THE UNIVERSITY OF CHICAGO

THE EVOLUTIONARY, BIOGEOGRAPHIC, AND GENETIC ORIGIN OF COLOR PATTERN DIVERSITY IN *PHYLLOBATES* POISON-DART FROGS

A DISSERTATION SUBMITTED TO
THE FACULTY OF THE DIVISION OF THE BIOLOGICAL SCIENCES
AND THE PRITZKER SCHOOL OF MEDICINE
IN CANDIDACY FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

DEPARTMENT OF ECOLOGY AND EVOLUTION

BY
ROBERTO MÁRQUEZ PIZANO

CHICAGO, ILLINOIS
JUNE 2020

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

A PhD is rarely the product of a single person's efforts. Mine is, clearly, not the exception. First and foremost I'd like to thank Marcus Kronforst, my advisor. Agreeing to advise a graduate student always entails some risk, and I am incredibly thankful that you took this risk, and for the seemingly endless support you gave me from then on. I have learned more than I can fit in these pages over the past years. The Kronforst lab has felt like home from the start. I have been fortunate to share this years with a fantastic group of individuals who have always made me feel welcome, comfortable, and part of a community. I have learned something from each and everyone of you, and I'm glad I get to stay for a little longer. I have also been fortunate to have three fantastic scientists in my committee. Every time I asked you were there to help. Trevor, thanks for always pointing out alternative explanations and for reminding me to look things from several viewpoints. John, thanks for encouraging me to never treat methods like black boxes, and to look under the hood to really understand what programs/models are doing. Marko, thanks for showing me a world where organisms and biological functions can be manipulated and studied experimentally. Having you three in my committee has certainly made me a better scientist.

In addition to my official mentors I have been lucky to have met many incredibly kind senior colleagues who have helped me along the way. Daniel Matute, thanks for making sure I survived moving countries to start a PhD. Your scientific and professional advice have since been instrumental in my career. Lauren, thank you for your support and trust. Working with you has and will certainly continue to enrich me and my work. I'm excited for what is to come. Rasmus,thanks for uninterestedly taking me into your lab when I was a student with some ideas about poison frogs and evolutionary genetics, and for so openly and generously sharing your own time and ideas on these subjects. Those months in your lab certainly left a mark. Bibiana, you introduced me to poison frogs, inspired me to go out and study them, and have supported me all along the way. Thanks.

This work would not exist without the many collaborators and colleagues that I've interacted with through these past years. None of it would have happened without Daniel Mejía and Pablo Palacios's immense knowledge of the Chocó and its frogs. Daniel, thanks for teaching me to walk quietly and listen to the forest, and to value the knowledge of the people who interact with it in their daily lives. Pablo, thanks for letting me into your home, treating me like your own, and helping me understand the beautiful place that is the Chocó. Field guides and assistants were also essential. Dago, Marino, and Don Manuel, thanks for sharing your knowledge, skills, and culture with me. I am honored that you have shared your own research and knowledge on poison frogs with me. You are the real experts in their natural history. My work in Colombia was greatly facilitated by Adolfo Amézquita, Andrew Crawford, and Alberto Farfán, who provided me with lab and work space at Universidad de los Andes. Museum work was made possible by the many curators and collection managers who hosted me and/or loaned me specimens. I am especially thankful to Alan Resetar for allowing me to use the Field Museum as my base for specimen work. I have certainly felt

at home at the collection. Tyler Linderoth's work was essential to the first chapter, and his help and advise improved all other chapters. Tyler, thanks for your friendship and for teaching me so much. Working with you was a blast. Nick VanKuren's advice on population genetics and bioinformatics also got me out of trouble more than once. Thanks Nick, for always being available and willing to answer all of my questions. Finally, I thank Mileidy Betancourth for her help and contributions to molecular work.

Having a robust support network was also essential to the completion of this work. All the great friends I made in Chicago make up the fabric of this network, and I am fortunate to have met you all. My Latin American friends have also made me feel closer to home. Having a "tintico" in Spanish certainly helps with the struggle of changing cultures, and moves science forward. Thank you all for being there for me whenever I've needed it. I'm especially thankful for my family's support. My parents have supported and encouraged me in this endeavor in every step from the very beginning (∼28 years ago). They even taught me how to catch frogs and raise tadpoles. My brother and sister are a safety net that I'm lucky to have and will always trust. My thanks to you four are infinite. Last, but not least (actually most), I cannot imagine going through a PhD without my immediate family. Lucía, thank you for reminding me that it is OK to just take a break and go play, that if there is a problem you can always just go fix it, and that "no matter what you step in, just keep walking along and singing your song, because it's all good" [1]. You have kept my priorities in a good place these five years. Valentina, thank you for being my co-conspirer. I am lucky to walk through life alongside a great friend, coparent, colleague, critic, and companion. You two make my life better.

Thank you all.

Roberto Márquez Pizano

CHAPTER 1

INTRODUCTION

Biological diversity can be characterized as variation along three main, loosely correlated axes: lineage (i.e. taxonomic) diversity, morphological disparity, and functional diversity. Lineage diversity refers to the number of groups (i.e. taxa) whose members share a closer genealogical or phylogenetic affinity among themselves than with members of other groups. Morphological disparity constitutes the variance exhibited by a group of entities, for example individuals within a species, or species in a community, at a measurable phenotype of interest. Finally, functional diversity represents the number of ecological functions that a set of organisms perform within a given ecosystem, or across ecosystems [2–6]. Variation in these three dimensions is governed by an array of interacting natural processes, namely evolution by natural selection [7–9], the production of phenotypic variation through changes in ontogeny [10–13], the focal lineage's previous history [14–17], and extrinsic abiotic processes, such as geological changes or astronomic impacts [18–20]. Understanding how these processes interact to produce (or destroy) biological variation is therefore important when characterizing the drivers of biodiversity.

The geographic landscape in which a lineage occurs plays a major role in its diversification dynamics, since several aspects of geography can influence the evolution of biodiversity. For instance, topography and distance have strong effects on the levels of gene flow between populations [9,21,22], and therefore affect the rate at which lineages diversify both taxonomically and phenotypically. Furthermore, environmental heterogeneity across the landscape generates variation in selective pressures that, in turn, can produce variation through adaptive evolution [7, 23, 24]. Thus, studying the historical processes that give rise to the geographic distribution of biological variation can provide great insights into the origins of biological

diversity

In this dissertation, I investigate the mechanisms driving the geographic distribution of lineages and phenotypes in the genus *Phyllobates* (Fig. 1.1), a recently diverged group of dendrobatid poison frogs where coloration has evolved dynamically and in a convergent fashion [25–27]. The first two chapters deal with the historical and biogeographic processes leading to two interesting instances of geographic variation: a "leapfrog" distribution [28], where phenotypically similar populations are disjunctly distributed, and a phenotypic cline [29], where populations exhibit a relatively gradual transition between two clearly distinct color patterns along a transect. Finally, the third chapter inquires on the genetic basis of color pattern differences among *Phyllobates* species, and investigates genetic variation at candidate genes to shed light on the evolutionary history of these genes in relation to the history of *Phyllobates* lineages.

1.1 The cellular and molecular basis of vertebrate coloration

Coloration is among the most widely studied phenotypes in animals. This is due to both its involvement in multiple aspects of animal biology, from thermoregulation and nutrition to inter and intraspecific communication (e.g. mate-choice or predator avoidance), as well as the ease with which many coloration phenotypes can be scored [30–33]. This has resulted in a profound understanding of the mechanisms through which pigmentation develops and evolves in vertebrate model systems [34–37], and increasingly in non-model systems [38–41].

In vertebrates, coloration is produced by chromatophores, a group of cells that originate in the neural crest [42,43], and can be classified in two broad groups based on the mechanisms through which they produce color [44]. First are pigmentary chromatophores, which produce color via the absorption of particular wavelengths of light by pigments within them. Three

kinds of pigmentary chromatophores are predominant in vertebrates: Melanophores, which contain melanin and are black or brown, and xanthophores and erythrophores, which are yellow or red, and contain carotenoid and pteridine pigments. The second type are structural chromatophores, which contain sub-cellular structures that produce color by scattering and reflecting certain wavelengths of light. Iridophores, which contain crystalline purine platelets, are the most widespread kind of structural chromatophore in vertebrates [42, 43, 45, 46]. In most vertebrates chromatophores are arranged three-dimensionally within the dermis, forming structures known as dermal chromatophore units, which consist of stacked xanthophores (or leucophores), iridophores and melanophores. Xanthophores are at the outermost end (ie. closest to the epidermis) followed by iridophores and finally melanophores, which are found near the base of the dermis. Melanophores, in addition, have dendritic processes that extend around and over iridophores (Fig. 1.2) [47]. A notable exception to this pattern are mammals and birds, in which pigment cells (mostly melanocytes, homologous to melanophores) are in the dermis, and transport pigments into neighboring keratinocyte cells, which may retain them or deposit them into hairs or feathers [48, 49].

Chromatophores dictate an animal's coloration in two ways: Through the actual colors produced by their pigments or nanostructures (ie. color), and through the spatial organization of these colors in the skin (ie. pattern). In pigmented chromatohpores, color is determined by the metabolic pathways involved in either the production of pigments, as is the case for melanins [50], or their uptake from the diet and posterior modification, as occurs with carotenoids [51]. In structural chromatophores, on the other hand, color is determined by several aspects of the reflective platelets, such as their width, separation, and orientation, or the height of platelet stacks within cells [42, 44].

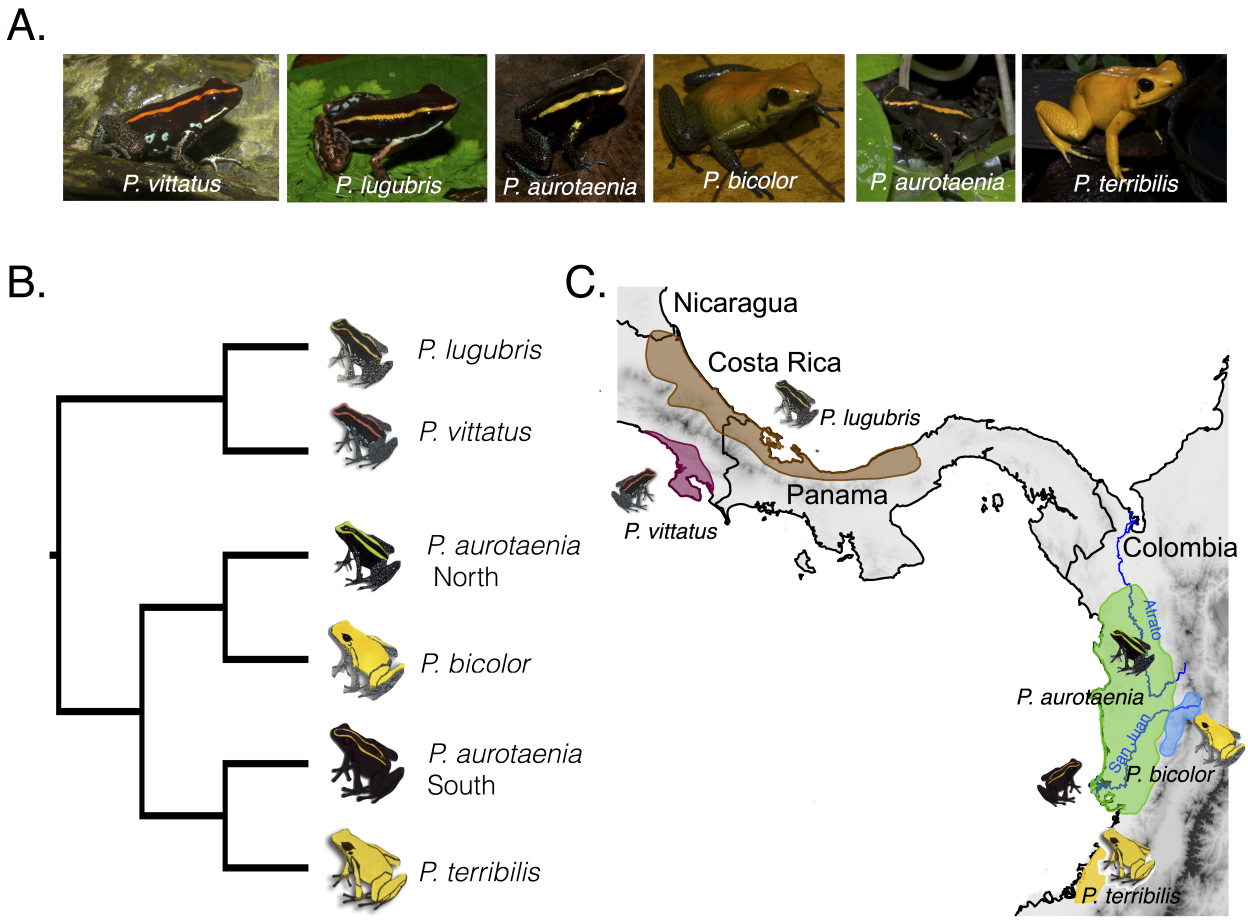

Figure 1.1: Color pattern diversity (A), currently accepted phylogenetic relationships [27] (B), and geographic distribution (C) of *Phyllobates* poison frogs. Species distribution polygons were obtained from the IUCN red list of threatened species website (https://www.iucnredlist.org/), and modified to fit natural history collection records and our own observations.

Melanic pigments are produced within melanophores, by organelles called melanosomes, through a pathway that is relatively conserved across vertebrates [50]. Melanosomes can produce different types of related pigments, whose relative quantities produce different colors. The best studied among them are eumelanin and pheomelanin, which are largely responsible for coat and plumage color variation in mammals and, to a lesser extent, birds [34, 39, 52]. Variants at multiple genes in this pathway have been associated to color variation in several vertebrate groups [34, 35, 39, 52], notably *mc1r*, and *agouti*, which interact to regulate

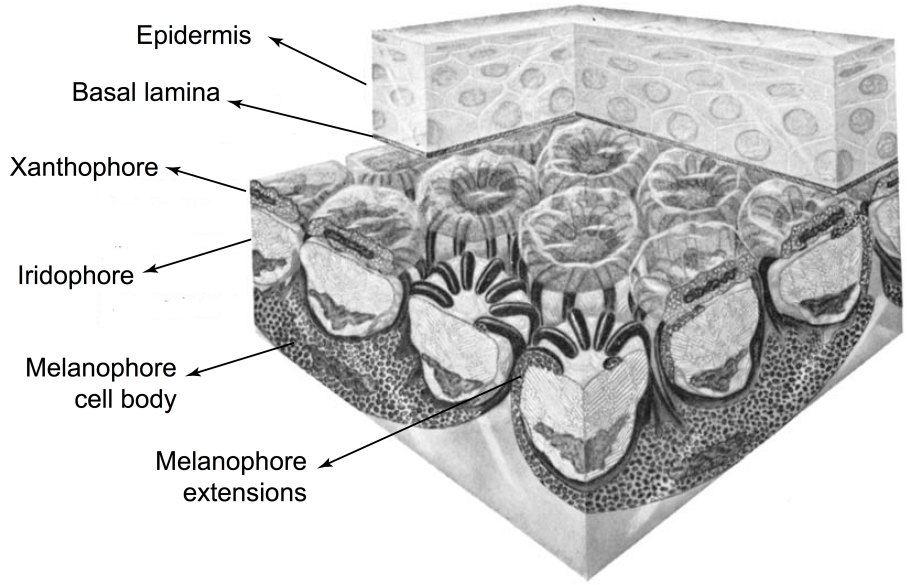

Figure 1.2: Three-dimensional structure of the dermal chromatophore unit, with melanophores are shown in the scattered state, extending over the iridophores. Figure modified from ref. 47.

the deposition of eumelanin and pheomelanin in mamalian hairs. *Mc1r* expression causes melanocytes to produce higher levels of the darker eumelanin, while *agouti*, which inhibits *mc1r* activity, increases the production of pheomelanin. Mutations that increase *mc1r* or decrease *agouti* activity therefore produce darker coat colors, and vice-versa. These two genes have also been associated to brown, black, and tan pigmentation in several other vertebrate groups, such as birds [53, 54], squamates [55, 56], and frogs [57]. Tyrosinase (*tyr*), which, in concert with other enzymes such as *TRP1* and *TRP2*, regulates the speed of melanogenesis [58], has also been shown to be associated with color variation in natural populations of vertebrates [35, 59, 60]. Finally, several genes involved in ion transport between melanosomes and the cellular environment of melanophores, notably *slc45a2*, *slc24a5*, and *oca2*, have also been associated to natural coloration variants [35,61,62]. These genes are involved in pH regulation, ion-mediated signaling, and tyrosinase trafficking into melanosomes, all of which modulate melanin synthesis [63]. An interesting additional pigment synthesized

in melanophores is pterorhodin, which produces a wine-red coloration. This pigment has only has been found in two distantly related groups of tree frogs, the Neotropical subfamily Phyllomedusinae and the Australo-Papuan family Pelodryadidae [64, 65]. Although pterorhodin-containing melanosomes display similar cellular and ontogenetic dynamics to melanin-containing ones, very little is known about pigment synthesis in these organelles [65].

Pteridines are also synthesized metabolically [66, 67], and recent work has revealed associations of genes involved in pteridin synthesis with color variation in lizards [68, 69]. Carotenoids, on the other hand cannot be synthesized by most animals (but see ref. 70), and are therefore obtained from dietary sources, sometimes being modified before deposition in chromatophores, hair, or feathers [51]. Variation in carotenoid-based coloration is usually achieved through either modulation of the levels or rates of pigment uptake and/or deposition, or through chemical modification of carotenoids to alter their light absorption properties [71]. Several genes involved in both processes have been shown to underlie pigmentation differences across vertebrate taxa, including birds [72–75], squamates [69], and fish [76, 77].

Color patterning, the spatial organization of an organism's color patches, is driven by differential expression of pigmentation pathway genes and chromatophore interactions in different regions of the skin. Although less well understood than the mechanisms governing color, the molecular basis of some common patterns have been characterized [36, 49]. For example, in several species of fish, frogs, and mice, ventrally-expressed inhibitors suppress ventral melanization, resulting in much lighter ventral than dorsal pigmentation [78–81]. In both mice and fish *agouti* (or one of its orthologs) has been identified as the melanization supressor [80, 81], and the same has been suggested for frogs [79].

The molecular mechanisms underlying striped color patterns in amphibians and fish have received considerable attention. In several species of salamanders, newts, and zebrafish

of the genus *Danio*, dark stripes are produced by dense aggregations of melanophores, which are absent or nearly absent in the lighter inter-stripes [49]. For some salamander species, the melanohpore-free region is formed as melanophores migrate away from the lateral line primordium during larval development [82, 83]. In zebrafish, stripe position is also achieved through chromatophore migration during development, and, in addition to interactions with the tissue environment, interactions between different chromatophore types, both over short [84, 85] and long [86] distances, play an important role in defining the spatial localization of color patches [36, 37, 87, 88]. Iridophores initially establish the location of the inter-stripe, which serves as a reference for posterior xanthophore and melanophore migration and differentiation to produce the final color pattern. Several genes involved in this process have been associated to pattern variation in zebrafish (see ref. 37 for a recent review).

In addition to ontogenetic change, vertebrate color patterns can also change over shorter time scales, and in a reversible manner. Famous examples of color change include the many polar birds and mammals that seasonally molt their coats or feathers to match their highly seasonal environment, or reptiles (e.g. chameleons) that can drastically change their coloration within seconds. In birds and mammals, most color change occurs through the shedding and posterior re-growth of hairs or feathers, which allows for the development of a new color pattern at each molt [89] In all other vertebrate groups, color change is achieved through the transport and rearrangement of pigments or platelets within chromatophores [44, 90]. The best studied example of this phenomenon is melanosome aggregation and scattering [91, 92], in which melanosomes are transported from the cell body to its periphery and vice-versa through a network of microtubules. When melanosomes are aggregated at the center of the cell, they appear paler than when scattered throughout the cytosol. This effect is accentuated if melanosomes travel up the cell's dendritic processes and occlude the iridiophores above them [90, 92, 93]. Another remarkable example of color change through

the movement of sub-cellular components was recently characterized in panther chameleons (*Furcifer pardalis*), where fine modulations of the spacing and orientation of guanine crystals within iridophores produce rapid and drastic changes in skin color and pattern [94].

The biological importance of pigmentation, as well as the relative ease with which color pattern elements can be visualized, both at macro and microscopic levels, have produced a great wealth of knowledge about the molecular, cellular, developmental, environmental, and evolutionary processes that generate color pattern variation in nature, making this trait an ideal workhorse to understand the origins of biological diversity across biological scales.

1.2 Study system

The genus *Phyllobates* is distributed in Southern Central America and Northwestern South America, from Nicaragua to Colombia (Fig.1.1). These frogs were first discovered by the Emberá in Western Colombians, who used their strong toxins to poison blowgun darts. This practice brought them to the attention of Western scientists through travelers and naturalists in the mid 19th century [95–97], and the potential biomedical uses for these toxins generated a sizable body of work on poison frog diversity and chemical ecology [98]. Myers and colleagues [25] restricted *Phyllobates* to a group of five species, characterized by the ability to secrete Batrachotoxin, a potent alkaloid neurotoxin that makes them distinctively more toxic than other frogs. Three of these species, *P. aurotaenia*, *P. vittatus*, and *P. lugubris*, are dark with two bright dorsolateral stripes, while the remaining two, *P. bicolor* and *P. terribilis*, display solid-yellow coloration [25, 99] (Fig 1.1A).

In all *Phyllobates* species, color pattern development begins in the late tadpole stages with the formation of bright dorsolateral stripe on a dark background. After this, striped species do not experience further substantial changes. In solid-colored species, on the other

hand, the stripes begin to broaden after metamorphosis, and the dark coloration between them is gradually replaced with a yellow suffusion, which eventually becomes solid yellow [25, R. Márquez *pers. obs.*] (Fig 1.3A). Based on these ontogenetic sequences, early systematsists inferred that the striped color pattern was ancestral to solid yellow [25], and proposed that *P. bicolor* and *P. terribilis* were derived sister species [25, 100]. However, further work based on DNA sequences suggested that *bicolor* and *terribilis* are not sister taxa, and that the populations grouped under *P. aurotaenia* may conform more than one species, [26, 27], pointing to convergent evolution of the solid-yellow color pattern. In view of their disjunct geographic distribution, the solid-yellow species of *Phyllobates* therefore represent both phylogenetic and geographic replicates for the occurrence of a derived phenotype.

In his revision of *Phyllobates* systematics Silverstone [99] reports that some individuals of *P. aurotaenia* from the upper San Juan River drainage, in Western Colombia (see map in Fig. 1.1C), display a coloration pattern intermediate between *P. bicolor* and other populations of *P. aurotaenia*, with broader dorsolateral stripes that are sometimes joined by a suffusion of yellow pigmentation towards the head, closely resembling the intermediate steps of coloration development in *P. bicolor* (Fig. 1.3). Furthermore, within some of these these populations color pattern varies from clearly defined, narrow dorsolateral stripes to a dorsum about 75% covered in yellow pigmentation (Fig. 1.3B). This highly variable intermediate zone, together with the phylogenetic and geographic variation in color pattern described above, and the gradual nature of color pattern development make *Phyllobates* an especially well suited system to study the processes guiding phenotypic evolution at multiple organismal and temporal scales. Below I give an overview of my three dissertation chapters, which constitute the first steps towards an integrative understanding of the evolution of color pattern in this group.

Figure 1.3: Variation in color pattern A) Throughout the development of a captive bred *P. bicolor*, and B) within the San Juan populations of *P. aurotaenia*

1.3 Overview of chapters

Chapter 2: Divergence, gene flow, and the origin of leapfrog geographic distributions: The history of color pattern variation in Phyllobates poison-dart frogs

My aim in this chapter was to understand the processes driving the origin and maintenance of the geographic distribution of solid-yellow *Phyllobates* populations in Western Colombia, especially focusing on divergence and gene flow between populations. Using a combination of phylogenetic and spatial population genetic analyses, I characterized the levels of genetic structure and gene flow, as well as the evolutionary relationships between populations. I found high levels of previously unrecognized genetic structure among populations, as well as an even more dynamic evolution of solid-yellow color patterns than previously thought. Furthermore, the data showed a strong signature of short-range gene flow between neighboring population, but not between distant populations. This indicates that gene flow between striped populations may be contributing to their phenotypic similarity, but the same cannot be said for solid-yellow ones.

Chapter 3: Disentangling the origin of a color pattern cline

In this chapter I focused on the color pattern cline present in the upper San Juan River drainage (see section 1.2 in pg. 8), which connects two recently diverged sister species, one of which is striped and the other solid-yellow. Based on patterns of genetic and phenotypic variation among populations in this area I aimed to understand the biogeographic origin of this color pattern cline. I found patterns consistent with either a recent parapatric range expansion or relatively old ongoing hybridization as the processes generating the cline. I favor the first explanation, as it conforms with our expectation of strong selection on these frogs'

aposematic coloration, which have been backed by recent work on closely related species.

Chapter 4: The genetic basis and evolutionary history of color pattern variation in Phyllobates poison frogs

Considering the convergent evolution of highly similar color patterns in *Phyllobates*, which even seem to share a common developmental basis, my last chapter aimed at identifying genomic regions associated with color pattern variation and investigating their evolutionary history to disentangle the genetic processes underlying this instance of convergent evolution. Using genome-wide divergence scans and association analyses I identified several loci associated with color pattern, out of which four stood out as candidate genes involved in color pattern differences. Genetic variation at these loci and their adjacent regions again suggested that introgression has not played a role in the convergent evolution of solid yellow coloration, pointing to independent evolution of these species' color patterns.

CHAPTER 2

DIVERGENCE, GENE FLOW, AND THE ORIGIN OF LEAPFROG GEOGRAPHIC DISTRIBUTIONS: THE HISTORY OF COLOR PATTERN VARIATION IN *PHYLLOBATES* POISON-DART FROGS

Roberto Márquez, Tyler P. Linderoth, Daniel Mejía-Vargas, Rasmus Nielsen, Adolfo Amézquita, Marcus R. Kronforst

2.1 Abstract

The geographic distribution of phenotypic variation among closely related populations is a valuable source of information about the evolutionary processes that generate and maintain biodiversity. Leapfrog distributions, in which phenotypically similar populations are disjunctly distributed and separated by phenotypically distinct populations, represent geographic replicates for the existence of a phenotype, and are therefore especially informative. These geographic patterns have mostly been studied from phylogenetic perspectives to understand how common ancestry and divergent evolution drive their formation. Other processes, such as gene flow between populations, have not received as much attention. Here we investigate the roles of divergence and gene flow between populations in the origin and maintenance of a leapfrog distribution in *Phyllobates* poison frogs. We found evidence for high levels of gene flow between neighboring populations but not over long distances, indicating that gene flow between central populations may have a homogenizing effect that maintains their similarity, and that introgression between "leapfroging" taxa has not played a prominent role as a driver of phenotypic diversity in *Phyllobates*. Although phylogenetic analyses suggest

that the leapfrog distribution was formed through independent evolution of the peripheral (i.e. leapfrogging) populations, the elevated levels of gene flow between geographically close populations poise alternative scenarios, such as the history of phenotypic change becoming decoupled from genome-averaged patterns of divergence, which we cannot rule out. These results highlight the importance of incorporating gene flow between populations into the study of geographic variation in phenotypes, both as a driver of phenotypic diversity and as a confounding factor of phylogeographic inferences.

Key Words: Phylogeography, spatial population genetics, convergent evolution, Dendrobatidae.

2.2 Introduction

Geography has a strong influence on the diversification of closely related lineages, since it largely mediates the level of gene flow between them [9, 21, 22]. Therefore, studying the geographic distribution of phenotypic and genetic variation among such lineages can generate valuable insights into the processes that generate biological diversity. An intriguing pattern of geographic variation is the "leapfrog" distribution, where phenotypically similar, closely related populations (of the same or recently diverged species) are disjunctly distributed and separated by phenotypically different lineages to which they are also closely related [28, 101]. Such patterns have been reported in multiple taxa, such as birds [28, 101–103], flowering plants [104, 105], and butterflies [106–109]. Since leapfrog patterns represent repeated instances of similar phenotypes in space, they provide a rich opportunity to understand the processes generating geographic variation in phenotypes.

