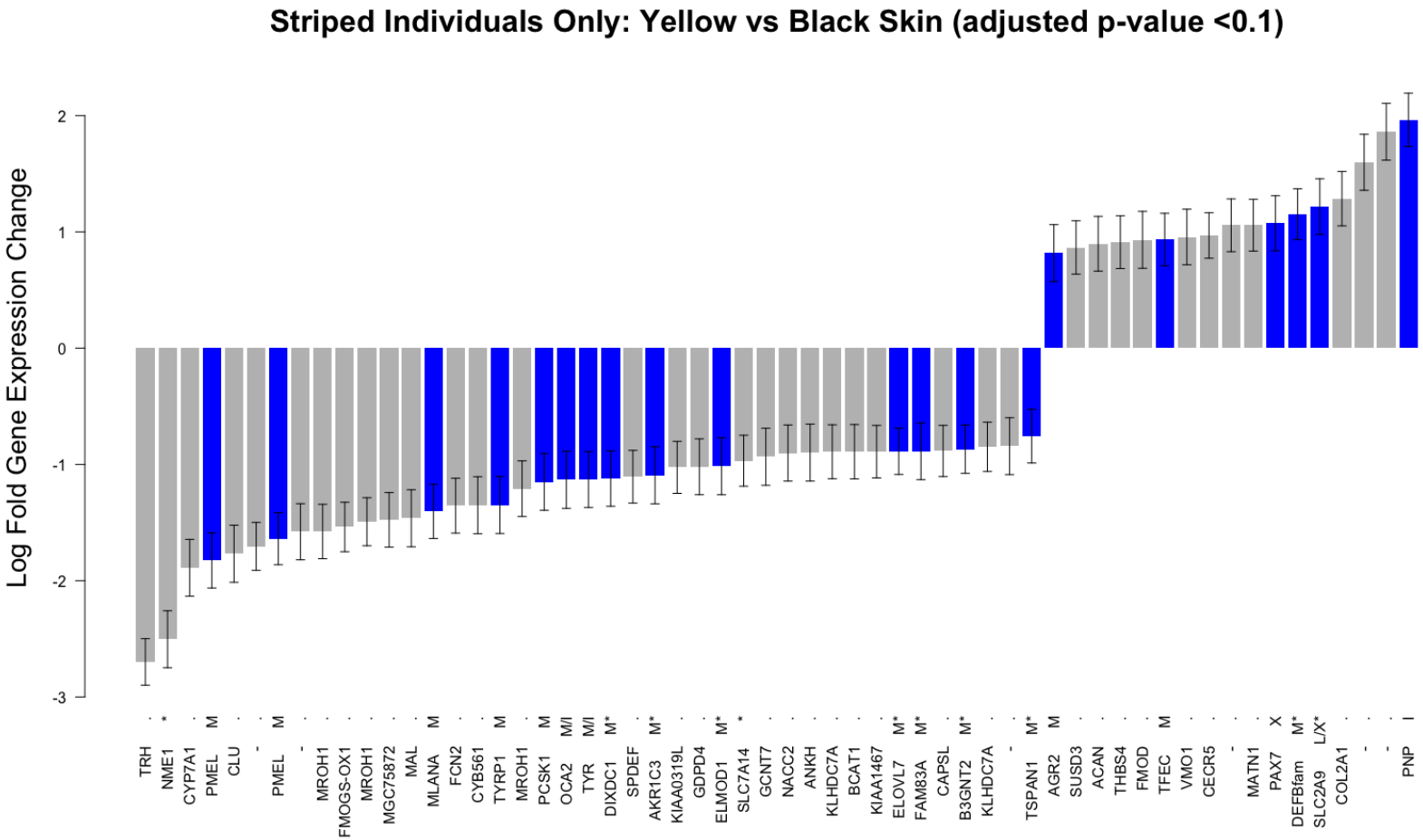
Supplementary Figure A5.8: Barplot showing the log fold change in expression between sDE clusters in the yellow-black skin comparison (with standard error bars). Genes known or suspected (*) to be involved in animal colouration (from post hoc-identifications) are highlighted in blue. Letters next to gene names indicate the associated chromatophore: M=melanophores; X=xanthophores; I=iridophores; L=leucophores; P=other or unknown, but pigmentation related.



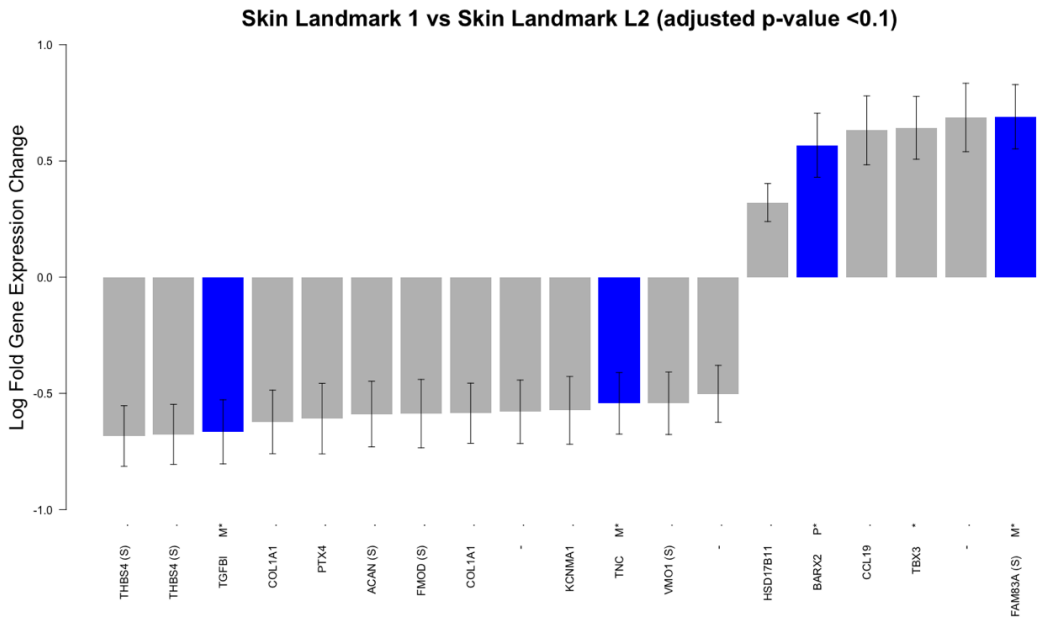
Supplementary Figure A5.9: Barplot showing the log fold change in expression between sDE clusters in the yellow-brown skin comparison (with standard error bars). Genes known or suspected (*) to be involved in animal colouration (from post hoc-identifications) are highlighted in blue. Letters next to gene names indicate the associated chromatophore: M=melanophores; X=xanthophores; I=iridophores; L=leucophores; P=other or unknown, but pigmentation related.



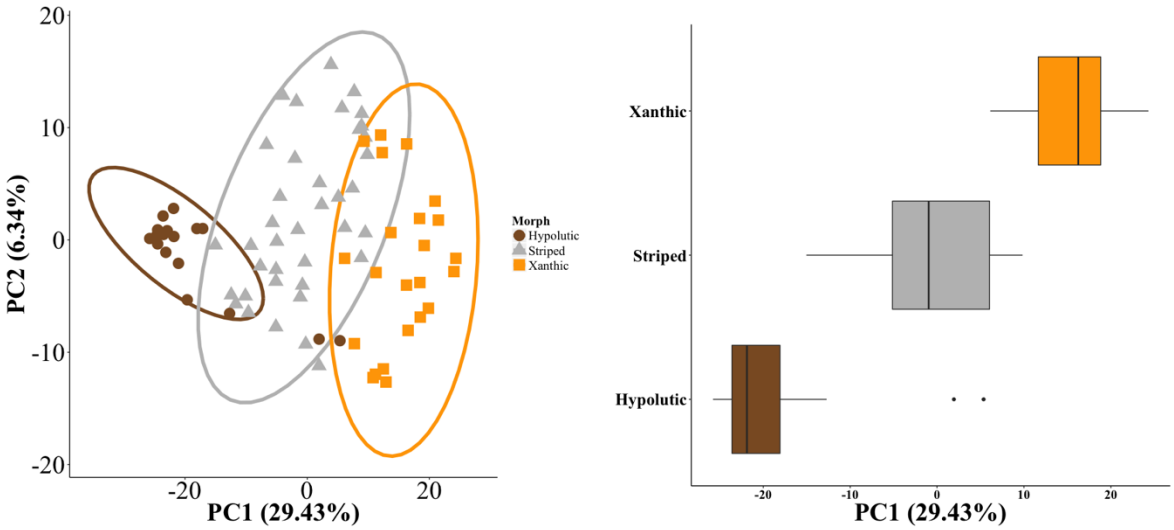
Supplementary Figure A5.10: Barplot showing the log fold change in expression between sDE clusters in the brown-black skin comparison (with standard error bars). Genes known or suspected (*) to be involved in animal colouration (from post hoc-identifications) are highlighted in blue. Letters next to gene names indicate the associated chromatophore: M=melanophores; X=xanthophores; I=iridophores; L=leucophores; P=other or unknown, but pigmentation related.



Supplementary Figure A5.11: Barplots showing the log fold change in expression between sDE clusters in the striped individual only yellow-black comparison (with standard error bars). Genes known or suspected (*) to be involved in animal colouration (from post hoc-identifications) are highlighted in blue. Letters next to gene names indicate the associated chromatophore: M=melanophores; X=xanthophores; I=iridophores; L=leucophores.



Supplementary Figure A5.12: Barplots showing the log fold change in expression between sDE clusters in the comparison of skin landmarks (with standard error bars). Genes known or suspected (*) to be involved in animal colouration (from post hoc-identifications) are highlighted in blue. Letters next to gene names indicate the associated chromatophore: M=melanophores; P=other or unknown, but pigmentation related. Genes overlapping with the striped-only yellow-black comparison are denoted by (S).



Supplementary Figure A5.13: *Patternize* analyses of the 79 salamanders used in genotype-phenotype association studies: **Left**: coloured by representative phenotype. **Right**: PC1 scores categorized by pre-assigned colour morph.

Appendix 6: Custom R scripts

Note: only presented are those scripts not associated to well documented packages.

A6.1. Creating RAD-loci whitelists from the Stacks catalogue

```
#Read in Stacks catalogue
snps.tags <- read.table("Stacks_batch.catalog.snps.tsv")

#Sub-set loci identifiers
snps.loci <- snps.tags[,c(3)]
head(snps.loci)

#Count the number of times an identifier appears (i.e. the number of SNPs per RAD-locus)
snps.counts <- aggregate(data.frame(count = snps.loci), list(value = snps.loci), length)

#Remove identifiers for loci with greater than X SNPs (here, 5)
maxSNP5 <- snps.counts[!rowSums(snps.counts[-1] >5),]

#Create whitelist of locus identifiers for maxSNPs
whitelist <- maxSNP5[,c(1)]
head(whitelist)

#Write whitelist to file
write.table(whitelist,"whitelist_maxSNP5.txt",sep="\t",row.names=FALSE,
col.names=FALSE)
```

A6.2. Calculating genotype error rates

```
#Import consensus sequence data (SNPs)
#Here: one individual/repeat per column, one base per row. SNP ID in column one.
replicates <- read.csv("replicates.csv", header=TRUE)
View(replicates)
```

#Markes ID	Sample-1_Rep1	Sample-1_Rep2	Sample-2_Rep1…
#1	A	A	A
#2	A	A	G
#3	C	C	C
#...			

```
#Subset the two technical replicates for an individual sample
Sample <- reps[,1:3]
summary(Sample)

#Identify missing data (i.e. loci not shared between replicates)
missing <- which(is.na(Sample), arr.ind=TRUE)
missing

#Remove SNP loci with missing data
rows <- missing[,1]
Sample <- Sample[ ! rownames(Sample) %in% rows, ]

#Calculate shared (same) and different (diff) SNP calls
Sample.counts <- data.frame(table(Sample$Sample-1_Rep1, Sample$Sample-1_Rep2) )
Sample.counts <- Sample.counts[ ! Sample.counts$Freq %in% 0, ]
summary(Sample.counts)
Sample.Same <- Sample.counts[Sample.counts$Var1==Sample.counts$Var2,]
Sample.diff <- Sample.counts[Sample.counts$Var1!=Sample.counts$Var2,]

#Calculate % genotype error rate
genotype.Error <- sum(Sample.diff$Freq)/sum(Sample.Same$Freq)*100
genotype.Error

#Repeat for all replicates and calculate average
```

Reference bibliography

Acord, M. A., C. D. Anthony, and C.-A. M. Hickerson. 2013. Assortative mating in a polymorphic salamander. Copeia 2013:676–683.

Acton, Q. A. (ed). 2013. Scleroproteins—Advances in research and application: 2013 edition. ScholarlyEditions.

Ainhoa Iraola, and M. Garcia-Paris. 2004. Molecular evidence of fraudulent transport and sale of *Salamandra salamandra* (Caudata: Salamandridae). Munibe 55:217–222.

Alexandrou, M. A., C. Oliveira, M. Maillard, R. A. R. McGill, J. Newton, S. Creer, and M. I. Taylor. 2011. Competition and phylogeny determine community structure in Müllerian co-mimics. Nature 469:84–88.

Alho, J. S., G. Herczeg, F. Söderman, A. Laurila, K. I. Jönsson, and J. Merilä. 2010. Increasing melanism along a latitudinal gradient in a widespread amphibian: local adaptation, ontogenic or environmental plasticity? BMC Evolutionary Biology 10:317.

Allendorf, F. W., P. A. Hohenlohe, and G. Luikart. 2010. Genomics and the future of conservation genetics. Nature reviews. Genetics 11:697–709.

Alves, R. N., A. S. Gomes, K. Stueber, M. Tine, M. A. S. Thorne, H. Småradóttir, R. Reinhard, M. S. Clark, I. Rønnestad, and D. M. Power. 2016. The transcriptome of metamorphosing flatfish. BMC Genomics 17:413.

Amemiya, C. T., J. Alfoldi, A. P. Lee, S. Fan, H. Philippe, I. MacCallum, I. Braasch, T. Manousaki, I. Schneider, N. Rohner, C. Organ, D. Chalopin, J. J. Smith, M. Robinson, R. A. Dorrington, M. Gerdol, B. Aken, M. A. Biscotti, M. Barucca, D. Baurain, A. M. Berlin, G. L. Blatch, F. Buonocore, T. Burmester, M. S. Campbell, A. Canapa, J. P. Cannon, A. Christoffels, G. De Moro, A. L. Edkins, L. Fan, A. M. Fausto, N. Feiner, M. Forconi, J. Gamieldien, S. Gnerre, A. Gnirke, J. V Goldstone, W. Haerty, M. E. Hahn, U. Hesse, S. Hoffmann, J. Johnson, S. I. Karchner, S. Kuraku, M. Lara, J. Z. Levin, G. W. Litman, E. Mauceli, T. Miyake, M. G. Mueller, D. R. Nelson, A. Nitsche, E. Olmo, T. Ota, A. Pallavicini, S. Panji, B. Picone, C. P. Ponting, S. J. Prohaska, D. Przybylski, N. R. Saha, V. Ravi, F. J. Ribeiro, T. Sauka-Spengler, G. Scapigliati, S. M. J. Searle, T. Sharpe, O. Simakov, P. F. Stadler, J. J. Stegeman, K. Sumiyama, D. Tabbaa, H. Tafer, J. Turner-Maier, P. van Heusden, S. White, L. Williams, M. Yandell, H. Brinkmann, J.-N. Volff, C. J. Tabin, N. Shubin, M. Schartl, D. B. Jaffe, J. H. Postlethwait, B. Venkatesh, F. Di Palma, E. S. Lander, A. Meyer, and K. Lindblad-Toh. 2013. The African coelacanth genome provides insights into tetrapod evolution. Nature 496:311–316.

Amsterdam, A., S. Burgess, G. Golling, W. Chen, Z. Sun, K. Townsend, S. Farrington, M. Haldi, and N. Hopkins. 1999. A large-scale insertional mutagenesis screen in zebrafish. Genes & development 13:2713–24.

Anderson, T. M., B. M. VonHoldt, S. I. Candille, M. Musiani, C. Greco, D. R. Stahler, D. W. Smith, B. Padhukasahasram, E. Randi, J. A. Leonard, C. D. Bustamante, E. A. Ostrander, H. Tang, R. K. Wayne, and G. S. Barsh. 2009. Molecular and evolutionary history of melanism in North American gray wolves. Science 323:1339–1343.

Andrews, K. R., J. M. Good, M. R. Miller, G. Luikart, and P. A. Hohenlohe. 2016. Harnessing the power of RADseq for ecological and evolutionary genomics. Nature Reviews Genetics 17:81.

Andrews, S. 2010. FastQC: a quality control tool for high throughput sequence data.

Anthony, C. D., M. D. Venesky, and C.-A. M. Hickerson. 2008. Ecological separation in a polymorphic terrestrial salamander. Journal of Animal Ecology 77:646–653.

April, C. S., G. S. Barsh, K. Marker, H. Rittenhouse, and R. Wolfert. 2007. Distinct pigmentary and melanocortin 1 receptor–dependent components of cutaneous defense against ultraviolet radiation. PLoS Genetics 3:e9.

Ashburner, M., C. A. Ball, J. A. Blake, D. Botstein, H. Butler, J. M. Cherry, A. P. Davis, K. Dolinski, S. S. Dwight, J. T. Eppig, M. A. Harris, D. P. Hill, L. Issel-Tarver, A. Kasarskis, S. Lewis, J. C. Matese, J. E. Richardson, M. Ringwald, G. M. Rubin, and G. Sherlock. 2000. Gene Ontology: tool for the unification of biology. Nature Genetics 25:25–29.

Aspengren, S., D. Hedberg, and M. Wallin. 2006. Studies of pigment transfer between *Xenopus laevis* melanophores and fibroblasts in vitro and in vivo1. Pigment Cell Research 19:136–145.

Axelsson, E., A. Ratnakumar, M.-L. Arendt, K. Maqbool, M. T. Webster, M. Perloski, O. Liberg, J. M. Arnemo, Å. Hedhammar, and K. Lindblad-Toh. 2013. The genomic signature of dog domestication reveals adaptation to a starch-rich diet. Nature 495:360–364.

Bagley, J. C. 2016. Setting DNA substitution models in BEAST. Availalbe at http://www.justinbagley.org/1058/setting-dna-substitution-models-beast. Date accessed: 2017-07-25.

Bagnara, J. T. 1972. Interrelationships of melanophores, iridophores, and xanthophores. Pp. 171–180 *in* V. Riley, ed. Pigmentation: Its genesis and biologic control. Appleton-Century-Crofts, New York.

Bagnara, J. T., M. E.Hadley, and J. D.Taylor. 1969. Regulation of bright-colored pigmentation of amphibians. General and Comparative Endocrinology 2:425–438.

Bagnara, J. T., P. J. Fernandez, and R. Fujii. 2007. On the blue coloration of vertebrates. Pigment Cell Research 20:14–26.

Bagnara, J. T., and W. R. Ferris. 1971. Interrelationship of chromatophores. Pp. 57–76 *in* T. Kawamura, T. B. Fitzpatrick, and M. Seiji, eds. Biology of the normal and abnormal melanocyte. University of Tokyo, Tokyo.

Bagnara, J. T., S. K. Frost, and J. Matsumoto. 1978. On the development of pigment patterns in amphibians. American Zoologist 18:301–312.

Bagnara, J. T., and J. H. Fukuzawa. 1990. Stimulation of cultured iridophores by amphibian ventral conditioned medium. Pigment Cell Research 3:243–250.

Bagnara, J. T., and M. E. Hadley. 1973. Chromatophores and color change: the comparative physiology of animal pigmentation. Prentice-Hall, Englewood eliffs, New Jersey.

Bagnara, J. T., J. D. Taylor, and M. E. Hadley. 1968. The dermal chromatophore unit. The Journal of cell biology 38:67–79.

Baird, N. A., P. D. Etter, T. S. Atwood, M. C. Currey, A. L. Shiver, Z. A. Lewis, E. U. Selker, W. A. Cresko, and E. A. Johnson. 2008. Rapid SNP discovery and genetic mapping using sequenced RAD markers. PloS one 3:e3376.

Bajger, J. 1980. Diversity of defensive responses in populations of fire toads (*Bombina bombina* and *Bombina variegata*). Herpetologica 36:133–137.

Balogová, M., and M. Uhrin. 2015. Sex-biased dorsal spot patterns in the fire salamander (*Salamandra salamandra*). Salamandra 51:12 – 18.

Bapteste, E., H. Brinkmann, J. A. Lee, D. V Moore, C. W. Sensen, P. Gordon, L. Duruflé, T. Gaasterland, P. Lopez, M. Müller, and H. Philippe. 2002. The analysis of 100 genes supports the grouping of three highly divergent amoebae: *Dictyostelium*, *Entamoeba*, and *Mastigamoeba*. Proceedings of the National Academy of Sciences of the United States of America 99:1414–1419.

Barbato, M., F. Hailer, P. Orozco-terWengel, J. Kijas, P. Mereu, P. Cabras, R. Mazza, M. Pirastru, and M. W. Bruford. 2017. Genomic signatures of adaptive introgression from European mouflon into domestic sheep. Scientific Reports 7:7623.

Beirl, A. J., T. H. Linbo, M. J. Cobb, and C. D. Cooper. 2014. *oca2* regulation of chromatophore differentiation and number is cell type specific in zebrafish. Pigment Cell & Melanoma Research 27:178–189.

Belden, L. K., and A. R. Blaustein. 2002. UV-B induced skin darkening in larval

salamanders does not prevent sublethal effects of exposure on growth. Copeia 2002:748–754.
Bell, R. C., and K. R. Zamudio. 2012. Sexual dichromatism in frogs: natural selection, sexual selection and unexpected diversity. Proceedings of the Royal Society B: Biological Sciences 279:4687–93.
Belur, V. D. 1991. Nearest neighbour (NN) norms, NN pattern classification techniques. IEEE Computer Society Press, Los Alamitos, CA.
Bennett, D. C., and M. L. Lamoreux. 2003. The color loci of mice – a genetic century. Pigment cell research 16:333–44.
Bergman, T. J., and J. C. Beehner. 2008. A simple method for measuring colour in wild animals: validation and use on chest patch colour in geladas (*Theropithecus gelada*). Biological Journal of the Linnean Society 94:231–240.
Bernatchez, S., M. Laporte, C. Perrier, P. Sirois, and L. Bernatchez. 2016. Investigating genomic and phenotypic parallelism between piscivorous and planktivorous lake trout (*Salvelinus namaycush*) ecotypes by means of RADseq and morphometrics analyses. Molecular Ecology 25:4773–4792.
Bernt, M., A. Donath, F. Jühling, F. Externbrink, C. Florentz, G. Fritzsch, J. Pütz, M. Middendorf, and P. F. Stadler. 2013. MITOS: Improved *de novo* metazoan mitochondrial genome annotation. Molecular Phylogenetics and Evolution 69:313–319.
Beukema, W. 2011. Ontogenetic pattern change in amphibians: the case of *Salamandra corsica*. Acta Herpetologica 6:169–174.
Beukema, W., P. De Pous, D. Donaire-Barroso, S. Bogaerts, J. Garcia-Porta, D. Escoriza, O. J. Arribas, E. Hassan, E. Mouden, and S. Carranza. 2013. Review of the systematics, distribution, biogeography and natural history of Moroccan amphibians. Zootaxa 3661:1–60.
Beukema, W., A. G. Nicieza, A. Lourenço, and G. Velo-Antón. 2016a. 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 54:127–136.
Beukema, W., P. D. E. Pous, D. Donaire, D. Escoriza, S. Bogaerts, A. G. Toxopeus, C. A. J. M. D. E. Bie, J. Roca, and S. Carranza. 2010. Biogeography and contemporary climatic differentiation among Moroccan *Salamandra algira*. Biological Journal of the Linnean Society 101:626–641.
Beukema, W., J. Speybroeck, G. Velo-Antón, D. Eikelmann, A. Rodríguez, S. Carranza, D. Donaire, M. Gehara, V. Helfer, S. Lötters, and E. Al. 2016b. *Salamandra*. Current Biology 26:R696–R697.
Biserkov, V. (ed). 2007. A field guide to amphibians and reptiles of Bulgaria. Green Balkans, Sofia.
Bogaerts, S. 2002. Farbkleidentwicklung bei einigen feuersalamanderarten und–unterarten. Amphibia 1:4–10.
Böhme, W., T. Hartmann, J. Fleck, and T. Schöttler. 2013. Miscellaneous notes on oriental fire salamanders (*Salamandra infraimmaculata* Martens, 1885) (Lissamphibia: Urodela: Salamandridae). Russian Journal of Herpetology 20:66–72.
Bolger, A. M., M. Lohse, and B. Usadel. 2014. Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120.
Bonato, L., and K. Grossenbacher. 2000. On the distribution and chromatic differentiation of the Alpine Salamander *Salamandra atra* Laurenti, 1768, between Val Lagarina and Val Sugana (Venetian Prealps): an updated review. Herpetozoa 13:171–180.
Bonato, L., and S. Steinfartz. 2005. Evolution of the melanistic colour in the Alpine salamander *Salamandra atra* as revealed by a new subspecies from the Venetian Prealps. Italian Journal of Zoology 72:253–260.
Bouckaert, R., J. Heled, D. Kühnert, T. Vaughan, C.-H. Wu, D. Xie, M. A. Suchard, A.

Rambaut, and A. J. Drummond. 2014. BEAST 2: A software platform for Bayesian evolutionary analysis. PLoS Computational Biology 10:e1003537.
Boulesteix, A.-L., S. Janitza, J. Kruppa, and I. R. König. 2012. Overview of random forest methodology and practical guidance with emphasis on computational biology and bioinformatics. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery 2:493–507.
Bouzid, S., L. Konecny, O. Grolet, C. J. Douady, P. Joly, and Z. Bouslama. 2017. Phylogeny, age structure, growth dynamics and colour pattern of the *Salamandra algira algira* population in the Edough Massif, northeastern Algeria. Amphibia-Reptilia, doi: 10.1163/15685381-00003127.
Braasch, I., M. Schartl, and J.-N. Volff. 2007. Evolution of pigment synthesis pathways by gene and genome duplication in fish. BMC Evolutionary Biology 7:74.
Brodie, E., and E. Brodie. 1990. Tetrodotoxin resistance in garter snakes: an evolutionary response of predators to dangerous prey. Evolution 44:651–659.
Browder, L. W. 1968. Pigmentation in *Rana pipiens*. I. Inheritance of the speckle mutation. Journal of Hereditry 59:162–166.
Bryant, D., R. Bouckaert, J. Felsenstein, N. A. Rosenberg, and R. A. Choudhury. 2012. Inferring species trees directly from biallelic genetic markers: Bypassing gene trees in a full coalescent analysis. Molecular Biology and Evolution 29:1917–1932.
Buckley, D. 2012. Evolution of viviparity in salamanders (Amphibia, Caudata). P. *in* eLS. John Wiley & Sons, Ltd, Chichester, UK.
Buckley, D., M. Alcobendas, M. García-París, and M. H. Wake. 2007. Heterochrony, cannibalism, and the evolution of viviparity in *Salamandra salamandra*. Evolution & Development 9:105–115.
Bull, J. J., J. P. Huelsenbeck, C. W. Cunningham, D. L. Swofford, and P. J. Waddell. 1993. Partitioning and combining data in phylogenetic analysis. Systematic Biology 42:384–397.
Burbach, J. P. H., M. P. Smidt, C. H. J. Asbreuk, J. J. Cox, H. Chen, and R. L. Johnson. 2000. A second independent pathway for development of mesencephalic dopaminergic neurons requires Lmx1b. Nature Neuroscience 3:337–341.
Cagan, A., and T. Blass. 2016. Identification of genomic variants putatively targeted by selection during dog domestication. BMC Evolutionary Biology 16:10.
Candille, S. I., C. B. Kaelin, B. M. Cattanach, B. Yu, D. A. Thompson, M. A. Nix, J. A. Kerns, S. M. Schmutz, G. L. Millhauser, and G. S. Barsh. 2007. A β-Defensin mutation causes black coat color in domestic dogs. Science 318:1418–1423.
Cao, K.-A. Le, F. Rohart, I. Gonzalez, S. Dejean, B. Gautier, F. Bartolo, P. Monget, J. Coquery, F. Yao, and B. Liquet. 2017. mixOmics: Omics data integration project. R package version 6.1.3.
Carbon, S., A. Ireland, C. J. Mungall, S. Shu, B. Marshall, and S. Lewis. 2009. AmiGO: online access to ontology and annotation data. Bioinformatics 25:288–289.
Cariou, M., L. Duret, and S. Charlat. 2013. Is RAD-seq suitable for phylogenetic inference? An in silico assessment and optimization. Ecology and Evolution 3:846–852.
Caro, T. 2005. The adaptive significance of coloration in mammals. BioScience 55:125–136.
Caspers, B. A., E. T. Krause, R. Hendrix, M. Kopp, O. Rupp, K. Rosentreter, and S. 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 23:239–250.
Caspers, B. A., S. Steinfartz, and E. T. Krause. 2015. Larval deposition behaviour and maternal investment of females reflect differential habitat adaptation in a genetically diverging salamander population. Behavioral Ecology and Sociobiology 69:407–413.
Catchen, J., P. A. Hohenlohe, S. Bassham, A. Amores, and W. A. Cresko. 2013. Stacks: an

analysis tool set for population genomics. Molecular Ecology 22:3124–3140.
Catchen, J. M., A. Amores, P. Hohenlohe, W. Cresko, J. H. Postlethwait, and D.-J. De Koning. 2011. Stacks: building and genotyping loci de novo from short-read sequences. G3 1:171–82.
Chen, L., G. Wang, Y.-N. Zhu, H. Xiang, and W. Wang. 2016. Advances and perspectives in the application of CRISPR/Cas9 in insects. Dong wu xue yan jiu = Zoological research 37:220–8.
Chen, M.-Y., D. Liang, and P. Zhang. 2015. Selecting question-specific genes to reduce incongruence in phylogenomics: A case study of jawed vertebrate backbone phylogeny. Systematic Biology 64:1104–1120.
Chen, Y., S. Znoiko, W. J. DeGrip, R. K. Crouch, and J. Ma. 2008. Salamander blue-sensitive cones lost during metamorphosis. Photochemistry and photobiology 84:855–62.
Chevreux, B., T. Wetter, and S. Suhai. 1999. Genome sequence assembly using trace signals and additional sequence information. German conference on bioinformatics 99:45–56.
Chiari, Y., V. Cahais, N. Galtier, and F. Delsuc. 2012. Phylogenomic analyses support the position of turtles as the sister group of birds and crocodiles (Archosauria). BMC Biology 10:65.
Chrétien, M., and M. Mbikay. 2016. 60 years of POMC: From the prohormone theory to pro-opiomelanocortin and to proprotein convertases (PCSK1 to PCSK9). Journal of molecular endocrinology 56:T49-62.
Clarke, K. R. 1993. Non-parametric multivariate analyses of changes in community structure. Austral Ecology 18:117–143.
Clusella-Trullas, S., J. S. Terblanche, T. M. Blackburn, and S. L. Chown. 2008. Testing the thermal melanism hypothesis: a macrophysiological approach. Functional Ecology 22:232–238.
Clusella-Trullas, S., J. H. van Wyk, and J. R. Spotila. 2007. Thermal melanism in ectotherms. Journal of Thermal Biology 32:235–245.
Cosentino, B. J., J.-D. Moore, N. E. Karraker, M. Ouellet, and J. P. Gibbs. 2017. Evolutionary response to global change: Climate and land use interact to shape color polymorphism in a woodland salamander. Ecology and Evolution 7:5426–5434.
Costa, C., C. Angelini, M. Scardi, P. Menesatti, and C. Utzeri. 2009. Using image analysis on the ventral colour pattern in *Salamandrina perspicillata* (Amphibia: Salamandridae) to discriminate among populations. Biological Journal of the Linnean Society 96:35–43.
Cott, H. 1940. Adaptive colouration in animals. Methuen and Co., London.
Crawford, N. G., D. E. Kelly, M. E. B. Hansen, M. H. Beltrame, S. Fan, S. L. Bowman, E. Jewett, A. Ranciaro, S. Thompson, Y. Lo, S. P. Pfeifer, J. D. Jensen, M. C. Campbell, W. Beggs, F. Hormozdiari, S. W. Mpoloka, G. G. Mokone, T. Nyambo, D. Wolde Meskel, G. Belay, J. Haut, H. Rothschild, L. Zon, Y. Zhou, M. A. Kovacs, M. Xu, T. Zhang, K. Bishop, J. Sinclair, C. Rivas, E. Elliot, J. Choi, S. A. Li, B. Hicks, S. Burgess, C. Abnet, D. E. Watkins-Chow, E. Oceana, Y. S. Song, E. Eskin, K. M. Brown, M. S. Marks, S. K. Loftus, W. J. Pavan, M. Yeager, S. Chanock, and S. Tishkoff. 2017. Loci associated with skin pigmentation identified in African populations. Science, doi: 10.1126/science.aan8433.
Cummings, M. E., X. E. Bernal, R. Reynaga, A. S. Rand, and M. J. Ryan. 2008. Visual sensitivity to a conspicuous male cue varies by reproductive state in *Physalaemus pustulosus* females. The Journal of experimental biology 211:1203–10.
Curran, K., J. A. Lister, G. R. Kunkel, A. Prendergas, D. M. Parichy, and D. W. Raible. 2010. Interplay between Foxd3 and Mitf regulates cell fate plasticity in the zebrafish neural crest. Developmental Biology 344:107–118.
Czypionka, T., T. Krugman, J. Altmüller, L. Blaustein, S. Steinfartz, A. R. Templeton, and