Two main hypotheses have been put forward to explain the origin of leapfrog distributions [28, 102]: First, the phenotypically similar, geographically disjunct populations can owe their resemblance to recent common ancestry (i.e. they are descendants of an ancestral population with the same phenotype), and the disjunct range of "leapfrogging" forms is due to biogeographic processes such as long-range migration or the extinction of geographically intermediate populations. Second, the distribution of phenotypes may be due to evolutionary convergence of populations with similar phenotypes, or divergence of the central (intervening) populations from the ancestral phenotype. Clear phylogenetic predictions can be drawn from these hypotheses: If phenotypic similarity among the leapfrogging populations is due solely to recent common ancestry, then such populations should be more closely related to one another than to geographically close populations that display the intervening phenotype. If the geographic distribution of phenotypes is due to convergent or divergent evolution then a correspondence between phylogeny and phenotypes is not expected. In this

case, however, ancestral state reconstructions can identify whether the central or peripheral populations exhibit derived (i.e. divergent) phenotypes. Accordingly, efforts to elucidate the evolutionary mechanisms behind leapfrog distributions have mainly focused on inferring the phylogenetic relationships among populations and using them to reconstruct the evolution of the phenotype in question [e.g. 102, 103, 105, 109, 110].

Although a cladogenetic description of population history can reveal a great deal about the origin of leapfrog distributions, it is unable to capture some important aspects of the diversification process. Among them is the extent of historical and contemporary gene flow between populations (or its absence), which can play an important role in the formation of leapfrog distributions. For instance, reduced levels of genetic exchange between populations with different phenotypes will promote the existence of such differences, while introgressive hybridization between populations can facilitate phenotypic convergence between them. Furthermore, if gene flow between geographically close populations with different phenotypes is pervasive, it can homogenize previous genetic divergence between these populations, decoupling the history of the phenotype from genome-wide patterns of divergence [111, 112], which can complicate inferences related to the origin of leapfrog distributions.

Here we examine the processes driving the origin of a leapfrog distribution present in *Phyllobates* poison-dart frogs. This genus is found from Southern Nicaragua to Western Colombia, and is composed of five nominal species [25, 99]: *P. vittatus*, *P. lugubris*, and *P. aurotaenia*, which exhibit a bright dorsolateral stripe on a dark background, and *P. terribilis* and *P. bicolor*, which display solid bright-yellow dorsal coloration (Fig. 1.1A; note that at a handful of localities *P. terribilis* exhibit solid mint-green color patterns). The two latter species exhibit a leapfrog distribution in Western Colombia, separated by *P. aurotaenia*: *P. bicolor* occurs on the slopes of the Western Andes, in the upper San Juan river basin, *P. aurotaenia* in the lowlands along the San Juan and Atrato Drainages and onto the Pacific

coast, and *P. terribilis* along the Pacific coast south of the San Juan's mouth (Fig. 1.1C).

Early systematic work grouped *P. terribilis* and bicolor as sister species based on morphological and ontogenetic characters [25, 100]. Although an early mitochondrial phylogeny supported these relationships [113], subsequent work has consistently recovered *P. terribilis* and *P. bicolor* as non-sister taxa [26, 27, 114, 115], and even suggested that *P. aurotaenia* may actually represent two distinct lineages, one sister to *P. bicolor* and the other to *P. terribilis* [26, 27]. Although these studies only included 1-4 samples per *Phyllobates* species, and were based on DNA sequences from a small number of markers (1-7 loci), their results are compatible with convergent evolution giving rise to the leapfrog distribution.

In this study we aim to shed light on the evolutionary genetic processes involved in the origin of the current geographic distribution of aposematic coloration in *Phyllobates* poison frogs. Based on substantially increased sampling across Colombian populations and thousands of genome-wide markers, we leverage phylogenetics and spatial population genetics to 1) elucidate the extent of genetic structure and evolutionary relationships among populations, and 2) evaluate the role of gene flow among populations in the formation of the leapfrog distribution.

2.3 Materials and Methods

To obtain a representative sample of Colombian *Phyllobates* populations, we conducted field expeditions to 23 localities throughout the genus's range (Fig. 2.1B), resulting in tissue (i.e. mouth swab, toe-clip, or liver) samples from 108 individuals (Table 2.2). In addition, we obtained eight samples of *P. vittatus* and *P. lugubris*, (four samples per species; Table 2.2) to serve as outgroups in our analyses. Both species are distributed in Central America, and have been consistently found to be the sister group of Colombian *Phyllobates* [26,27,113].

mtDNA Sequencing and analysis

To gain initial insight into the levels of genetic variation and structure among populations we sequenced fragments of three mtDNA markers: 16S rRNA (16S; 569bp), Cytochrome Oxidase I (COI barcoding fragment; 658bp), and Cytochrome b (Cytb; 699bp) from 74 individuals. We extracted DNA using either Qiagen DNeasy spin columns or a salt precipitation protocol [116], and used primers 16Sar and 16Sbr [117], Chmf4 and Chmr4 [118], and CytbDen3-L and CytbDen1-H [119] to amplify the 16S, COI, and Cytb loci, respectively. Thermal cycling protocols consisted of 2 min at 95^{o}C, 30-35 cycles of 30 sec at 95^{o}C, 1 min at 45^{o}C and 1.5 min at 72^{o}C, and a final 5 min at 72^{o}C. PCR products were purified with ExoSAP (Affymetrix) and sequenced in both directions using an ABI 3500 Genetic Analyzer (Applied Biosystems). Chromatograms were assembled and visually inspected in Geneious R9 [120] to produce finalized consensus sequences.

We aligned our sequences and those available in GenBank (Table 2.2) using MUSCLE [121], and built mtDNA trees with PhyML 3.3 [122] and MrBayes 3.2.6 [123, 124]. MrBayes analyses consisted of 10 million iterations (two runs with four chains each), sampling every 1,000 iterations, and discarding the first 2,500 trees (25%) as burnin. PhyML runs started from five different random trees, and used SPR moves to search the tree space. Nodal support was evaluated using aBayes scores [125]. To obtain an estimate of divergence times between mtDNA haplotypes, we inferred a time-calibrated tree using BEAST v. 2.5.0. [126]. Based on results from previous work [127], we set a log-normal prior with mean 8.13 million years (MY) and standard deviation 1.2 MY (i.e. $log(mean) = 2.12$, $log(s.d.) = 0.1$) for the root age of *Phyllobates*. We used a Calibrated Yule tree prior, and set default priors for all other parameters, except for the clock rate mean and the Yule birth rate, which were set to $gamma(0.01, 1000)$. We ran the MCMC sampler for 100 million iterations, sampling

every 10,000, and generated a maximum clade credibility (MCC) tree using Tree Annotator (distributed with BEAST) after discarding the first 5% of trees as burnin. Mixing and stationarity of BEAST and MrBayes runs were evaluated visually and based on effective sample sizes (ESS) using Tracer v. 1.5 [128]. All mtDNA analyses were performed under partitioning schemes and molecular evolution models chosen with PartitionFinder2 [129].

Transcriptome-enabled exon capture

Based on the results of mtDNA analyses we chose 63 samples (60 ingroup, 3 outgroup) from 17 localities (Table 2.2) representing the range of observed mtDNA variation among Colombian populations, and used them to perform transcriptome-enabled exon capture [130, 131]. Briefly, we designed a set of DNA capture probes based on a transcriptome assembly and used them to enrich standard Illumina libraries for a subset of the genome.

Transcriptome sequencing. We generated a transcriptome assembly from liver, muscle, skin, and heart tissue of a single *P. bicolor* juvenile. RNA was extracted using Qiagen RNeasy spin columns, and pooled in equimolar ratios by tissue type to build a single cDNA library, which was sequenced on an Illumina HiSeq 2000. We filtered and trimmed reads using Trimmomatic v. 0.25 [132], and used Trinity (release 2013-02-25) [133] to assemble them under default parameters, except for the minimum contig length, which was increased to 250bp. Finally we collapsed redundant contigs (e.g. alternative isoforoms) with CD-HIT-EST V.4.5.3 [134].

Enrichment probe design. We annotated our transcriptome using blastx [135] against *Xenopus tropicalis* proteins (JGI 4.2.72), and used Exonerate [136] to identify intron-exon boundaries in order to split transcripts into individual exons. We then chose a final set of exons to enrich in the following way: First we discarded those under 100bp, with GC

content below 40% and above 70%, or which overlapped by more than 10bp based on Exonerate annotations. Next, we identified putatively repetitive elements and RNA-coding genes (e.g. rRNAs) in our transcriptome assembly with RepeatMasker v. 4.0 [137] and BLASTn, respectively, and removed exons overlapping them. Finally, we blasted our exon set against itself with BLASTn under default parameters, and whenever two or more exons matched each other (e-value $< 10^{-10}$), we retained only one of them. This resulted in 38,888 exons (7.57Mb) that passed filters, which were used to design 1,943,120 100bp probes that were tiled at 3bp and printed on two Agilent SureSelect custom 1M-feature microarrays.

DNA library preparation, target enrichment, and sequencing. We extracted DNA as described above, and used a Diagenode Bioruptor to shear each extraction to a 100-500bp fragment distribution by performing 3-4 rounds of sonication (7min of 30s on/off cycles per round). DNA libraries were built following ref. 138, except for bead cleanups, which were done using a 1.6:1 ratio of beads to library (instead of the recommended 1.8:1) to obtain a slightly larger final fragment size distribution. Finished libraries were combined in equimolar ratios into two 22.5 μg pools (one per array) for target enrichment. Array hybridization was performed largely following ref. 139 with minor modifications: Each library pool was mixed with xGen Universal P5 and P7 blocking oligonucleotides and a mixture of chicken, human, and mouse COT-1 DNA. The two capture eluates were amplified separately by 18 cycles of PCR. To reduce the propagation of PCR-induced errors, each eluate was amplified in four parallel reactions. PCR products were pooled so that both captures were equally represented, and sequenced on Ilumina HiSeq 2500 and 4000 machines.

Bioinformatic pipeline. De-multiplexed read files were filtered by 1) collapsing PCR-duplicate reads with SuperDeduper [140], 2) trimming low quality bases and removing adapter contamination with Trimmonatic [132] and Skewer [141] under default parameters, except for the minimum read length, which was increased to 36 bp, and 3) merging

overlapping read pairs with FLASH [142]. To generate a reference for read mapping, we combined all cleaned reads from the ingroup species (i.e. *P. terribilis*, *aurotaenia*, and *bicolor*), and generated six de novo assemblies with different kmer sizes ($k = 21, 31, 41, 51, 61,$ and 71) using ABySS [143]. We then merged the six assemblies using CD-HIT-EST [134] and Cap3 [144]. Finally, we identified contigs that matched our target exons using BLASTn, and retained only these for further analyses.

Reads from each sample were mapped to the reference using Bowtie2 v. 2.1.0 [145], and outputs were sorted with Samtools v. 1.0 [146], de-duplicated with Picard v.1.8.4 (http://broadinstitute.github.io/picard), and re-aligned around indels with GATK v. 3.3.0 [147]. We filtered our data in the following ways: First, we performed a reciprocal blast using the methods described above and removed any contigs with more than one match (e-value $< 10^{-10}$). Second, we used ngsParalog (https://github.com/tplinderoth/ngsParalog) to identify contigs with variants stemming from read mismapping due to paralogy and/or incorrect assembly. This program uses allele frequencies to calculate a likelihood ratio for whether the reads covering a site are derived from more than one locus in the genome, while incorporating the uncertainty inherent in NGS genotyping. We calculated p-values for these likelihood ratios based on a 50:50 mixed χ^2 distribution with one and zero degrees of freedom under the null, and removed any contigs with significantly paralogous sites after Bonferronni correction ($\alpha = 0.05$). Third, we restricted all analyses to contigs covered by at least one read in at least 20 individuals, bases with quality above 30, and read pairs mapping uniquely to the same contig (i.e. proper pairs) with mapping quality above 20. Finally, we removed samples with less than 2.5 million sites covered by at least one read after filtering. This resulted in a dataset of 32,516 contigs (12.95 Mb) and 57 samples, which were used in all downstream analyses.

Population Structure

To characterize genome-wide patterns of population differentiation we used our exon capture dataset to perform Principal Component Analysis (PCA) of genetic covariances calculated in PCangsd v.0.94 [148], to estimate admixture proportions (k = 2-9) in ngsAdmix v.32 [149], and to build a minimum-evolution tree in FastME v.2.1.5 [150], using genetic distances estimated with ngsDist [151]. Nodal support for this tree was evaluated using 500 bootstrapped distance matrices produced in ngsDist by sampling blocks of 10 SNPs. These three analyses used genotype likelihoods (GL) as input, which were estimated in Angsd v.0.9.18 [152] at sites covered by at least one read in at least 50% of the samples. PCA and ngsAdmix analyses excluded sites with minor allele frequencies below 0.05, and genetic distance estimation for the ME tree was restricted to variable sites (i.e. SNP p-value < 0.05).

Finally, we reconstructed a population graph to evaluate the relationships among our sampling localities using Treemix [153]. We called genotypes using the HaplotypeCaller and GenotypeGVCFs tools of GATK v.3.3.0 under default parameters, except for the heterozygozity prior, minimum base quality, and minimum variant-calling confidence, which were increased to 0.005, 30, and 20, respectively, to accommodate for the multi-species nature of our dataset. We then obtained allele counts for biallelic SNPs that were at least 1kb apart within each contig (usually resulting in a single SNP per contig, since most contigs were under 1kb), and with at least 50% genotyping (20,275 SNPs), using Plink v.1.90 [154]. In two cases, two nearby populations of the same color pattern (16.4 and 19.7 Km apart; Fig 2.5), which clustered closely in all other genetic structure analyses, were merged into single demes for allele count estimation due to small sample sizes. In addition, since we only had exon capture data for one *P. vittatus* individual, only *P. lugubris* was used as outgroup in this analysis. We ran Treemix v.1.13 assuming $m = 0$–6 migration edges, and chose the optimal number of migration edges by performing likelihood ratio tests in which we compared

each value of m to the one immediately smaller. P-values were calculated based on a χ^2 distribution with two degrees of freedom, since adding an extra edge adds two parameters (weight and direction of migration) to the model. This approach recovered $m = 2$ as the most likely scenario (Table 2.3); results for $m = 0$–6 are presented in Fig. 2.6.

Phylogenetic relationships between lineages

We reconstructed a species tree under the multispecies coalescent model, assuming independent sites, as implemented in SNAPP [155]. SNAPP requires individuals to be assigned to operational taxonomic units (OTUs) a priori. Given our small sample sizes for some localities, as well as the evidence of gene flow between localities (see Results section), we took an ad-hoc approach and grouped our sampling localities into eight geographically and phenotypically coherent groups that showed evidence of being genetically distinct entities (see locality colors in Fig. 2.1B). Further details on our OTU selection criteria can be found in the supplementary materials on page 43. For computational efficiency, SNAPP was run on a reduced version of the Treemix dataset described above, restricted to SNPs genotyped for at least 75% of individuals and at least one member of each OTU (5,938 SNPs). We ran the MCMC sampler under default priors for 1,000,000 iterations, sampling every 250, and discarded the first 150,000 as burnin. Stationarity and mixing were evaluated in Tracer as detailed above, and the posterior tree distribution was summarized as a maximum clade credibility (MCC) tree in TreeAnnotator. To obtain estimates of divergence times between OTUs, we assumed a mutation rate of $\mu = 1 \times 10^{-9}$ mutations per year based on previous estimates [156,157], and a generation time of one year (*Phyllobates* frogs are sexually mature at $\sim 10-18$ months after hatching [25, R. Márquez *pers. obs.*], and converted branch lengths to time units as $T = \tau g/\mu$, where T is the divergence time in years, τ the branch length in coalescent units, g the generation time, and μ the mutation rate [155].

Phylogenetic Comparative Analyses

We reconstructed ancestral color patterns along the SNAPP MCC tree using maximum parsimony [158] in the R package phangorn [159]. Aposematic coloration has been shown to co-evolve with several other traits, such as body size, toxicity, and diet specialization in dendrobatid frogs [119, 160, 161]. Therefore, we investigated the extent of correlated evolution between color pattern, body size, and toxicity. We used the snout-to-vent length (SVL) as a proxy for body size, and the average amount of batrachotoxin (BTX) in a frog's skin as a proxy for toxicity. BTX is the most abundant and toxic alkaloid found in *Phyllobates* skins [25, 162]. BTX levels were obtained from Table 2 of ref. 163, and SVL was measured from specimens in natural history collections (193 specimens; Table 2.4). We used mean SVL values for each lineage in analyses, and log-transformed BTX levels to attain normality of residuals. Correlations between traits were evaluated using phylogenetic generalized least squares regression (pGLS [164, 165] with either Brownian motion [166], Lambda [167], or Ornstein–Uhlenbeck [165] correlation structures. The best correlation structure was chosen by performing pGLS with the three correlation structures and comparing the fit of each model based on the AIC. Correlation structures were generated using the R package ape [168] and regressions were performed in the nlme package [169]. In addition to the highest clade credibility tree, we also conducted tests of phylogenetic correlations on 1,000 randomly selected trees from the post-burnin SNAPP posterior distribution to account for phylogenetic uncertainty.

Spatial population genetics

To investigate the extent of divergence and gene flow between species in an explicitly

spatial context we first we generated a geo-genetic map of the Colombian *Phyllobates* populations in SpaceMix [170]. This consists of a bidimensional plot where the distance between two populations represents their expected geographic distance under stationary isolation by distance, accounting for the fact that a fraction of a population's alleles may have been acquired through migration from another region of the map, whose location is also estimated. As input we used allele counts generated as detailed above (see *Population Structure* section), for SNPs that were variable among Colombian individuals (8,093 sites). We then parameterized the full ("source_and_target") SpaceMix model with an MCMC run comprised of 10 initial exploratory chains (500,000 iterations each), followed by a 500,000,000 iteration "long" run, which was sampled every 10,000 iterations. We used default prior settings, and centered spatial (i.e. location) priors for each population at their sampling location. For the two demes composed of individuals from nearby localities we used the midpoint of the segment connecting both localities (Fig. 2.5).

SpaceMix models long-distance gene flow between populations in terms of admixture proportions, which represent the probability that an allele in a given population migrated recently from a different location of the geo-genetic map. Since leapfrog distributions can arise through introgressive hybridization between disjunct populations, we evaluated the evidence for gene flow between distant populations in our data by comparing the fit of models where admixture proportions were either fixed to 0 or allowed to vary. We did so by estimating the Bayes Factor [171] between both models using the Savage-Dickey density ratio [172], which approximates the Bayes Factor between nested models. Further details on this estimator and our implementation for SpaceMix models can be found in the Appendix.

Next, we used EEMS [173] to identify areas of the landscape where gene flow between populations is especially prevalent or reduced. Briefly, this algorithm estimates the rate at which genetic similarity decays with distance (i.e. the effective migration). Regions

where this decay is quick or slow can be interpreted as barriers or corridors of migration, respectively. We estimated mean squared genetic differences between samples from genotype likelihoods in ATLAS [174], and used them as input for EEMS. We set the number of demes to 500, and averaged across 10 independent 10,000,000-step MCMC runs logged every 1,000 steps (20% burnin). Since our genetic dissimilarity matrix was inferred from genotype likelihoods, specifying the number of SNPs used to compute the matrix (required by EEMS) was not straightforward. We used the number of sites with SNP p-value below 0.05, as calculated with Angsd (221,825 sites).

Finally, we assessed how attributes of the landscape influence genetic divergence between populations. Based on the results of EEMS and SpaceMix analyses, we evaluated the effect of three landscape features on genetic divergence: geographic distance, differences in elevation, and the presence of the San Juan River as a potential corridor of gene flow. To do so, we used the multiple matrix regression with randomization (MMRR) approach [175], which is an extension of multiple linear regression for distance matrices.

Our regression model consisted of genetic distance as a response variable and geographic distance, difference in elevation, and the effect of the San Juan river as a dispersal corridor as explanatory variables. As a proxy for genetic distance, we used the linearized genome-wide weighted F_{ST} ($F_{ST}/[1 - F_{ST}]$) [176], estimated using Angsd based on 2D-site frequency spectra (SFS). To maximize the number of sites used to estimate each SFS, we included contigs with data for less than 20 individuals (but that passed all other filters) in this analysis. We estimated geodesic distances among populations based on GPS coordinates taken in the field using the pointDistance() function of the raster R package [177], and calculated elevation differences based on measurements taken in the field or extracted from Google Earth. To generate a proxy for the San Juan river as a dispersal corridor we built a resistance layer where every pixel overlapping the San Juan river had a value of 1 and all others had values

of 100, and used this layer to calculate least cost distances between populations with the costDistance() function of the gdistance R package [178,179]. Finally, we regressed the least cost distance against the geodesic distance, and saved the model residuals as a measure of the component of the resistance distance not explained by geographic distance. These residuals were used as an explanatory variable in our model. The MMRR analysis was run using the script archived by Wang [175] (https://doi.org/10.5061/dryad.kt71r) with 10,000 permutations to estimate p-values.

2.4 Results

Population structure among Colombian Phyllobates

As expected from a multi-species dataset, we found multiple genetically structured clusters of individuals, which were largely concordant across our exon-enrichment and mtDNA analyses (Fig. 2.1, 2.6-2.7). However, these clusters align much more closely with geography than either coloration or the current taxonomy: All populations of *P. terribilis* grouped with the southern populations of *P. aurotaenia*, while the northeastern populations of *P. aurotaenia* clustered closely with the northern populations of *P. bicolor*. The southern *P. bicolor* and the *P. aurotaenia* populations east and west of the Baudó mountains also formed independent clusters, but their relationship to other lineages was less clear. Finally, two sequences from captive-bred *P. aurotaenia* of unknown origin (sequenced by refs. 26 and 114) were sister to those from the southern populations of *P. bicolor* in our mtDNA genealogy (Fig. 2.1A). These results highlight the existence of several previously unrecognized (i.e. cryptic) lineages. Notably, they reveal the existence of three independent solid-yellow lineages, instead of two as previously thought, since the populations currently classified as *P. bicolor* clustered as two clearly separate and independent lineages. This points to an even

greater discordance between coloration phenotypes and genetic similarity than previously thought.

Phylogenetic relationships and divergence times

The inferred species tree was generally consistent with our genetic structure results, since tree topologies largely mirrored geography (Fig. 2.1-2.2). The three yellow lineages were recovered as sister to a striped lineage. The topology of the SNAPP tree was largely concordant with those obtained in mtDNA and Treemix analyses. We only found inconsistencies in the placement of the *P. aurotaenia* populations from the eastern and western flanks of the Serranía del Baudó: Mitochondrial haplotypes from these two populations were part of a closely-related clade that also included all *P. aurotaenia* sequences from the Atrato river. This clade was sister to another one containing sequences from the northern *P. bicolor* and the San Juan *P. aurotaenia* (Fig. 2.1A). Treemix also recovered the eastern and western Baudó populations as sister taxa, but they were sister to the rest of the Colombian populations (Fig. 2.1D and 2.6). Finally SNAPP recovered only the western Baudó *P. aurotaenia* as sister to all other Colombian populations, while the eastern Baudó *P. aurotaenia* was sister to the southern *P. bicolor* (Fig. 2.2). Treemix inferred a migration edge from the base of the clade containing the populations of *P. bicolor* and *P. aurotaenia* from the San Juan and Atrato drainages into the eastern Baudó *P. aurotaenia* (Fig. 2.1D). Since Treemix reconciles instances where a bifurcating tree model, such as the one used by SNAPP, does not fit the data well by incorporating migration edges between branches of the tree, this result suggests that these differences may be due to gene flow among populations.

Divergence time estimation based on the SNAPP tree revealed a Plio-Pleistocene diversification of *Phyllobates*, and were generally concordant with previous estimates [26, 127], indicating that our mutation rate and generation time assumptions are reasonable. The most

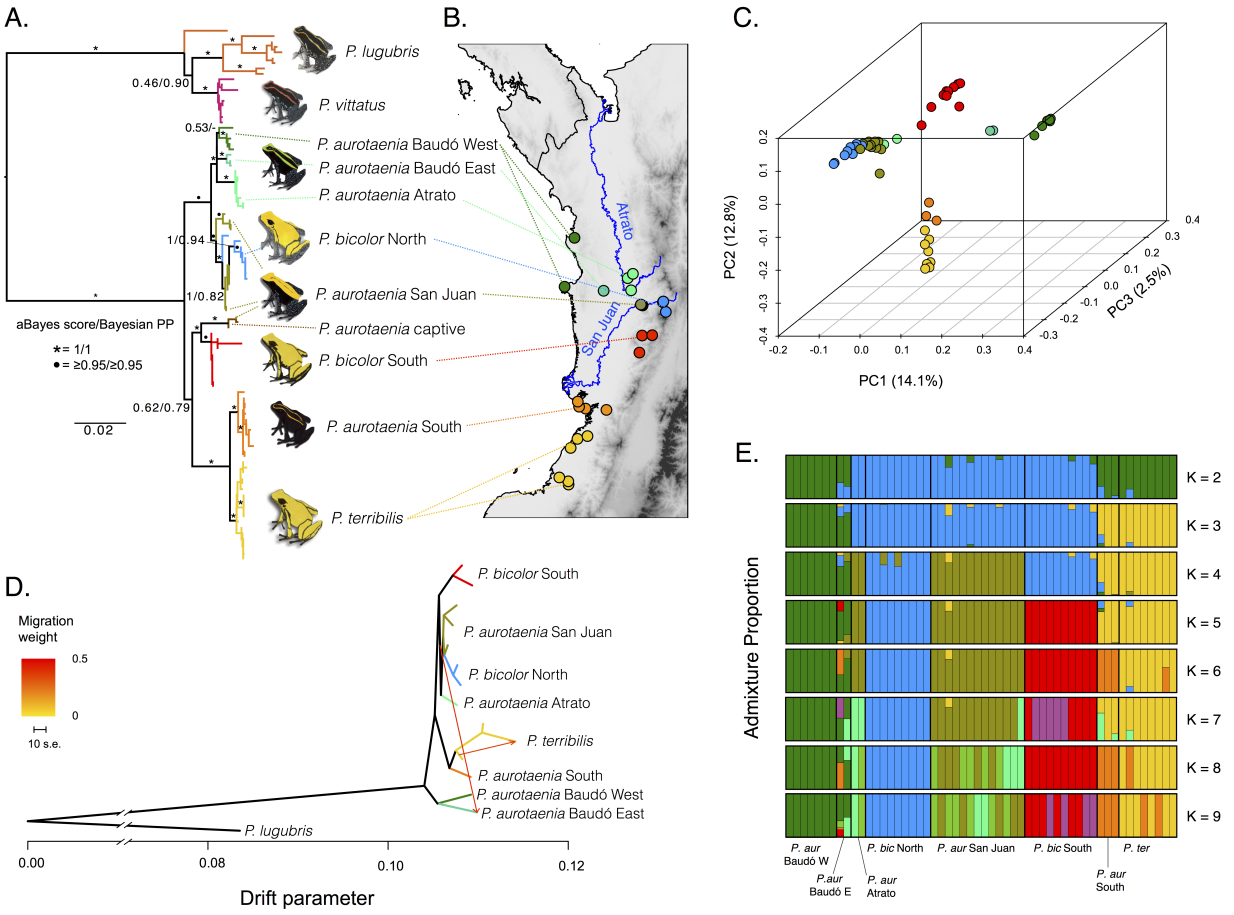

Figure 2.1: Genetic structure among *Phyllobates* populations in western Colombia. A) Maximum likelihood mtDNA genealogy inferred from 1926bp. B) Sampling localities for this study. C) Principal component analysis plot based on the first three components accounting for 29.4% of the variance. D) Treemix population graph assuming 2 migration edges. E) Individual admixture proportions assuming 2-9 ancestral populations. Colors in A-D correspond to the operational taxonomic units (OTU) used for phylogenetic analyses. Colors in E were chosen to loosely represent these clusters.