A. W. Nolte. 2015. Ecological transcriptomics—a non-lethal sampling approach for endangered fire salamanders. Methods in Ecology and Evolution 6:1417–1425.
da Fonseca, R. R., A. Albrechtsen, G. E. Themudo, J. Ramos-Madrigal, J. A. Sibbesen, L. Maretty, M. L. Zepeda-Mendoza, P. F.Campos, R. Heller, and R. J. Pereira. 2016. Next-generation biology: Sequencing and data analysis approaches for non-model organisms. Marine Genomics 30:3–13.
Danecek, P., A. Auton, G. Abecasis, C. A. Albers, E. Banks, M. A. DePristo, R. E. Handsaker, G. Lunter, G. T. Marth, S. T. Sherry, G. McVean, and R. Durbin. 2011. The variant call format and VCFtools. Bioinformatics 27:2156–2158.
Darias, M. J., K. B. Andree, A. Boglino, I. Fernández, A. Estévez, and E. Gisbert. 2013. Coordinated regulation of chromatophore differentiation and melanogenesis during the ontogeny of skin pigmentation of *Solea senegalensis* (Kaup, 1858). PLoS ONE 8:e63005.
Darriba, D., G. L. Taboada, R. Doallo, and D. Posada. 2012. jModelTest 2: more models, new heuristics and parallel computing. Nature Methods 9:772–772.
Darst, C. R., and M. E. Cummings. 2006. Predator learning favours mimicry of a less-toxic model in poison frogs. Nature 440:208.
Darst, C. R., M. E. Cummings, and D. C. Cannatella. 2006. A mechanism for diversity in warning signals: conspicuousness versus toxicity in poison frogs. Proceedings of the National Academy of Sciences of the United States of America 103:5852–7.
Darwell, C. T., D. M. Rivers, and D. M. Althoff. 2016. RAD-seq phylogenomics recovers a well-resolved phylogeny of a rapid radiation of mutualistic and antagonistic yucca moths. Systematic Entomology 41:672–682.
Davey, J. W. L. W., and M. W. L. Blaxter. 2010. RADSeq: next-generation population genetics. Briefings in Functional Genomics 9:416–23.
Davidson, N. M., and A. Oshlack. 2014. Corset: enabling differential gene expression analysis for de novoassembled transcriptomes. Genome Biology 15:410.
Davis, A. K., B. Farrey, and S. Altizer. 2004. Quantifying monarch butterfly larval pigmentation using digital image analysis. Entomologia Experimentalis et Applicata 113:145–147.
Davis, A. K., B. D. Farrey, and S. Altizer. 2005. Variation in thermally induced melanism in monarch butterflies (Lepidoptera: Nymphalidae) from three North American populations. Journal of Thermal Biology 30:410–421.
Davis, A. K., and J. R. Milanovich. 2010. Lead-phase and red-stripe color morphs of red-backed salamanders *Plethodon cinereus* differ in hematological stress indices: A consequence of differential predation pressure? Current Zoology 56:238–243.
Davis, A., and J. Maerz. 2007. Spot symmetry predicts body condition in spotted salamanders, *Ambystoma maculatum*. Applied Herpetology 4:195–205.
Degani, G. 1986. Plasma proteins and morphology of *Salamandra salamandra* in Israel. Amphibia-Reptilia 7:105–114.
Degnan, J. H., and N. A. Rosenberg. 2009. Gene tree discordance, phylogenetic inference and the multispecies coalescent. Trends in Ecology & Evolution 24:332–340.
Dell'Ampio, E., K. Meusemann, N. U. Szucsich, R. S. Peters, B. Meyer, J. Borner, M. Petersen, A. J. Aberer, A. Stamatakis, M. G. Walzl, B. Q. Minh, A. von Haeseler, I. Ebersberger, G. Pass, and B. Misof. 2014. Decisive data sets in phylogenomics: Lessons from studies on the phylogenetic relationships of primarily wingless insects. Molecular Biology and Evolution 31:239–249.
Delsuc, F., G. Tsagkogeorga, N. Lartillot, and H. Philippe. 2008. Additional molecular support for the new chordate phylogeny. genesis 46:592–604.
Díaz-Arce, N., H. Arrizabalaga, H. Murua, X. Irigoien, and N. Rodríguez-Ezpeleta. 2016. RAD-seq derived genome-wide nuclear markers resolve the phylogeny of tunas. Molecular Phylogenetics and Evolution 102:202–207.
Diepeveen, E. T., and W. Salzburger. 2011. Molecular characterization of two endothelin

pathways in East African cichlid fishes. Journal of Molecular Evolution 73:355–368.
Dikow, R. B., and W. L. Smith. 2013. Complete genome sequences provide a case study for the evaluation of gene-tree thinking. Cladistics 29:672–682.
Donaire-Barroso, D., and S. Bogaerts. 2001. Observations on viviparity of Salamandra algira in North Morocco. Pp. 147–151 *in* P. Lymberakis, E. Valakos, P. Pafilis, and M. Mylonas, eds. Herpetologia Candiana. Societas Europaea Herpetologica, Irakleio.
Donaire, D., and S. Bogaerts. 2003. A new sub-species of *Salamandra algira* Bedriaga, 1883 from northern Morocco. Podarcis 4:84–100.
Doucet, S. M., and D. J. Mennill. 2010. Dynamic sexual dichromatism in an explosively breeding Neotropical toad. Biology Letters 6:63–66.
Doucet, S. M., M. D. Shawkey, M. K. Rathburn, H. L. Mays, and R. Montgomerie. 2004. Concordant evolution of plumage colour, feather microstructure and a melanocortin receptor gene between mainland and island populations of a fairy-wren. Proceedings of the Royal Society B: Biological Sciences 271:1663–1670.
Drake, C. A., K. L. Magerowski, L. E. Rupprecht, M. D. Wagner, and S. C. Waltos. 2009. An allometric analysis of *Ambystoma maculatum* spot size against body length and weight. Journal of Ecological Research 11:31–35.
Dray, S., and A.-B. Dufour. 2007. The ade4 Package: Implementing the duality diagram for ecologists. JSS Journal of Statistical Software 22.
Dreher, C. E., M. E. Cummings, and H. Pröhl. 2015. An analysis of predator selection to affect aposematic coloration in a poison frog species. PLOS ONE 10:e0130571.
Dreher, C. E., A. Rodríguez, M. E. Cummings, and H. Pröhl. 2017. Mating status correlates with dorsal brightness in some but not all poison frog populations. Ecology and Evolution, doi: 10.1002/ece3.3531.
Drummond, A. J., and A. Rambaut. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evolutionary Biology 7:214.
Dubois, A., and J. Raffaëlli. 2009. A new ergotaxonomy of the family Salamandridae Goldfuss, 1820 (Amphibia, Urodela). Alytes 26:1–85.
Duellman, W. E., and L. Trueb. 1994. Biology of amphibians. The John Hopkins University Press, London.
Dufresnes, C., and N. Perrin. 2015. Effect of biogeographic history on population vulnerability in European amphibians. Convervation Biology 29:1235–1241.
Duggen, S., K. Hoernle, P. van den Bogaard, L. Rüpke, and J. Phipps Morgan. 2003. Deep roots of the Messinian salinity crisis. Nature 422:602–606.
Earl, D. A., and B. M. VonHoldt. 2012. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conservation Genetics Resources 4:359–361.
Eaton, D. A. R. 2014. PyRAD: assembly of de novo RADseq loci for phylogenetic analyses. Bioinformatics 30:1844–1849.
Eaton, D. A. R., and R. H. Ree. 2013. Inferring Phylogeny and Introgression using RADseq Data: An Example from Flowering Plants (Pedicularis: Orobanchaceae). Systematic Biology 62:689–706.
Eaton, D. A. R., E. L. Spriggs, B. Park, and M. J. Donoghue. 2017. Misconceptions on missing data in RAD-seq phylogenetics with a deep-scale example from flowering plants. Systematic Biology 66:399–412.
Ebanks, J. P., A. Koshoffer, R. R. Wickett, S. Schwemberger, G. Babcock, T. Hakozaki, and R. E. Boissy. 2011. Epidermal keratinocytes from light vs. dark skin exhibit differential degradation of melanosomes. Journal of Investigative Dermatology 131:1226–1233.
Edgar, R. C. 2010. Search and clustering orders of magnitude faster than BLAST. Bioinformatics 26:2460–2461.
Edwards, S. V. 2009. Is a new and general theory of molecular systematics emerging? Evolution 63:1–19.

Eiselt, J. 1958. Der feuersalamander, *Salamandra salamandra* (L) Beitrage zu einer taxonomischen Synthese. Abhandlungen und Berichte fur Naturkunde und Vorgeschichte, Magdeburg 10:77–154.

Eizirik, E., N. Yuhki, W. E. Johnson, M. Menotti-Raymond, S. S. Hannah, and S. J. O'Brien. 2003. Molecular genetics and evolution of melanism in the cat family. Current Biology 13:448–453.

Elewa, A., H. Wang, C. Talavera-López, A. Joven, G. Brito, A. Kumar, L. S. Hameed, M. Penrad-Mobayed, Z. Yao, N. Zamani, Y. Abbas, I. Abdullayev, R. Sandberg, M. Grabherr, B. Andersson, and A. Simon. 2017. Reading and editing the *Pleurodeles waltl* genome reveals novel features of tetrapod regeneration. Nature Communications 8:2286.

Elmer, K. R., and A. Meyer. 2011. Adaptation in the age of ecological genomics: insights from parallelism and convergence. Trends in Ecology and Evolution 26:298–306.

Emerson, K. J., C. R. Merz, J. M. Catchen, P. A. Hohenlohe, W. A. Cresko, W. E. Bradshaw, and C. M. Holzapfel. 2010. Resolving postglacial phylogeography using high-throughput sequencing. Proceedings of the National Academy of Sciences of the United States of America 107:16196–200.

Emerson, S. B., T. A. Cooper, and J. R. Ehleringer. 1990. Convergence in reflectance spectra among tree frogs. Functional Ecology 4:47–51.

Endler, J. A. 1980. Natural selection on color patterns in *Poecilia reticulata*. Evolution 34:76–91.

Endler, J. A. 1990. On the measurement and classification of color in studies of animal coloration. Biological Journal of the Linnean Society 41:315–352.

Endler, J. A., D. A. Westcott, J. R. Madden, and T. Robson. 2005. Animal visual systems and the evolution of color patterns: sensory processing illuminates signal evolution. Evolution; international journal of organic evolution 59:1795–818.

Escoriza, D. 2016. *Salamandra algira spelaea* (Beni Snassen fire salamander); new distributional records. The Herpetological Bulletin 136:40–41.

Escoriza, D., and M. Del Mar Comas. 2007. Description of a new subspecies of *Salamandra algira* Bedriaga, 1883 (Amphibia: Salamandridae) from the Beni Snassen massif (Northeast Morocco). Salamandra 43:77–90.

Evanno, G., S. Regnaut, and J. Goudet. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. Molecular Ecology 14:2611–2620.

Evans, T., A. D. Johnson, and M. Loose. 2018. Virtual Genome Walking across the 32 Gb *Ambystoma mexicanum* genome; assembling gene models and intronic sequence. Scientific Reports 8:618.

Excoffier, L., I. Dupanloup, E. Huerta-Sánchez, V. C. Sousa, and M. Foll. 2013. Robust demographic inference from genomic and SNP data. PLOS Genetics 9:e1003905.

Farnesi, L. C., H. C. M. Vargas, D. Valle, and G. L. Rezende. 2017. Darker eggs of mosquitoes resist more to dry conditions: Melanin enhances serosal cuticle contribution in egg resistance to desiccation in *Aedes*, *Anopheles* and *Culex* vectors. PLOS Neglected Tropical Diseases 11:e0006063.

Feldmann, R. 1987. Überwinterung, Ortstreue und Lebensalter des Feuersalamanders, *Salamandra salamandra terrestris*. Schlussbericht einer Langzeituntersuchung. Jahrbuch für Feldherpetologie 1:33–44.

Ferrie, G. M., V. C. Alford, J. Atkinson, E. Baitchman, D. Barber, W. S. Blaner, G. Crawshaw, A. Daneault, E. Dierenfeld, M. Finke, G. Fleming, R. Gagliardo, E. A. Hoffman, W. Karasov, K. Klasing, E. Koutsos, J. Lankton, S. R. Lavin, A. Lentini, S. Livingston, B. Lock, T. Mason, A. McComb, C. Morris, A. P. Pessier, F. Olea-Popelka, T. Probst, C. Rodriguez, K. Schad, K. Semmen, J. Sincage, M. A. Stamper, J. Steinmetz, K. Sullivan, S. Terrell, N. Wertan, C. J. Wheaton, B. Wilson, and E. V Valdes. 2014. Nutrition and health in amphibian husbandry. Zoo biology 33:485–501.