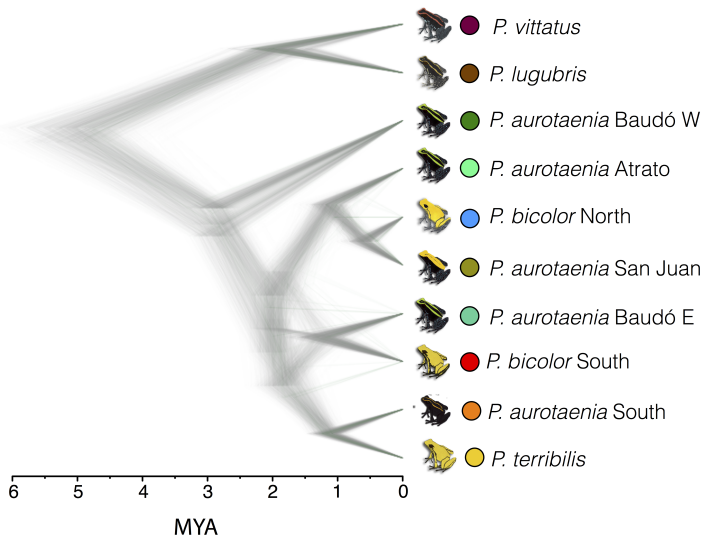

Figure 2.2: Phylogenetic relationships and divergence times among *Phyllobates* lineages inferred using SNAPP. Divergence times assume a mutation rate of 10-9 mutations per year and a generation time of one year. Each individual tree represents one sample from the SNAPP posterior distribution. Clades present in more posterior trees have higher posterior probabilities. The color scheme is as in Fig. 2.1.

recent common ancestor (MRCA) of *Phyllobates* was placed at 5.1 million years ago (MYA), with subsequent cladogenesis events from the late Pliocene to the Pleistocene (2.9-0.6 MYA; Fig. 2.2). These divergence times were slightly older but within the 95% HPD intervals of those estimated from mtDNA sequences (Fig. 2.8).

Comparative analyses

Ancestral state reconstructions found the striped phenotype to be ancestral to solid-yellow (Fig. 2.3A). Phylogenetic regressions revealed a strong relationship between color pattern and size, with solid-yellow lineages being significantly larger than striped ones (Brownian Motion: $\beta = 11.95, t = 9.92, df = 10, p = 9.03 \times 10^{-6}$; Fig. 2.3A), but a much weaker relationship between coloration and toxicity (Ornstein–Uhlenbeck: $\beta = 1.19, t = 2.48, df = 5, p = 0.089$; Fig. 2.3B). These results, suggest that at least two co-evolving traits (solid

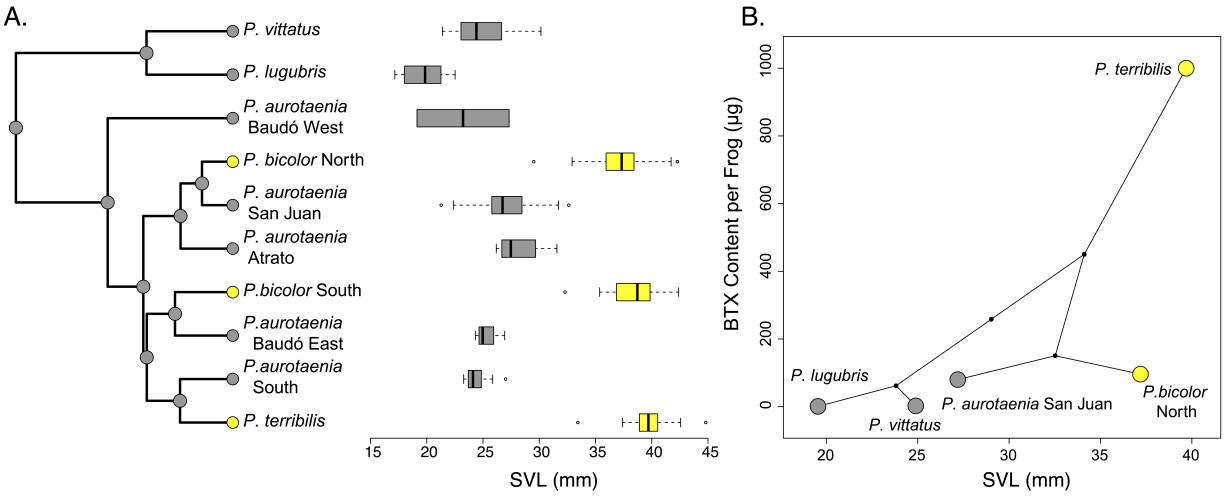

Figure 2.3: Evolutionary patterns of body size, color pattern, and toxicity in *Phyllobates*. A) Maximum clade credibility tree derived from the SNAPP posterior distribution and phylogenetic distribution of color pattern and snout-to-vent length (SVL) values among lineages. B) Phylogenetic biplot depicting the relationship between mean SVL and mean batrachotoxin concentration. Grey boxes/points represent striped lineages while yellow ones represent solid-yellow lineages.

yellow coloration and larger size), possibly related to predator avoidance, are distributed in a leapfrog fashion in *Phyllobates*. Regressions performed over a set of posterior trees instead of the summary tree resulted in effect sizes and p-values centered around and qualitatively equivalent to those estimated using the summary tree, showing that the above conclusions are robust to the phylogenetic uncertainty present in our species tree reconstruction (Fig. 2.9).

Spatial Population Genetics

The effective migration surface estimated by EEMS revealed a clear corridor of migration that matches the course of the San Juan river to a remarkable degree, considering that this method is completely agnostic to the topography of the landscape (Fig. 2.4A). This corridor connects most of the sampled *P. aurotaenia* populations and the northern *P. bicolor*, and could explain the discordance between mtDNA and exon capture datasets in the phylogenetic

Table 2.1: Results from the multiple matrix regression with randomization (MMRR) analysis. P-values were estimated using 10,000 permutations.

Predictor	**Coefficient**	**t-statistic**	**p-value**
Intercept	0.0325	0.5286	1.00000
Geodesic Distance	0.6419	10.0547	<0.00001
San Juan LC Distance	0.2386	4.0094	0.01370
Elevation Difference	0.2886	4.5994	0.00320
	$r^2 = 0.62, F = 48.13, p < 0.00001$		

placement of *P. aurotaenia* populations from the Eastern and Western Baudó mountains. Concordantly, SpaceMix estimated geo-genetic locations of populations along the San Juan corridor that were much closer to one another than their actual geographic positions: *P. aurotaenia* populations from the upper San Juan and Atrato drainages and the northern *P. bicolor* converged to very close locations in the upper/mid San Juan, overlapping considerably. The Baudó (east and west) and southern populations of *P. aurotaenia* were also shifted towards this area, but to a lesser extent (Fig. 2.4B).

In addition EEMS estimated very low levels of migration in the area enclosing the two southern *P. bicolor* populations, suggesting the existence of barriers to gene flow around these populations. Interestingly, the geo-genetic location of these populations was inferred north of its geographic location, past the mid San Juan cluster, and only slightly overlapping with other populations. The estimated long-distance admixture proportions were minimal for all populations (Fig. 2.4C), and the model with these proportions fixed at 0 was overwhelmingly supported over one where they were allowed to vary (Bayes factor = 1748).

In agreement with EEMS and SpaceMix results, the MMRR analysis found significant effects of geographic distance, elevation differences, and the San Juan as a migration corridor on genetic divergence between localities (Table 2.1). Geographic distance was, expectedly, the strongest predictor ($\beta = 0.64, t = 10.05, p < 0.00001$), but elevation differences ($\beta =$

$0.29, t = 4.60, p = 0.003$), and the San Juan as a barrier ($\beta = 0.24, t = 4.01, p = 0.013$) still had appreciable effects on genetic divergence.

2.5 Discussion

Our main goal for this study was to understand the evolutionary and biogeoographic processes that have shaped the leapfrog distribution of color pattern among *Phyllobates* populations, focusing on the roles of genetic divergence and gene flow. We found patterns of genetic structure and phylogenetic affinity between populations that closely match geography, and a strong signal of short-range gene flow, especially along the San Juan river, together with compelling evidence against gene flow between distant populations.

These results provide strong evidence against the hypothesis that introgression of color pattern alleles between disjunct populations has played a role in generating the geographic distribution of this trait. Instead, they suggest an important role for short-range gene flow between neighboring populations. The high level of migration among the central striped populations along the San Juan river suggests that allele movement between these populations can have a homogenizing effect that maintains their phenotypic similarity. In addition, we find evidence for a barrier to gene flow that encloses the two sampled populations of the southern *P. bicolor* lineage, probably associated with differences in elevation, which could be helping maintain the phenotypic distinctiveness of this population. Conversely, the northern *P. bicolor* populations showed a strong signature of gene flow with their neighboring striped populations, suggesting that other forces, possibly selection, are maintaining the phenotypic differences between these populations in the face of recurrent gene flow. Nevertheless, to fully reject or accept these hypotheses, the history the alleles underlying color pattern differences must me taken into account [111].

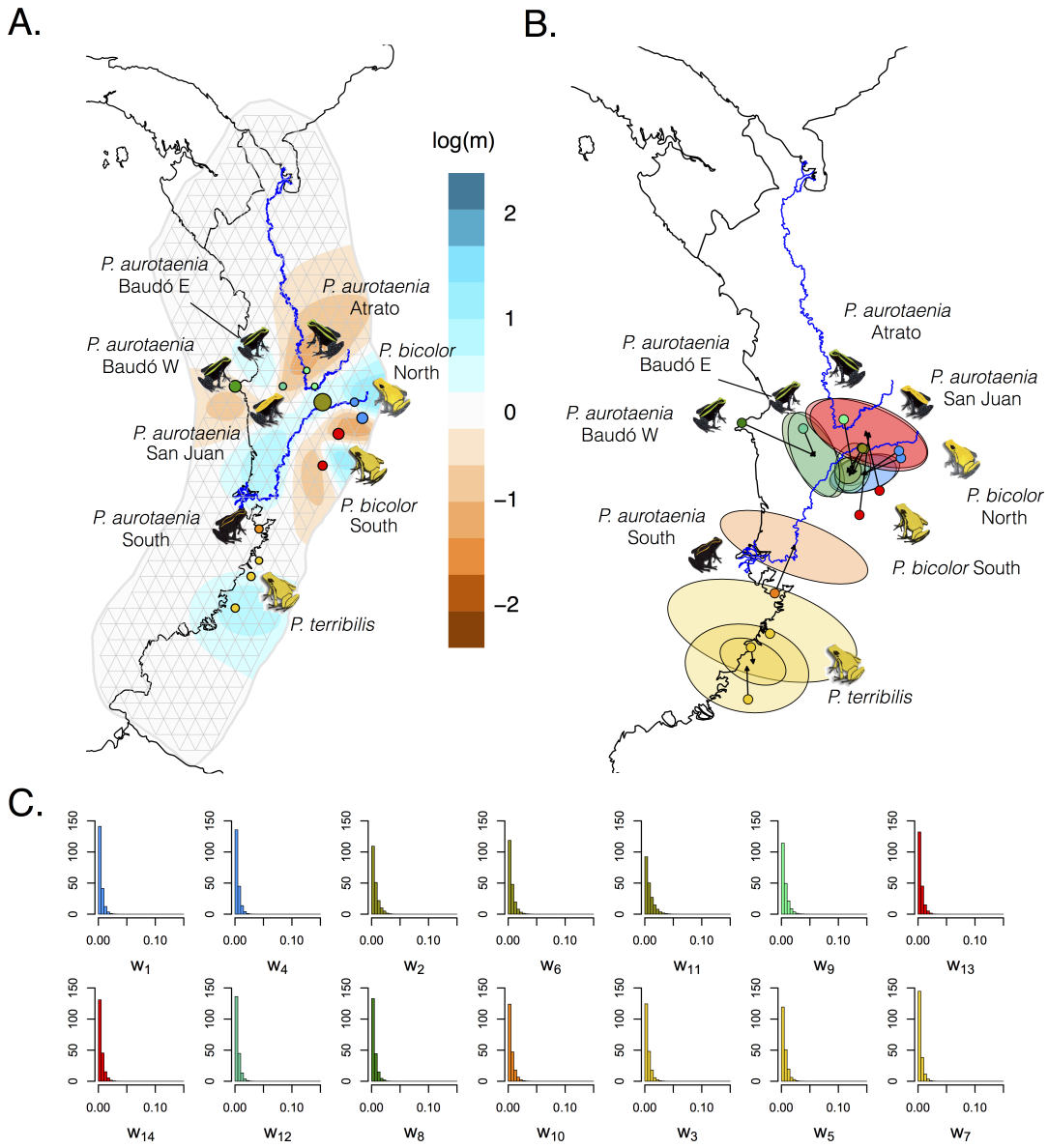

Figure 2.4: Gene flow among *Phyllobates* species across Western Colombia. A) EEMS effective migration surface; cyan and brown areas of the map are those with higher or lower migration between demes, respectively, than expected under isolation by distance. Grey lines depict the population grid and habitat outline used by EEMS. B) SpaceMix Geogenetic map. Ellipses represent 95% Bayesian credible intervals around geogenetic locations, and colored dots represent actual sampling locations. Arrows connect sampling and geogenetic locations. C) Posterior distributions of the admixture proportion parameters from the SpaceMix model for each population. Histograms, points, and ellipses are colored by OTU as in Fig. 2.

Our phylogenetic reconstructions are consistent with a scenario in which the disjunct solid-yellow populations evolved their color patterns independently. However, the high levels of gene flow between geographically proximal populations and the close correspondence between phylogeny and geography lead us to suspect that, at least to an extent, the recovered phylogenetic relationships are a product of prevalent gene flow between neighboring populations, and therefore do not reflect the history of color pattern. A recent simulation study showed that even moderate levels of gene flow between geographic neighbors can confound phylogenetic analyses [112]. This scenario seems especially likely in the case of the northern populations of *P. bicolor*, given the signature of gene flow with their nearby striped populations (e.g. the San Juan and Atrato *P. aurotaenia*), but less so in the case of the southern *P. bicolor*, considering the strong barriers to gene flow inferred around the populations of this lineage. For *P. terribilis* we cannot favor either scenario, since we did not find strong evidence for or against gene flow with its sister *P. aurotaenia* populations.

It is therefore plausible that convergent evolution and common ancestry have both played a role in the origin of this leapfrog distribution. Teasing apart these two scenarios is challenging if gene flow between neighboring populations is prevalent, since high levels of genetic exchange between geographically close populations can erode existing differentiation between them, leading to patterns of genetic/phylogenetic affinity across the genome that mirror geography, regardless of their previous history. In the case of leapfrog distributions, this means that, in the face of persistent gene flow, peripheral populations will be closest to their phenotypically distinct neighbors, even if their phenotypic similarity stems from common ancestry. However, admixture between lineages is seldom uniform across the genome, since selection (see below) can restrict gene flow at certain genomic regions [180–182]. Such regions can therefore preserve historic signatures that have been erased by gene flow elsewhere in the genome. This is likely to be the case for loci underlying color pattern variation in *Phyllobates*, especially in cases such as the Northern *P. bicolor*, where phenotypic differences persist in

spite of gene flow. Hence, the history of alleles at these loci should provide unique insights into the history of this phenotype. Future studies to identify such loci and understand their evolutionary history in relation to our current results will be instrumental to uncover the demographic processes leading to the current geographic and phylogenetic distribution of solid-yellow color pattern in *Phyllobates*, since they will allow for much more explicit tests of the hypotheses presented here.

Regardless of whether common ancestry or convergent evolution are at play in this system, it seems clear that differential selective pressures on the striped and solid-yellow populations have been involved in the origin and/or maintenance of the geographic distribution of color patterns. Independent evolution of similar phenotypes is many times promoted by similar changes in selective regimes [7, 183], and selection is required to maintain phenotypic differences between populations in the face of gene flow [184]. Two of the three solid-yellow lineages (the northern and southern *P. bicolor*) occur at higher elevations ($\sim 600 - 1000$m.a.s.l) than the rest of the genus ($\sim 0 - 500$m.a.s.l). It is therefore possible that these mid-elevation habitats pose selective pressures (e.g. predator communities or light environments) different from those of lowland forests, which favor solid-yellow patterns over striped ones. The many known examples of variation in coloration across altitudinal gradients lend support to this idea [e.g. 185–189]. A similar situation could also be the case with *P. terribilis*, given its distribution at the southern edge of the genus's range, where it may also experience different selective pressures from those faced by its closely related striped lineages.

The nature of color pattern variation in *Phyllobates* (i.e. solid vs striped) suggests that differential predation pressures may be important for the origin/maintenance of solid-yellow patterns. In aposematic species, advertisement signals with complex pattern elements, such as stripes, have been shown to serve a distance-dependent purpose, acting as conspicuous signals at short distances, while providing camouflage at long distances

The fact that we find a signature of correlated evolution between size and color pattern is compatible with this idea, since larger aposematic signals have been shown to be more detectable and memorable for predators [190, 191]. Accordingly, size and conspicuousness are positively correlated among Dendrobatid poison frog species [119, 192]. However, we do not find a comparable pattern for toxicity, which has also been shown to co-vary with conspicuousness in poison frogs [119, 161]. This could be an artifact of low statistical power, since data are available only for one striped and two plain yellow Colombian lineages, but we cannot rule out the possibility that toxicity is indeed comparable between solid and striped populations. Furthermore, considering that aposematism relies on avoidance learning, it is possible that, despite similar levels of BTX, solid and striped populations differ in levels of palatability to predators. In any case, a scenario where solid and striped populations are similarly toxic and/or palatable is still compatible with predation pressures driving evolutionary convergence, since all species are, in any case, considerably toxic [25, 163]. However, other explanations, such as geographic variation in mate preference [193–196], could also explain our results and cannot be ruled out.

It is worth noting, however, that the correlated evolution of body size and color pattern could also be due to ontogenetic integration [197]. Tadpoles of all *Phyllobates* species are dark grey, and all of them develop a dorsolateral stripe shortly before metamorphosis, which remains unchanged until adulthood in striped lineages. Solid-yellow frogs, on the other hand, gradually lose dark pigmentation, until the solid adult pattern is attained a few months after metamorphosis [25] (Fig. 1.3A). Therefore it is possible that, for example, the evolution of an extended growth period could generate changes in both body size and color pattern. If this is the case, then the concerted evolution of advertisement signal and body size would not necessarily be evidence of striped and solid patterns representing alternative predator avoidance strategies.

Our divergence time estimates indicate that the diversification of *Phyllobates* has followed the Plio-Pleistocene history of the Central American and the Chocó bioregions. The first cladogenesis event in our tree, which divides the Central American and Chocoan taxa was reconstructed between 4.5-5.9 MYA, which coincides with previously identified increases in faunal migration between Central and South America at 6 MYA [26,198]. Further branching within South American lineages occurred later than 3 MYA, after both the Atrato [199, 200] and Tumaco [201] basins emerged above sea level to form the current Chocoan landscape. The Pleistocene was characterized by recurrent climatic and environmental fluctuations, which have been proposed as major drivers of neotropical rainforest biodiversity [202–205]. Although the central Chocó has traditionally been regarded as a relatively stable Pleistocene forest refuge throughout the Quaternary [204,206,207], a notion supported by multiple palynological studies [208–213], there is some evidence of fluctuations in sea level, temperature, fluvial discharge, and, to a lesser extent, precipitation throughout the Quaternary in this region [212,213]. Despite being less dramatic than those experienced by other tropical forests (e.g. Amazonia), these fluctuations appear to have been related to changes in vegetation, especially the extent of mangrove forests [212]. This may have promoted periodic retractions of *Phyllobates* populations towards the San Juan, perhaps resulting in increased rates of gene flow among them. Future work to understand how climatic fluctuations over the Quaternary have shaped the distribution of suitable habitat for *Phyllobates* frogs should shed further light on the biogeographic history of this genus in Northern South America.

Finally, our findings have broad implications for the systematics of *Phyllobates*. First and foremost, this study provides definitive evidence that the populations currently grouped under *P. aurotaenia* represent multiple independently-evolving lineages, many of which have probably been reproductively isolated for enough time to warrant recognition as separate species. Furthermore, we find that *P. bicolor* is comprised of two well-structured lineages that may have evolved similar phenotypes independently. Third, we find highly variable

levels of mtDNA divergence within *P. lugubris* (0-4% 16S, 0.2-8% COI, and 0-5.7% Cytb uncorrected p-distances), which could also be due to the existence of cryptic species.

It is, therefore, evident that a thorough revision of *Phyllobates* systematics is warranted. At this point, however, we refrain from modifying the group's current taxonomy for several reasons. First, our results clearly illustrate that, as previously suggested [114,214], coloration alone is not effective for diagnosing species of Dendrobatid frogs. Therefore, integrating multiple lines of evidence (e.g. coloration, genetic variation, alkaloid profiles, bioacoustic data, larval and adult morphology) is needed to disentangle species limits. Second, although our study represents a substantial increase in geographic sampling, there are still considerable gaps, such as the lower San Juan drainage, or the mid-elevation forests south of the distribution of *P. bicolor*, that need to be considered. The fact that the two captive-bred *P. aurotaenia* included in mtDNA analyses are sister to the southern *P. bicolor* (Fig. 2.1A) suggests that we have not yet sampled the full diversity of *Phyllobates* lineages in Colombia. Finally the holotype of *P. aurotaenia* was collected in Condoto, Chocó, which is considerably distant from any of our sampling localities (Fig. 2.10), and the type locality of *P. bicolor* is unknown [25]. This situation poses nomenclature issues, since, even if a robust species delimitation were available, naming these species would not be straightforward until the type specimens of *P. bicolor* and *P. aurotaenia* can be confidently assigned to one of them. Further work with increased sampling, including type specimens, and integrating multiple lines of evidence is therefore still needed to generate a taxonomy for *Phyllobates* that more accurately represents the genus's evolutionary history.

2.6 Concluding Remarks

Leapfrog distributions constitute geographic replicates for the occurrence of a phenotype, and therefore provide important information about the origins of phenotypic diversity among closely related lineages. Here we show that, despite marked genetic structure and differentiation, there is considerable gene flow between phenotypically similar populations at the center of a poison-dart frog leapfrog distribution. This has probably been important for the origin and maintenance of the geographic distribution of color patterns in this group. Furthermore, we found instances of both reduced and increased levels of gene flow between neighboring populations with different phenotypes, suggesting that in some cases reduced gene exchange can contribute to the maintenance of phenotypic differences between populations in a leapfrog distribution, while in others these differences actually persist in the face of gene flow, probably due to local adaptation of different forms.

However, we are unable to answer a commonly addressed question about leapfrog distributions: whether phenotypic differences between populations stem from common ancestry or independent evolution. Even though our phylogenetic reconstructions unambiguously suggest the latter on their own, our finding of extensive gene flow among neighboring populations casts doubt on this conclusion. Several other studies on the history of leapfrog distributions have obtained similar phylogenies that align with geography instead of phenotypic similarity [102, 103, 110, 215, 216], leading to the view that leapfrog distributions are often due to independent evolution. Our results, therefore, add to the notion that alternative explanations such as pervasive gene flow [111, 112] or incomplete lineage sorting [217] can decouple the history of phenotypic change at a given trait from genome-wide patterns of divergence, possibly leading to erroneous inferences of convergent evolution [218].

2.7 Acknowledgements

We thank Pablo Palacios-Rodríguez, José Alfredo Hernández, Carolina Esquivel, Diana Galindo, Mabel González, and Fernando Vargas-Salinas for assistance in the field, Lydia Smith, Valeria Ramírez-Castañeda, Alvaro Hernández, and Ke Bi for help with molecular and bioinformatic procedures, Andrea Paz for advice on MMRR analyses, and Alan Resetar (FMNH), Andrew Crawford and Alberto Farfán (ANDES), Rayna Bell and Addison Wynn (USNM), Andrés Acosta and Carlos Montaña (IAvH), John Taylor Rengifo (UTCh), Greg Schneider (UMMZ), and David Kizirian (AMNH) for facilitating access to preserved specimens. Comments and suggestions from Trevor Price, John Novembre, Valentina Gómez-Bahamón, John Bates, Daniel Matute, the Bates/Hackett lab, the Kronforst lab, and four anonymous reviewers greatly improved this paper. We sincerely thank Lina M. Arenas for allowing us to use her beautiful frog illustrations for our figures. This work was funded by a Basic Sciences Grant from the Vice Chancellor of Research at Universidad de los Andes, a Colombia Biodiversa Scholarship from the Alejandro Angel Escobar foundation, a Pew Biomedical Scholarship, Neubauer Family funds and a Steiner Award from the University of Chicago, and NSF grant DEB-1655336. RM was partially supported by a Fellowship for Young Researchers and Innovators (Otto de Greiff) from COLCIENCIAS. Computations were performed on the University of Chicago's Gardner HPC cluster, funded by NIH grant TR000430. Tissue collections were authorized by permits No. 2194 and 1380 from the Colombian Ministry of Environment and Authority for Environmental Licenses (ANLA).

2.8 Author Contributions

R.M., A.A., and M.R.K. conceived the project, R.M., T.P.L, R.N., M.R.K, and A.A. designed the research, R.M., A.A., R.N. and M.R.K. acquired funding, A.A., D.M-V., and R.M.

collected samples, R.M. and T.P.L. generated the data, and R.M. analyzed the data and wrote the paper with input from M.K. and edits from all authors.

2.9 Supplementary information

Supplementary Note:

Criteria used to select Operational Taxonomic Units for SNAPP

Being a species tree inference algorith, SNAPP requires all gentoyped individuals to be assigned to an operational taxonomic unit (OTU), and infers the relationships and divergence times between these OTUs. Due to our low and non-uniform sample sizes across populations, the presence of some potentially important sampling gaps, and the strong evidence for gene flow across the landscape, we took an ad-hoc approach to OTU selection. We defined OTUs as geographically, genetically, and phenotypically cohesive groups that appeared to be independently evolving units.

More specifically, we assigned a group geographically close individuals to an OTU if they met at least three of the following criteria:

1. All individuals were either striped or solid-colored.
2. All individuals/localities were monophyletic in mtDNA, Treemix (Tmx), and minimum-evolution (ME) tree analyses.
3. Alternatively, all individuals/localities were paraphyletic with respect to members of at most one other OTU in mtDNA, Treemix, and minimum-evolution tree analyses.
4. All individuals formed a discrete cluster on the first three axes of the PCA, which did not overlap with individuals from any other group.
5. All individuals formed a distinct, unique block in ngsAdmix analyses for at least two values of K.

Below we detail criteria fulfilled by each of our eight OTUs:

OTU	Color code	Criteria fulfilled
P. aurotaenia Baudó West		1, 2a (mtDNA, ME), 3, 4
P. aurotaenia Baudó East		1, 2a (mtDNA), 2b (ME), 3
P. aurotaenia Atrato		1, 2a (mtDNA, ME), 3
P. bicolor North		1, 2a (Tmx, ME), 2b (mtDNA), 4
P. aurotaenia San Juan		1, 2b (Tmx, mtDNA, ME), 4
P. bicolor South		1, 2a (Tmx, mtDNA, ME), 3, 4
P. aurotaenia South		1, 2a (mtDNA), 2b (ME), 3
P. terribilis		1, 2a (Tmx, mtDNA), 2b (ME), 3

Finally, we note that, since we obtained largely congruent phylogenetic results between our SNAPP analysis using these OTUs and other analyses that used individuals or sampling localities as the unit of analysis (ie. the mtDNA, ME, and Treemix trees), we are confident that our SNAPP results, and those derived from them (ie. phylogenetic comparative analyses) are biologically sound.

Supplementary Figures

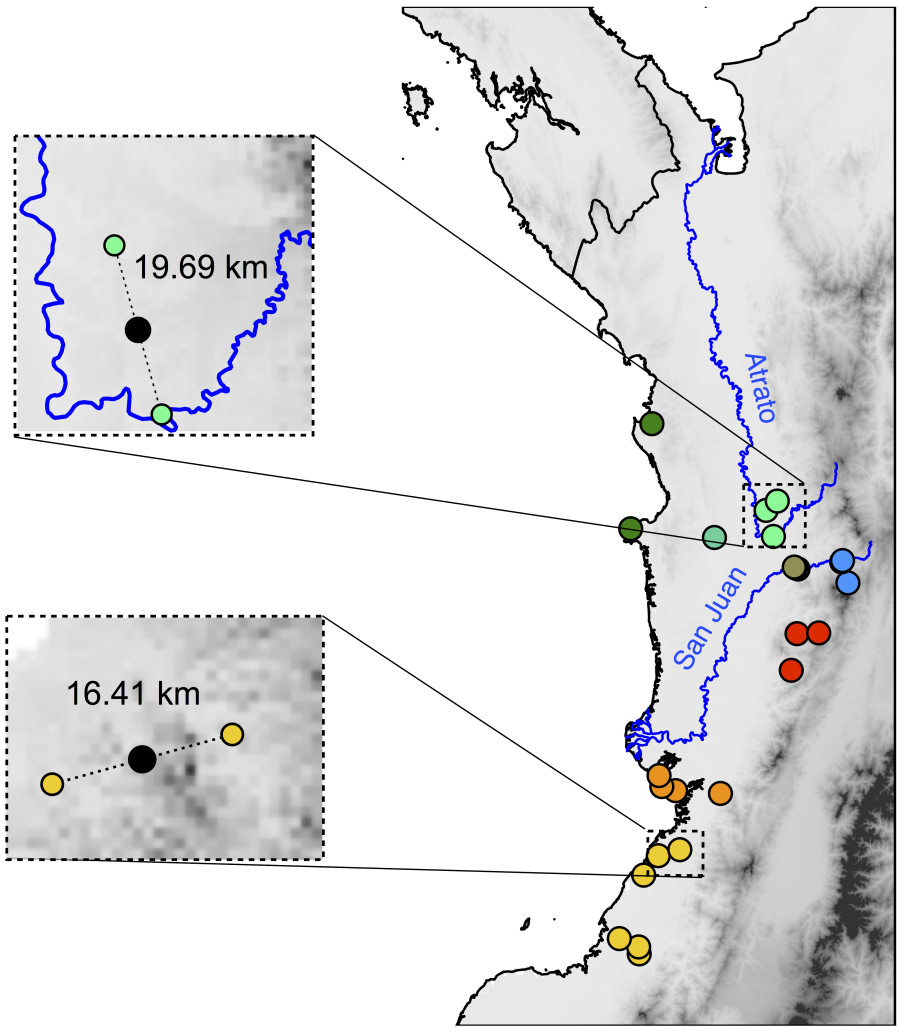

Figure 2.5: Localities joined into a single deme for locality-level analyses.

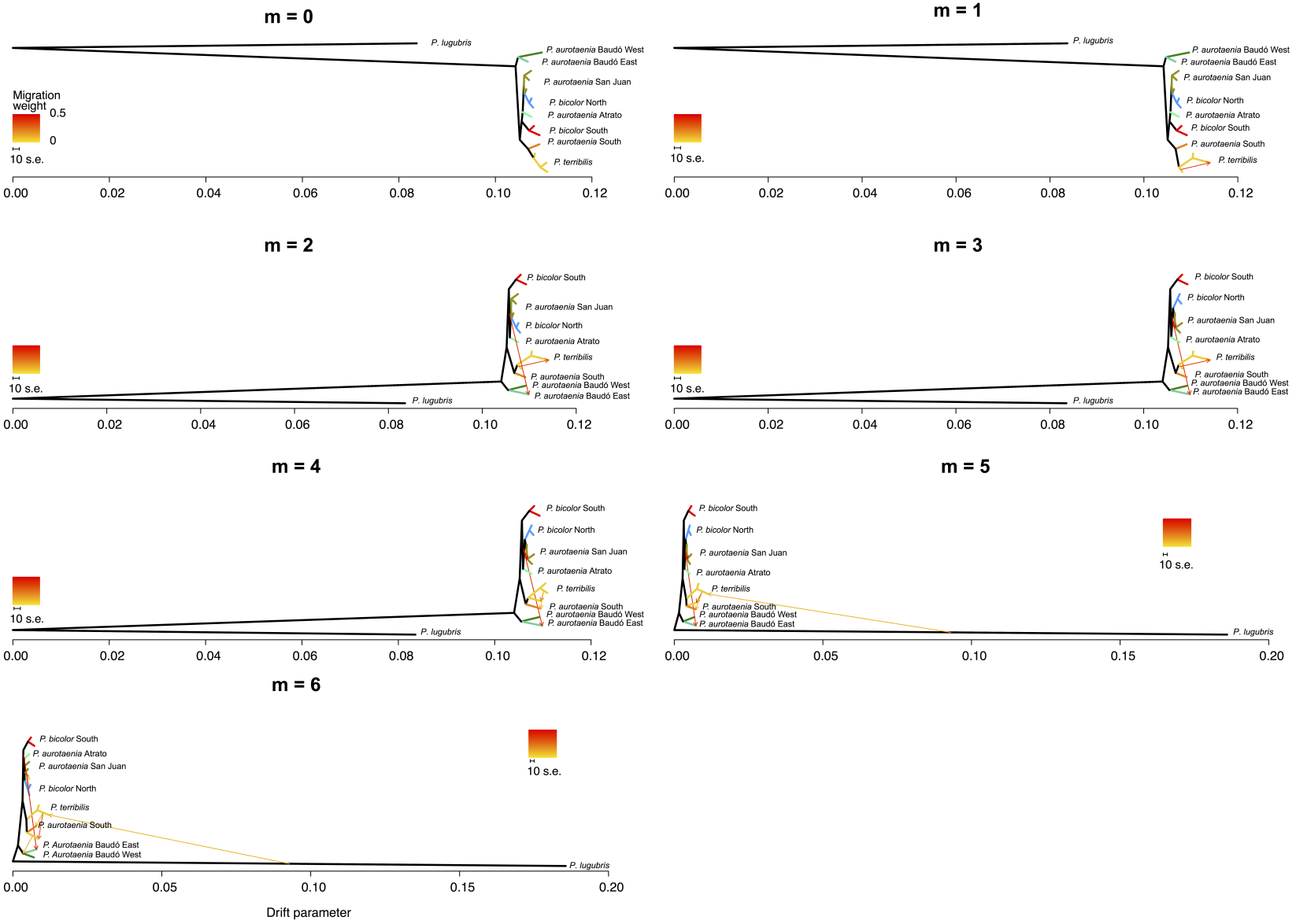

Figure 2.6: Results of Treemix analyses run with $m = 0 - 6$ migration edges.

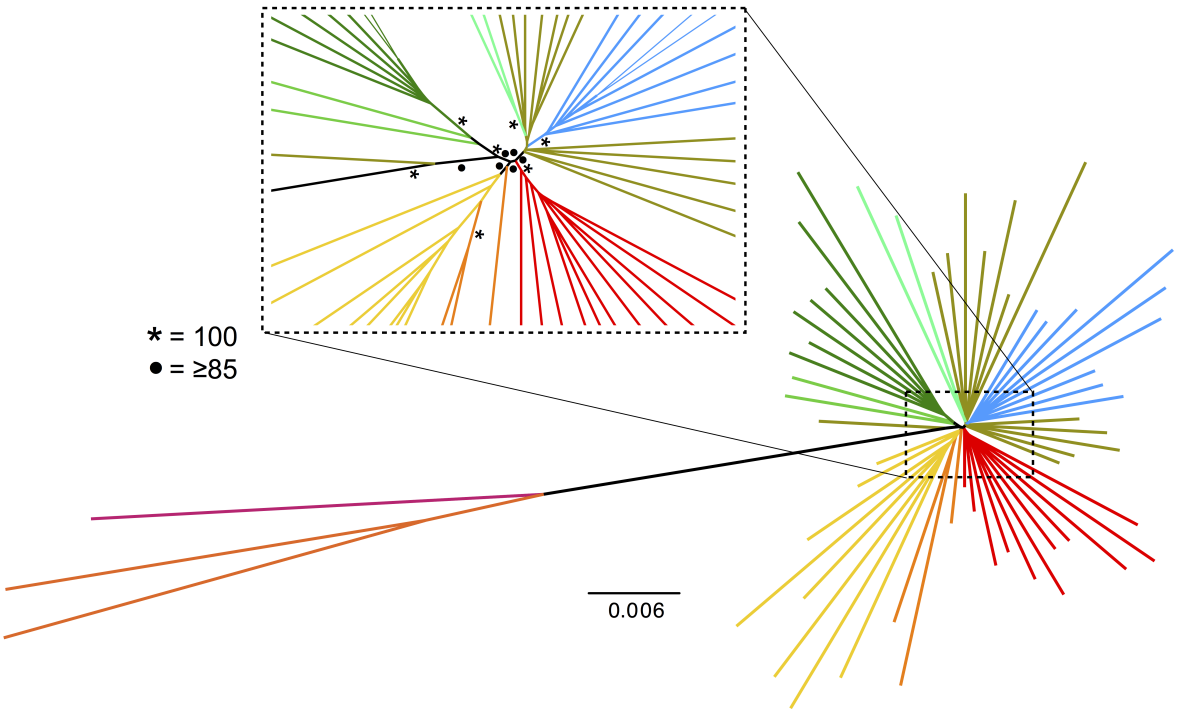

Figure 2.7: Minimum-evolution tree based on genetic distances.

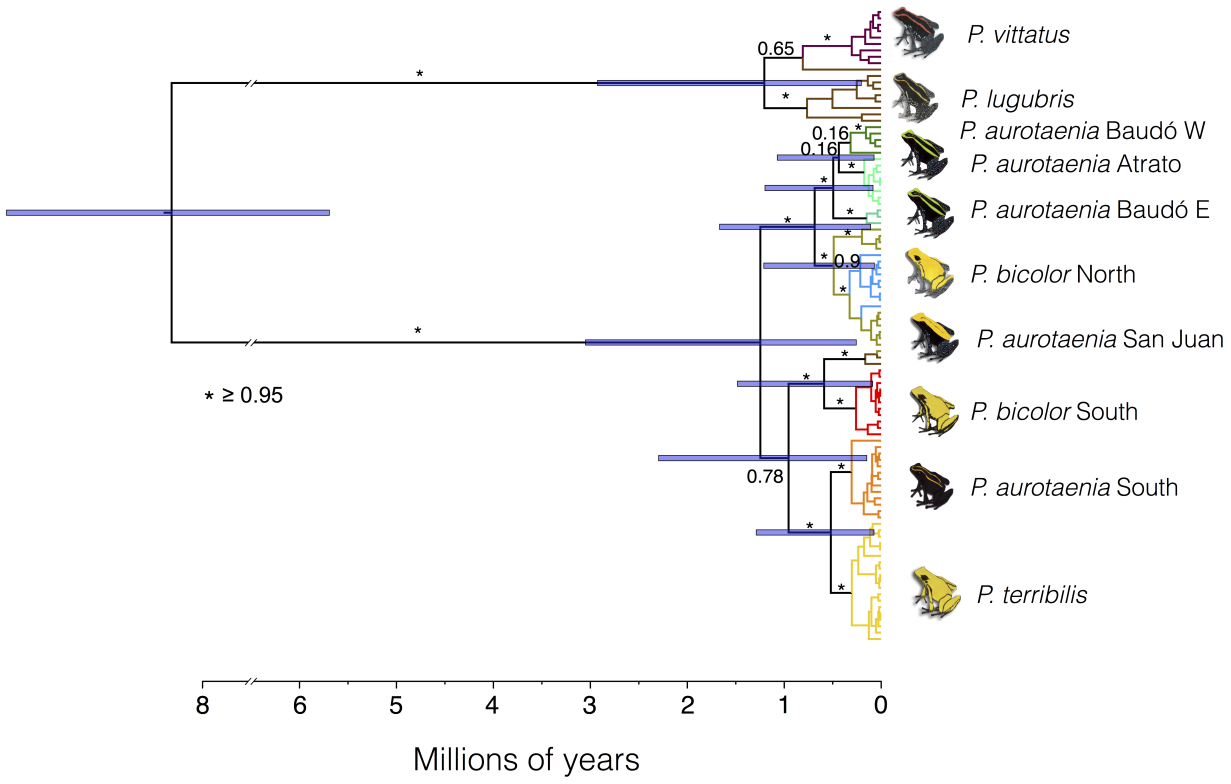

Figure 2.8: Mitochondrial DNA time tree inferred using BEAST 2.

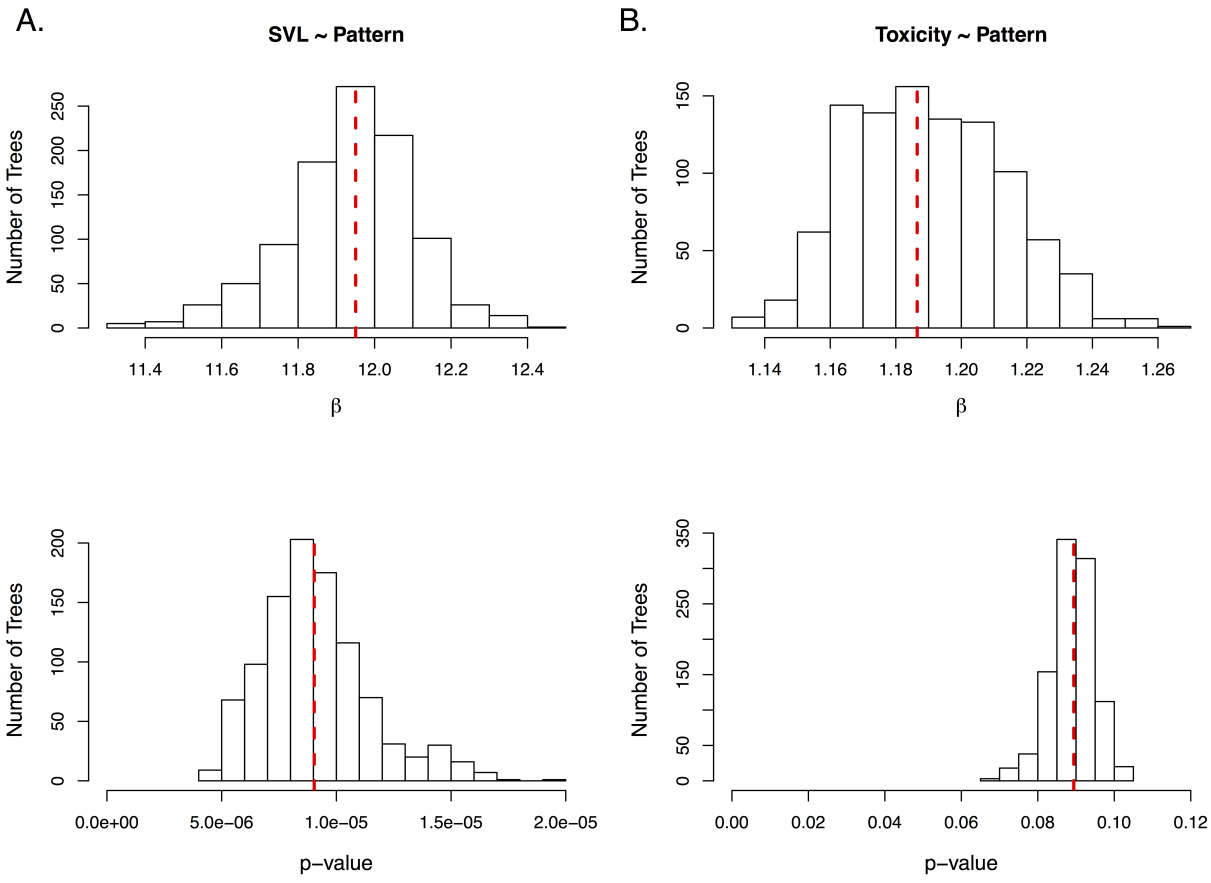

Figure 2.9: Results of phylogenetic comparative analyses run on 1000 randomly-drawn trees from the SNAPP posterior tree distribution.

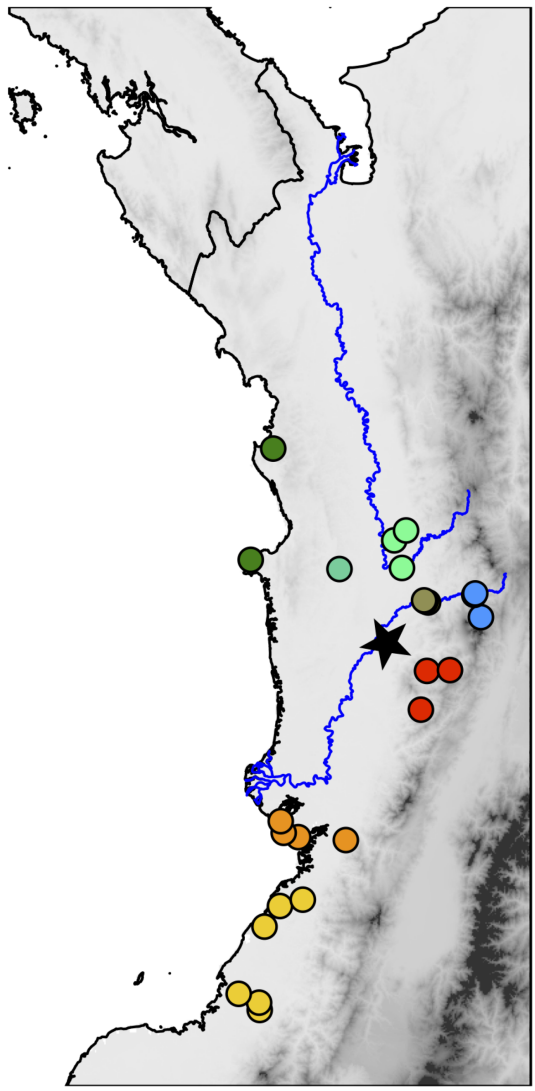

Figure 2.10: Map of the type locality of *P. aurotaenia*.

Supplementary Tables

Table 2.2: Information and accession numbers for samples used in this study. Locality data is limited to prevent illegal traffic. TBD = To be determined upon publication.

Voucher	**Nominal Species**	**OTU**	**Locality**	**mtDNA**			**Exon enrichment**	
				16S	**COI**	**Cytb**	**Number of sites**	**Mean Coverage**
NA	*P. aurotaenia*	NA	Pet Trade	DQ502153	DQ502855	DQ502586	-	-
TNHC64419	*P. aurotaenia*	NA	Pet Trade	EU342640	-	-	-	-
GECOH1596	*P. aurotaenia*	P. aurotaeina Baudo East	Colombia, Chocó, Cantón de San Pablo, La Victoria	-	-	-	4444086	1.65
GECOH1598	*P. aurotaenia*	P. aurotaeina Baudo East	Colombia, Chocó, Cantón de San Pablo, La Victoria	TBD	TBD	TBD	-	-
GECOH1599	*P. aurotaenia*	P. aurotaeina Baudo East	Colombia, Chocó, Cantón de San Pablo, La Victoria	-	-	-	4945758	1.74
GECOH1600	*P. aurotaenia*	P. aurotaeina Baudo East	Colombia, Chocó, Cantón de San Pablo, La Victoria	TBD	TBD	TBD	-	-
GECOH1602	*P. aurotaenia*	P. aurotaeina Baudo East	Colombia, Chocó, Cantón de San Pablo, La Victoria	TBD	TBD	TBD	-	-
GECOH1307	*P. aurotaenia*	*P. aurotaenia* Atrato	Colombia, Chocó, Lloró	TBD	TBD	TBD	5587321	1.99
GECOH1309	*P. aurotaenia*	*P. aurotaenia* Atrato	Colombia, Chocó, Lloró	TBD	TBD	TBD	-	-

Table 2.2 continued from previous page

Voucher	Nominal Species	OTU	Locality	mtDNA			Exon enrichment	
				16S	COI	Cytb	Number of sites	Mean Coverage
TNHCFS4990	*P. aurotaenia*	*P. aurotaenia* Atrato	Colombia, Chocó, Quibdó	HQ291005	-	HQ290582	-	-
GECOH1065	*P. aurotaenia*	*P. aurotaenia* Atrato	Colombia, Chocó, Quibdó, Pacurita	TBD	TBD	TBD	-	-
GECOH1066	*P. aurotaenia*	*P. aurotaenia* Atrato	Colombia, Chocó, Quibdó, Pacurita	TBD	TBD	TBD	-	-
GECOH1067	*P. aurotaenia*	*P. aurotaenia* Atrato	Colombia, Chocó, Quibdó, Pacurita	TBD	TBD	TBD	-	-
GECOH1068	*P. aurotaenia*	*P. aurotaenia* Atrato	Colombia, Chocó, Quibdó, Pacurita	TBD	TBD	TBD	-	-
TNHCFS4495	*P. aurotaenia*	*P. aurotaenia* Atrato	Colombia, Chocó, Quibdó, Pandó	-	-	-	5731874	2.17
GECOHF144	*P. aurotaenia*	*P. aurotaenia* Atrato	NA	TBD	TBD	TBD	-	-
NA	*P. aurotaenia*	*P. aurotaenia* Baudo West	Colombia, Chocó, Bahía Solano	AY263240	-	AF324033	-	-
GECOH1144	*P. aurotaenia*	*P. aurotaenia* Baudo West	Colombia, Chocó, Nuquí, Arusí	TBD	TBD	TBD	-	-
GECOH1145	*P. aurotaenia*	*P. aurotaenia* Baudo West	Colombia, Chocó, Nuquí, Arusí	TBD	TBD	TBD	-	-
GECOH1155	*P. aurotaenia*	*P. aurotaenia* Baudo West	Colombia, Chocó, Nuquí, Arusí	TBD	TBD	TBD	6495913	2.38
GECOH849	*P. aurotaenia*	*P. aurotaenia* Baudo West	Colombia, Chocó, Nuquí, Arusí	-	-	-	4386449	1.64

Table 2.2 continued from previous page

Voucher	Nominal Species	OTU	Locality	mtDNA			Exon enrichment	
				16S	COI	Cytb	Number of sites	Mean Coverage
GECOH850	*P. aurotaenia*	*P. aurotaenia* Baudo West	Colombia, Chocó, Nuquí, Arusí	-	-	-	5471012	1.83
GECOH851	*P. aurotaenia*	*P. aurotaenia* Baudo West	Colombia, Chocó, Nuquí, Arusí	-	-	-	5014550	1.81
GECOH852	*P. aurotaenia*	*P. aurotaenia* Baudo West	Colombia, Chocó, Nuquí, Arusí	-	-	-	4696628	1.73
GECOH855	*P. aurotaenia*	*P. aurotaenia* Baudo West	Colombia, Chocó, Nuquí, Arusí	-	-	-	4744002	1.72
GECOH856	*P. aurotaenia*	*P. aurotaenia* Baudo West	Colombia, Chocó, Nuquí, Arusí	-	-	-	7393605	2.84
GECOH858	*P. aurotaenia*	*P. aurotaenia* Baudo West	Colombia, Chocó, Nuquí, Arusí	TBD	TBD	TBD	-	-
GECOH704	*P. aurotaenia*	*P. aurotaenia* San Juan	Colombia, Chocó, Playa de Oro	TBD	TBD	TBD	5995802	2.21
GECOH705	*P. aurotaenia*	*P. aurotaenia* San Juan	Colombia, Chocó, Playa de Oro	TBD	TBD	TBD	4426676	1.68
GECOH706	*P. aurotaenia*	*P. aurotaenia* San Juan	Colombia, Chocó, Playa de Oro	TBD	TBD	TBD	4968702	1.8
GECOH2102	*P. aurotaenia*	*P. aurotaenia* San Juan	Colombia, Chocó, Tadó, Bochoroma	-	-	-	5454509	1.96
GECOH1482	*P. aurotaenia*	*P. aurotaenia* San Juan	Colombia, Chocó, Tadó, Bochoromá	TBD	TBD	TBD	-	-
GECOH1483	*P. aurotaenia*	*P. aurotaenia* San Juan	Colombia, Chocó, Tadó, Bochoromá	TBD	TBD	TBD	-	-

Table 2.2 continued from previous page

Voucher	Nominal Species	OTU	Locality	mtDNA			Exon enrichment	
				16S	COI	Cytb	Number of sites	Mean Coverage
GECOH1484	*P. aurotaenia*	*P. aurotaenia* San Juan	Colombia, Chocó, Tadó, Bochoromá	TBD	TBD	TBD	-	-
GECOH1610	*P. aurotaenia*	*P. aurotaenia* San Juan	Colombia, Chocó, Tadó, Bochoromá	-	-	-	4854908	1.75
GECOH2155	*P. aurotaenia*	*P. aurotaenia* San Juan	Colombia, Chocó, Tadó, Bochoromá	-	-	-	3639367	1.5
GECOH2156	*P. aurotaenia*	*P. aurotaenia* San Juan	Colombia, Chocó, Tadó, Bochoromá	-	-	-	4344879	1.67
GECOH2157	*P. aurotaenia*	*P. aurotaenia* San Juan	Colombia, Chocó, Tadó, Bochoromá	-	-	-	3541775	1.48
GECOH2158	*P. aurotaenia*	*P. aurotaenia* San Juan	Colombia, Chocó, Tadó, Bochoromá	-	-	-	3324869	1.5
GECOH2159	*P. aurotaenia*	*P. aurotaenia* San Juan	Colombia, Chocó, Tadó, Bochoromá	-	-	-	3786184	1.52
GECOH707	*P. aurotaenia*	*P. aurotaenia* San Juan	Colombia, Chocó, Tadó, Bochoromá	TBD	TBD	TBD	6567520	2.41
GECOH708	*P. aurotaenia*	*P. aurotaenia* San Juan	Colombia, Chocó, Tadó, Bochoromá	TBD	TBD	TBD	4366779	1.64
GECOH709	*P. aurotaenia*	*P. aurotaenia* San Juan	Colombia, Chocó, Tadó, Bochoromá	TBD	TBD	TBD	3704324	1.57
GECOH738	*P. aurotaenia*	*P. aurotaenia* San Juan	Colombia, Chocó, Tadó, Bochoromá	TBD	TBD	TBD	-	-
NA	*P. aurotaenia*	*P. aurotaenia* San Juan	Colombia, Chocó, Tadó, Bochoromá	AY263241	-	AF324034	-	-

Table 2.2 continued from previous page

Voucher	Nominal Species	OTU	Locality	mtDNA			Exon enrichment	
				16S	COI	Cytb	Number of sites	Mean Coverage
GECOH1189	*P. aurotaenia*	*P. aurotaenia* south	Colombia, Valle del Cauca, Buenaventura	TBD	TBD	TBD	-	-
GECOH1190	*P. aurotaenia*	*P. aurotaenia* south	Colombia, Valle del Cauca, Buenaventura, Maguipi	TBD	TBD	TBD	-	-
GECOH1191	*P. aurotaenia*	*P. aurotaenia* south	Colombia, Valle del Cauca, Buenaventura, Maguipi	TBD	TBD	TBD	-	-
GECOH1192	*P. aurotaenia*	*P. aurotaenia* south	Colombia, Valle del Cauca, Buenaventura, Maguipi	TBD	TBD	TBD	-	-
GECOH1193	*P. aurotaenia*	*P. aurotaenia* south	Colombia, Valle del Cauca, Buenaventura, Maguipi	TBD	TBD	TBD	-	-
GECOH1194	*P. aurotaenia*	*P. aurotaenia* south	Colombia, Valle del Cauca, Buenaventura, Maguipi	TBD	TBD	TBD	-	-

Table 2.2 continued from previous page

Voucher	Nominal Species	OTU	Locality	mtDNA			Exon enrichment	
				16S	COI	Cytb	Number of sites	Mean Coverage
GECOH1168	*P. aurotaenia*	*P. aurotaenia* south	Colombia, Valle del Cauca, Buenaventura, Pianguita	TBD	TBD	TBD	-	-
GECOH1169	*P. aurotaenia*	*P. aurotaenia* south	Colombia, Valle del Cauca, Buenaventura, Pianguita	TBD	TBD	TBD	-	-
GECOH1170	*P. aurotaenia*	*P. aurotaenia* south	Colombia, Valle del Cauca, Buenaventura, Pianguita	TBD	TBD	TBD	-	-
GECOH1171	*P. aurotaenia*	*P. aurotaenia* south	Colombia, Valle del Cauca, Buenaventura, Pianguita	TBD	TBD	TBD	-	-
GECOH1172	*P. aurotaenia*	*P. aurotaenia* south	Colombia, Valle del Cauca, Buenaventura, Pianguita	TBD	TBD	TBD	-	-
GECOH1173	*P. aurotaenia*	*P. aurotaenia* south	Colombia, Valle del Cauca, Buenaventura, Pianguita	TBD	TBD	TBD	1538131	1.26

Table 2.2 continued from previous page

Voucher	Nominal Species	OTU	Locality	mtDNA			Exon enrichment	
				16S	COI	Cytb	Number of sites	Mean Coverage
GECOH664	*P. aurotaenia*	*P. aurotaenia* south	Colombia, Valle del cauca, Buenaventura, Pianguita	-	-	-	2645045	1.4
GECOH667	*P. aurotaenia*	*P. aurotaenia* south	Colombia, Valle del cauca, Buenaventura, Pianguita	-	-	-	2361394	1.36
GECOH668	*P. aurotaenia*	*P. aurotaenia* south	Colombia, Valle del cauca, Buenaventura, Pianguita	-	-	-	5264012	2.05
GECOH672	*P. aurotaenia*	*P. aurotaenia* south	Colombia, Valle del cauca, Buenaventura, Pianguita	TBD	TBD	TBD	4782244	1.8
NA	*P. bicolor*	NA	Pet trade	DQ502181	DQ502884	DQ502617	-	-
TNHC62488	*P. bicolor*	NA	Pet Trade	AY364580	-	-	-	-
GECOH1174	*P. bicolor*	*P. bicolor* North	Colombia, Risaralda, Pueblo Rico	TBD	TBD	TBD	-	-
GECOH1176	*P. bicolor*	*P. bicolor* North	Colombia, Risaralda, Pueblo Rico	TBD	TBD	TBD	-	-