Figuet, E., M. Ballenghien, J. Romiguier, and N. Galtier. 2014. Biased gene conversion and GC-content evolution in the coding sequences of reptiles and vertebrates. Genome biology and evolution 7:240–50.
Fisher-Reid, M. C., T. N. Engstrom, C. A. Kuczynski, P. R. Stephens, and J. J. Wiens. 2013. Parapatric divergence of sympatric morphs in a salamander: incipient speciation on Long Island? Molecular Ecology 22:4681–4694.
Fitzpatrick, B. M., K. Shook, and R. Izally. 2009. Frequency-dependent selection by wild birds promotes polymorphism in model salamanders. BMC Ecology 9:12.
Frichot, E., and O. François. 2015. LEA : An R package for landscape and ecological association studies. Methods in Ecology and Evolution 6:925–929.
Frichot, E., S. D. Schoville, G. Bouchard, and O. François. 2013. Testing for Associations between Loci and Environmental Gradients Using Latent Factor Mixed Models. Molecular Biology and Evolution 30:1687–1699. Princeton University Press, Princeton (NJ).
Frost, D. R. 2017. Amphibian Species of the World: an Online Reference. Version 6.0.
Frost, D. R., T. Grant, J. Faivovich, R. H. Bain, A. Haas, C. F. B. Haddad, R. O. De Sá, A. Channing, M. Wilkinson, S. C. Donnellan, C. J. Raxworthy, J. A. Campbell, B. L. Blotto, P. Moler, R. C. Drewes, R. A. Nussbaum, J. D. Lynch, D. M. Green, and W. C. Wheeler. 2006. The amphibian tree of life. Bulletin of the AMNH 297:1–370.
Frost, S. K., and S. J. Robinson. 1984. Pigment cell differentiation in the Fire-Bellied Toad, *Bombina orientalis*. I: structural, chemical, and physical aspect of the adult pigment pattern. Journal of Morphology 179:229–242.
Fukuzawa, T., and J. T. Bagnara. 1989. Control of melanoblast differentiation in amphibia by alpha-melanocyte stimulating hormone, a serum melanization factor, and a melanization inhibiting factor. Pigment Cell Research 2:171–181.
Fukuzawa, T., and H. Ide. 1988. A ventrally localized inhibitor of melanization in *Xenopus laevis* skin. Developmental Biology 129:25–36.
Fukuzawa, T., P. Samaraweera, F. T. Mangano, J. H. Law, and J. T. Bagnara. 1995. Evidence that *MIF* plays a role in the development of pigmentation patterns in the frog. Developmental Biology 167:148–158.
García-París, A. M., M. Alcobendas, D. Buckley, and D. 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 57:129–143.
Garcia, T. S., J. Stacy, and A. Sih. 2004. Larval salamander response to UV radiation and predation risk: color change and microhabitat use. Ecological Applications 14:1055–1064.
Gasser, F. 1978. Le polytypisme de léspéce paléarctique Salamandra salamandra (L.) (Amphibien, Urodéle). II. Systématique, relations génétiques et tendences évolutives dans l'aire de répartition. Archives de zoologie expérimentale et générale 119:635–668.
Gatesy, J., and M. S. Springer. 2014. Phylogenetic analysis at deep timescales: Unreliable gene trees, bypassed hidden support, and the coalescence/concatalescence conundrum. Molecular Phylogenetics and Evolution 80:231–266.
Geneva, A. J., C. A. Muirhead, S. B. Kingan, and D. Garrigan. 2015. A new method to scan genomes for introgression in a secondary contact model. PLOS ONE 10:e0118621.
Gomez, D., C. Richardson, T. Lengagne, S. Plenet, P. Joly, J.-P. Léna, and M. Théry. 2009. The role of nocturnal vision in mate choice: females prefer conspicuous males in the European tree frog (*Hyla arborea*). Proceedings of the Royal Society of London B: Biological Sciences 276:2351–8.
Gordon, A., and G. J. Hannon. 2008. Fastx-toolkit: fastx_barcode_splitter.pl.
Gordon, S. P., A. Es Opez-Sepulcre, and D. N. Reznick. 2012. Predation-associated

differences in the sex linkage of guppy male coloration. Evolution 66:912–918.
Grabherr, M. G., B. J. Haas, M. Yassour, J. Z. Levin, D. A. Thompson, I. Amit, X. Adiconis, L. Fan, R. Raychowdhury, Q. Zeng, Z. Chen, E. Mauceli, N. Hacohen, A. Gnirke, N. Rhind, F. di Palma, B. W. Birren, C. Nusbaum, K. Lindblad-Toh, N. Friedman, and A. Regev. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotech 29:644–652.
Graham, L., and J. M. Orenstein. 2007. Processing tissue and cells for transmission electron microscopy in diagnostic pathology and research. Nature protocols 2:2439–50.
Grant, A. H., and E. B. Liebgold. 2017. Color-Biased Dispersal Inferred by Fine-Scale Genetic Spatial Autocorrelation in a Color Polymorphic Salamander. Journal of Heredity 108:588–593.
Gregory, T. R. 2016. Animal genome size database. http://www.genomesize.com (accessed:December 2016).
Gregory, T. R. 2017. Animal Genome Size Database. http://www.genomesize.com (accessed:December 2017).
Grether, G. F., G. R. Kolluru, and K. Nersissian. 2004. Individual colour patches as multicomponent signals. Biological reviews of the Cambridge Philosophical Society 79:583–610.
Greven, H. 2003a. Larviparity and pueriparity. Pp. 447–475 *in* D. M. Sever, ed. Reproductive Biology and Phylogeny, Reproductive Biology and Phylogeny of Urodela. Science Publisher Inc, Enfield USA.
Greven, H. 2003b. Larviparity and pueriparity. Pp. 447–475 *in* D. M. Sever, ed. Reproductive Biology and Phylogeny, Reproductive Biology and Phylogeny of Urodela. Science Publisher Inc, Enfield USA.
Greven, H. 1998. Survey of the oviduct of salamandrids with special reference to the viviparous species. Journal of Experimental Zoology 282:507–525.
Greven, H., and G. D. Guex. 1994. Structural and physiological aspects of viviparity in *Salamandra salamandra*. Mertensiella 4:139–160.
Gross, J. B., R. Borowsky, and C. J. Tabin. 2009. A novel role for Mc1r in the parallel evolution of depigmentation in independent populations of the cavefish *Astyanax mexicanus*. PLoS genetics 5:e1000326.
Grossenbacher, K. 1997. Salamandra atra. Pp. 68–69 *in* J.-P. Gasc, ed. Atlas of Amphibians and Reptiles in Europe. Muséum national d'histoire naturelle, Paris.
Grossenbacher, K. 1994. Zur Systematik und Verbreitung der Alpensalamander (*Salamandra atra atra*, *Salamandra atra aurorae*, *Salamandra lanzai*). Abhandlungen und Berichte für Naturkunde 17:75–81.
Guerrero, F., V. Pérez-Mellado, M. Gil, and M. Lizana. 1990. Food habits and trophic availability in the high mountain population of the Spotted Salamander from Spain (*Salamandra salamandra almanzoris*)(Caudata: Salamandridae). Folia Zoologica 39:341–353.
Guindon, S., O. Gascuel, B. Rannala, A. Meyer, Y. Chen, M. Clamp, L. Clarke, G. Coates, T. Cox, F. Cunningham, and K. Okubo. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Systematic Biology 52:696–704. BioMed Central.
Haas, B. J., A. Papanicolaou, M. Yassour, M. Grabherr, P. D. Blood, J. Bowden, M. B. Couger, D. Eccles, B. Li, M. Lieber, M. D. MacManes, M. Ott, J. Orvis, N. Pochet, F. Strozzi, N. Weeks, R. Westerman, T. William, C. N. Dewey, R. Henschel, R. D. LeDuc, N. Friedman, and A. Regev. 2013. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nature Protocols 8:1494–1512.
Habermehl, G., and G. Spiteller. 1967. Massenspektren der Salamander-Alkaloide. Justus Liebigs Annalen der Chemie 706:213–222.

Haddad, C. F. B., and C. P. A. Prado. 2005. Reproductive modes in frogs and their unexpected diversity in the Atlantic forest of Brazil. BioScience 55:207–217.
Hahn, C., L. Bachmann, and B. Chevreux. 2013. Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads—a baiting and iterative mapping approach. Nucleic Acids Research 41:e129–e129.
Hall, B., A. Limaye, and A. B. Kulkarni. 2009. Overview: generation of gene knockout mice. Current Protocols in Cell Biology Chapter 19:Unit 19.1217.
Hammond, S. A., R. L. Warren, B. P. Vandervalk, E. Kucuk, H. Khan, E. A. Gibb, P. Pandoh, H. Kirk, Y. Zhao, M. Jones, A. J. Mungall, R. Coope, S. Pleasance, R. A. Moore, R. A. Holt, J. M. Round, S. Ohora, B. V Walle, N. Veldhoen, C. C. Helbing, and I. Birol. 2017. The North American bullfrog draft genome provides insight into hormonal regulation of long noncoding RNA. Nature Communications 8:1433.
Harkey, G. A., and R. D. Semlitsch. 1988. Effects of temperature on growth, development, and color polymorphism in the ornate chorus frog *Pseudacris ornata*. Copeia 4:1001–1007.
Hawkes, J. W. 1974. The structure of fish skin. Cell and Tissue Research 149:159–172.
Heinen, J. T. 1986. The significance of color change in newly metamorphosed American toads (*Bufo a. americanus*). Journal of Herpetology 28:87–93.
Heled, J., and A. J. Drummond. 2010. Bayesian inference of species trees from multilocus data. Molecular Biology and Evolution 27:570–580.
Hellsten, U., R. M. Harland, M. J. Gilchrist, D. Hendrix, J. Jurka, V. Kapitonov, I. Ovcharenko, N. H. Putnam, S. Shu, L. Taher, I. L. Blitz, B. Blumberg, D. S. Dichmann, I. Dubchak, E. Amaya, J. C. Detter, R. Fletcher, D. S. Gerhard, D. Goodstein, T. Graves, I. V Grigoriev, J. Grimwood, T. Kawashima, E. Lindquist, S. M. Lucas, P. E. Mead, T. Mitros, H. Ogino, Y. Ohta, A. V Poliakov, N. Pollet, J. Robert, A. Salamov, A. K. Sater, J. Schmutz, A. Terry, P. D. Vize, W. C. Warren, D. Wells, A. Wills, R. K. Wilson, L. B. Zimmerman, A. M. Zorn, R. Grainger, T. Grammer, M. K. Khokha, P. M. Richardson, and D. S. Rokhsar. 2010. The genome of the Western clawed frog *Xenopus tropicalis*. Science 328:633–636.
Henning, F., J. C. Jones, P. Franchini, and A. Meyer. 2013. Transcriptomics of morphological color change in polychromatic Midas cichlids. BMC Genomics 14:171.
Hensel, J. L., and E. D. Brodie. 1976. An experimental study of aposematic coloration in the salamander *Plethodon jordani*. Copeia 1976:59–65.
Herczeg, G., C. Matsuba, and J. Merilä. 2010. Sequence variation in the melanocortin-1 receptor gene (*Mc1r*) does not explain variation in the degree of melanism in a widespread amphibian. Annales Zoologici Fennici 47:37–45.
Higdon, C. W., R. D. Mitra, and S. L. Johnson. 2013. Gene expression analysis of zebrafish melanocytes, iridophores, and retinal pigmented epithelium reveals indicators of biological function and developmental origin. PloS one 8:e67801.
Hill, G. E. 2016. Mitonuclear coevolution as the genesis of speciation and the mitochondrial DNA barcode gap. Ecology and Evolution 6:5831–5842.
Hirobe, T., K. Wakamatsu, and S. Ito. 2007. Excess Tyrosine Stimulates Eumelanin and Pheomelanin Synthesis in Cultured Slaty Melanocytes from Neonatal Mouse Epidermis. Zoological Science 24:209–217.
Hoekstra, H. E. 2006. Genetics, development and evolution of adaptive pigmentation in vertebrates. Heredity 97:222–234.
Hoekstra, H. E., K. E. Drumm, and M. W. Nachman. 2004. Ecological genetics of adaptive color polymorphism in pocket mice: geographic variation in selected and neutral genes. Evolution 58:1329–1341.
Hoffman, E. A., and M. S. Blouin. 2000. A review of colour and pattern polymorphisms in anurans. Biological Journal of the Linnean Society 70:633–665.
Honkavaara, J., M. Koivula, E. Korpimaki, H. Siitari, and J. Viitala. 2002. Ultraviolet

vision and foraging in terrestrial vertebrates. Oikos 98:505–511.
Hou, Y., M. D. Nowak, V. Mirré, C. S. Bjorå, C. Brochmann, and M. Popp. 2015. Thousands of RAD-seq Loci Fully Resolve the Phylogeny of the Highly Disjunct Arctic-Alpine Genus Diapensia (Diapensiaceae). PLOS ONE 10:e0140175.
Houde, A. E. 1997. Sex, color, and mate choice in guppies. Princeton University Press, Princeton (NJ).
Huang, H., and L. L. Knowles. 2014. Unforeseen consequences of excluding missing data from next-generation sequences: Simulation study of RAD sequences. Systematic Biology 65:357–365.
Hubbard, J. K., J. A. C. Uy, M. E. Hauber, H. E. Hoekstra, and R. J. Safran. 2010. Vertebrate pigmentation: from underlying genes to adaptive function. Trends in Genetics 26:231–239.
Hurlbert, S. H. 1970. Predator responses to the vermilion-spotted newt (*Notophdhabnus viridescens*). Journal of Herpetology 4:47–55.
Ide, H., and T. Hama. 1972. Guanine formation in isolated iridophores from bullfrog tadpoles. Biochimica et Biophysica Acta (BBA) - General Subjects 286:269–271.
Irisarri, I., D. Baurain, H. Brinkmann, F. Delsuc, J.-Y. Sire, A. Kupfer, J. Petersen, M. Jarek, A. Meyer, M. Vences, and H. Philippe. 2017. Phylotranscriptomic consolidation of the jawed vertebrate timetree. Nature Ecology and Evolution 1:1370–1378.
Irisarri, I., and A. Meyer. 2016. The identification of the closest living relative(s) of tetrapods: Phylogenomic lessons for resolving short ancient internodes. Systematic Biology 65:1057–1075.
Ito, S., and K. Wakamatsu. 2011. Diversity of human hair pigmentation as studied by chemical analysis of eumelanin and pheomelanin. Journal of the European Academy of Dermatology and Venereology 25:1369–1380.
IUCN. 2008. An Analysis of Amphibians on the 2008 IUCN Red List.
Jablonski, N. G. 1998. Ultraviolet light-induced neural tube defects in amphibian larvae and their implications for the evolution of melanized pigmentation and declines in amphibian populations. Journal of Herpetology 32:455–457.
Jacobs, G. H. 1992. Ultraviolet Vision in Vertebrates. Oxford University Press.
Jacobson, S. K., and J. J. Vandenberg. 1991. Reproductive ecology of the endangered golden toad (*Bufo periglenes*). Journal of Herpetology 25:321–327.
Jarvis, E. D., S. Mirarab, A. J. Aberer, B. Li, P. Houde, C. Li, S. Y. W. Ho, B. C. Faircloth, B. Nabholz, J. T. Howard, A. Suh, C. C. Weber, R. R. da Fonseca, J. Li, F. Zhang, H. Li, L. Zhou, N. Narula, L. Liu, G. Ganapathy, B. Boussau, M. S. Bayzid, V. Zavidovych, S. Subramanian, T. Gabaldón, S. Capella-Gutiérrez, J. Huerta-Cepas, B. Rekepalli, K. Munch, M. Schierup, B. Lindow, W. C. Warren, D. Ray, R. E. Green, M. W. Bruford, X. Zhan, A. Dixon, S. Li, N. Li, Y. Huang, E. P. Derryberry, M. F. Bertelsen, F. H. Sheldon, R. T. Brumfield, C. V. Mello, P. V. Lovell, M. Wirthlin, M. P. C. Schneider, F. Prosdocimi, J. A. Samaniego, A. M. V. Velazquez, A. Alfaro-Núñez, P. F. Campos, B. Petersen, T. Sicheritz-Ponten, A. Pas, T. Bailey, P. Scofield, M. Bunce, D. M. Lambert, Q. Zhou, P. Perelman, A. C. Driskell, B. Shapiro, Z. Xiong, Y. Zeng, S. Liu, Z. Li, B. Liu, K. Wu, J. Xiao, X. Yinqi, Q. Zheng, Y. Zhang, H. Yang, J. Wang, L. Smeds, F. E. Rheindt, M. Braun, J. Fjeldsa, L. Orlando, F. K. Barker, K. A. Jønsson, W. Johnson, K.-P. Koepfli, S. O'Brien, D. Haussler, O. A. Ryder, C. Rahbek, E. Willerslev, G. R. Graves, T. C. Glenn, J. McCormack, D. Burt, H. Ellegren, P. Alström, S. V. Edwards, A. Stamatakis, D. P. Mindell, J. Cracraft, E. L. Braun, T. Warnow, W. Jun, M. T. P. Gilbert, and G. Zhang. 2014. Whole-genome analyses resolve early branches in the tree of life of modern birds. Science 346:1320–1331.
Joger, U., and S. Steinfartz. 1994. Zur subspezifischen Gliederung der su¨diberischen Feuersalamander (*Salamandra salamandra*, Komplex). Abhandlungen und Berichte

fur Naturkunde und Vorgeschichte, Magdeburg 17:83–98.
Johnson, J. A., and E. D. Brodie. 1975. The selective advantage of the defensive posture of the newt, *Taricha granulosa*. The American Midland Naturalist 93:139–148.
Jombart, T. 2008. adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics 24:1403–1405.
Jombart, T., and I. Ahmed. 2011. adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. Bioinformatics 27:3070–3071.
Kaelin, C. B., X. Xu, L. Z. Hong, V. A. David, K. A. McGowan, A. Schmidt-Küntzel, M. E. Roelke, J. Pino, J. Pontius, G. M. Cooper, H. Manuel, W. F. Swanson, L. Marker, C. K. Harper, A. van Dyk, B. Yue, J. C. Mullikin, W. C. Warren, E. Eizirik, L. Kos, S. J. O'Brien, G. S. Barsh, and M. Menotti-Raymond. 2012. Specifying and sustaining pigmentation patterns in domestic and wild cats. Science 337:1536–41.
Katoh, K., and D. M. Standley. 2013. MAFFT Multiple Sequence Alignment Software version 7: improvements in performance and usability. Molecular Biology and Evolution 30:772–780.
Kats, L. B., and R. G. Vandragt. 1986. Background color-matching in the spring peeper, *Hyla crucifer*. Copeia 1986:109–115.
Keinath, M. C., V. A. Timoshevskiy, N. Y. Timoshevskaya, P. A. Tsonis, S. R. Voss, and J. J. Smith. 2015. Initial characterization of the large genome of the salamander *Ambystoma mexicanum* using shotgun and laser capture chromosome sequencing. Scientific Reports 5:16413.
Kelber, A., M. Vorobyev, and D. Osorio. 2003. Animal colour vision – behavioural tests and physiological concepts. Biological Reviews of the Cambridge Philosophical Society 78:S1464793102005985.
Kelsh, R. N. 2004. Genetics and Evolution of Pigment Patterns in Fish. Pigment Cell Research 17:326–336.
Kijas, J. M., R. Wales, A. Törnsten, P. Chardon, M. Moller, and L. Andersson. 1998. Melanocortin receptor 1 (*MC1R*) mutations and coat color in pigs. Genetics 150:1177–1185.
Kim, D., G. Pertea, C. Trapnell, H. Pimentel, R. Kelley, and S. L. Salzberg. 2013. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biology 14:R36.
Kimura, T., Y. Nagao, H. Hashimoto, Y. Yamamoto-Shiraishi, S. Yamamoto, T. Yabe, S. Takada, M. Kinoshita, A. Kuroiwa, and K. Naruse. 2014. Leucophores are similar to xanthophores in their specification and differentiation processes in medaka. Proceedings of the National Academy of Sciences of the United States of America 111:7343–8.
King, R. B., S. Hauff, and J. . Phillips. 1994. Physiological color change in the green treefrog: responses to background brightness and temperature. Copeia 1994:422–432.
Kodric-Brown, A., and S. C. Johnson. 2002. Ultraviolet reflectance patterns of male guppies enhance their attractiveness to females. Animal Behaviour 63:391–396.
Köhler, G., and S. Steinfartz. 2006. A new subspecies of the fire salamander, *Salamandra salamandra* (Linnaeus, 1758) from the Tendi valley, Asturias, Spain. Salamandra 42:13–20.
Kraemer, A. C., J. Kissner, and D. C. Adams. 2012. Morphological color-change in the red-backed salamander (*Plethodon cinereus*) while kept in captivity. Copeia 2012:748–755.
Krauss, J., P. Astrinidis, P. Astrinides, H. G. Frohnhöfer, B. Walderich, and C. Nüsslein-Volhard. 2013. transparent, a gene affecting stripe formation in Zebrafish, encodes the mitochondrial protein *Mpv17* that is required for iridophore survival. Biology open 2:703–10.
Kronforst, M. R., G. S. Barsh, A. Kopp, J. Mallet, A. Monteiro, S. P. Mullen, M. Protas, E. B. Rosenblum, C. J. Schneider, H. E. Hoekstra, and A. Nia Monteiro. 2012.

Unraveling the thread of nature's tapestry: the genetics of diversity and convergence in animal pigmentation. Pigment Cell and Melanoma Research 25:411–433.
Kronforst, M. R., and R. Papa. 2015. The functional basis of wing patterning in Heliconius butterflies: the molecules behind mimicry. Genetics 200:1–19.
Kubatko, L. S., J. H. Degnan, and T. Collins. 2007. Inconsistency of phylogenetic estimates from concatenated data under coalescence. Systematic Biology 56:17–24.
Kuchta, S. R., A. H. Krakauer, and B. Sinervo. 2008. Why does the yellow-eyed Ensatina have yellow eyes? Batesian mimicry of Pacific newts (genus *Taricha*) by the salamander *Ensatina eschscholtzii xanthoptica*. Evolution; international journal of organic evolution 62:984–90.
Kumar, S., G. Stecher, and K. Tamura. 2016. MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. Molecular Biology and Evolution 33:1870–1874.
Kumazawa, T., and R. Fujii. 1984. Concurrent releases of norepinephrine and purines by potassium from adrenergic melanosome-aggregating nerve in tilapia. Comparative Biochemistry and Physiology Part C: Comparative Pharmacology 78:263–266.
Kwon, T. 2017. AmphiBase: A new genomic resource for non-model amphibian species. genesis 55:e23010.
Lamason, R. L., M.-A. P. K. Mohideen, J. R. Mest, A. C. Wong, H. L. Norton, M. C. Aros, M. J. Jurynec, X. Mao, V. R. Humphreville, J. E. Humbert, S. Sinha, J. L. Moore, P. Jagadeeswaran, W. Zhao, G. Ning, I. Makalowska, P. M. McKeigue, D. O'donnell, R. Kittles, E. J. Parra, N. J. Mangini, D. J. Grunwald, M. D. Shriver, V. A. Canfield, and K. C. Cheng. 2005. SLC24A5, a putative cation exchanger, affects pigmentation in zebrafish and humans. Science (New York, N.Y.) 310:1782–6.
Lambert, S. M., T. W. Reeder, and J. J. Wiens. 2015. When do species-tree and concatenated estimates disagree? An empirical analysis with higher-level scincid lizard phylogeny. Molecular Phylogenetics and Evolution 82, Part A:146–155.
Lanfear, R., B. Calcott, S. Y. W. Ho, and S. Guindon. 2012. PartitionFinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. Molecular Biology and Evolution 29:1695–1701.
Langhelle, A., M. J. Lindell, and P. Nyström. 1999. Effects of ultraviolet radiation on amphibian embryonic and larval development. Journal of Herpetology 33:449–456.
Laporte, M., S. a. Pavey, C. Rougeux, F. Pierron, M. Lauzent, H. Budzinski, P. Labadie, E. Geneste, P. Couture, M. Baudrimont, and L. Bernatchez. 2016. RAD sequencing reveals within-generation polygenic selection in response to anthropogenic organic and metal contamination in North Atlantic Eels. Molecular Ecology 25:219–237.
Le, S. Q., and O. Gascuel. 2008. An improved general amino acid replacement matrix. Molecular Biology and Evolution 25:1307–1320.
Leaché, A. D., B. L. Banbury, J. Felsenstein, A. N. M. De Oca, and A. Stamatakis. 2015a. Short tree, long tree, right tree, wrong tree: New acquisition bias corrections for inferring SNP phylogenies. Systematic Biology 64:1032–1047.
Leaché, A. D., A. S. Chavez, L. N. Jones, J. A. Grummer, A. D. Gottscho, and C. W. Linkem. 2015b. Phylogenomics of Phrynosomatid lizards: Conflicting signals from sequence capture versus restriction site associated DNA sequencing. Genome Biology and Evolution 7:706–719.
Leaché, A. D., and B. Rannala. 2011. The accuracy of species tree estimation under simulation: A comparison of methods. Systematic Biology 60:126–137.
Lemmon, A. R., S. A. Emme, and E. M. Lemmon. 2012. Anchored hybrid enrichment for massively high-throughput phylogenomics. Systematic Biology 61:727–744.
Lenzi-Mattos, R., M. M. Antoniazzi, C. F. B. Haddad, D. V. Tambourgi, M. T. Rodrigues, and C. Jared. 2005. The inguinal macroglands of the frog *Physalaemus nattereri* (Leptodactylidae): structure, toxic secretion and relationship with deimatic behaviour. Journal of Zoology 266:385–394.

Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin, and 1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25:2078–2079.
Li, L., C. J. Stoeckert, and D. S. Roos. 2003. OrthoMCL: identification of ortholog groups for eukaryotic genomes. Genome research 13:2178–89.
Liaw, A., and M. Wiener. 2002. Classification and Regression by randomForest. R News 2:18–22.
Liebau, A. 2013. Visual spectral sensitivity and adaptation in nocturnal treefrogs (Dr. rer. Nat. thesis). Center for Systems Neuroscience, Hannover.
Lind, O., M. Mitkus, P. Olsson, and A. Kelber. 2013. Ultraviolet vision in birds: the importance of transparent eye media. Proceedings of the Royal Society of London B: Biological Sciences 281.
Lischer, H. E. L., and L. Excoffier. 2012. PGDSpider: an automated data conversion tool for connecting population genetics and genomics programs. Bioinformatics 28:298–299.
Lister, J. A., C. P. Robertson, T. Lepage, S. L. Johnson, and D. W. Raible. 1999. *nacre* encodes a zebrafish microphthalmia-related protein that regulates neural-crest-derived pigment cell fate. Development 126:3757–3767.
Liu, L., L. Yu, L. Kubatko, D. K. Pearl, and S. V. Edwards. 2009. Coalescent methods for estimating phylogenetic trees. Molecular Phylogenetics and Evolution 53:320–328.
Lopes, R. J., J. D. Johnson, M. B. Toomey, M. S. Ferreira, P. M. Araujo, J. Melo-Ferreira, L. Andersson, G. E. Hill, J. C. Corbo, and M. Carneiro. 2016. Genetic basis for red coloration in birds. Current Biology 26:1427–1434.
Losos, J. B. 2011. Convergence, adaptation, and constraint. Evolution 65:1827–1840.
Love, M. I., W. Huber, and S. Anders. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology 15:550.
Lowe, T. M., and P. P. Chan. 2016. tRNAscan-SE On-line: integrating search and context for analysis of transfer RNA genes. Nucleic Acids Research 44:W54–W57.
Lyerla, T. A., and D. L. Jameson. 1968. Development of color in chimeras of Pacific tree frogs. Copeia 1968:113–128.
Maan, M. E., and M. E. Cummings. 2012. Poison frog colors are honest signals of toxicity, particularly for bird predators. The American naturalist 179:E1-14.
Maan, M. E., and K. M. Sefc. 2013. Colour variation in cichlid fish: developmental mechanisms, selective pressures and evolutionary consequences. Seminars in cell & developmental biology 24:516–28.
Macedonia, J. M., S. James, L. W. Wittle, and D. L. Clark. 2000. Skin Pigments and Coloration in the Jamaican Radiation of Anolis Lizards. Journal of Herpetology 34:99–109.
Machado, D. J., M. L. Lyra, and T. Grant. 2016. Mitogenome assembly from genomic multiplex libraries: comparison of strategies and novel mitogenomes for five species of frogs. Molecular Ecology Resources 16:686–693.
Madison, D. M. 1998. Habitat-contingent reproductive behaviour in radio-implanted salamanders: a model and test. Animal Behaviour 55:1203–1210.
Madison, D. M. 1997. The Emigration of Radio-Implanted Spotted Salamanders, *Ambystoma maculatum*. Journal of Herpetology 31:542–551.
Madison, D. M., and L. Farrand. 1998. Habitat use during breeding and emigration in radio-implanted tiger salamanders, Ambystoma tigrinum. Copeia 1998:402–410.
Maia, R., C. M. Eliason, P.-P. Bitton, S. M. Doucet, and M. D. Shawkey. 2013. pavo : an R package for the analysis, visualization and organization of spectral data. Methods in Ecology and Evolution 4:906–913.
Mallarino, R., C. Henegar, M. Mirasierra, M. Manceau, C. Schradin, M. Vallejo, S. Beronja, G. S. Barsh, and H. E. Hoekstra. 2016. Developmental mechanisms of stripe patterns in rodents. Nature 539:518–523.

Mantione, K., C. Kim, F. Casares, and G. Stefano. 2012. Microarray validation of molecular and cellular signaling in *Homarus americanus* and *Penaeus monodon*. ISJ 9:212–222.
Marin, Y. E., and C. B. Lin. 2008. CS12-2 Expression and regulation of Aldo-Keto Reductase (AKR) 1C subfamily genes in skin by UV Abstracts of the Conjoint Meeting of XXth International Pigment Cell Conference & Vth International Melanoma Research Congress. Abstracts of the Conjoint Meeting of XXth International Pigment Cell Conference & Vth International Melanoma Research Congress 21:243–336.
Martin, S. H., J. W. Davey, and C. D. Jiggins. 2015. Evaluating the Use of ABBA–BABA Statistics to Locate Introgressed Loci. Molecular Biology and Evolution 32:244–257.
Martínez-Solano, I., M. Alcobendas, D. Buckley, and M. Garcia-Paris. 2005. Molecular characterisation of the endangered *Salamandra salamandra almanzoris* (Caudata, Salamandridae). Annales Zoologici Fennici 42:57–68.
Martins, M. 1989. Deimatic behavior in *Pleurodema brachyops*. Journal of Herpetology 23:305–307.
Massatti, R., A. A. Reznicek, and L. L. Knowles. 2016. Utilizing RADseq data for phylogenetic analysis of challenging taxonomic groups: A case study in *Carex* sect. *Racemosae*. American journal of botany 103:337–47.
Mayor, R., and E. Theveneau. 2013. The neural crest. Development (Cambridge, England) 140:2247–51.
Mayr, E. 1963. Animal species and evolution. Belknap Press of Harvard University Press, Cambridge (MA).
Mayr, E., and P. D. Ashlock. 1991. Principles of systematic biology. McGraw-Hill, Inc., New York.
McCormack, J. E., S. M. Hird, A. J. Zellmer, B. C. Carstens, and R. T. Brumfield. 2013. Applications of next-generation sequencing to phylogeography and phylogenetics. Molecular Phylogenetics and Evolution 66:526–38.
McLean, C. A., A. Lutz, K. J. Rankin, D. Stuart-Fox, and A. Moussalli. 2017. Revealing the biochemical and genetic basis of color variation in a polymorphic lizard. Molecular Biology and Evolution 34:1924–1935.
Mebs, D., and W. Pogoda. 2005. Variability of alkaloids in the skin secretion of the European fire salamander (Salamandra salamadra terrestris). Toxicon: official journal of the International Society on Toxinology 45:603–606.
Medugorac, I., A. Graf, C. Grohs, S. Rothammer, Y. Zagdsuren, E. Gladyr, N. Zinovieva, J. Barbieri, D. Seichter, I. Russ, A. Eggen, G. Hellenthal, G. Brem, H. Blum, S. Krebs, and A. Capitan. 2017. Whole-genome analysis of introgressive hybridization and characterization of the bovine legacy of Mongolian yaks. Nature Genetics 49:470–475.
Meirmans, P. G., and P. H. van Tienderen. 2004. genotype and genodive: two programs for the analysis of genetic diversity of asexual organisms. Molecular Ecology Notes 4:792–794.
Merabet, K., M. Karar, A. Dahmana, and A. Moali. 2016. New locality of *Salamandra algira* in Algeria. Herpetological Bulletin 135:32–33.
Meredith, P., and T. Sarna. 2006. The physical and chemical properties of eumelanin. Pigment Cell Research 19:572–594.
Metzker, M. L. 2010. Sequencing technologies—the next generation. Nature Reviews Genetics 11:31–46.
Miaud, C., F. Andreone, A. Ribéron, S. De Michelis, V. Clima, J. Castanet, H. Francillon-Vieillot, and R. Guyétant. 2001. Variations in age, size at maturity and gestation duration among two neighbouring populations of the alpine salamander (*Salamandra lanzai*). Journal of Zoology 254:251–260.
Miller, M. A., W. Pfeiffer, and T. Schwartz. 2010. Creating the CIPRES Science Gateway