Table 2.2 continued from previous page

Voucher	Nominal Species	OTU	Locality	mtDNA			Exon enrichment	
				16S	COI	Cytb	Number of sites	Mean Coverage
GECOH1177	*P. bicolor*	*P. bicolor* North	Colombia, Risaralda, Pueblo Rico	TBD	TBD	TBD	-	-
GECOH1178	*P. bicolor*	*P. bicolor* North	Colombia, Risaralda, Pueblo Rico	TBD	TBD	TBD	-	-
GECOH1179	*P. bicolor*	*P. bicolor* North	Colombia, Risaralda, Pueblo Rico	-	-	-	2287765	1.31
GECOH1485	*P. bicolor*	*P. bicolor* North	Colombia, Risaralda, Pueblo Rico	TBD	TBD	TBD	-	-
GECOH2083	*P. bicolor*	*P. bicolor* North	Colombia, Risaralda, Pueblo Rico	-	-	-	6931235	2.51
GECOH2085	*P. bicolor*	*P. bicolor* North	Colombia, Risaralda, Pueblo Rico	-	-	-	4536535	1.67
GECOH2086	*P. bicolor*	*P. bicolor* North	Colombia, Risaralda, Pueblo Rico	-	-	-	4578766	1.69
GECOH2095	*P. bicolor*	*P. bicolor* North	Colombia, Risaralda, Pueblo Rico	-	-	-	4270892	1.63

Table 2.2 continued from previous page

Voucher	Nominal Species	OTU	Locality	mtDNA			Exon enrichment	
				16S	COI	Cytb	Number of sites	Mean Coverage
GECOH2096	*P. bicolor*	*P. bicolor* North	Colombia, Risaralda, Pueblo Rico	-	-	-	7289836	2.92
GECOH522	*P. bicolor*	*P. bicolor* North	Colombia, Risaralda, Pueblo Rico	TBD	TBD	TBD	-	-
GECOH523	*P. bicolor*	*P. bicolor* North	Colombia, Risaralda, Pueblo Rico	TBD	TBD	TBD	-	-
GECOH701	*P. bicolor*	*P. bicolor* North	Colombia, Risaralda, Pueblo Rico	TBD	TBD	TBD	-	-
GECOH732	*P. bicolor*	*P. bicolor* North	Colombia, Risaralda, Pueblo Rico	-	-	-	6256585	2.17
GECOH733	*P. bicolor*	*P. bicolor* North	Colombia, Risaralda, Pueblo Rico	-	-	-	4846323	1.8
GECOH734	*P. bicolor*	*P. bicolor* North	Colombia, Risaralda, Pueblo Rico	-	-	-	4122907	1.61
GECOH735	*P. bicolor*	*P. bicolor* North	Colombia, Risaralda, Pueblo Rico	-	-	-	3826662	1.56

Table 2.2 continued from previous page

Voucher	Nominal Species	OTU	Locality	mtDNA			Exon enrichment	
				16S	COI	Cytb	Number of sites	Mean Coverage
NA	*P. bicolor*	*P. bicolor* south	Colombia, Chocó	AF128577	-	AF128579	-	-
GECOH1039	*P. bicolor*	*P. bicolor* south	Colombia, Chocó, San José del Palmar	TBD	TBD	TBD	-	-
GECOH2062	*P. bicolor*	*P. bicolor* south	Colombia, Chocó, San José del Palmar	TBD	TBD	TBD	5227721	1.96
GECOH2063	*P. bicolor*	*P. bicolor* south	Colombia, Chocó, San José del Palmar	-	-	-	5642128	2.03
GECOH2064	*P. bicolor*	*P. bicolor* south	Colombia, Chocó, San José del Palmar	-	-	-	5193444	1.95
GECOH2065	*P. bicolor*	*P. bicolor* south	Colombia, Chocó, San José del Palmar	-	-	-	3126237	1.58
GECOH2066	*P. bicolor*	*P. bicolor* south	Colombia, Chocó, San José del Palmar	TBD	TBD	TBD	-	-
GECOH2067	*P. bicolor*	*P. bicolor* south	Colombia, Chocó, San José del Palmar	-	-	-	1802599	1.25

Table 2.2 continued from previous page

Voucher	Nominal Species	OTU	Locality	mtDNA			Exon enrichment	
				16S	COI	Cytb	Number of sites	Mean Coverage
GECOH2068	*P. bicolor*	*P. bicolor* south	Colombia, Chocó, San José del Palmar	-	-	-	1149310	1.25
GECOH2069	*P. bicolor*	*P. bicolor* south	Colombia, Chocó, San José del Palmar	-	-	-	4625511	1.76
GECOH2070	*P. bicolor*	*P. bicolor* south	Colombia, Chocó, San José del Palmar	TBD	TBD	TBD	-	-
GECOH2074	*P. bicolor*	*P. bicolor* south	Colombia, Chocó, San José del Palmar	-	-	-	1603599	1.24
GECOH2076	*P. bicolor*	*P. bicolor* south	Colombia, Chocó, San José del Palmar	-	-	-	4789516	1.73
NA	*P. bicolor*	*P. bicolor* south	Colombia, Chocó, San José del Palmar	-	-	AF324035	-	-
GECOH2128	*P. bicolor*	*P. bicolor* south	Colombia, Chocó, Sipí	TBD	TBD	TBD	2777725	1.38
GECOH2129	*P. bicolor*	*P. bicolor* south	Colombia, Chocó, Sipí	TBD	TBD	TBD	5830851	2.06
GECOH2130	*P. bicolor*	*P. bicolor* south	Colombia, Chocó, Sipí	TBD	TBD	TBD	4318890	1.62

Table 2.2 continued from previous page

Voucher	Nominal Species	OTU	Locality	mtDNA			Exon enrichment	
				16S	COI	Cytb	Number of sites	Mean Coverage
GECOH2131	*P. bicolor*	*P. bicolor* south	Colombia, Chocó, Sipí	-	-	-	4935062	1.77
GECOH2132	*P. bicolor*	*P. bicolor* south	Colombia, Chocó, Sipí	TBD	TBD	TBD	-	-
GECOH963	*P. lugubris*	*P. lugubris*	Costa Rica, Limón, Puerto Viejo de Talamanca	TBD	TBD	TBD	6832241	2.46
GECOH964	*P. lugubris*	*P. lugubris*	Costa Rica, Limón, Puerto Viejo de Talamanca	TBD	TBD	TBD	6552110	2.54
GECOH965	*P. lugubris*	*P. lugubris*	Costa Rica, Limón, Puerto Viejo de Talamanca	TBD	TBD	TBD	-	-
GECOH966	*P. lugubris*	*P. lugubris*	Costa Rica, Limón, Puerto Viejo de Talamanca	TBD	TBD	TBD	-	-
KRL1735	*P. lugubris*	*P. lugubris*	El Cope, Panama	FJ784587	FJ766769	-	-	-
OMNH33325	*P. lugubris*	*P. lugubris*	Nicaragua, San Juan River, Isla de Diamante	DQ502061	DQ502783	DQ502492	-	-
ZFMK69881	*P. lugubris*	*P. lugubris*	Panama	AF124105		AF324038	-	-
USNMFS195116	*P. lugubris*	*P. lugubris*	Panamá, Bocas del Toro, Isla Popa	DQ283043	DQ502753	DQ502456	-	-

Table 2.2 continued from previous page

Voucher	Nominal Species	OTU	Locality	mtDNA			Exon enrichment	
				16S	COI	Cytb	Number of sites	Mean Coverage
NA	*P. lugubris*	*P. lugubris*	Panama, Panama, Bocas del Toro	AF128575	-	AF128576	-	-
GECOH301	*P. terribilis*	*P. terribilis*	Captive bred	TBD	TBD	TBD	-	-
GECOH1180	*P. terribilis*	*P. terribilis*	Colombia, Cauca, Timbiquí, La Brea	TBD	TBD	TBD	-	-
GECOH1181	*P. terribilis*	*P. terribilis*	Colombia, Cauca, Timbiquí, La Brea	TBD	TBD	TBD	6577024	2.39
GECOH1182	*P. terribilis*	*P. terribilis*	Colombia, Cauca, Timbiquí, La Brea	TBD	TBD	TBD	2858150	1.36
GECOH1183	*P. terribilis*	*P. terribilis*	Colombia, Cauca, Timbiquí, La Brea	TBD	TBD	TBD	3681267	1.48
NA	*P. terribilis*	*P. terribilis*	Colombia, Cauca, Timbiquí, La Brea	-	-	AF324037	-	-
AMNHA118566	*P. terribilis*	*P. terribilis*	Colombia, Cauca, Timbiquí, Quebrada Guanguí	DQ502157	DQ502861	DQ502593	-	-
NA	*P. terribilis*	*P. terribilis*	Colombia, Cauca, Timbiquí, Saija river mouth	AF124133	-	AF324036	-	-
GECOH1163	*P. terribilis*	*P. terribilis*	Colombia, Valle del Cauca, Buenaventura, Río Cajambre	TBD	JX887787	TBD	-	-

Table 2.2 continued from previous page

Voucher	Nominal Species	OTU	Locality	mtDNA			Exon enrichment	
				16S	COI	Cytb	Number of sites	Mean Coverage
GECOH1164	*P. terribilis*	*P. terribilis*	Colombia, Valle del Cauca, Buenaventura, Río Cajambre	TBD	TBD	TBD	-	-
GECOH1165	*P. terribilis*	*P. terribilis*	Colombia, Valle del Cauca, Buenaventura, Río Cajambre	TBD	TBD	TBD	-	-
GECOH646	*P. terribilis*	*P. terribilis*	Colombia, Valle del Cauca, Buenaventura, Río Cajambre	TBD	TBD	TBD	-	-
GECOH697	*P. terribilis*	*P. terribilis*	Colombia, Valle del Cauca, Buenaventura, Río Cajambre	-	-	-	6518925	2.77
GECOH355	*P. terribilis*	*P. terribilis*	Colombia, Valle del Cauca, Buenaventura, Río Naya	TBD	JX887786	TBD	-	-
GECOH356	*P. terribilis*	*P. terribilis*	Colombia, Valle del Cauca, Buenaventura, Río Naya	TBD	TBD	TBD	-	-

Table 2.2 continued from previous page

Voucher	Nominal Species	OTU	Locality	mtDNA			Exon enrichment	
				16S	COI	Cytb	Number of sites	Mean Coverage
GECOH432	*P. terribilis*	*P. terribilis*	Colombia, Valle del Cauca, Buenaventura, Río Naya	-	-	-	7090596	2.42
GECOH434	*P. terribilis*	*P. terribilis*	Colombia, Valle del Cauca, Buenaventura, Río Naya	TBD	JX887784	TBD	-	-
GECOH435	*P. terribilis*	*P. terribilis*	Colombia, Valle del Cauca, Buenaventura, Río Naya	TBD	TBD	TBD	4805522	1.8
GECOH436	*P. terribilis*	*P. terribilis*	Colombia, Valle del Cauca, Buenaventura, Río Naya	-	-	-	7125232	3.04
GECOH643	*P. terribilis*	*P. terribilis*	Colombia, Valle del Cauca, Buenaventura, Río Yurumanguí	TBD	JX887785	TBD	5894325	2.09
NA	*P. terribilis*	*P. terribilis*	Pet Trade	DQ502180	DQ502883	DQ502616	-	-
TNHC64420	*P. terribilis*	*P. terribilis*	Pet Trade	HQ291006	-	HQ290583	-	-
ZMFK69882	*P. vittatus*	*P. vittatus*	Costa Rica	AF124134	-	AF324039	-	-

Table 2.2 continued from previous page

Voucher	Nominal Species	OTU	Locality	mtDNA			Exon enrichment	
				16S	COI	Cytb	Number of sites	Mean Coverage
NA	*P. vittatus*	*P. vittatus*	Costa Rica, Corcovado	AF128580	-	AF128582	-	-
GECOH970	*P. vittatus*	*P. vittatus*	Costa Rica, Puntarenas, Dos Brazos	TBD	TBD	TBD	-	-
GECOH971	*P. vittatus*	*P. vittatus*	Costa Rica, Puntarenas, Dos Brazos	TBD	TBD	TBD	-	-
GECOH976	*P. vittatus*	*P. vittatus*	Costa Rica, Puntarenas, Dos Brazos	TBD	TBD	TBD	6070250	2.27
GECOH979	*P. vittatus*	*P. vittatus*	Costa Rica, Puntarenas, Dos Brazos	TBD	TBD	TBD	-	-
1026	*P. vittatus*	*P. vittatus*	NA	EF107173			-	-
NA	*P. vittatus*	*P. vittatus*	Pet Trade	DQ502152	DQ502854	DQ502585	-	-
TNHC64421	*P. vittatus*	*P. vittatus*	Puntarenas, Costa Rica	EU342635	-	-	-	-

Table 2.3: Results of likelihood ratio tests performed on Treemix analyses run with m = 0-6 migration edges.

m	**Log-likelihood**	**Likelihood ratio statistic**	**P-value** ($df = 2$)
0	935.85900	-	-
1	944.09800	16.478	**0.00026**
2	952.81000	17.424	**0.00016**
3	955.80100	5.982	0.05024
4	958.31800	5.034	0.08070
5	960.25300	3.870	0.14442
6	962.12200	3.738	0.15428

Table 2.4: Information and snout-to-vent lengths of museum specimens measured for comparative analyses. Collection abbreviations are as follows: AMNH: American Museum of Natural History; ANDES-A: Museo de Historia Natural ANDES, COLZOOCH: Colección de Zoología, Universidad Tecnológica del Chocó; FMNH: field Museum of Natural History; IAvH: Instituto Alexander von Humboldt; USNM: Smithsonian National Museum of Natural History; DMV: Daniel Mejía-Vargas field Series, GECOH: Adolfo Amézquita Field Series; RMP: Roberto Márquez field series.

Voucher	**Species**	**OTU**	**SVL (mm)**
COLZOOCH-H0379	*P. aurotaenia*	*P. aurotaenia* Atrato	26.18
COLZOOCH-H0432	*P. aurotaenia*	*P. aurotaenia* Atrato	27.17
COLZOOCH-H0561	*P. aurotaenia*	*P. aurotaenia* Atrato	31.57
TNHC-FS4996	*P. aurotaenia*	*P. aurotaenia* Atrato	27.78
COLZOOCH-H0096	*P. aurotaenia*	*P. aurotaenia* Baudó East	24.97
COLZOOCH-H0097	*P. aurotaenia*	*P. aurotaenia* Baudó East	24.32
COLZOOCH-H0839	*P. aurotaenia*	*P. aurotaenia* Baudó East	26.93
ANDES-A2434	*P. aurotaenia*	*P. aurotaenia* Baudó West	19.13

Table 2.4 continued from previous page

Voucher	**Species**	**OTU**	**SVL (mm)**
IAvH-Am-6570	*P. aurotaenia*	*P. aurotaenia* Baudó West	27.32
AMNH102684	*P. aurotaenia*	*P. aurotaenia* San Juan	29.89
AMNH102685	*P. aurotaenia*	*P. aurotaenia* San Juan	32.63
AMNH114040	*P. aurotaenia*	*P. aurotaenia* San Juan	31.64
AMNH118532	*P. aurotaenia*	*P. aurotaenia* San Juan	27.81
AMNH118533	*P. aurotaenia*	*P. aurotaenia* San Juan	29.39
AMNH118534	*P. aurotaenia*	*P. aurotaenia* San Juan	27.83
AMNH118535	*P. aurotaenia*	*P. aurotaenia* San Juan	30.42
AMNH118537	*P. aurotaenia*	*P. aurotaenia* San Juan	26.79
AMNH118538	*P. aurotaenia*	*P. aurotaenia* San Juan	26.66
AMNH118539	*P. aurotaenia*	*P. aurotaenia* San Juan	27.55
AMNH118540	*P. aurotaenia*	*P. aurotaenia* San Juan	26.03
AMNH161109	*P. aurotaenia*	*P. aurotaenia* San Juan	28.6
AMNH161110	*P. aurotaenia*	*P. aurotaenia* San Juan	28.22
AMNH161111	*P. aurotaenia*	*P. aurotaenia* San Juan	27.17
AMNH85238	*P. aurotaenia*	*P. aurotaenia* San Juan	26.34
AMNH85239	*P. aurotaenia*	*P. aurotaenia* San Juan	26.69
AMNH85240	*P. aurotaenia*	*P. aurotaenia* San Juan	26.22
AMNH85241	*P. aurotaenia*	*P. aurotaenia* San Juan	24.65
AMNH85242	*P. aurotaenia*	*P. aurotaenia* San Juan	21.28
AMNH85244	*P. aurotaenia*	*P. aurotaenia* San Juan	26.18
AMNH85245	*P. aurotaenia*	*P. aurotaenia* San Juan	25.25
AMNH87153	*P. aurotaenia*	*P. aurotaenia* San Juan	25.07

Table 2.4 continued from previous page

Voucher	**Species**	**OTU**	**SVL (mm)**
AMNH87154	*P. aurotaenia*	*P. aurotaenia* San Juan	26.08
AMNH87155	*P. aurotaenia*	*P. aurotaenia* San Juan	24.82
AMNH87156	*P. aurotaenia*	*P. aurotaenia* San Juan	25.41
AMNH87157	*P. aurotaenia*	*P. aurotaenia* San Juan	29.51
AMNH87158	*P. aurotaenia*	*P. aurotaenia* San Juan	26.23
AMNH87159	*P. aurotaenia*	*P. aurotaenia* San Juan	25.42
AMNH87160	*P. aurotaenia*	*P. aurotaenia* San Juan	29.65
AMNH87161	*P. aurotaenia*	*P. aurotaenia* San Juan	28.29
AMNH87162	*P. aurotaenia*	*P. aurotaenia* San Juan	30.56
AMNH87163	*P. aurotaenia*	*P. aurotaenia* San Juan	26.68
AMNH87164	*P. aurotaenia*	*P. aurotaenia* San Juan	27.16
AMNH87165	*P. aurotaenia*	*P. aurotaenia* San Juan	28.78
AMNH87166	*P. aurotaenia*	*P. aurotaenia* San Juan	28.14
AMNH87169	*P. aurotaenia*	*P. aurotaenia* San Juan	28.61
AMNH87170	*P. aurotaenia*	*P. aurotaenia* San Juan	31.53
AMNH87171	*P. aurotaenia*	*P. aurotaenia* San Juan	26.64
AMNH87172	*P. aurotaenia*	*P. aurotaenia* San Juan	27.02
AMNH87173	*P. aurotaenia*	*P. aurotaenia* San Juan	27.15
ANDES-A3727	*P. aurotaenia*	*P. aurotaenia* San Juan	24.92
ANDES-A3728	*P. aurotaenia*	*P. aurotaenia* San Juan	29.18
COLZOOCH-H1021	*P. aurotaenia*	*P. aurotaenia* San Juan	29.07
DMV018	*P. aurotaenia*	*P. aurotaenia* San Juan	27.28
DMV030	*P. aurotaenia*	*P. aurotaenia* San Juan	30.19

Table 2.4 continued from previous page

Voucher	**Species**	**OTU**	**SVL (mm)**
GECOH2663	*P. aurotaenia*	*P. aurotaenia* San Juan	26.56
IAvH-Am-4134	*P. aurotaenia*	*P. aurotaenia* San Juan	24.59
IAvH-Am-4135	*P. aurotaenia*	*P. aurotaenia* San Juan	25.17
IAvH-Am-4136	*P. aurotaenia*	*P. aurotaenia* San Juan	26.31
IAvH-Am-4137	*P. aurotaenia*	*P. aurotaenia* San Juan	24.52
IAvH-Am-4138	*P. aurotaenia*	*P. aurotaenia* San Juan	25.16
IAvH-Am-4139	*P. aurotaenia*	*P. aurotaenia* San Juan	25.39
IAvH-Am-4141	*P. aurotaenia*	*P. aurotaenia* San Juan	25.85
IAvH-Am-4142	*P. aurotaenia*	*P. aurotaenia* San Juan	26.51
RMP001	*P. aurotaenia*	*P. aurotaenia* San Juan	25.71
RMP016	*P. aurotaenia*	*P. aurotaenia* San Juan	25.03
USNM146630	*P. aurotaenia*	*P. aurotaenia* San Juan	28.08
USNM146631	*P. aurotaenia*	*P. aurotaenia* San Juan	27.8
USNM147204	*P. aurotaenia*	*P. aurotaenia* San Juan	31.42
USNM147205	*P. aurotaenia*	*P. aurotaenia* San Juan	27.34
USNM147207	*P. aurotaenia*	*P. aurotaenia* San Juan	26.22
USNM150632	*P. aurotaenia*	*P. aurotaenia* San Juan	28.33
USNM150633	*P. aurotaenia*	*P. aurotaenia* San Juan	26.65
USNM150634	*P. aurotaenia*	*P. aurotaenia* San Juan	27.06
USNM150635	*P. aurotaenia*	*P. aurotaenia* San Juan	26.03
USNM150636	*P. aurotaenia*	*P. aurotaenia* San Juan	25.45
USNM150637	*P. aurotaenia*	*P. aurotaenia* San Juan	22.37
USNM150638	*P. aurotaenia*	*P. aurotaenia* San Juan	29.84

Table 2.4 continued from previous page

Voucher	Species	OTU	SVL (mm)
USNM150639	*P. aurotaenia*	*P. aurotaenia* San Juan	31.7
USNM150640	*P. aurotaenia*	*P. aurotaenia* San Juan	26.28
USNM150641	*P. aurotaenia*	*P. aurotaenia* San Juan	26.78
USNM158881	*P. aurotaenia*	*P. aurotaenia* San Juan	24.62
ANDES-A3666	*P. aurotaenia*	*P. aurotaenia* South	23.76
ANDES-A3667	*P. aurotaenia*	*P. aurotaenia* South	23.99
ANDES-A3668	*P. aurotaenia*	*P. aurotaenia* South	25.31
ANDES-A3669	*P. aurotaenia*	*P. aurotaenia* South	24.37
ANDES-A3670	*P. aurotaenia*	*P. aurotaenia* South	27.01
ANDES-A3671	*P. aurotaenia*	*P. aurotaenia* South	24.17
ANDES-A3672	*P. aurotaenia*	*P. aurotaenia* South	23.49
USNM145105	*P. aurotaenia*	*P. aurotaenia* South	24.39
USNM145106	*P. aurotaenia*	*P. aurotaenia* South	23.28
USNM149725	*P. aurotaenia*	*P. aurotaenia* South	24.05
USNM149726	*P. aurotaenia*	*P. aurotaenia* South	23.62
USNM149727	*P. aurotaenia*	*P. aurotaenia* South	25.86
AMNH114038	*P. bicolor*	*P. bicolor* North	41.72
AMNH114039	*P. bicolor*	*P. bicolor* North	37.97
AMNH9224	*P. bicolor*	*P. bicolor* North	34.39
AMNH98209	*P. bicolor*	*P. bicolor* North	42.28
AMNH98210	*P. bicolor*	*P. bicolor* North	34.74
AMNH98211	*P. bicolor*	*P. bicolor* North	35.81
AMNH98212	*P. bicolor*	*P. bicolor* North	35.18

Table 2.4 continued from previous page

Voucher	Species	OTU	SVL (mm)
AMNH98213	*P. bicolor*	*P. bicolor* North	37.41
AMNH98214	*P. bicolor*	*P. bicolor* North	37.24
AMNH98215	*P. bicolor*	*P. bicolor* North	36.07
AMNH98216	*P. bicolor*	*P. bicolor* North	40.2
AMNH98217	*P. bicolor*	*P. bicolor* North	38.16
AMNH98218	*P. bicolor*	*P. bicolor* North	38.13
AMNH98219	*P. bicolor*	*P. bicolor* North	39.23
AMNH98220	*P. bicolor*	*P. bicolor* North	34.47
AMNH98221	*P. bicolor*	*P. bicolor* North	36.31
AMNH98222	*P. bicolor*	*P. bicolor* North	35.36
AMNH98223	*P. bicolor*	*P. bicolor* North	36.16
AMNH98225	*P. bicolor*	*P. bicolor* North	34.81
AMNH98226	*P. bicolor*	*P. bicolor* North	36.9
AMNH98227	*P. bicolor*	*P. bicolor* North	37.81
AMNH98228	*P. bicolor*	*P. bicolor* North	39.33
AMNH98229	*P. bicolor*	*P. bicolor* North	36.7
AMNH98230	*P. bicolor*	*P. bicolor* North	33.96
AMNH98231	*P. bicolor*	*P. bicolor* North	39.35
AMNH98232	*P. bicolor*	*P. bicolor* North	36.15
AMNH98233	*P. bicolor*	*P. bicolor* North	35.07
AMNH98234	*P. bicolor*	*P. bicolor* North	29.51
AMNH98235	*P. bicolor*	*P. bicolor* North	38.33
AMNH98236	*P. bicolor*	*P. bicolor* North	39.22

Table 2.4 continued from previous page

Voucher	**Species**	**OTU**	**SVL (mm)**
AMNH98239	*P. bicolor*	*P. bicolor* North	37.5
AMNH98240	*P. bicolor*	*P. bicolor* North	37.9
AMNH98241	*P. bicolor*	*P. bicolor* North	37.04
AMNH98242	*P. bicolor*	*P. bicolor* North	36.82
AMNH98243	*P. bicolor*	*P. bicolor* North	37.6
AMNH98244	*P. bicolor*	*P. bicolor* North	38.12
GECOH1174	*P. bicolor*	*P. bicolor* North	38.52
GECOH1177	*P. bicolor*	*P. bicolor* North	37.12
IAvH-Am-2827	*P. bicolor*	*P. bicolor* North	40.23
IAvH-Am-2828	*P. bicolor*	*P. bicolor* North	39.74
IAvH-Am-2829	*P. bicolor*	*P. bicolor* North	37.73
IAvH-Am-2830	*P. bicolor*	*P. bicolor* North	40.49
USNM147225	*P. bicolor*	*P. bicolor* North	36.66
USNM147226	*P. bicolor*	*P. bicolor* North	32.92
GECOH2661	*P. bicolor*	*P. bicolor* South	36.18
GECOH2665	*P. bicolor*	*P. bicolor* South	32.29
GECOH2711	*P. bicolor*	*P. bicolor* South	38.43
GECOH2712	*P. bicolor*	*P. bicolor* South	35.37
IAvH-Am-2826	*P. bicolor*	*P. bicolor* South	37.16
IAvH-Am-4271	*P. bicolor*	*P. bicolor* South	40.99
IAvH-Am-4272	*P. bicolor*	*P. bicolor* South	42.39
IAvH-Am-4273	*P. bicolor*	*P. bicolor* South	39.85
IAvH-Am-4274	*P. bicolor*	*P. bicolor* South	37.11

Table 2.4 continued from previous page

Voucher	**Species**	**OTU**	**SVL (mm)**
IAvH-Am-4275	*P. bicolor*	*P. bicolor* South	39.89
IAvH-Am-4276	*P. bicolor*	*P. bicolor* South	41.38
IAvH-Am-4277	*P. bicolor*	*P. bicolor* South	39.04
IAvH-Am-4278	*P. bicolor*	*P. bicolor* South	39.01
IAvH-Am-4279	*P. bicolor*	*P. bicolor* South	36.87
IAvH-Am-4280	*P. bicolor*	*P. bicolor* South	39.17
IAvH-Am-4281	*P. bicolor*	*P. bicolor* South	39.04
USNM137366	*P. bicolor*	*P. bicolor* South	36.3
USNM139777	*P. bicolor*	*P. bicolor* South	37.89
FMNH101098	*P. lugubris*	*P. lugubris*	17.29
FMNH152016	*P. lugubris*	*P. lugubris*	18.08
FMNH153314	*P. lugubris*	*P. lugubris*	19.85
FMNH153659	*P. lugubris*	*P. lugubris*	19.99
FMNH153756	*P. lugubris*	*P. lugubris*	18.5
FMNH153757	*P. lugubris*	*P. lugubris*	21.62
FMNH153762	*P. lugubris*	*P. lugubris*	17.6
FMNH154407	*P. lugubris*	*P. lugubris*	22.24
FMNH154413	*P. lugubris*	*P. lugubris*	19.94
FMNH174520	*P. lugubris*	*P. lugubris*	21.26
FMNH67971	*P. lugubris*	*P. lugubris*	22.52
FMNH83468	*P. lugubris*	*P. lugubris*	17.16
FMNH83470	*P. lugubris*	*P. lugubris*	18.02
AMNH88865	*P. terribilis*	*P. terribilis*	40.53

Table 2.4 continued from previous page

Voucher	Species	OTU	SVL (mm)
AMNH88866	*P. terribilis*	*P. terribilis*	40.31
AMNH88867	*P. terribilis*	*P. terribilis*	38.9
AMNH88868	*P. terribilis*	*P. terribilis*	40.29
AMNH88869	*P. terribilis*	*P. terribilis*	39.76
AMNH88870	*P. terribilis*	*P. terribilis*	38.93
AMNH88871	*P. terribilis*	*P. terribilis*	42.57
AMNH88876	*P. terribilis*	*P. terribilis*	41.12
IAvH-Am-4538	*P. terribilis*	*P. terribilis*	38.41
IAvH-Am-4539	*P. terribilis*	*P. terribilis*	37.39
IAvH-Am-4540	*P. terribilis*	*P. terribilis*	39.63
IAvH-Am-4541	*P. terribilis*	*P. terribilis*	39.64
IAvH-Am-4563	*P. terribilis*	*P. terribilis*	44.8
USNM151381	*P. terribilis*	*P. terribilis*	33.43
UMMZ129261	*P. vittatus*	*P. vittatus*	23.04
UMMZ129262	*P. vittatus*	*P. vittatus*	21.4
UMMZ131885	*P. vittatus*	*P. vittatus*	24.65
UMMZ142966	*P. vittatus*	*P. vittatus*	24.18
UMMZ142966	*P. vittatus*	*P. vittatus*	27.54
UMMZ142966	*P. vittatus*	*P. vittatus*	30.15
UMMZ193401	*P. vittatus*	*P. vittatus*	26.65
UMMZ193402	*P. vittatus*	*P. vittatus*	22.36
UMMZ193403	*P. vittatus*	*P. vittatus*	24.1
UMMZ238485	*P. vittatus*	*P. vittatus*	24.79

2.10 Appendix

SpaceMix model comparison using the Savage-Dickey ratio to estimate Bayes Factors

Bayes Factors

A commonly used metric for model comparison in a Bayesian framework is the Bayes Factor [171, 219], which is the ratio of the marginal likelihoods of two competing models. From Bayes's theorem, the ratio of the posterior probabilities of such models, M_0 and M_1, after observing data D (i.e. the posterior odds) can be expressed as

$$\frac{P(M_0|D)}{P(M_1|D)} = \frac{P(M_0)}{P(M_1)} \frac{P(D|M_0)}{P(D|M_1)} \tag{2.1}$$

This expression shows how the Bayes Factor $[P(D|M_0)/P(D|M_1)]$ drives the change from prior odds $[P(M_0)/P(M_1)]$ to posterior odds $[P(M_0|D)/P(M_1|D)]$, providing an intuitive metric of how well each model explains the data.