for inference of large phylogenetic trees. Proceedings of the Gateway Computing Environments Workshop (GCE) 1–8.
Mills, M. G., and L. B. Patterson. 2009. Not just black and white: Pigment pattern development and evolution in vertebrates. Seminars in Cell & Developmental Biology 20:72–81.
Milograna, S. R., M. R. Ribeiro, F. T. Bell, and J. C. McNamara. 2016. Pigment Translocation in Caridean Shrimp Chromatophores: Receptor Type, Signal Transduction, Second Messengers, and Cross Talk Among Multiple Signaling Cascades. Journal of Experimental Zoology Part A: Ecological Genetics and Physiology 325:565–580.
Mirarab, S., M. S. Bayzid, and T. Warnow. 2016. Evaluating summary methods for multilocus species tree estimation in the presence of incomplete lineage sorting. Systematic Biology 65:366–380.
Mirarab, S., and T. Warnow. 2015. ASTRAL-II: coalescent-based species tree estimation with many hundreds of taxa and thousands of genes. Bioinformatics 31:i44–i52.
Misof, B., S. Liu, K. Meusemann, R. S. Peters, A. Donath, C. Mayer, P. B. Frandsen, J. Ware, T. Flouri, R. G. Beutel, O. Niehuis, M. Petersen, F. Izquierdo-Carrasco, T. Wappler, J. Rust, A. J. Aberer, U. Aspöck, H. Aspöck, D. Bartel, A. Blanke, S. Berger, A. Böhm, T. R. Buckley, B. Calcott, J. Chen, F. Friedrich, M. Fukui, M. Fujita, C. Greve, P. Grobe, S. Gu, Y. Huang, L. S. Jermiin, A. Y. Kawahara, L. Krogmann, M. Kubiak, R. Lanfear, H. Letsch, Y. Li, Z. Li, J. Li, H. Lu, R. Machida, Y. Mashimo, P. Kapli, D. D. McKenna, G. Meng, Y. Nakagaki, J. L. Navarrete-Heredia, M. Ott, Y. Ou, G. Pass, L. Podsiadlowski, H. Pohl, B. M. von Reumont, K. Schütte, K. Sekiya, S. Shimizu, A. Slipinski, A. Stamatakis, W. Song, X. Su, N. U. Szucsich, M. Tan, X. Tan, M. Tang, J. Tang, G. Timelthaler, S. Tomizuka, M. Trautwein, X. Tong, T. Uchifune, M. G. Walzl, B. M. Wiegmann, J. Wilbrandt, B. Wipfler, T. K. F. Wong, Q. Wu, G. Wu, Y. Xie, S. Yang, Q. Yang, D. K. Yeates, K. Yoshizawa, Q. Zhang, R. Zhang, W. Zhang, Y. Zhang, J. Zhao, C. Zhou, L. Zhou, T. Ziesmann, S. Zou, Y. Li, X. Xu, Y. Zhang, H. Yang, J. Wang, J. Wang, K. M. Kjer, and X. Zhou. 2014. Phylogenomics resolves the timing and pattern of insect evolution. Science 346:763–767.
Moali, A., S. Steinfartz, D. Donaire, E. Sanchez, K. Merabet, M. Vences, U. Joger, S. Bogaerts, M. Karar, and A. Dahmana. 2016. Phylogeographic relationships and shallow mitochondrial divergence of Algerian populations of *Salamandra algira*. Amphibia-Reptilia 37:1–8.
Modesti, A., S. Aguzzi, and R. Manenti. 2011. A case of complete albinism in *Lissotriton vulgaris meridionalis*. 4:395–396.
Mohlhenrich, E. R., and R. L. Mueller. 2016. Genetic drift and mutational hazard in the evolution of salamander genomic gigantism. Evolution 70:2865–2878.
Montero-Mendieta, S., M. Grabherr, H. Lantz, I. De la Riva, J. A. Leonard, M. T. Webster, and C. Vilà. 2017. A practical guide to build *de-novo* assemblies for single tissues of non-model organisms: the example of a Neotropical frog. PeerJ 5:e3702. PeerJ Inc.
Montieth, K. E., and P. W. C. Paton. 2006. Emigration behavior of spotted salamanders on golf courses in southern Rhode Island. Journal of Herpetology 40:195–205.
Montoliu, L., W. S. Oetting, and D. C. Bennett. 2017. European Society for Pigment Cell Research.
Moustafa, A., and D. Bhattacharya. 2008. PhyloSort: a user-friendly phylogenetic sorting tool and its application to estimating the cyanobacterial contribution to the nuclear genome of Chlamydomonas. BMC Evolutionary Biology 8:6.
Mundy, N. I., N. S. Badcock, T. Hart, K. Scribner, K. Janssen, and N. J. Nadeau. 2004. Conserved genetic basis of a quantitative plumage trait involved in mate choice. Science 303:1870–1873.
Mundy, N. I., J. Stapley, C. Bennison, R. Tucker, H. Twyman, K.-W. Kim, T. Burke, T. R.

Birkhead, S. Andersson, and J. Slate. 2016. Red carotenoid coloration in the zebra finch is controlled by a Cytochrome P450 gene cluster. Current Biology 26:1435–1440.
Murisier, F., and F. Beermann. 2006. Genetics of pigment cells: lessons from the tyrosinase gene family. Histology and histopathology 21:567—578.
Mutz, T. 1992. Vergleich der farbkleidentwicklung bei unterarten des feuersalamanders. Urodela Info 3:4–5.
Myers, C. W., and J. W. Daly. 1976. Preliminary evaluation of skin toxins and vocalizations in taxonomic and evolutionary studies of poison-dart frogs (Dendrobatidae). Bulletin of the American Museum of Natural History 157:173–262.
Nater, A., R. Burri, T. Kawakami, L. Smeds, and H. Ellegren. 2015. Resolving evolutionary relationships in closely related species with whole-genome sequencing data. Systematic Biology 64:1000–1017.
NCBI Resource Coordinators. 2017. Database Resources of the National Center for Biotechnology Information. Nucleic acids research 45:D12–D17.
Nelms, B. L., and P. A. Labosky. 2010. Pax Genes. Morgan & Claypool Life Sciences.
Nelson, T. C., and W. A. Cresko. 2017. Ancient genomic variation underlies repeated ecological adaptation in young 1 stickleback populations. bioRxiv Pre-Print:http://dx.doi.org/10.1101/167981.
Nery, L. E. M., and A. M. de Lauro Castrucci. 1997. Pigment cell signalling for physiological color change. Comparative Biochemistry and Physiology Part A: Physiology 118:1135–1144.
Ng'oma, E., M. Groth, R. Ripa, M. Platzer, and A. Cellerino. 2014. Transcriptome profiling of natural dichromatism in the annual fishes *Nothobranchius furzeri* and *Nothobranchius kadleci*. BMC Genomics 15:754.
Nixon, K. C., and Q. D. Wheeler. 1990. An amplification of the phylogenetic species concept. Cladistics 6:211–223.
Nuckels, R. J., A. Ng, T. Darland, and J. M. Gross. 2009. The Vacuolar-ATPase Complex Regulates Retinoblast Proliferation and Survival, Photoreceptor Morphogenesis, and Pigmentation in the Zebrafish Eye. Investigative Opthalmology & Visual Science 50:893. The Association for Research in Vision and Ophthalmology.
O'Neill, E. M., and K. H. Beard. 2010. Genetic basis of a color pattern polymorphism in the coqui frog *Eleutherodactylus coqui*. Journal of Heredity 101:703–709.
Obst, F. J. 1981. Der Feuersalamander des Pirin-Gebirge in Bulgarien als *Salamandra salamandra beschkovi* subsp. n. eine vorlaufige Mitteilung (Amphibia, Urodela, Salamandridae). Faunistische Abhandlungen. Staatliches Museum für Tierkunde in Dresden 8:197–201.
Odenthal, J., K. Rossnagel, P. Haffter, R. N. Kelsh, E. Vogelsang, M. Brand, F. J. van Eeden, M. Furutani-Seiki, M. Granato, M. Hammerschmidt, C. P. Heisenberg, Y. J. Jiang, D. A. Kane, M. C. Mullins, and C. Nusslein-Volhard. 1996. Mutations affecting xanthophore pigmentation in the zebrafish, Danio rerio. Development 123:391–398.
Olsson, M., D. Stuart-Fox, and C. Ballen. 2013. Genetics and evolution of colour patterns in reptiles. Seminars in Cell & Developmental Biology 24:529–541.
Orr, H. A. 2005. The genetic theory of adaptation: a brief history. Nature Reviews Genetics 6:119.
Oshima, N., E. Nakata, and M. S. Kamagata. 1998. Light-induced pigment aggregation in xanthophores of the medaka, *Oryzias latipes*. Pigment Cell Research 11:362–367.
Pante, E., J. Abdelkrim, A. Viricel, D. Gey, S. C. France, M. C. Boisselier, and S. Samadi. 2014. Use of RAD sequencing for delimiting species. Heredity 114:450.
Parichy, D. M., and J. M. Turner. 2003. Temporal and cellular requirements for Fms signaling during zebrafish adult pigment pattern development. Development 130:817 LP-833.

Pederzoli, A., A. Gambarelli, and C. Restani. 2003. Xanthophore migration from the dermis to the epidermis and dermal remodeling during <i>Salamandra salamandra salamandra<i/> (L .) larval development. Pigment Cell Research 16:50–58.
Pederzoli, A., and P. Trevisan. 1990. Pigmentary system of the adult alpine salamander Salamandra atra aurorae (Trevisan, 1982). Pigment cell research 3:80–9.
Pei, W., L. Xu, G. K. Varshney, B. Carrington, K. Bishop, M. Jones, S. C. Huang, J. Idol, P. R. Pretorius, A. Beirl, L. A. Schimmenti, K. S. Kindt, R. Sood, and S. M. Burgess. 2016. Additive reductions in zebrafish *PRPS1* activity result in a spectrum of deficiencies modeling several human *PRPS1*-associated diseases. Scientific reports 6:29946.
Pereira, R. J., I. Martínez-Solano, and D. Buckley. 2016. Hybridization during altitudinal range shifts: nuclear introgression leads to extensive cyto-nuclear discordance in the fire salamander. Molecular Ecology 25:1551–1565.
Peterson, B. K., J. N. Weber, E. H. Kay, H. S. Fisher, and H. 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:e37135.
Pezaro, N., V. Rovelli, O. Segev, A. R. Templeton, and L. Blaustein. 2017. Suspected rat predation on the Near Eastern Fire Salamander (Salamandra infraimmaculata) by selective consumption of non-toxic tissue. Zoology in the Middle East 1–3.
Philippe, H., H. Brinkmann, D. V. Lavrov, D. T. J. Littlewood, M. Manuel, G. Wörheide, and D. Baurain. 2011. Resolving difficult phylogenetic questions: why more sequences are not enough. PLoS Biology 9:e1000602.
Philippe, H., F. Delsuc, H. Brinkmann, and N. Lartillot. 2005. Phylogenomics. Annual Review of Ecology, Evolution, and Systematics 36:541–562.
Phillimore, A. B., and I. P. F. Owens. 2006. Are subspecies useful in evolutionary and conservation biology? Proceedings of the Royal Society B: Biological Sciences 273:1049–1053.
Poelstra, J. W., H. Ellegren, and J. B. W. Wolf. 2013. An extensive candidate gene approach to speciation: diversity, divergence and linkage disequilibrium in candidate pigmentation genes across the European crow hybrid zone. Heredity 111:467.
Poelstra, J. W., N. Vijay, M. P. Hoeppner, and J. B. W. Wolf. 2015. Transcriptomics of colour patterning and coloration shifts in crows. Molecular Ecology 24:4617–4628.
Posada, D. 2016. Phylogenomics for Systematic Biology. Systematic biology 65:353–6.
Posso-Terranova, A., and J. A. Andres. 2017. Diversification and convergence of aposematic phenotypes: Truncated receptors and cellular arrangements mediate rapid evolution of coloration in harlequin poison frogs. Evolution, doi: 10.1111/evo.13335.
Pough, F. H., R. M. Andrews, J. E. Cadle, M. L. Crump, A. H. Savitky, and K. D. Wells. 2003. Herpetology. 3rd ed. Benjamin Cummings, New York.
Pritchard, J. K., M. Stephens, and P. Donnelly. 2000. Inference of population structure using multilocus genotype data. Genetics 155:945–959.
Protas, M. E., C. Hersey, D. Kochanek, Y. Zhou, H. Wilkens, W. R. Jeffery, L. I. Zon, R. Borowsky, and C. J. Tabin. 2006. Genetic analysis of cavefish reveals molecular convergence in the evolution of albinism. Nature genetics 38:107–11.
Prudic, K. L., A. K. Skemp, and D. R. Papaj. 2007. Aposematic coloration, luminance contrast, and the benefits of conspicuousness. Behavioral Ecology 18:41–46.
Prum, R. O., J. S. Berv, A. Dornburg, D. J. Field, J. P. Townsend, E. M. Lemmon, and A. R. Lemmon. 2015. A comprehensive phylogeny of birds (Aves) using targeted next-generation DNA sequencing. Nature 526:569–573.
Przyrembel, C., B. Keller, and C. Neumeyer. 1995. Trichromatic color vision in the salamander (*Salamandra salamandra*). Journal of Comparative Physiology A 176:575–586.
Purcell, S., B. Neale, K. Todd-Brown, L. Thomas, M. A. R. Ferreira, D. Bender, J. Maller, P. Sklar, P. I. W. de Bakker, M. J. Daly, and P. C. Sham. 2007. PLINK: A Tool Set