The marginal likelihood of a model corresponds to the likelihood for that model, weighed by its priors, and integrated over all the parameter space:

$$P(D|M) = \int P(\theta|M)P(D|\theta, M)d\theta \tag{2.2}$$

Except for some simple models where analytical solutions are possible, estimating marginal likelihoods can easily become a large computational endeavor requiring complex MCMC sam-

pling, especially for complex models with many parameters (see refs 220 and 221 for recent reviews of available approaches).

The Savage-Dickey ratio

An interesting exception to the problem described above are comparisons between nested models, where one of the models being compared can be obtained by setting a subset of parameters of the other model to fixed values. More formally, if model M_1 has parameters $\theta = (\psi, \phi)$, where ψ and ϕ are parameter vectors, model M_0 is considered to be nested within M_1 if M_0 has parameters $(\psi, \phi = \phi_0)$, where ϕ_0 is a vector of fixed values, and if the prior and marginal likelihood of M_1 are equal to those of M_0 when ϕ is equal (or tends to) ϕ_0. That is: $lim_{\phi \to \phi_0} P(\psi|\phi, M_1) = P(\psi|M_0)$ and $P(D|\langle \psi, \phi_0 \rangle, M_1) = P(D|\psi, M_0)$.

Consider the marginal likelihood of M_0

$$P(D|M_0) = \int P(\psi|M_0)P(D|\psi, M_0)d\psi \tag{2.3}$$

If the conditions described above hold, this can be re-written as

$$P(D|M_0) = \int P(\psi|\phi_0, M_1)P(D|\langle \psi, \phi_0 \rangle, M_1)d\psi = P(D|M_1, \phi = \phi_0) \tag{2.4}$$

Applying Bayes's theorem

$$P(D|M_0) = \frac{P(\phi = \phi_0|D, M_1)P(D|M_1)}{P(\phi = \phi_0|M_1)} \tag{2.5}$$

From this we can obtain the Bayes factor in favor of M_0

$$\frac{P(D|M_0)}{P(D|M_1)} = \frac{P(\phi = \phi_0|D, M_1)}{P(\phi = \phi_0|M_1)} \tag{2.6}$$

This constitutes the ratio of the posterior and prior distributions of ϕ under M_1 evaluated at ϕ_0, also known as the Savage-Dickey density ratio [172]. Intuitively this ratio can be interpreted as the change from prior to posterior probability of $\phi = \phi_0$ after allowing parameters ϕ to vary.

Application to SpaceMix models

One of the main advantages of SpaceMix is that it explicitly incorporates admixture between populations in an isolation by distance context [170]. This, among other things, allows users to test the extent to which long distance admixture explains the genetic covariance observed among populations by comparing models where populations draw admixture with those where the admixture proportions are fixed to 0. Since the second kind of model is nested within the first, the Savage-Diceky ratio can readily be used for this comparison as detailed below.

The parameters of interest for this comparison are the admixture proportions for each population $w = (w_1, w_2, \ldots w_k)$ which represent the probability $(0 < w_k < 0.5)$ that an

allele sampled from population k is of admixed origin (i.e. it originated at a geogenetic location different from that of k). The Bayes factor in favor of a model with fixed admixture proportions (M_0) versus a model where admixture proportions are free to vary (M_1) can be calculated using the Savage-Dickey ratio as

$$\frac{P(D|M_0)}{P(D|M_1)} = \frac{P(w = z|D, M_1)}{P(w = z|M_1)} \tag{2.7}$$

Where $w = (w_1, w_2, \ldots w_k)$ is the vector of admixture proportions for k populations and $z = (z_1, z_2, \ldots z_k)$ the set of fixed admixture proportions defined under M_0. If one is comparing admixture versus no admixture then z is composed entirely of zeros, but this approach can be used with any set of fixed values.

$P(w = z|M_1)$ represents the prior probability of observing z under M_1. SpaceMix uses a prior distribution of w_k where $w_k \sim Beta(\alpha = 1, \beta = 100)$ [170]. Therefore, under the assumption that admixture proportions are independent from one another, this probability is simply the product of the $Beta(1, 100)$ distribution evaluated at each value of $2z$

$$P(w = z|M_1) = \prod_{i=1}^{k} Beta(1, 100)\Big|_{2z_i} \tag{2.8}$$

Following the same logic, the posterior probability of observing z given M_1 and data D is

$$P(w = z|D, M_1) = \prod_{i=1}^{k} P(w_i = z_i|D, M_1)\Big|_{z_i} \tag{2.9}$$

Since there is no analytical solution for $P(w_i = z_i|D, M_1)$ it must approximated from the posterior distribution of w_i obtained through MCMC sampling. Below we illustrate how this approach was used to evaluate the extent of long-distance admixture in *Phyllobates* frogs, including the R code used.

Our goal was to evaluate the extent to which events of long-distance admixture have influenced the genetic covariance among 14 populations of *Phyllobates* frogs distributed across the biogeographic Chocó of Western Colombia (see main text for further details). In statistical terms, we aimed to estimate the Savage-Dickey ration between the full SpaceMix model, which allows for populations to draw long distance admixture (M_1), and one where all 14 admixture proportions (z in Eq. 2.7) were fixed at 0 (M_0).

The denominator of the Savage-Dickey ratio is straightforward to calculate using equation 2.8:

```
z <- rep(0,14)
prior <- prod(dbeta(2*z, shape1=1, shape2=100))
prior
#[1] 1e+28
```

To estimate the numerator we parameterized the full (i.e."source_and_target") SpaceMix model to obtain well-sampled estimates of the posterior distributions of the admixture proportions for our 14 populations (i.e. $w = (w_1, w_2, \ldots w_{14})$). We then approximated the posterior density of each w parameter at 0 by fitting a logspline density function [222] and

evaluating it at $w_k = 0$. Finally, we applied Eq. 2.9 to obtain the numerator and calculate the Savage-Dickey ratio.

```
library(SpaceMix)
library(polspline)

mcmc <- load("Spacemix_Counts_Full_space_MCMC_output1.Robj")

w_eval=numeric(14)

for(i in 1:14){
        w <- admix.proportions[i,]
        spline <- logspline(w, lbound=0) # Fit logspline density
        w_at0 <- dlogspline(0,spline) # Evaluate at w=0
        w_eval[i] <- w_at0
        }

posterior <- prod(w_eval)

posterior
#[1] 1.748146e+31
```

Briefly, the above code loops through the sets of posterior samples for $w_1 - w_{14}$, fits a logspline density and evaluates it at zero, generating a set set of posterior probability estimates whose product constitutes an approximation of $P(w = z|D, M_1)$.

Finally, we can use Eq. 2.7 to get the Bayes factor in support of M_0

$$\frac{P(D|M_0)}{P(D|M_1)} = \frac{P(w = z|D, M_1)}{P(w = z|M_1)} = \frac{1.748 \times 10^{31}}{1 \times 10^{28}} = 1748.16 \tag{2.10}$$

The above indicates that M_0 is overwhelmingly favored by the data over M_1, indicating strong support in for a scenario where gene flow between distant populations is very low or non existent.

A complementary way to visualize this result is plotting the prior and posterior distributions of all w parameters (Fig. 2.11). This allows us to see that in all but one case the posterior density of w at $w = 0$ is much higher than the prior, indicating that observing the data greatly increases support in favor of a scenario with no long-distance admixture. The only exception is w_1, there the posterior density at 0 is about one half as high as the prior. However, the posterior density very quickly increases and surpasses the prior at $w_1 = 1.86 \times 10^{-7}$, and has 95% of its mass below 0.012, so, although it doesn't provide support in favor of $w = 0$, the estimate for w_1 is very low and compatible with a scenario of minimal long-distance admixture.

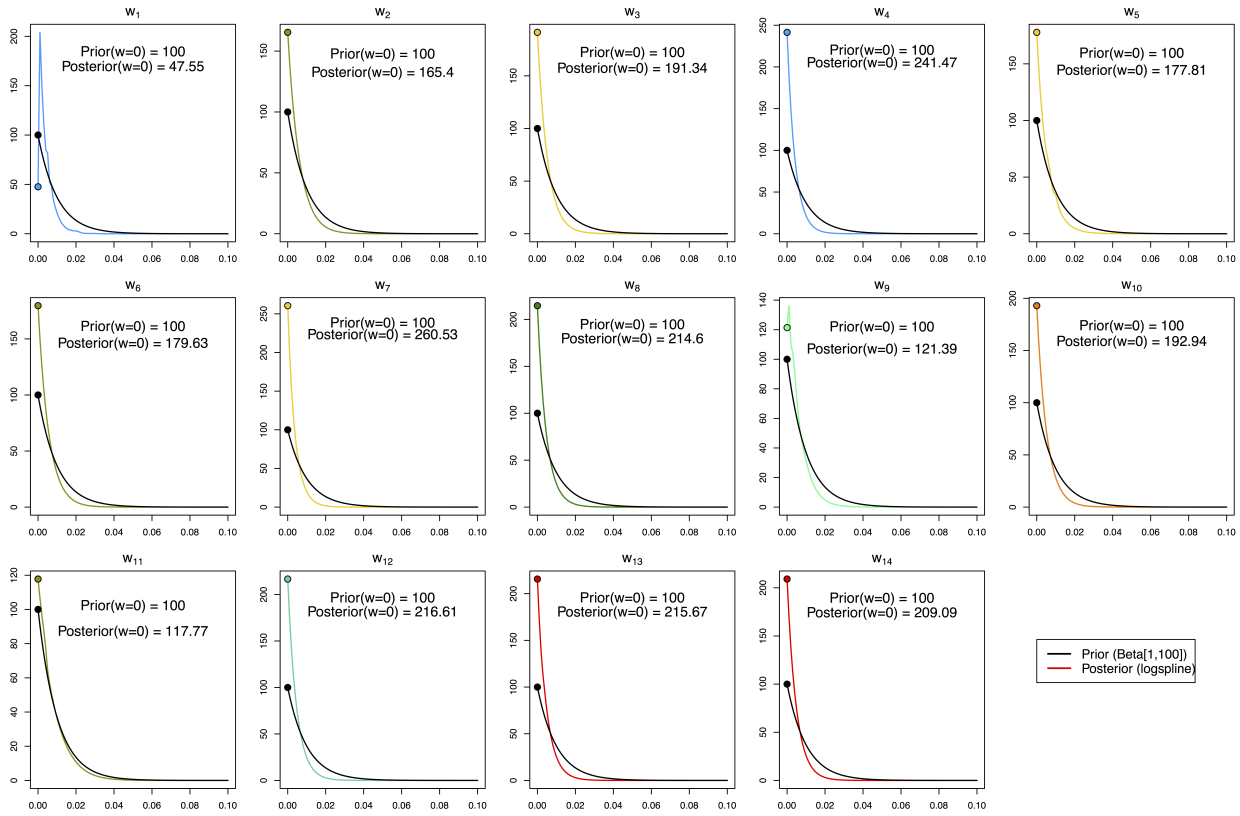

Figure 2.11: Prior and posterior (logspline) distributions of the admixture proportions for the 14 populations of *Phyllobates* analyzed. Prior distributions are colored black and posterior distributions as lines colored by OTU in accordance to Fig. 2.1. Dots represent the point estimate of each distribution at $w_k = 0$.

CHAPTER 3

DISENTANGLING THE ORIGIN OF A COLOR PATTERN CLINE IN *PHYLLOBATES* POISON FROGS

Roberto Márquez, Daniel Mejía-Vargas, Pablo Palacios-Rodríguez, Marcus R. Kronforst

3.1 Abstract

Geographic clines in phenotypic traits or allele frequencies can form either due to hybridization between previously isolated populations, or due to isolation by distance and divergent selection acting on parapatric populations. Distinguishing between these processes is challenging since the properties of a cline are mainly determined by the interplay between migration and selection regardless of its biogeographic history. Here we use spatial population genetic and geographic cline analyses to shed light on the biogeographic origin of a color pattern cline present in *Phyllobates* poison-dart frogs (Anura: Dendrobatiade). We find evidence for a recent range expansion from one tail of the cline to the other, coupled with signatures of moderately reduced gene flow along the cline, but very gradual patterns of variation in color pattern and body size, which suggest a scenario of ongoing parapatric divergence. This work highlights how complementary approaches can shed light on the processes that produce geographic clines.

Key Words: Hybrid zone, parapatry, geographic range expansion, aposematism.

3.2 Introduction

Gradual variation in phenotypes and allele frequencies across the landscape is ubiquitous in nature [29, 184, 223]. Patterns of phenotypic or genotypic change along such clines are thought to reflect the interplay between selection and gene flow, since adaptation to the local environment or genomic background will promote differentiation between populations, while gene flow between them will erode such differences [184, 224]. Therefore, geographic clines have been widely studied to gain insight on multiple aspects of natural populations, such as the magnitude and nature of selective forces acting on genotypes and/or phenotypes [224–227] or generating reproductive isolation between populations [195, 228, 229], the extent of gene flow experienced by different regions of the genome [75, 230, 231], or the genetic basis of phenotypic traits [232–234]. Despite the volume of research on geographic clines, some essential aspects are still challenging to resolve. Among them are the biogeographic processes that produce clines.

Two main scenarios have been proposed to result in a cline [21, 184, 223]: First, if the range of a species is large enough for isolation by distance to operate, the reduced gene flow between distant populations will generate clines in the frequencies of alleles under weak selection. Furthermore, if selective regimes are not uniform across the species' range, local adaptation of parapatric populations that are still connected by gene flow will generate phenotypic and allele frequency clines. This scenario is known as primary contact (or primary intergradation) [184]. Second, when two allopatric populations that have not evolved complete reproductive isolation come into secondary contact, hybridization between them can generate a phenotypically and genetically intermediate contact zone, resulting in a clinal geographic pattern [21]. In both cases the interplay between selection and migration determines the patterns of clinal variation (e.g. the shape of the cline). Therefore, although sometimes distinguishable at early stages [235], at equilibrium both processes generate very

similar or even identical patterns [184,236]. This makes distinguishing between clines formed under primary or secondary contact on the basis of present variation a difficult task. Alternative sources of historical data can be used to circumvent this limitation, such as the fossil or pollen records [237], or the projection of climatic habitat suitability models onto paleo-climatic data [eg. 238]. However, detailed paleontological data are not available for most taxa, and inferences based on such data or reconstructions of ancient climate are limited in their accuracy and spatiotemporal resolution. Therefore, additional, complementary approaches to investigate the origin of geographic clines are desirable. Recent advances in spatial population genetics and historical demography, such as the ability to parameterize complex demographic models [e.g. 239–241], to identify regions of unusually high or low migration across a landscape [173, 242], or to explicitly incorporate geography into inferences regarding discrete population structure [243, 244] provide new and promising tools to tackle this question.

Poison frogs of the genus *Phyllobates* offer a compelling system to study the origin of geographic clines. Across most of their range, populations of this genus exhibit either a solid-yellow color pattern, or brightly colored stripes on a dark background (Fig. 1.1) [25]. An interesting exception is the upper San Juan river drainage in Western Colombia, where populations of *Phyllobates* transition gradually from the solid-yellow *P. bicolor* at mid-elevations to the striped and smaller *P. aurotaenia* in the lowlands (Fig. 3.1A), [25, 99]. Recent work has shown that the populations of *P. bicolor* at the Eastern end of the cline are very closely related to their neighboring *P. aurotaenia*, and identified a corridor of increased gene flow that connects the clinally-varying populations along the San Juan river [245]. Considering *P. bicolor* is the only montane species of *Phyllobates*, the San Juan cline could therefore have formed either in parapatry if highland and lowland populations became differentiated in the face of gene flow (primary contact), or due to secondary contact if these populations were isolated during differentiation.

In this study we aim to elucidate the biogeographic history of *Phyllobates* populations along the upper and mid San Juan drainage, as well as the evolutionary processes leading to color pattern variation among them. We evaluate the evidence of a recent range expansion from the Chocoan lowlands into the Western Andes, describe genotypic and phenotypic variation across the cline, and ask whether the observed patterns support primary or secondary contact between *P. bicolor* and *P. aurotaenia* as the process responsible for the upper San Juan cline.

3.3 Materials and Methods

Field sampling

We sampled *Phyllobates* frogs at nine localities (2-17 frogs per locality) along the San Juan and Atrato drainages in the Chocoan forests of Western Colombia (Fig. 3.1A; Table 3.1). Frogs were located visually or acoustically and captured manually. After capture we took dorsal and ventral photographs against a plain, opaque background alongside a Tiffen Q-13 Color Separation Guide, and collected toe-clippings for DNA extraction. Some individuals were euthanized through an overdose of topical lidocaine and deposited at the amphibian collection of the ANDES Natural History Museum (Universidad de los Andes, Bogotá, Colombia). In these cases liver and/or muscle samples were obtained for DNA extraction. Tissue samples from five individuals of *P. terribilis* (Table 3.1) were obtained as detailed above, and used as outgroups in some analyses.

Genotype data

Whole-genome resequencing. We obtained genome-wide genotypic information for 66 individuals along the cline through whole-genome re-sequencing. DNA was extracted using DNeasy spin columns (Qiagen) following the manufacturer's protocol, except for the final elution step, in which the elution buffer was heated to 60°C, and allowed to incubate on the spin column membrane for 10 min before centrifugation. This step was performed twice to increase overall DNA yield. Following extraction, 142-1000μg of DNA per sample were sheared on a Covaris S-series ultrasonicator to obtain fragment size distributions peaked at $\sim$ 300bp. The sheared DNA was used to build barcoded paired-end (PE) Illumina libraries with either NEBNext Ultra II (New England Biolabs) or Kappa HyperPrep (Kapa Biosystems) kits. Finalized libraries were amplified by 5-7 cycles of PCR, and sequenced on an Illumina NovaSeq 6000 S4 150bp PE flowcell. Prior to running the S4 flowcell a 50bp PE SP flowcell was run for quality control and to refine sample pooling. Libraries for *P. terribilis* outgroup samples were prepared as detailed above but sequenced in a 100bp PE Ilumina HiSeq 4000 flowcell.

Read mapping and data filtering. Raw reads were trimmed and quality-filtered using Trimmomatic v.0.36 [132] and Skewer v.0.2.2 [141] under default parameters, except for the minimum read length, which was increased to 64bp for 150bp reads and 36bp for 100bp reads, and the leading and trailing qualities, which were increased to 15. Reads were mapped to a draft genome assembly of *P. terribilis* generated for an ongoing project [Machado, Márquez, Reed, Janies, Kronforst & Grant, *Unpublished*]. Briefly, we used paired-end ($\sim$50X coverage) and mate-paired (3, 5, 8, and 10Kb insert sizes; $\sim$30X coverage) Illumina sequences to generate an assembly using Masurca [246, 247], which we then scaffolded with in vivo Hi-C data ($\sim$20X coverage) using SALSA [248]. This resulted in 442,103 scaffolds spanning 3.8Gb with an N50 of 11Kb. Prior to mapping we used RepeatMasker [137] to identify and mask repetitive elements. Reads were mapped using bwa-mem v.0.7.17 [249] with default parameters, sorted with Samtools v.1.6 [146], and de-duplicated with Picard v.2.8.1

(http://broadinstitute.github.io/picard).

After mapping, we calculated multiple statistics for data filtering (see below) individually for each of the six sampling localities with five or more sampled individuals, and only used sites that passed all filters in all of these localities for further analyses. We used Angsd [152] to identify sites with base and mapping qualities above 20, that were covered by at least one read in 50% of individuals, and that passed tests for strand bias (p-value > 0.0001), base quality bias ($p > 1 \times 10^{-6}$), mapping quality bias ($p > 1 \times 10^{-6}$), Hardy-Weinberg equilibrium ($p > 1 \times 10^{-6}$), and distance-from-edge-of-read bias ($p > 0.0001$). Next we used ngsParalog (https://github.com/tplinderoth/ngsParalog) [250] to identify regions of our assembly exhibiting signs of read mismapping due to paralogy or misassembly. To do so, we first calculated likelihood ratios for paralogous mapping with the calcLR command, based on a pileup of reads and their corresponding quality scores generated with Samtools. Next we used the HMM approach described by Linderoth [250], which uses depth and mismapping likelihood ratios to identify stretches of the genome that exhibit paralogous mapping. Given the fragmented nature of our assembly, we parameterized the transition probabilities of the HMM usinng genome-wide likelihood ratios and depths before running the HMM on each scaffold. We ran the HMM on scaffolds with at least 20 apparently variable sites (i.e. SNP p-value < 0.05). On scaffolds with a smaller number of SNPs we calculated p-values associated to paralogy likelihood ratios based on a 50:50 mixed χ^2 distribution with one and zero degrees of freedom, and excluded sites with significant likelihood ratios after Bonferronni correction ($\alpha = 0.1$). After filtering, we retained 210,991,438 sites, out of which 6,913,384 were variable (i.e SNP p-value < 0.05).

Phenotypic data

We examined phenotypic variation in two traits that are likely to vary adaptively across

the cline: color pattern and body size. As a proxy for color pattern we used the proportion of a frog's dorsum covered in bright pigmentation (p_{bright}). Field photographs of each specimen were imported in the camera's RAW format to the program RawTherapee v 5.5 (www.rawtherapee.com), where they were white-balanced based on the photographed color standard, and exported to 8-bit JPEG format with the highest possible resolution. We then used GIMPv 2.10.10 (www.gimp.org) to crop JPEG images so that they included only the frog's dorsum on a white background, and to manually remove flash glare and debris using the "healing" and "clone stamp" tools. Finally, we used the R package patternize [251] to identify pixels corresponding to dark and brightly colored areas using the patRegk() function, which uses k-means clustering to assign pixels to a preset number of color groups, and calculated p_{bright} as the number of brightly colored pixels divided by the total number of non-white pixels in the image. Given the variability of shooting conditions between photographs, we analyzed each photograph separately, and visually chose the number of color groups that best allowed us to extract the frog's color pattern. For four images for which pattern segmentation was especially challenging we increased color saturation in GIMP to improve the contrast between brightly colored and dark pixels.

To quantify body size we measured the snout-to-vent length (SVL) of all photographed frogs using Image J [252]. Measurements were calibrated either with a millimeter ruler photographed alongside the frog, or the centimeter scale on the Tiffen Q-13 color card. To evaluate the precision of this procedure, we also measured the SVL of nine individuals that were preserved at the ANDES-A collection using a digital caliper with 0.01mm resolution. Measurements were consistent between methods ($r^2 = 0.88$, slope = 1.10, $F_{1,9} = 69.25, p = 0.000016$; Fig. 3.5).

Population structure, effective migration, and genetic diversity

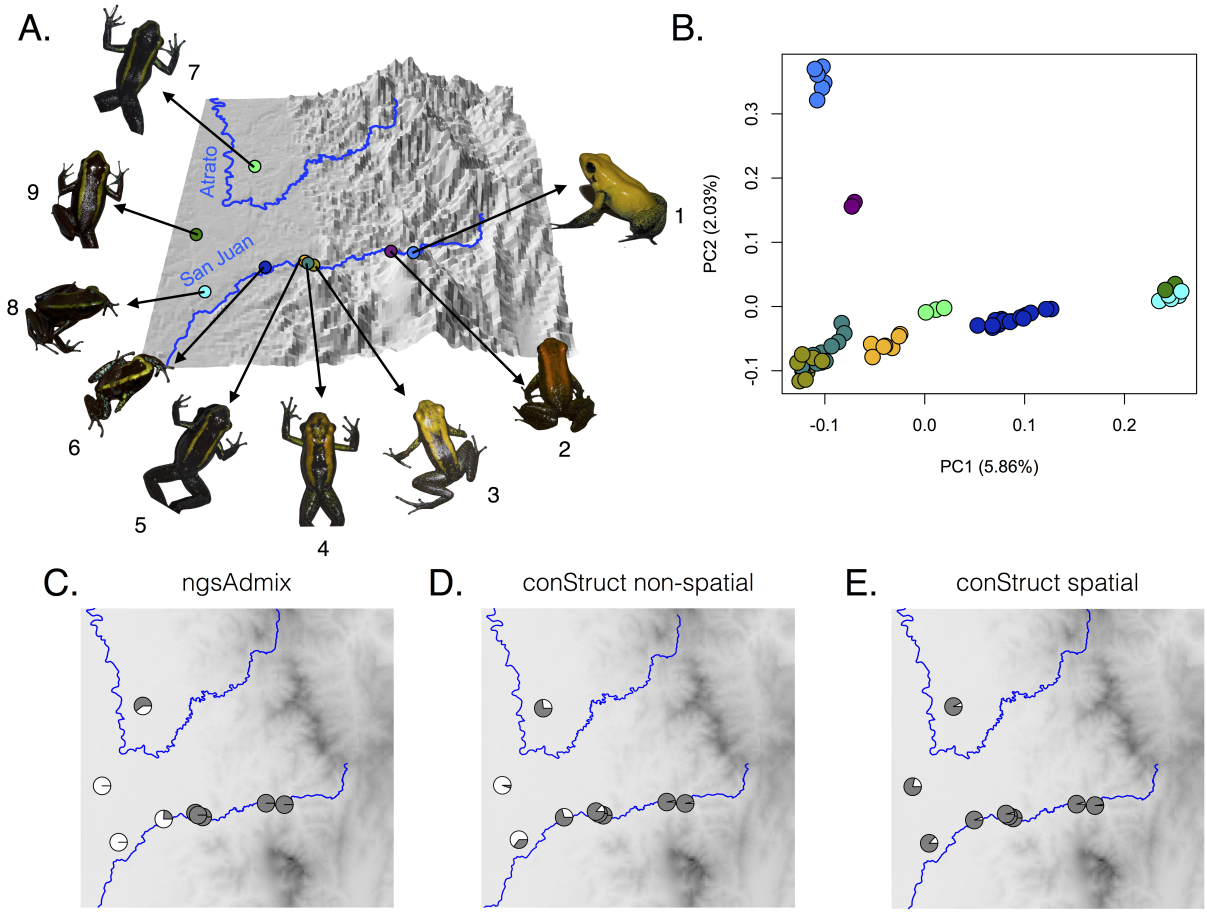

Figure 3.1: Geographic distribution of phenotypic (A) and genetic (B-E) variation of *Phyllobates* populations in the upper San Juan river of Western Colombia. Panels C-E display admixture proportions inferred assuming $k = 2$ clusters with ngsAdmix (C), and conStruct assuming spatially aware (E) and non-spatially aware models (D).

To get an overview of genetic structure in the San Juan, we performed a principal component analysis (PCA) in PCAngsd [148] and estimated admixture proportions (k = 2–4 clusters) using ngsAdmix [149]. Both analyses were performed on genotype likelihoods estimated in Angsd. In addition we used conStruct [243] to simultaneously evaluate continuous and discrete genetic structure. This method consists of a clustering algorithm similar to that implemented in STRUCTURE [253] or ngsAdmix, with the added feature of incorporating isolation by distance in addition to discrete structure to explain observed genetic variation. We ran conStruct assuming k = 2–4 layers (comparable to clusters), on allele frequencies estimated in Angsd for each population. MCMC runs consisted of two independent 50,000 iteration chains (20% burnin) that were logged every 50 samples and averaged to estimate parameters. In addition to the spatially-aware conStruct model, we ran analyses under a non-spatial model, comparable to the STRUCTURE model. Comparing the results of spatially-aware and unaware analyses allowed us to assess the extent to which population structure can be attributed to discrete (e.g. allopatry) or continuous (e.g. isolation by distance) processes. To aid this comparison, we estimated the contribution of each layer to the total allelic covariance between populations.

Finally, to gain insight into historical levels of gene flow among the sampled populations, we used EEMS [173] to estimate an effective migration surface. Briefly EEMS estimates the rate of decay in genetic similarity with geographic distance across the landscape. Regions with slower decay can be interpreted as areas with high levels of historical migration and vice versa. We estimated mean squared pairwise genetic distances between all samples in ATLAS v. 0.9 [174], and ran five parallel EEMS MCMC runs for 10,000,000 steps (20% burnin), logging every 1,000 steps, and setting the number of demes to 1,000. The model used by EEMS includes a set of parameters that describe within-population genetic diversity across the range. Since spatial variation in genetic diversity can bear valuable biogeographic information, we also plotted an effective genetic diversity surface.

Detecting a range expansion

To gain insight into the historical patterns of migration of *Phyllobates* populations along the San Juan drainage we calculated the directionality index Ψ, proposed by Peter and Slatkin [254], which quantifies the asymmetry in the two-dimensional site frequency spectrum (2D SFS) of two populations. Such asymmetries are thought to arise due to the founder events that occur during a range expansion [254, 255]. We estimated Ψ between all pairs of populations based on the unfolded 2D SFS, using equation 1b from ref. 254. To reduce the effect of incorrect polarization in 2D SFS estimation, we first estimated unfolded allele frequencies for each population using the *P. terribilis* reference sequence as the ancestral allele, then estimated the 3D SFS between each pair of populations and *P. terribilis*, and finally extracted the 2D SFS restricted to sites that were fixed in *P. terribilis*. Since the estimation of Ψ requires 2D SFSs with equal sample sizes for both populations, we randomly down-sampled populations to a diploid sample size of two before SFS estimation. To evaluate the robustness of our results to downsampling, we estimated pairwise Ψ based on five different randomly downsampled datasets. Estimates of Ψ for the five datasets were highly consistent (Fig 3.6), so we randomly chose one set of results for further analyses. Statistical significance for estimates of Ψ was evaluated using 100 bootstrapped SFSs generated in Angsd. Finally, we used the 36 pairwise estimates of Ψ to test for a range expansion versus stationary isolation by distance and to estimate the most likely origin of expansion using the single.origin() function of the rangeExpansion R package [254, 256].

Geographic cline characterization

To characterize phenotypic variation along the cline, we used the sigmoidal function pro-

posed by Szymura and Barton [225, 257] to describe trait variation along a one-dimensional cline:

$$y = \frac{1 + tanh(\frac{2(x-c)}{w})}{2(\frac{1}{y_{max}-y_{min}})} + y_{min} \tag{3.1}$$

Where x is the position along the cline, c is the cline center, w the cline width, and y_{max} and y_{min} are the maximum and minimum trait values at the tails of the cline. The distance along the cline was defined as the linear distance from the easternmost sampled population (Pop. 1), obtained using the pointDistance() function of the raster R package [177], and parameters were optimized using non-linear least squares regression as implemented in the nlsLM() function of the R package minpack.lm [258]. Prior to model fitting SVL measurements were scaled between zero and one as

$$SVL_{scaled} = \frac{SVL_i - min(SVL)}{max(SVL) - min(SVL)} \tag{3.2}$$

Since p_{bright} measurements are proportions, and therefore already between zero and one, they were not scaled. Given its geographic location, assigning a position for population 7 along a one-dimensional cline was ambiguous so this population was excluded from all geographic cline analyses.

To describe genetic variation along the San Juan we took a twofold approach. First, we fit a sigmoid function as detailed above to the admixture proportions obtained with ngsAdmix, and conStruct assuming $k = 2$. Second, we identified SNPs with derived allele frequency differences greater than 0.9 between the two ends of the cline (ie. Pop. 1 vs Pops. 8 and

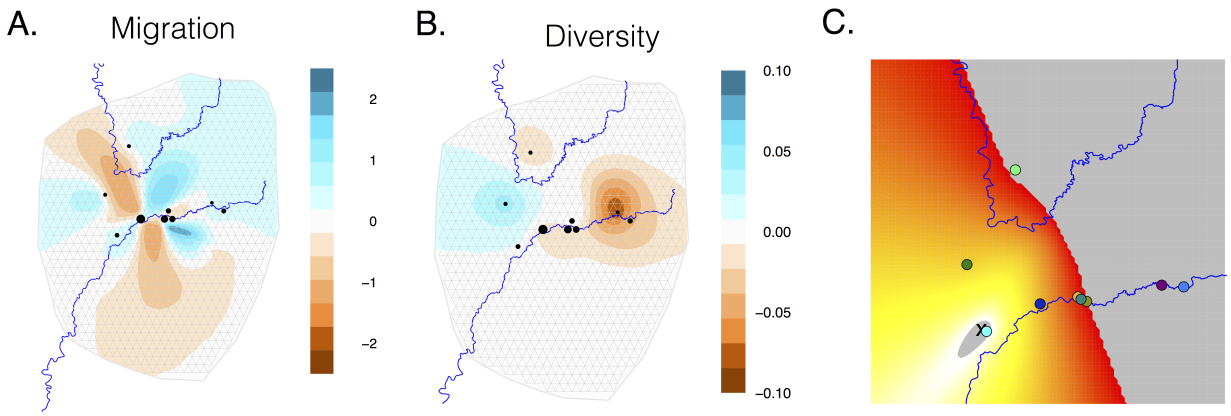

Figure 3.2: Effective migration (A) and genetic diversity (B) surfaces estimated in EEMS, and most likely origin of a range expansion (C) estimated based on directionality indices in the rangeExpansion R package. Blue colors in panels A-B indicate areas of higher migration/diversity and brown colors indicate areas of lower migration/diversity. In panel C the most likely origin is marked by an X, and brighter yellow regions indicate a more likely origin of expansion.

9), and fit individual sigmoid functions to the derived allele frequencies at each SNP. Allele frequencies were estimated from genotype likelihoods in Angsd, and derived alleles were polarized based on the *P. terribilis* reference sequence. In order to avoid mispolarization, we restricted this analysis to sites where the reference allele was fixed among the five outgroup *P. terribilis* samples, and where the derived allele frequency in population 1 was higher than in population 9. Sites where model fitting did not converge were removed from the analysis.

3.4 Results

Population structure and history

Population structure analyses showed clear genetic structuring among populations, which also closely matched geography. Individuals from each sampling locality clustered together and had similar admixture proportions, and populations were genetically closer to their geographic neighbors (Fig. 3.1B-E). Admixture proportions estimated with ngsAdmix and the

non-spatial conStruct analysis showed similar patterns, with admixture proportions transitioning from "*P. bicolor*-like" to "*P. aurotaenia*-like" from east to west, and populations $5-7$ exhibiting intermediate admixture proportions (Fig. 3.1C-D). In the spatially aware conStruct analyses, a predominant layer contributed the majority of allelic covariance across populations (100-78% with $k = 2$), and this contribution decreased slightly from east to west populations (Fig. 3.1E, 3.7).

In concordance with previous work [245], the EEMS migration surface (Fig. 3.2A) recovered a gene flow corridor along the San Juan river that connects most populations along the cline. However, our denser sampling also revealed clear areas of reduced migration along the San Juan (e.g. between populations 5 and 6), suggesting some discontinuity in gene flow along the cline. The genetic diversity surface (Fig. 3.2B) showed a clear eastward decline in within-population genetic dissimilarity, with montane populations being less diverse than lowland ones. Concordantly, the estimated directionality indices (Table 3.2) strongly supported a range expansion over stationary isolation by distance ($r^2 = 0.94, p = 2 \times 10^{-18}$), whose most likely center was inferred to be in the Chocoan lowlands, at the western tail of the cline (Fig. 3.2C).

Geographic cline analyses

Both scored phenotypes (color pattern and body size) displayed gradual, eastwardly increasing trends across the sampled area (Fig 3.3A,C). Although admixture proportions estimated with ngsAdmix showed a sharp break between populations 5 and 6, those estimated with conStruct showed much more gradual patterns (Fig. 3.3B). Individual allele frequency clines were highly variable in shape (Fig. 33.3D, Fig. 3.4), with cline centers extending across the eastern half of the cline (Fig. 3.4A), and widths ranging from 0.58Km to 140.39 Km (Fig. 3.4B). The average cline (ie. the cline corresponding to the means of parameter

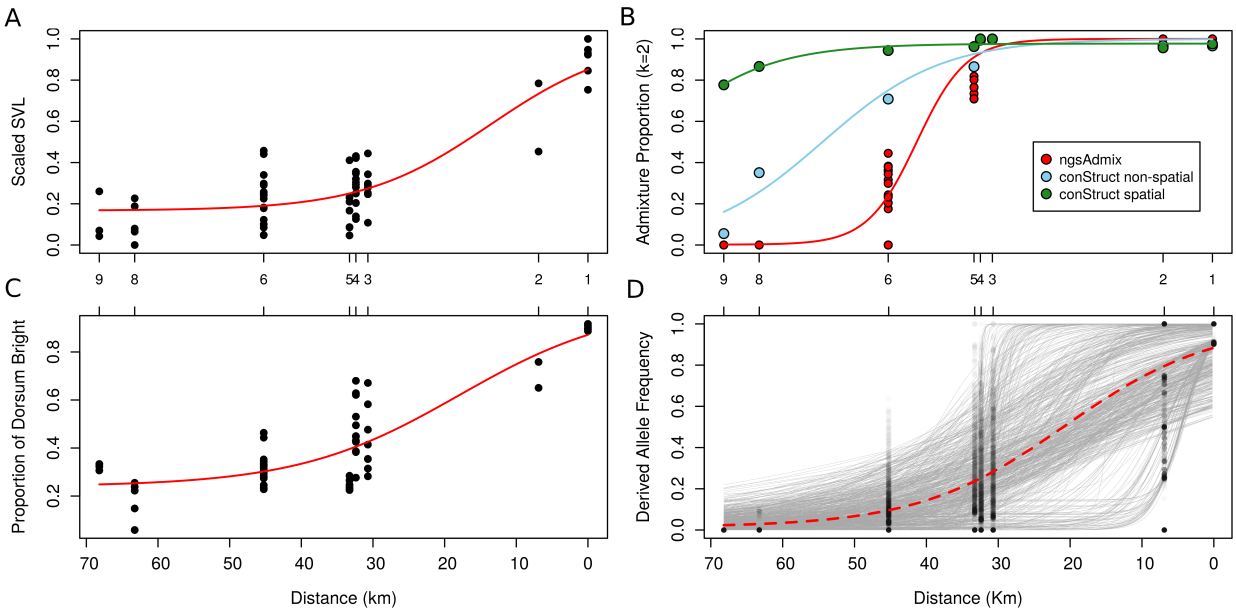

Figure 3.3: One-dimensional geographic clines fit to body size (A), color pattern (B), admixture proportions (C), and frequencies of alleles highly differentiated between tails of the cline (D). In A-C dots represent individual frogs or population-wide admixture proportions (from conStruct analyses), and solid lines depict the best fit sigmoidal function. In panel D dots represent allele frequencies at each population, grey lines represent the individual sigmoid functions fit to each allele's frequencies, and the red broken line represents the sigmoid function corresponding to the mean parameter estimates across all SNPs.

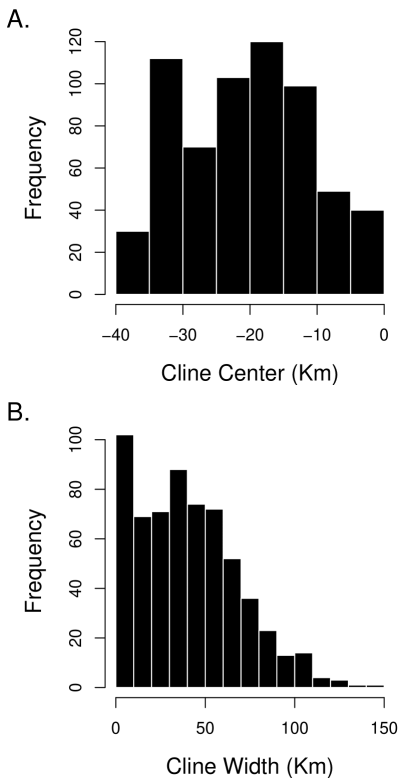

Figure 3.4: Inferred allele frequency cline centers (A) and widths (B) for highly differentiated alleles between tails of the cline.

estimates across loci) was centered at 20.64Km, and displayed a gradual pattern (width = 41.02Km) similar to those observed for SVL and p_{bright}. Nevertheless, a fraction (7%) of loci displayed sharp clines with centers between populations 3 and 5 (ie. 30.71 – 33.27Km) and widths equal or less than the width of the admixture proportion cline.

3.5 Discussion

Our spatial population genetic analyses indicate that a recent range expansion from the Chocoan lowlands into the foothills of the Western Andes gave rise to the current distribution of *Phyllobates* populations along the upper San Juan river (Fig. 3.2B-C). Furthermore, we found evidence for relatively high levels of historical migration along the San Juan (Fig.

3.2A), except for a region of reduced gene flow, which coincides with a breakpoint in population structure (Fig 3.3B). It therefore seems likely that the gradual variation in color pattern and body size observed across the cline (Fig. 3.3A,C) arose during or shortly after this range expansion, with local adaptation of the montane populations playing an important role in the differentiation between populations.

These gradual phenotypic clines, together with signatures of genetic discontinuity recovered by EEMS and ngsAdmix, and the wide variation in individual allele frequency cline shapes could have arisen under both primary and secondary contact in the face of a range expansion [21,184]. Under secondary contact, the ancestors of *P. bicolor* would have become isolated from populations in the lowlands after colonizing the Andean foothills. Following a period of allopatric differentiation marked by local adaptation of the highland populations, hybridization upon secondary contact would have given rise to phenotypic intermediates. Initially, these intermediate phenotypes would be restricted to the contact zone, and the wide, gradual clines observed in the present would be due to the diffusion of introgressed alleles along the San Juan [225, 236]. Similarly, individual allele frequency clines would be similar upon secondary contact, acquiring different shapes as selection restricts or allows their movement along the cline. In this case, the signatures of genetic discontinuity would correspond to the period of allopatric isolation between highland and lowland populations. Alternatively, if connectivity across the San Juan was not interrupted during or after the range expansion, parapatric divergence due to local adaptation and isolation by distance would lead to the phenotypic intermediates [184]. As highland populations became locally adapted, the ongoing movement of alleles along the San Juan would generate the current patterns of gradual phenotypic (and allele frequency) variation. Divergent selection favoring solid yellow patterns in the highlands and striped ones in the lowlands, would gradually purge intermediate phenotypes, eventually forming a transition zone. Considering color pattern is probably a polygenic trait in *Phyllobates*, the buildup of associations between loci across the

genome should also contribute to divergence [259]. The signatures of genetic discontinuity would, in this case, reflect this ongoing reduction of migration due to divergent selection.

Our data do not allow us to unambiguously discard either scenario. However, they favor a primary contact scenario, since secondary contact would require dispersal of color pattern alleles along the cline to be much stronger than divergent selection on this trait in order for a gradual transition between solid and striped patterns to form [184, 236, 260]. Examples of divergent selection as a driver of variation in defensive coloration have been described in a wide variety of taxa [e.g. 261–264], including several species of dendrobatid poison frogs [265–267]. In multiple cases, these systems exhibit sharp transitions between morphs at contact zones [e.g. 195, 261, 268, 269]. Although a few poison frog studies have not found differences in predation pressures between local and foreign colorations [e.g. 270, 271], these studies have attributed the existence of different morphs to other selective pressures, for example sexual selection [270], or to methodological shortcomings, such as not targeting relevant predator groups [272]. No studies have, to our knowledge, assessed directly whether divergent selection exists between striped and solid color patterns in *Phyllobates*. However, recent work on *P. vittatus* (striped) found that predation attempts were significantly more frequent on solid-orange clay models than on their striped counterparts at a site inhabited only by *P. vittatus* [273], indicating selection against solid-colored patterns by predators that only interact with striped *Phyllobates*. In view of the evidence above, we consider it unlikely for divergent selection on color pattern between opposite ends of the cline to be weak.

Under primary intergradation the gradual phenotypic clines would also indicate low levels of divergent selection on color pattern if a balance between migration and selection has been reached. However, if this assumption of equilibrium is dropped, then the observed patterns can be interpreted as signatures of a nascent transition zone, which is in the process of becoming steeper and narrower due to divergent selection. This falls in line with the results

found using the spatially-aware conStruct model (Fig. 3.1E, 3.3B, 3.7), which suggest that the majority of population structure can be explained by a single gene pool under isolation by distance, rather than discrete genetic structure. The strong evidence in favor of a recent range expansion further adds to this notion, since it indicates that our focal populations do not behave as an island model under equilibrium isolation by distance. Therefore, we favor a scenario of parapatric differentiation to explain our results. Future studies investigating the selective pressures faced by both color morphs and, especially, their intermediates are needed to further evaluate the likelihood of gradual color pattern clines under primary and secondary contact.

It is worth noting that a third scenario that involves secondary contact could also explain our results: If striped frogs are overall more fit than yellow ones across the whole cline, then upon secondary contact this morph could sweep into the range of *P. bicolor*, generating a broad "tail" of intermediate morphs that would eventually become extinct, along with the solid yellow pattern [274, 275]. Although there are examples of poison frog color morphs that seem to be overall better predator deterrents [276] or preferred during mate choice [277], the fact that there are three independent solid-yellow lineages of *Phyllobates*, two of which inhabit higher elevations than the rest of the genus [245], lends little support to this scenario.

Phylogenetic studies of *Phyllobates* frogs have consistently found that solid-yellow lineages are not sister to one another, suggesting that this trait has evolved convergently in the genus [27, 26, 245]. However, given the concordance of phylogeny and geography in this system, together with high levels of gene flow recently uncovered, it has been suggested that the reconstructed phylogenies of *Phyllobates* may be a reflection of recent gene flow rather than the history of populations [245], casting doubt on inferences of convergent evolution. If, as suggested here, the upper San Juan cline arose through parapatric differentiation, then our results would indicate that the close phylogenetic affinity found between the northern

populations of *P. bicolor* and the upper San Juan and Atrato *P. aurotaenia* are due to common ancestry and not recent gene flow, supporting the convergent origins of solid-yellow patterns.

Admixture proportions estimated with ngsAdmix displayed a much more pronounced and narrow break between ends of the cline (Fig. 3.3B). Taken at face value, these results would indicate that there is an appreciable degree of reproductive isolation between ends of the cline, much higher than expected in the early stages of parapatric differentiation. Furthermore, despite the reduced levels of migration, alleles controlling color pattern and body size variation would be able to move relatively freely along the cline. However, population assignment methods that do not explicitly incorporate spatial information, such as ngsAdmix, have been shown to over estimate discrete genetic structure in the face of isolation by distance [278, 279]. Considering that the admixture proportions obtained with the spatially-aware conStruct model show a much less pronounced pattern than those from ngsAdmix or the spatial conStruct (Fig. 3.1, 3.3B), it is likely that ngsAdmix is overestimating the degree of discrete genetic structure in the system. Admixture proportions inferred using the STRUCTURE model (which is the base of ngsAdmix) are often used in geographic cline analyses to represent genome-wide patterns of clinal variation, which serve as a baseline to interpret patterns observed in traits or loci of interest, for example to evaluate the extent of selection on such traits or loci [260]. As demonstrated here, using the STRUCTURE model for this purpose can exaggerate the strength of stepped clines, which can in turn bias inferences about selection or reproductive isolation. Therefore, we suggest that this procedure may not be appropriate for cases in which isolation by distance is thought to contribute appreciably to genetic variation.

Disentangling the historical processes that give rise to geographic clines remains a challenging endeavor. This work provides evidence for parapatric divergence as the main process

behind a color pattern cline in *Phyllobates* poison frogs, and exemplifies how the integration of recent advances in spatial population genetics with traditional cline analyses can provide useful insights into the history of geographic clines.

3.6 Acknowledgements

We would like to thank our guides, hosts, and field assistants while sampling in Western Colombia: Dago, Marino, Don Manuel, Guachiné, and Santiago Casas-Cardona. We also thank Mileidy Betancourth for help with molecular work, Nick VanKuren and Tyler Linderoth for advise and help with analyses, Andrew J. Crawford and Alberto Farfán for facilitating specimen access and deposition at the ANDES museum, and Trevor Pirce and John Novembre for valuable suggestions and discussion. This work was supported by the National Science Foundation (award numbers DEB-1702014, IOS-1922624, and IOS-1827333), the National Geographic Society (award number 9786-15), The National Institutes of Health (award number R35GM131828), the Pew Biomedical Foundation, and the Committee on Evolutionary Biology at the University of Chicago. Tissue collections were authorized by permit 1177 from the Colombian National Authority for Environmental Licenses (ANLA), and animal use procedures were approved by the University of Chicago's Institutional Animal Care and Use Committee (ACUP No. 72416).

3.7 Author Contributions

R.M. conceived the study. R.M. and M.R.K. designed the study and acquired funding. R.M. D.M-V, and P.P-R performed field work. R.M. generated and analyzed the data. R.M. wrote the manuscript with input from M.R.K. All authors read and approved the manuscript.

3.8 Supplementary information

Supplementary Figures

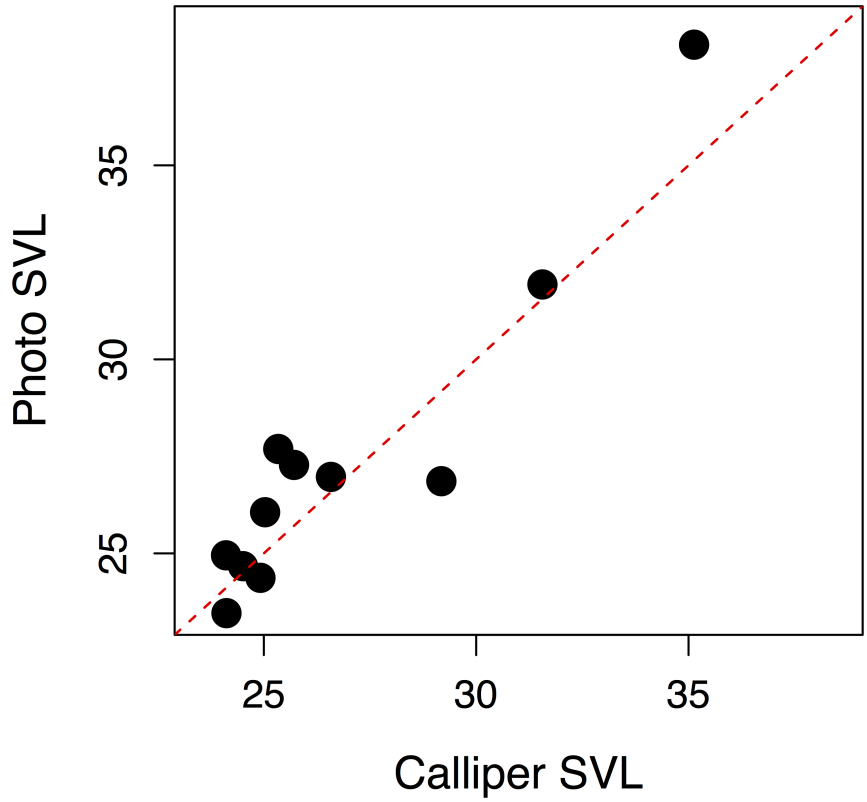

Figure 3.5: Comparison of snout-to-vent length measurements performed on dorsal photographs and museum specimens of the same individual. The dashhed red line represents $y = x$.

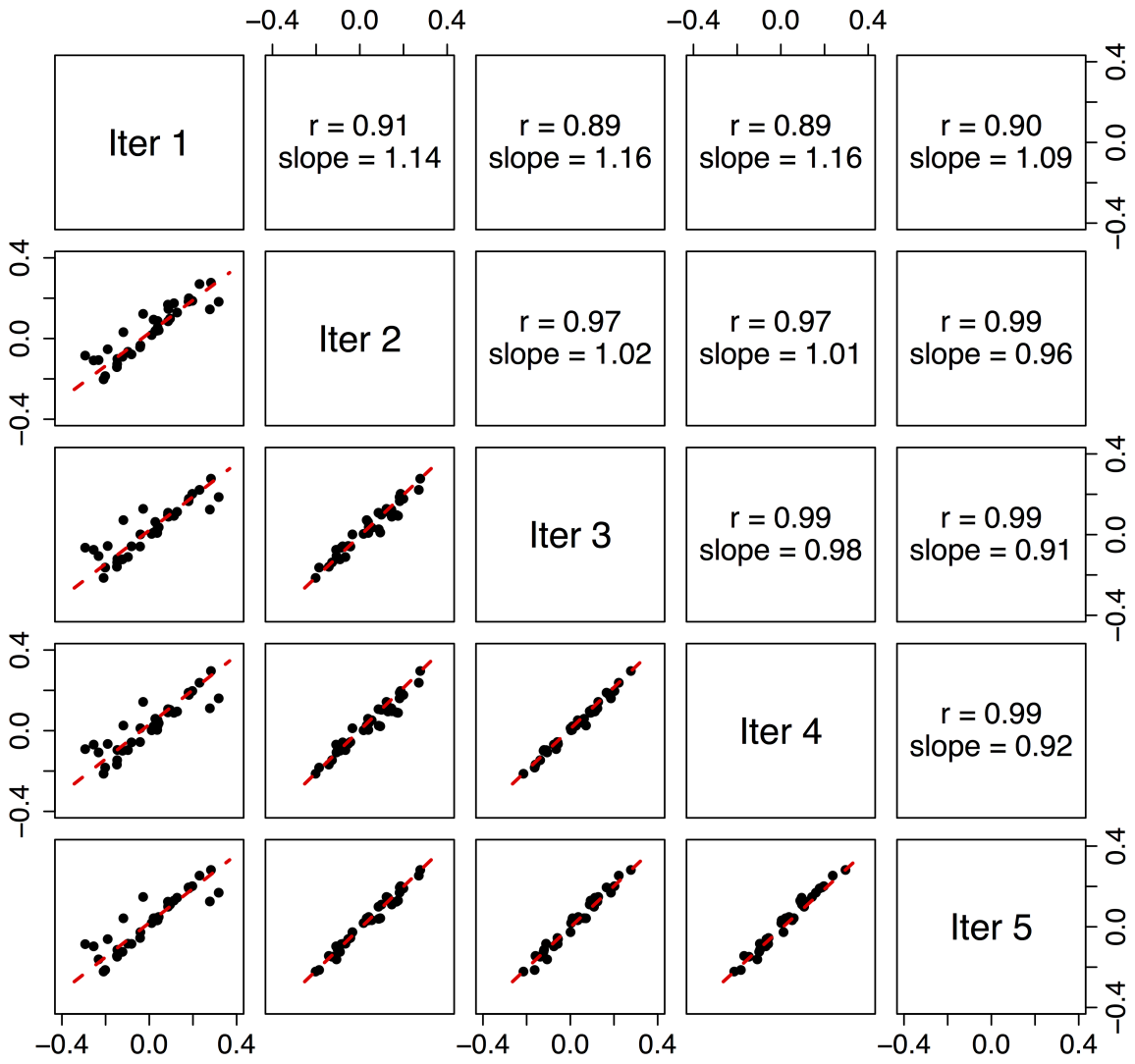

Figure 3.6: Comparisons of directionality indices (Ψ) obtained with five different sub-samples of individuals from each population. The dashed red line represents $y = x$. Analyses presented in the main text where done on Iteration 5, which was randomly selected.