for Whole-Genome Association and Population-Based Linkage Analyses. The American Journal of Human Genetics 81:559–575.
Puritz, J. B., C. M. Hollenbeck, and J. R. Gold. 2014. dDocent : a RADseq, variant-calling pipeline designed for population genomics of non-model organisms. PeerJ 2:e431.
Pyron, R. A. 2014. Biogeographic analysis reveals ancient continental vicariance and recent oceanic dispersal in amphibians. Systematic Biology 63:779–797.
Pyron, R. A., F. W. Hsieh, A. R. Lemmon, E. M. Lemmon, and C. R. Hendry. 2016. Integrating phylogenomic and morphological data to assess candidate species-delimitation models in brown and red-bellied snakes (*Storeria*). Zoological Journal of the Linnean Society 177:937–949.
R Core Team. 2013. A language and environment for statistical computing. Foundation for Statistical Computing, Vienna, Austria. http://www.r-project.org/.
Rajpurohit, S., and O. Nedved. 2013. Clinal variation in fitness related traits in tropical drosophilids of the Indian subcontinent. Journal of Thermal Biology 38:345–354.
Rajpurohit, S., O. Nedved, and A. G. Gibbs. 2013. Meta-analysis of geographical clines in desiccation tolerance of Indian drosophilids. Comparative Biochemistry and Physiology Part A: Molecular and Integrative Physiology 164:391–398.
Rajpurohit, S., R. Parkash, and S. Ramniwas. 2008. Body melanization and its adaptive role in thermoregulation and tolerance against desiccating conditions in drosophilids. Entomological Research 38:49–60.
Rajpurohit, S., L. M. Peterson, A. J. Orr, A. J. Marlon, and A. G. Gibbs. 2016. An experimental evolution test of the relationship between melanism and desiccation survival in insects. PLOS ONE 11:e0163414.
Ramniwas, S., B. Kajla, K. Dev, and R. Parkash. 2013. Direct and correlated responses to laboratory selection for body melanisation in *Drosophila melanogaster*: support for the melanisation-desiccation resistance hypothesis. The Journal of experimental biology 216:1244–54.
Rebhan, M., V. Chalifa-Caspi, J. Prilusky, and D. Lancet. 1997. GeneCards: integrating information about genes, proteins and diseases. Trends in genetics : TIG 13:163.
Recknagel, H., A. Jacobs, P. Herzyk, and K. R. Elmer. 2015. Double-digest RAD sequencing using Ion Proton semiconductor platform (ddRADseq-ion) with nonmodel organisms. Molecular Ecology Resources 15:1316–1329.
Reilein, A. R., I. S. Tint, N. I. Peunova, G. N. Enikolopov, and V. I. Gelfand. 1998. Regulation of organelle movement in melanophores by protein kinase A (PKA), protein kinase C (PKC), and protein phosphatase 2A (PP2A). The Journal of cell biology 142:803–13.
Revell, L. J. 2012. phytools: an R package for phylogenetic comparative biology (and other things). Methods in Ecology and Evolution 3:217–223.
Rheindt, F. E., M. K. Fujita, P. R. Wilton, and S. V. Edwards. 2014. Introgression and Phenotypic Assimilation in Zimmerius Flycatchers (Tyrannidae): Population Genetic and Phylogenetic Inferences from Genome-Wide SNPs. Systematic Biology 63:134–152. Oxford University Press.
Riberon, A., C. Miaud, K. Grossenbacher, and P. Taberlet. 2002a. Phylogeography of the Alpine salamander, *Salamandra atra* (Salamandridae) and the influence of the Pleistocene climatic oscillations on population divergence. Molecular Ecology 10:2555–2560.
Riberon, A., E. Sotiriou, C. Miaud, F. Andreone, and P. Taberlet. 2002b. Lack of genetic diversity in *Salamandra lanzai* revealed by cytochrome b gene sequences. Copeia 2002:229–232.
Richards-Zawacki, C. L., I. J. Wang, and K. Summers. 2012. Mate choice and the genetic basis for colour variation in a polymorphic dart frog: inferences from a wild pedigree. Molecular Ecology 21:3879–3892.
Richards, P. M., M. M. Liu, N. Lowe, J. W. Davey, M. L. Blaxter, and A. Davison. 2013.

RAD-Seq derived markers flank the shell colour and banding loci of the *Cepaea nemoralis* supergene. Molecular ecology 22:3077–89.
Ries, C., J. Spaethe, M. Sztatecsny, C. Strondl, and W. Hödl. 2008. Turning blue and ultraviolet: sex-specific colour change during the mating season in the Balkan moor frog. Journal of Zoology 276:229–236.
Rittenhouse, T. A. G., and R. D. Semlitsch. 2006. Grasslands as movement barriers for a forest-associated salamander: Migration behavior of adult and juvenile salamanders at a distinct habitat edge. Biological Conservation 131:14–22.
Rivers, D. M., C. T. Darwell, and D. M. Althoff. 2016. Phylogenetic analysis of RAD-seq data: examining the influence of gene genealogy conflict on analysis of concatenated data. Cladistics 32:672–681.
Roberts, J. L., J. L. Brown, R. Schulte, W. Arizabal, and K. Summers. 2007. Rapid diversification of colouration among populations of a poison frog isolated on sky peninsulas in the central cordilleras of Peru. Journal of Biogeography 34:417–426.
Robinson, R. 2014. Using molecular techniques for rank designation: a reassessment and discussion of *Salamandra salamandra alfredschmidti*'s subspecies ranking using mitochondrial DNA. Masters thesis. University of Glasgow.
Rochette, N. C., and J. M. Catchen. 2017. Deriving genotypes from RAD-seq short-read data using Stacks. Nature Protocols 12:2640–2659.
Rodríguez, A., J. D. J. D. Burgon, M. Lyra, I. Irisarri, D. Baurain, L. Blaustein, B. Göçmen, S. Künzel, B. K. B. K. Mable, A. W. A. W. Nolte, M. Veith, S. Steinfartz, K. R. K. R. Elmer, H. Philippe, and M. Vences. 2017. Inferring the shallow phylogeny of true salamanders (*Salamandra*) by multiple phylogenomic approaches. Molecular Phylogenetics and Evolution 115:16–26.
Rohlf, F. J. 2003. TpsSuper, Version 1.12. Department of Ecology and Evolution, State University of New York at Stony Brook, Stony Brook, NY.
Rohlf, F. J., and D. Slice. 1990. Extensions of the Procrustes method for the optimal superimposition of landmarks. Systematic Zoology 39:40–59.
Rokas, A., and P. Abbot. 2009. Harnessing genomics for evolutionary insights. Trends in Ecology & Evolution 24:192–200.
Ronquist, F., M. Teslenko, P. van der Mark, D. L. Ayres, A. Darling, S. Höhna, B. Larget, L. Liu, M. A. Suchard, and J. P. Huelsenbeck. 2012. MrBayes 3.2: Efficient Bayesian phylogenetic inference and model choice across a large model space. Systematic Biology 61:539–542.
Rosenblum, E. B. 2006. Convergent evolution and divergent selection: lizards at the White Sands ecotone. The American naturalist 167:1–15.
Rosenblum, E. B., H. E. Hoekstra, and M. W. Nachman. 2004. Adaptive reptile color variation and the evolution of the *Mc1r* gene. Evolution 58:1794–1808.
Roulin, A., and A.-L. Ducrest. 2013. Genetics of colouration in birds. Seminars in Cell & Developmental Biology 24:594–608.
Roure, B., N. Rodriguez-Ezpeleta, and H. Philippe. 2007. SCaFoS: a tool for Selection, Concatenation and Fusion of Sequences for phylogenomics. BMC Evolutionary Biology 7:S2.
Rubin, B. E. R., R. H. Ree, C. S. Moreau, N. Lartillot, and P. Holland. 2012. Inferring phylogenies from RAD sequence data. PLoS ONE 7:e33394.
Rudh, A., and A. Qvarnström. 2013. Adaptive colouration in amphibians. Seminars in Cell and Developmental Biology 24:553–561.
Rudh, A., B. Rogell, and J. Höglund. 2007. Non-gradual variation in colour morphs of the strawberry poison frog *Dendrobates pumilio*: genetic and geographical isolation suggest a role for selection in maintaining polymorphism. Molecular Ecology 16:4284–4294.
Salvador, A. 1974. Guía de los anfibios y reptiles Espa ñoles. ICONA, Madrid.
Salzburger, W., I. Braasch, and A. Meyer. 2007. Adaptive sequence evolution in a color

gene involved in the formation of the characteristic egg-dummies of male haplochromine cichlid fishes. BMC Biology 5:51.
Saporito, R. A., R. Zuercher, M. Roberts, K. G. Gerow, and M. A. Donnelly. 2007. Experimental evidence for aposematism in the Dendrobatid poison frog *Oophaga pumilio*. Copeia 2007:1006–1011.
Sardiello, M., S. Cairo, B. Fontanella, A. Ballabio, and G. Meroni. 2008. Genomic analysis of the *TRIM* family reveals two groups of genes with distinct evolutionary properties. BMC Evolutionary Biology 8:225.
Sayyari, E., and S. Mirarab. 2016. Fast coalescent-based computation of local branch support from quartet frequencies. Molecular Biology and Evolution 33:1654–1668.
Scheet, P., and M. Stephens. 2006. A Fast and Flexible Statistical Model for Large-Scale Population Genotype Data: Applications to Inferring Missing Genotypes and Haplotypic Phase. The American Journal of Human Genetics 7878:629–644.
Schmidt, B. R., R. Feldmann, and M. Schaub. 2005. Demographic processes underlying population growth and decline in *Salamandra salamandra*. Conservation Biology 19:1149–1156.
Schneider, C. A., W. S. Rasband, and K. W. Eliceiri. 2012. NIH Image to ImageJ: 25 years of image analysis. Nature Methods 9.
Schulte, L. M., R. A. Saporito, I. Davison, and K. Summers. 2017. The palatability of Neotropical poison frogs in predator-prey systems: do alkaloids make the difference? Biotropica 49:23–26.
Schwalm, P. A., and P. H. Starret. 1977. Infrared Reflectance in Leaf-Sitting Neotropical Frogs. Science 196:1225–1226.
Session, A. M., Y. Uno, T. Kwon, J. A. Chapman, A. Toyoda, S. Takahashi, A. Fukui, A. Hikosaka, A. Suzuki, M. Kondo, S. J. van Heeringen, I. Quigley, S. Heinz, H. Ogino, H. Ochi, U. Hellsten, J. B. Lyons, O. Simakov, N. Putnam, J. Stites, Y. Kuroki, T. Tanaka, T. Michiue, M. Watanabe, O. Bogdanovic, R. Lister, G. Georgiou, S. S. Paranjpe, I. van Kruijsbergen, S. Shu, J. Carlson, T. Kinoshita, Y. Ohta, S. Mawaribuchi, J. Jenkins, J. Grimwood, J. Schmutz, T. Mitros, S. V Mozaffari, Y. Suzuki, Y. Haramoto, T. S. Yamamoto, C. Takagi, R. Heald, K. Miller, C. Haudenschild, J. Kitzman, T. Nakayama, Y. Izutsu, J. Robert, J. Fortriede, K. Burns, V. Lotay, K. Karimi, Y. Yasuoka, D. S. Dichmann, M. F. Flajnik, D. W. Houston, J. Shendure, L. DuPasquier, P. D. Vize, A. M. Zorn, M. Ito, E. M. Marcotte, J. B. Wallingford, Y. Ito, M. Asashima, N. Ueno, Y. Matsuda, G. J. C. Veenstra, A. Fujiyama, R. M. Harland, M. Taira, and D. S. Rokhsar. 2016. Genome evolution in the allotetraploid frog *Xenopus laevis*. Nature 538:336.
Shafer, A. B. A., C. R. Peart, S. Tusso, I. Maayan, A. Brelsford, C. W. Wheat, and J. B. W. Wolf. 2017. Bioinformatic processing of RAD-seq data dramatically impacts downstream population genetic inference. Methods in Ecology and Evolution 8:907–917.
Shakhnarovich, G., T. Darrell, and P. Indyk. 2006. Nearest-neighbor methods in learning and vision. MIT Press, Cambridge, MA.
Shin, S.-H., and Y.-M. Lee. 2013. Glyceollins, a novel class of soybean phytoalexins, inhibit SCF-induced melanogenesis through attenuation of *SCF*/*c-kit* downstream signaling pathways. Experimental & Molecular Medicine 45.
Siedel, U., E. Hartmann, and A. Hein. 2012. Farb- und Zeichnungsanomalien beim Feuersalamander (*Salamandra salamandra*). Amphibia 11:4–19.
Sillero, N., J. Campos, A. Bonardi, C. Corti, R. Creemers, P.-A. Crochet, J. C. Isailović, M. Denoël, G. F. Ficetola, J. Gonçalves, S. Kuzmin, P. Lymberakis, P. de Pous, A. Rodríguez, R. Sindaco, J. Speybroeck, B. Toxopeus, D. R. Vieites, and M. Vences. 2014. Updated distribution and biogeography of amphibians and reptiles of Europe. Amphibia-Reptilia 35:1–31.
Singh, A. P., and C. Nüsslein-Volhard. 2015. Zebrafish stripes as a model for vertebrate

colour pattern formation. Current Biology 25:R81–R92.
Siva-Jothy, M. T. 1999. Male wing pigmentation may affect reproductive success via female choice in a calopterigid damselfly (Zygoptera). Behaviour 136:1365–1377.
Skoglund, P., S. Mallick, M. C. Bortolini, N. Chennagiri, T. Hünemeier, M. L. Petzl-Erler, F. M. Salzano, N. Patterson, and D. Reich. 2015. Genetic evidence for two founding populations of the Americas. Nature 525:104.
Smith-Gill, S. J., C. M. Richards, and G. W. Nace. 1972. Genetic and metabolic bases of two "albino" phenotypes in the leopard frog, *Rana pipiens*. The Journal of Experimental Zoology 180:157–167.
Smith, J. J., S. Putta, J. A. Walker, D. K. Kump, A. K. Samuels, J. R. Monaghan, D. W. Weisrock, C. Staben, and S. R. Voss. 2005. Sal-Site: Integrating new and existing ambystomatid salamander research and informational resources. BMC Genomics 6:181.
Snyder, M., J. Du, and M. Gerstein. 2010. Personal genome sequencing: current approaches and challenges. Genes & Development 24:423–431.
Sparreboom, M., and J. W. Arntzen. 2014. Salamanders of the Old World: The Salamanders of Europe, Asia and Northern Africa. KNNV Publishing, Zeist.
Spencer, S. A. 2015. The role of tfec in zebrafish neural crest cell and RPE development. Virginia Commonwealth University.
Speybroeck, J., W. Beukema, B. Bok, J. Van der Voort, and I. Velikov. 2016. Field guide to the amphibians and reptiles of Britain and Europe. Bloomsbury Natural History.
Speybroeck, J., W. Beukema, and P.-A. Crochet. 2010. A tentative species list of the European herpetofauna (amphibia and reptilia) — An update. Zootaxa 2492:1–27.
Springer, M. S., and J. Gatesy. 2016. The gene tree delusion. Molecular Phylogenetics and Evolution 94:1–33.
Stamatakis, A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30:1312–1313.
Stapley, J., J. Reger, P. G. D. Feulner, C. Smadja, J. Galindo, R. Ekblom, C. Bennison, A. D. Ball, A. P. Beckerman, and J. Slate. 2010. Adaptation genomics: the next generation. Trends in ecology & evolution 25:705–12.
Steinfartz, S., M. Veith, and D. 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 9:397–410.
Steinfartz, S., S. Vicario, J. W. Arntzen, and A. Caccone. 2007a. A Bayesian approach on molecules and behavior: reconsidering phylogenetic and evolutionary patterns of the Salamandridae with emphasis on *Triturus* newts. Journal of Experimental Zoology Part B: Molecular and Developmental Evolution 308B:139–162.
Steinfartz, S., M. Weitere, and D. Tautz. 2007b. Tracing the first step to speciation: ecological and genetic differentiation of a salamander population in a small forest. Molecular ecology 16:4550–61.
Stevens, M., I. C. Cuthill, A. M. M. Windsor, and H. J. Walker. 2006. Disruptive contrast in animal camouflage. Proceedings of the Royal Society B: Biological Sciences 273:2433–2438. London.
Stevens, M., C. A. Párraga, I. C. Cuthill, J. C. Partridge, and T. O. M. S. Troscianko. 2007. Using digital photography to study animal coloration. Biological Journal of the Linnean Society 90:211–237.
Stuart-Fox, D., and A. Moussalli. 2009. Camouflage, communication and thermoregulation: lessons from colour changing organisms. Philosophical transactions of the Royal Society of London. Series B, Biological sciences 364:463–70.
Stuczka, A., C.-A. Hickerson, and C. Anthony. 2016. Niche partitioning along the diet axis in a colour polymorphic population of Eastern red-backed salamanders, *Plethodon cinereus*. Amphibia-Reptilia 37:283–290.

Summers, K., and M. E. Clough. 2001. The evolution of coloration and toxicity in the poison frog family (Dendrobatidae). Proceedings of the National Academy of Sciences of the United States of America 98:6227–6232.
Summers, K., M. P. Speed, J. D. Blount, and A. M. M. Stuckert. 2015. Are aposematic signals honest? A review. Journal of Evolutionary Biology 28:1583–1599.
Sun, Y.-B., Z.-J. Xiong, X.-Y. Xiang, S.-P. Liu, W.-W. Zhou, X.-L. Tu, L. Zhong, L. Wang, D.-D. Wu, B.-L. Zhang, C.-L. Zhu, M.-M. Yang, H.-M. Chen, F. Li, L. Zhou, S.-H. Feng, C. Huang, G.-J. Zhang, D. Irwin, D. M. Hillis, R. W. Murphy, H.-M. Yang, J. Che, J. Wang, and Y.-P. Zhang. 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:E1257–E1262.
Takahashi, T., N. Nagata, and T. Sota. 2014. Application of RAD-based phylogenetics to complex relationships among variously related taxa in a species flock. Molecular Phylogenetics and Evolution 80:137–144.
Tattersall, G. J., P. C. Eterovick, and D. V de Andrade. 2006. Tribute to R. G. Boutilier: skin colour and body temperature changes in basking *Bokermannohyla alvarengai* (Bokermann 1956). The Journal of experimental biology 209:1185–96.
Teng, H., Y. Zhang, C. Shi, F. Mao, W. Cai, L. Lu, F. Zhao, Z. Sun, and J. Zhang. 2017. Population Genomics Reveals Speciation and Introgression between Brown Norway Rats and Their Sibling Species. Molecular Biology and Evolution 34:2214–2228.
ter Braak, C. J. F., and S. de Jong. 1993. The objective function of partial least squares regression. Journal of Chemometrics 12:41–54.
Thayer, A. 1909. Concealing Coloration in the Animal Kingdom. Macmillan, New York.
The Gene Ontology Consortium. 2015. Gene Ontology Consortium: going forward. Nucleic Acids Research 43:D1049–D1056.
The UniProt Consortium. 2017. UniProt: the universal protein knowledgebase. Nucleic Acids Research 45:D158–D169.
Theron, E., K. Hawkins, E. Bermingham, R. E. Ricklefs, and N. I. Mundy. 2001. The molecular basis of an avian plumage polymorphism in the wild: A melanocortin-1-receptor point mutation is perfectly associated with the melanic plumage morph of the bananaquit, *Coereba flaveola*. Current Biology 11:550–557.
Thiesmeier, B. 2004. Der Feuersalamander. Laurenti publishing house, Bielefeld.
Thompson, J. J. W., S. A. O. Armitage, and M. T. Siva-Jothy. 2002. Cuticular colour change after imaginal eclosion is time-constrained: blacker beetles darken faster. Physiological Entomology 27:136–141.
Thorn, R., and J. Raffaëlli. 2001. Les salamandres de l'ancien monde. Boubée, Paris.
Titus, T. A., and A. Larson. 1995. A molecular phylogenetic perspective on the evolutionary radiation of the salamander family Salamandridae. Systematic Biology 44:125–151.
Tonini, J., A. Moore, D. Stern, M. Shcheglovitova, and G. Ortí. 2015. Concatenation and species tree methods exhibit statistically indistinguishable accuracy under a range of simulated conditions. PLOS Currents Tree of Life, doi: 10.1371/currents.tol.34260cc27551a527b124ec5f6334b6be.
Tracy, C. R., K. A. Christian, G. Betts, and C. R. Tracy. 2008. Body temperature and resistance to evaporative water loss in tropical Australian frogs. Comparative Biochemistry and Physiology Part A: Molecular and Integrative Physiology 150:102–108.
Van Belleghem, S. M., R. Papa, H. Ortiz-Zuazaga, F. Hendrickx, C. D. Jiggins, W. Owen McMillan, and B. A. Counterman. 2017. patternize : An R package for quantifying colour pattern variation. Methods in Ecology and Evolution, doi: 10.1111/2041-210X.12853.
Veith, M. 1992. The fire salamander, *Salamandra salamandra* L., in central Europe: subspecies distribution and intergradation. Amphibia-Reptilia 13:297–313.

Veith, M., and S. Steinfartz. 2004. When non-monophyly results in taxonomic consequences - the case of *Mertensiella* within the Salamandridae (Amphibia: Urodela). Salamandra 40:67–80.
Veith, M., S. 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 36:7–16.
Velo-Antón, G., and D. Buckley. 2015. Common Salamander - *Salamandra salamandra*. P. *in* A. Salvador and I. Martínez-Solano, eds. Virtual Encyclopedia of the Spanish Vertebrates. National Museum of Natural Sciences, Madrid.
Velo-antón, G., and A. Cordero-rivera. 2011. Predation by invasive mammals on an insular viviparous population of Salamandra salamandra. 4:299–301.
Velo-Antón, G., M. 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 45:345–352.
Velo-Antón, G., K. R. Zamudio, and A. Cordero-Rivera. 2012. Genetic drift and rapid evolution of viviparity in insular fire salamanders (*Salamandra salamandra*). Heredity 108:410–8.
Vences, M., P. Galán, D. R. Vieites, M. Puente, K. Oetter, and S. Wanke. 2002. Field body temperatures and heating rates in a montane frog population: the importance of black dorsal pattern for thermoregulation. Annales Zoologici Fennici 39:209–220.
Vences, M., E. Sanchez, J. S. Hauswaldt, D. Eikelmann, A. Rodríguez, S. Carranza, D. Donaire, M. Gehara, V. Helfer, S. Lötters, P. Werner, S. Schulz, and S. Steinfartz. 2014. Nuclear and mitochondrial multilocus phylogeny and survey of alkaloid content in true salamanders of the genus *Salamandra* (Salamandridae). Molecular Phylogenetics and Evolution 73:208–216.
Venesky, M. D., and C. D. Anthony. 2007. Antipredator adaptations and predator avoidance by two color morphs of the eastern red-backed salamander, *Plethodon cinereus*. Herpetologica 63:450–458.
Venesky, M. D., A. Hess, J. A. DeMarchi, A. Weil, J. Murone, C.-A. M. Hickerson, and C. D. Anthony. 2015. Morph-specific differences in disease prevalence and pathogen-induced mortality in a terrestrial polymorphic salamander. Journal of Zoology 295:279–285.
Veysey, J. S., K. J. Babbitt, and A. Cooper. 2009. An experimental assessment of buffer width: Implications for salamander migratory behavior. Biological Conservation 146:2227–2239.
Vignieri, S. N., J. G. Larson, and H. E. Hoekstra. 2010. The selective advantage of crypsis in mice. Evolution 64:2153–2158.
Villafuerte, F., and J. J. Negro. 1998. Digital imaging for colour measurement in ecological. Ecology Letters 1:151–154.
von Lintig, J. 2010. Colors with functions: Elucidating the biochemical and molecular basis of carotenoid metabolism. Annual Review of Nutrition 30:35–56.
Wake, M. H. 1993. Evolution of oviductal gestation in amphibians. Journal of Experimental Zoology 266:394–413.
Wake, M. H. 2015. Fetal adaptations for viviparity in amphibians. Journal of Morphology 276:941–960.
Walsh, N., J. Dale, K. J. McGraw, M. A. Pointer, and N. I. Mundy. 2012. Candidate genes for carotenoid coloration in vertebrates and their expression profiles in the carotenoid-containing plumage and bill of a wild bird. Proceedings. Biological sciences 279:58–66.
Wang, H.-J., W.-T. Li, Y.-N. Liu, F.-S. Yang, and X.-Q. Wang. 2017. Resolving interspecific relationships within evolutionarily young lineages using RNA-seq data:

An example from *Pedicularis* section *Cyathophora* (Orobanchaceae). Molecular Phylogenetics and Evolution 107:345–355.
Wang, Z., M. Gerstein, and M. Snyder. 2009. RNA-Seq: a revolutionary tool for transcriptomics. Nature Reviews Genetics 10:57–63.
Warburg, M. 2007a. Longevity in *Salamandra infraimmaculata* from Israel with a partial review of life expectancy in urodeles. Salamandra 43:21.
Warburg, M. R. 2006. Breeding site tenacity in the fire Salamander Salamandra salamandra: a quarter of a century observations in a xeric-inhabiting isolated metapopulation. Bulletin de la Société Herpétologique de France 118:1–18.
Warburg, M. R. 2008a. Changes in recapture rate of a rare salamander in an isolated metapopulation studied for 25-years. Russian Journal of Herpetology 15:11–18.
Warburg, M. R. 2008b. Growth-rate in free-roaming salamanders *Salamandra infraimmaculata*: a long-term study. 27:61–69.
Warburg, M. R. 2007b. The phenology of a rare salamander (*Salamandra infraimmaculata*) in a population breeding under unpredictable ambient conditions: a 25 year study. Acta Herpetologica 2:147–157.
Watahiki, A., K. Waki, N. Hayatsu, T. Shiraki, S. Kondo, M. Nakamura, D. Sasaki, T. Arakawa, J. Kawai, M. Harbers, Y. Hayashizaki, and P. Carninci. 2004. Libraries enriched for alternatively spliced exons reveal splicing patterns in melanocytes and melanomas. Nature Methods 1:233–239.
Weisrock, D. W., T. J. Papenfuss, J. R. Macey, S. N. Litvinchuk, R. Polymeni, I. H. Ugurtas, E. Zhao, H. Jowkar, and A. Larson. 2006. A molecular assessment of phylogenetic relationships and lineage accumulation rates within the family Salamandridae (Amphibia, Caudata). Molecular Phylogenetics and Evolution 41:368–383.
Weisrocka, D. W., J. R. Macey, I. H. Ugurtas, A. Larson, and T. J. Papenfuss. 2001. Molecular phylogenetics and historical biogeography among Salamandrids of the "True" salamander clade: Rapid branching of numerous highly divergent lineages in *Mertensiella luschani* associated with the rise of Anatolia. Molecular Phylogenetics and Evolution 18:434–448.
Wells, K. D. 2007. The ecology and behavior of amphibians. University of Chicago Press.
Wen, J., A. Egan, R. B. Dikow, and E. Zimmer. 2015. Utility of transcriptome sequencing for phylogenetic inference and character evolution. International Association for Plant Taxonomy (IAPT).
Wente, W. H., and J. B. Phillips. 2005. Microhabitat selection by the Pacific treefrog, *Hyla regilla*. Animal Behaviour 70:279–287.
Wetterstrand, K. A. 2017. DNA Sequencing Costs: Data from the NHGRI Genome Sequencing Program (GSP).
Whitlock, M. C., and K. Lotterhos. 2014. OutFLANK: Fst outliers with trimming. R package version 0.1.
Wickett, N. J., S. Mirarab, N. Nguyen, T. Warnow, E. Carpenter, N. Matasci, S. Ayyampalayam, M. S. Barker, J. G. Burleigh, M. A. Gitzendanner, B. R. Ruhfel, E. Wafula, J. P. Der, S. W. Graham, S. Mathews, M. Melkonian, D. E. Soltis, P. S. Soltis, N. W. Miles, C. J. Rothfels, L. Pokorny, A. J. Shaw, L. DeGironimo, D. W. Stevenson, B. Surek, J. C. Villarreal, B. Roure, H. Philippe, C. W. dePamphilis, T. Chen, M. K. Deyholos, R. S. Baucom, T. M. Kutchan, M. M. Augustin, J. Wang, Y. Zhang, Z. Tian, Z. Yan, X. Wu, X. Sun, G. K.-S. Wong, and J. Leebens-Mack. 2014. Phylotranscriptomic analysis of the origin and early diversification of land plants. Proceedings of the National Academy of Sciences 111:E4859–E4868.
Wiener, P., and S. Wilkinson. 2011. Deciphering the genetic basis of animal domestication. Proceedings of the Royal Society B: Biological Sciences 278:3161 LP-3170.
Wilson, K., S. C. Cotter, A. F. Reeson, and J. K. Pell. 2001. Melanism and disease

resistance in insects. Ecology Letters 4:637–649.
Winters, A. E., N. F. Green, N. G. Wilson, M. J. How, M. J. Garson, N. J. Marshall, and K. L. Cheney. 2017. Stabilizing selection on individual pattern elements of aposematic signals. Proceedings of the Royal Society B: Biological Sciences 284.
Withers, P. C. 1995. Evaporative water loss and colour change in the Australian desert tree frog *Litoria rubella* (Amphibia Hylidae). Records of the Western Australian Museum 17:277–281.
Wong, A. K., A. Krishnan, V. Yao, A. Tadych, and O. G. Troyanskaya. 2015. IMP 2.0: a multi-species functional genomics portal for integration, visualization and prediction of protein functions and networks. Nucleic acids research 43:W128-33.
Woodcock, M. R., J. Vaughn-Wolfe, A. Elias, D. K. Kump, K. D. Kendall, N. Timoshevskaya, V. Timoshevskiy, D. W. Perry, J. J. Smith, J. E. Spiewak, D. M. Parichy, and S. R. Voss. 2017. Identification of mutant genes and introgressed tiger salamander DNA in the laboratory axolotl, *Ambystoma mexicanum*. Scientific Reports 7:6.
Wright, A. N., and K. R. Zamudio. 2002. Color pattern asymmetry as a correlate of habitat disturbance in spotted salamanders (*Ambystoma maculatum*). Journal of Herpetology 36:129–133.
Xu, X., G.-X. Dong, X.-S. Hu, L. Miao, X.-L. Zhang, D.-L. Zhang, H.-D. Yang, T.-Y. Zhang, Z.-T. Zou, T.-T. Zhang, Y. Zhuang, J. Bhak, Y. S. Cho, W.-T. Dai, T.-J. Jiang, C. Xie, R. Li, and S.-J. Luo. 2013. The Genetic Basis of White Tigers. Current Biology 23:1031–1035.
Yasutomi, M., and S. Yamada. 1998. Formation of the dermal chromatophore unit (DCU) in the tree frog <i>Hyla arborea<i/>. Pigment Cell Research 11:198–205.
Yoshida, K., R. Miyagi, S. Mori, A. Takahashi, T. Makino, A. Toyoda, A. Fujiyama, and J. Kitano. 2016. Whole-genome sequencing reveals small genomic regions of introgression in an introduced crater lake population of threespine stickleback. Ecology and evolution 6:2190–204.
Young, J. E., K. A. Christian, S. Donnellan, C. R. Tracy, and D. Parry. 2015. Comparative analysis of cutaneous evaporative water loss in frogs demonstrates correlation with ecological habits. Physiological and biochemical zoology : PBZ 78:847–56.
Zhang, P., T. J. Papenfuss, M. H. Wake, L. Qu, and D. B. Wake. 2008. Phylogeny and biogeography of the family Salamandridae (Amphibia: Caudata) inferred from complete mitochondrial genomes. Molecular Phylogenetics and Evolution 49:586–597.
Zhu, W., L. Wang, Z. Dong, X. Chen, F. Song, N. Liu, H. Yang, and J. Fu. 2016. Comparative transcriptome analysis identifies candidate genes related to skin color differentiation in red tilapia. Scientific Reports 6:31347.
Zimin, A., K. A. Stevens, M. W. Crepeau, A. Holtz-Morris, M. Koriabine, G. Marcais, D. Puiu, M. Roberts, J. L. Wegrzyn, P. J. de Jong, D. B. Neale, S. L. Salzberg, J. A. Yorke, and C. H. Langley. 2014. Sequencing and Assembly of the 22-Gb Loblolly Pine Genome. Genetics 196:875–890.