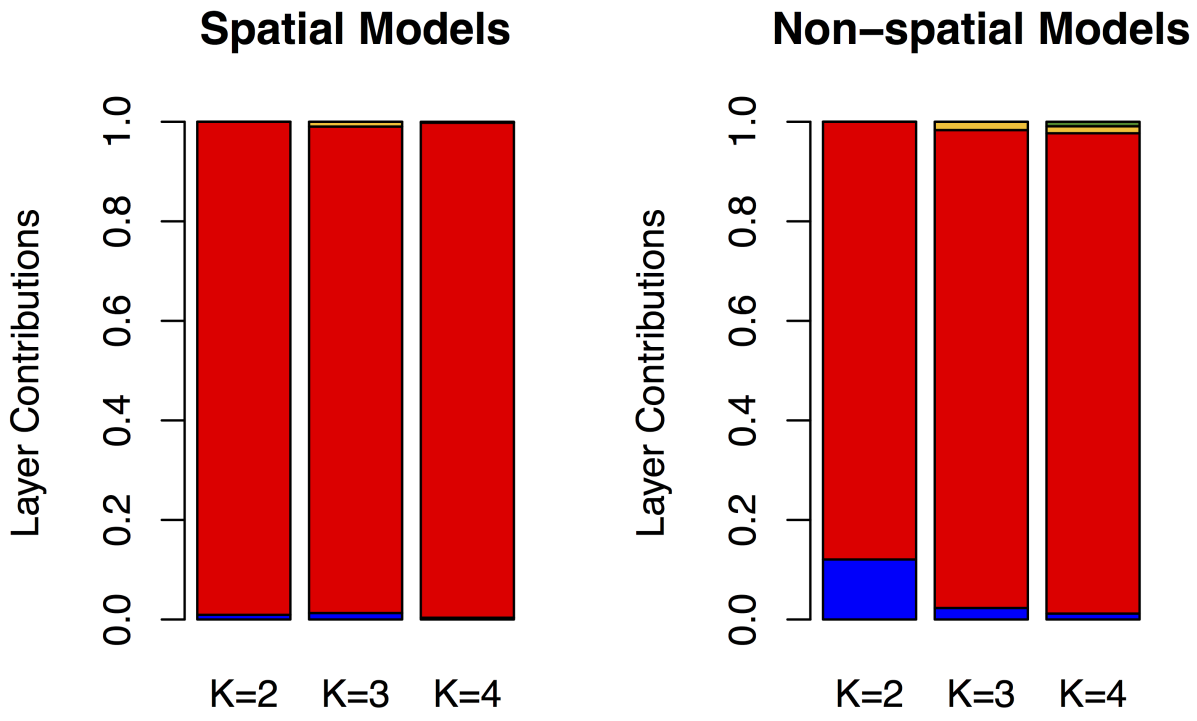

Figure 3.7: Layer contributions for spatial and non-spatial conStruct models fit assuming $k = 2 - 4$ layers.

Supplementary Tables

Table 3.1: Information and phenotypic data for samples used in this study. Population numbers follow Fig 3.1. Species names follow the current taxonomy, which we consider requires modification (see Chpater 2). Pop. refers to the population number in Fig. 3.1; OG stands for Outgroup.

Field No.	**Species**	**Locality**	**Pop.**	**SVL**	p_{bright}
RMP001	*P. aurotaenia*	Tadó, Chocó, Colombia	4	27.28	0.531
RMP011	*P. aurotaenia*	Tadó, Chocó, Colombia	3	25.48	0.315
RMP012	*P. aurotaenia*	Tadó, Chocó, Colombia	3	28.06	0.354
RMP013	*P. aurotaenia*	Tadó, Chocó, Colombia	3	28.16	0.476
RMP015	*P. aurotaenia*	Tadó, Chocó, Colombia	3	29.03	0.282
RMP016	*P. aurotaenia*	Tadó, Chocó, Colombia	4	26.06	NA
RMP017	*P. aurotaenia*	Tadó, Chocó, Colombia	4	29.95	0.386
RMP018	*P. aurotaenia*	Tadó, Chocó, Colombia	4	25.81	0.431
RMP019	*P. aurotaenia*	Tadó, Chocó, Colombia	4	27.26	0.426
RMP020	*P. aurotaenia*	Tadó, Chocó, Colombia	4	31.52	0.448
RMP021	*P. aurotaenia*	Tadó, Chocó, Colombia	4	30.1	0.628
RMP022	*P. aurotaenia*	Tadó, Chocó, Colombia	4	29.14	0.621
RMP023	*P. aurotaenia*	Tadó, Chocó, Colombia	4	20.34	0.282
RMP024	*P. aurotaenia*	Tadó, Chocó, Colombia	4	22.57	0.364
RMP025	*P. aurotaenia*	Tadó, Chocó, Colombia	4	28.13	0.277
RMP026	*P. aurotaenia*	Tadó, Chocó, Colombia	4	28.95	0.383
RMP027	*P. aurotaenia*	Tadó, Chocó, Colombia	4	28.7	0.450
RMP028	*P. aurotaenia*	Tadó, Chocó, Colombia	4	31.34	0.495
RMP029	*P. aurotaenia*	Tadó, Chocó, Colombia	4	29.44	0.680
RMP038	*P. bicolor*	Pueblo Rico, Risaralda, Colombia	1	37.53	0.909
RMP039	*P. bicolor*	Pueblo Rico, Risaralda, Colombia	1	39.26	0.913

Table 3.1 continued from previous page

Field No.	Species	Locality	Pop.	SVL	p_{bright}
RMP040	*P. bicolor*	Pueblo Rico, Risaralda, Colombia	1	40.81	0.917
RMP041	*P. bicolor*	Pueblo Rico, Risaralda, Colombia	1	42.14	0.887
RMP042	*P. bicolor*	Pueblo Rico, Risaralda, Colombia	1	41.16	0.899
RMP043	*P. bicolor*	Pueblo Rico, Risaralda, Colombia	1	40.71	0.894
RMP056	*P. aurotaenia*	Cantón de San Pablo, Chocó, Colombia	9	24.26	0.323
RMP057	*P. aurotaenia*	Cantón de San Pablo, Chocó, Colombia	9	24.77	0.333
RMP058	*P. aurotaenia*	Cantón de San Pablo, Chocó, Colombia	9	28.32	0.306
RMP062	*P. aurotaenia*	Quibdó, Chocó, Colombia	7	26.6	0.289
RMP063	*P. aurotaenia*	Quibdó, Chocó, Colombia	7	26.75	0.246
RMP064	*P. aurotaenia*	Quibdó, Chocó, Colombia	7	25.4	0.249
RMP069	*P. aurotaenia*	Tadó, Chocó, Colombia	3	28.92	0.414
RMP074	*P. aurotaenia*	Tadó, Chocó, Colombia	3	31.76	0.582
RMP075	*P. aurotaenia*	Tadó, Chocó, Colombia	3	29.87	0.671
RMP076	*P. aurotaenia*	Tadó, Chocó, Colombia	3	28.67	0.355
RMP077	*P. aurotaenia*	Tadó, Chocó, Colombia	5	25.05	0.247
RMP078	*P. aurotaenia*	Tadó, Chocó, Colombia	5	26.57	0.231
RMP079	*P. aurotaenia*	Tadó, Chocó, Colombia	5	27.39	0.265
RMP080	*P. aurotaenia*	Tadó, Chocó, Colombia	5	25.08	0.225
RMP081	*P. aurotaenia*	Tadó, Chocó, Colombia	5	24.33	0.236
RMP082	*P. aurotaenia*	Tadó, Chocó, Colombia	5	27.77	0.285
RMP083	*P. aurotaenia*	Tadó, Chocó, Colombia	5	31.14	0.265
RMP084	*P. aurotaenia*	Tadó, Chocó, Colombia	6	28.89	0.314
RMP085	*P. aurotaenia*	Tadó, Chocó, Colombia	6	24.37	0.326
RMP086	*P. aurotaenia*	Tadó, Chocó, Colombia	6	26.86	0.352
RMP087	*P. aurotaenia*	Unión Panamericana, Chocó, Colombia	8	23.46	0.238

Table 3.1 continued from previous page

Field No.	Species	Locality	Pop.	SVL	p_{bright}
RMP089	*P. aurotaenia*	Unión Panamericana, Chocó, Colombia	8	24.67	0.255
RMP091	*P. aurotaenia*	Unión Panamericana, Chocó, Colombia	8	24.95	0.148
RMP117	*P. aurotaenia*	Unión Panamericana, Chocó, Colombia	8	27.69	0.059
RMP118	*P. aurotaenia*	Unión Panamericana, Chocó, Colombia	8	26.97	0.223
RMP120	*P. aurotaenia*	Tadó, Chocó, Colombia	6	25.34	0.342
RMP121	*P. aurotaenia*	Tadó, Chocó, Colombia	6	31.7	0.464
RMP122	*P. aurotaenia*	Tadó, Chocó, Colombia	6	25.75	0.315
RMP124	*P. aurotaenia*	Tadó, Chocó, Colombia	6	29.03	0.443
RMP125	*P. aurotaenia*	Tadó, Chocó, Colombia	6	29.81	0.303
RMP126	*P. aurotaenia*	Tadó, Chocó, Colombia	6	24.35	0.274
RMP127	*P. aurotaenia*	Tadó, Chocó, Colombia	6	28.26	0.228
RMP128	*P. aurotaenia*	Tadó, Chocó, Colombia	6	26.79	0.294
RMP129	*P. aurotaenia*	Tadó, Chocó, Colombia	6	28	0.234
RMP130	*P. aurotaenia*	Tadó, Chocó, Colombia	6	27.66	0.327
RMP131	*P. aurotaenia*	Tadó, Chocó, Colombia	6	27.96	NA
RMP132	*P. aurotaenia*	Tadó, Chocó, Colombia	6	21.77	0.268
RMP133	*P. aurotaenia*	Tadó, Chocó, Colombia	6	25.04	0.246
RMP134	*P. aurotaenia*	Tadó, Chocó, Colombia	6	32.01	0.331
RMP135	*P. aurotaenia*	Tadó, Chocó, Colombia	6	28.19	0.282
RMP141	*P. bicolor*	Pueblo Rico, Risaralda, Colombia	2	38.11	0.758
RMP142	*P. bicolor*	Pueblo Rico, Risaralda, Colombia	2	31.93	0.651
RMP140	*P. aurotaenia*	Tadó, Chocó, Colombia	6	NA	NA
RMP930	*P. aurotaenia*	Tadó, Chocó, Colombia	3	NA	NA
RMP918	*P. aurotaenia*	Tadó, Chocó, Colombia	3	NA	NA
MGG359	*P. terribilis*	Timbiqui, Cauca, Colombia	OG	NA	NA

Table 3.1 continued from previous page

Field No.	**Species**	**Locality**	**Pop.**	**SVL**	p_{bright}
MGG371	*P. terribilis*	Timbiqui, Cauca, Colombia	OG	NA	NA
MGG391	*P. terribilis*	Timbiqui, Cauca, Colombia	OG	NA	NA
MGG395	*P. terribilis*	Timbiqui, Cauca, Colombia	OG	NA	NA
MGG397	*P. terribilis*	Timbiqui, Cauca, Colombia	OG	NA	NA

Table 3.2: Directionality indices (Ψ) between localities used for range expansion analyses. Population numbers follow Fig. 3.1

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**
2	0.0425							
3	0.1105	0.0995						
4	0.1244	0.1099	0.0186					
5	0.1299	0.1441	0.0320	-0.0267				
6	0.1903	0.2014	-0.1243	-0.1141	-0.0839			
7	0.0395	0.0486	-0.0554	0.0427	-0.0848	-0.1484		
8	0.2537	0.2818	0.1951	-0.2142	-0.1443	0.0326	-0.2219	
9	0.1255	0.1690	-0.0861	-0.0969	-0.0612	0.0422	-0.1615	0.1479

CHAPTER 4

THE GENETIC BASIS AND EVOLUTIONARY HISTORY OF COLOR PATTERN VARIATION IN *PHYLLOBATES* POISON FROGS

Roberto Márquez, Mileidy Betancourth-Cundar, Manuel Guayara, Marcus R. Kronforst

4.1 Abstract

Understanding how heritable variation arises in a population is instrumental to the study of phenotypic evolution and adaptation, since such variation is the substrate upon which selective pressures act to drive evolution. New character states can enter a population in three ways: *de novo* mutation, selection and recombination of preexisting alleles into new genotypes, or introgression from another population. Here we evaluate the roles of these three processes in the evolution of color pattern in *Phyllobates* poison-dart frogs, where this trait has evolved dynamically in the genus's recent evolutionary history. Using genome-wide association and divergence scans, we identified genomic regions associated with color pattern variation within and between natural populations of *Phyllobates*, and investigated the evolutionary history of alleles at these loci in order to shed light on the processes through which new color-pattern variants have entered and evolved in populations of this group. We found several genes putatively involved in melanosome movement to be associated with color pattern variation, suggesting a role for this process in the formation of the different color patterns observed in *Phyllobates*. Genetic variation at these loci and their adjacent genomic regions was consistent with independent evolution of solid-yellow alleles and not introgressive hybridization as the driver of color pattern convergence. However, our data do not allow us

to address whether these alleles originated *de novo* in each lineage, or were selected from preexisting ancestral variants. This work constitutes a first step towards understanding the functional and evolutionary mechanisms behind convergent evolution of coloration in poison-dart frogs.

Key Words: Convergent evolution, divergence scan, GWAS, melanophore, melanosome.

4.2 Introduction

A phenotypic trait evolves when it is variable among members of a population and this variation is heritable. If variation is also correlated with differential fitness, then natural selection will influence the direction of evolutionary change [8, 280], increasing the frequency of traits with higher fitness. Therefore, a fundamental goal of evolutionary biology is understanding how heritable phenotypic variation arises and is selected upon within populations. *De novo* mutation is the ultimate source of genetic (and therefore heritable phenotypic) variation. However, it is not the only way in which new phenotypic variants can enter a population. Trait variation can also arise as selection brings preexisting genetic variants together into new, recombinant genotypes that code for new phenotypes, and alleles can originate in other populations (or species) and get introgressed into the focal population through hybridization and backcrossing. Although all three mechanisms are known to contribute to adapttation in natural populations [eg. 281–286], their relative importance is not well understood. Adaptation is expected to occur faster under the two latter scenarios, since selection will act upon variation that is already present, rather than "waiting" for random mutations that generate beneficial alleles [287–290]. However, the specific alleles that are beneficial under a certain selective regime may not be present within a population's standing genetic variation [288, 289], and hybridization between species or highly diverged populations is often, in itself, disadvantageous [21, 287, 291].

Traits that evolve dynamically in young clades provide a promising opportunity to study the origins of beneficial genetic variants. In poison frogs of the family Dendrobatidae, conspicuous coloration has evolved independently in multiple clades [292,293], and is known to be involved in several aspects of these frogs' biology [294], such as predator avoidance [295–297], mate choice [193–195, 298, 299], and territoriality [300]. Within Dendrobatidae, the genus *Phyllobates* is composed of five nominal species that exhibit either a solid-yellow color pattern

or a brightly-colored dorsolateral stripe on a dark background [25,99] (Fig. 1.1, 4.1). Recent work has shown that the solid-yellow populations can be grouped into three distinct lineages, all of which are more closely related to striped lineages than to other yellow ones [245] (Fig. 4.1A). Furthermore, Márquez et al. [245] identified a corridor of gene flow along the upper San Juan river, which connects several striped and yellow populations, including some with intermediate phenotypes, which transition clinally between striped and solid-yellow [25, 99, see Chapter 3] (Fig. 4.1B). The dynamic evolution of color pattern in the recent history of *Phyllobates* (∼2.5 million years [245]) revealed by these studies makes this system especially well-suited to study the origin of heritable phenotypic variation.

Our goal in this paper is to investigate the roles of *de novo* mutation, introgression, and selection of standing ancestral variation in the evolution of color pattern in *Phyllobates* frogs. To this end, we take advantage of natural variation in color pattern, both between and within populations of this genus, in order to identify loci associated with color pattern variation, and interrogate the patterns of genetic variation at these loci to shed light on the processes through which the current diversity of color patterns present in *Phyllobates* populations arose.

4.3 Materials and Methods

Identifying loci associated to color pattern

To pinpoint loci associated with color pattern, we took advantage of two instances of variation in recently diverged natural populations (Fig. 4.1): First, we conducted a genome-wide divergence scan between *P. terribilis* and the southern populations of *P. aurotaenia*, to which *P. terribilis* is closely related [27, 245]; Fig. 4.1A, C). Second, we integrated genome-wide association (GWA) analysis with divergence scans in closely related populations along

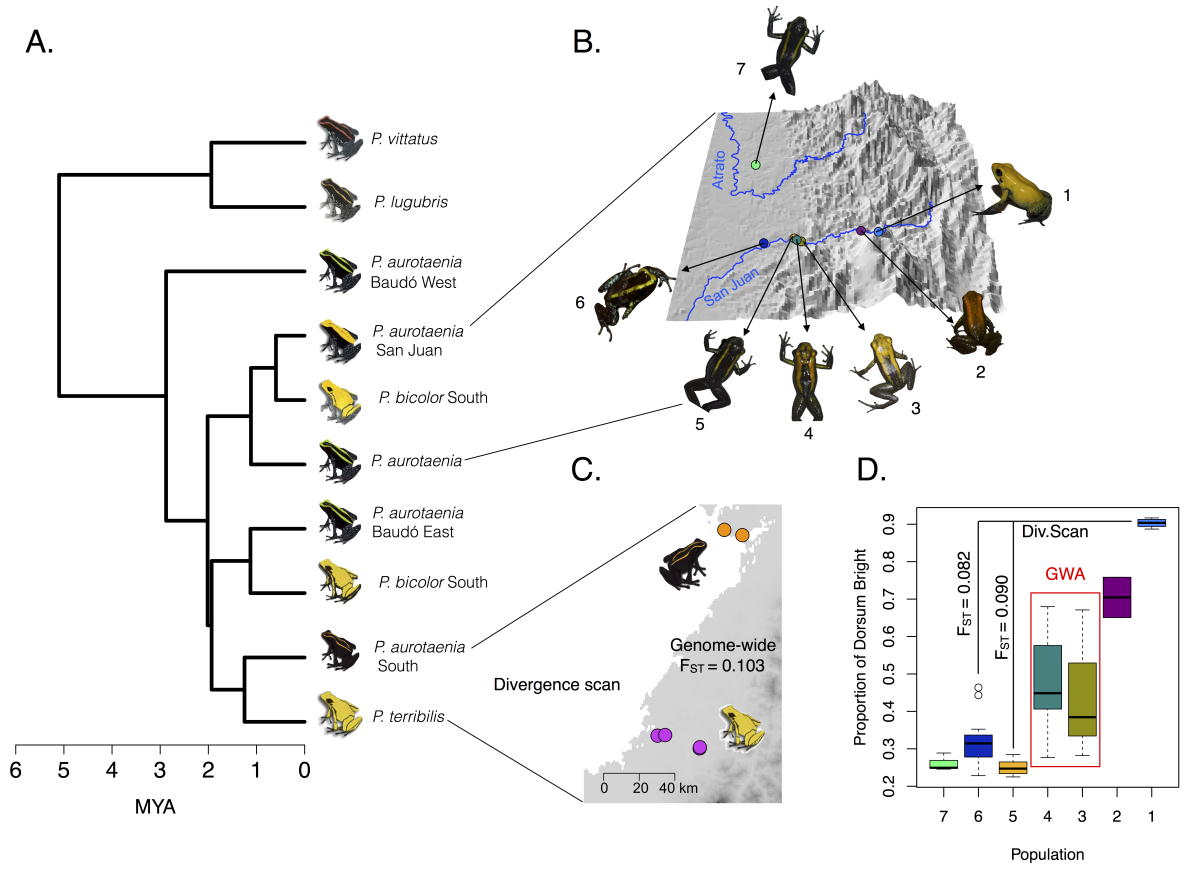

Figure 4.1: Phylogenetic relationships (A) and natural variation in color pattern among *Phyllobates* poison frogs (B-D). Phylogenetic relationships and divergence times follow ref. 245, and measurements in Panel D correspond to the proportion of the dorsum covered in yellow pigmentation, obtained from Chapter 3. F_{ST} values are genome-wide weighted averages calculated as detailed in the Materials and Methods section. In panels C-D we indicate the analyses used to identify loci associated to coloration in each instance.

the San Juan River color pattern cline (Fig. 4.1B, D).

For the first divergence scan we generated whole-genome re-sequencing data for 9 southern *P. aurotaenia* and 7 *P. terribilis* (Table 4.1), and added the 5 *P. terribilis* re-sequencing datasets generated in Chapter 3 (Table 3.2). DNA was extracted from toe-clippings, mouth-swabs or liver samples using Qiagen DNeasy Blood and Tissue kits, following the manufacturer's guidelines, and sheared on a Covaris S-series ultrasonicator to a fragment size distribution peaked at $\sim$ 200 bp. Paired-end (PE) Illumina libraries were built using the Kappa HyperPrep kit (Roche), amplified through 5-7 cycles of PCR, and sequenced on a HiSeq 4000 100bp PE flowcell. The resulting reads were quality-filtered and trimmed using Trimmomatic v. 0.36 [132] and Skewer v. 0.22 [141] under default parameters, save for minimum read lengths and base qualities, which were set to 36 and 15, respectively. We then used the bioinformatic pipeline detailed in Chapter 3 to map reads to the *P. terribilis* reference genome (440,103 scaffolds, 3.8Gb, N50 = 11Kb; Machado, Márquez, Reed, Janies, Kronforst & Grant, *Unpublished*) and to filter the resulting mappings. Briefly the reference genome was assembled and scaffolded using Illumina paired-end, mate-paired, and in vivo Hi-C libraries, and the bioinformatic pipeline consisted of mapping reads, identifying PCR duplicates, and filtering for genotyping rate, strand bias, base quality bias, Hardy-Weinberg equilibrium, and read mismapping. Filtering was performed separately for each species, and only sites that passed all filters in both species were used. At this point, three *P. terribilis* samples, all derived from mouth-swabs, were discarded due to low levels of genotyping and coverage.

We identified divergent sites between *P. terribilis* and its sister *P. aurotaenia* populations (Fig. 4.1C) using the likelihood ratio test developed by Kim et al. [301], which evaluates whether allele frequencies at a given site differ between two sets of samples. Tests were performed in ngsAssociation (https://github.com/tplinderoth/ngsAssociation) [250], excluding

sites with sequencing depth per individual above 120X. To at least partially account for the effect of population stratification we estimated the genome-wide score inflation factor λ [302] using the R package GenABEL [303] (function estlambda, method = "regression"), and divided the likelihood ratio (LR) for each site by λ. Finally, we calculated p-values for the weighed LRs based on a χ^2 distribution with one degree of freedom. Associated regions for further analysis were identified by plotting log-transformed p-values along the genome assembly (ie. a Manhattan plot), and searching for "tower-like" peaks of association. Since the highly fragmented nature of our genome assembly poses a challenge for peak identification, we also retained 5Kb windows containing at least three significantly differentiated SNPs ($\alpha = 0.05$ after FDR adjustment [304]).

For the San Juan cline, we used the mapped whole-genome re-sequencing data (ie. bam files) and color pattern measurements (i.e. proportion of the dorsum brightly colored, p_{bright}) generated in Chapter 3 (Table 3.2). We searched for loci that were 1) highly differentiated between solid-yellow (*P. bicolor*) and striped (*P. aurotaenia*) populations at both ends of the cline, and 2) statistically associated with p_{bright} in populations showing intermediate color patterns (Fig. 4.1B, D). Differentiated loci were identified based on the population branch statistic (PBS), which quantifies the degree of divergence of a lineage from its common ancestor with two other lineages, based on pairwise F_{ST} estimates [305]. We estimated F_{ST} for non-overlapping 5Kb windows along the genome using the version of Hudson's estimator [306] derived by Bhatia et al. [307], which estimates F_{ST} between two populations at a single, biallelic marker:

$$F_{ST} = \frac{(\hat{p}_1 - \hat{p}_2)^2 - \frac{\hat{p}_1(1-\hat{p}_1)}{n_1-1} - \frac{\hat{p}_2(1-\hat{p}_2)}{n_2-1}}{\hat{p}_1(1-\hat{p}_2) + \hat{p}_2(1-\hat{p}_1)} \tag{4.1}$$

Where $\hat{p}_i$ and n_i are the sample allele frequency and sample size for population i, respectively. The weighted average F_{ST} for a given genomic window with m SNPs was calculated as:

$$\bar{F}_{ST} = \frac{\sum_{i=1}^{m} N_m}{\sum_{i=1}^{m} D_m} \tag{4.2}$$

Where N_m is the numerator and D_m the denominator of equation 4.1 applied to marker m. Finally, for each window PBS was calculated as

$$PBS_{Pop1} = \frac{T_{12} + T_{13} - T_{23}}{2} \tag{4.3}$$

Where T_{ij} is the log-transformed $\bar{F}_{ST}$ ($-log[1.05 - \bar{F}_{ST}]$) between populations i and j. Note that our transformation is a slightly modified version of the commonly used $-log(1 - F_{ST})$, which we use to account for the fact that $-log(1 - F_{ST})$ tends to infinity as F_{ST} approaches 1. We estimated the PBS for *P. bicolor* against three different sets of striped taxa, which involved all combinations of two closely related and geographically proximal populations of *P. aurotaenia* (populations 5 and 6 in Fig. 4.1B), and the Southern populations of *P. aurotaenia* (sister to *P. terribilis*, see Fig. 4.1A, C). Two individuals from one of the San Juan striped populations (pop. 6 in Fig. 4.1B) had noticeably higher p_{bright} values than the rest (see Fig. 4.1D) and were therefore excluded from PBS calculations. Windows that were among the 100 with the highest PBS in all three comparisons were retained as for further analysis.

GWA analysis was done using the univariate linear mixed model approach implemented

in GEMMA [308], which uses a relatedness matrix to account for population structure. As input for GEMMA we used posterior mean genotypes, obtained by estimating genotype likelihoods in Angsd v.0.918 [152], and using BEAGLE v. 3.3.2 [309] to impute genotypes and calculate posterior mean genotypes. Posterior mean genotypes range from 0 to 2, and represent a continuous measure of the number of copies of the derived allele in a single diploid individual, weighed by genotyping uncertainty. These values are calculated as $0(P[AA]) + 1(P[Aa]) + 2(P[aa])$, where A is the major allele, a is the minor allele, and $P[xy]$ is the posterior probability of genotype xy. GWA analysis was limited to polymorphic sites (SNP p-value < 0.001) with at least 80% call rate prior to imputation, and minor allele frequencies above 0.05 after imputation. Given our small sample size for this analysis (17 individuals) no sites displayed statistically significant associations with p_{bright} after FDR adjustment of the 1,517,120 p-values ($\alpha = 0.05$), so we retained 5Kb regions with at least three of the 100 most highly associated SNPs as candidates. P-values in this subset ranged from 3.4×10^{-5} to 4.0×10^{-7}.

Annotation of associated loci

We annotated the scaffolds containing associated regions using AUGUSTUS [310,311] to perform *de novo* coding sequence prediction, and querying the predicted peptide sequences against the online version of the PANTHER HMM scoring tool [312]. For loci where AUGUSTUS or PANTHER were unable to predict or annotate coding sequences we queried the scaffold containing each locus against the anuran (taxid: 8342) and human (taxid: 9606) proteins in the RefSeq database using blastx [135] with a 10^{-5} e-value cutoff, and used the highest scoring hit as a provisional annotation. We then re-queried the translated coding sequence that matched the best blastx hit against PANTHER for a final annotation.

Genetic variation at candidate loci

Finally, we selected a set of candidate color pattern loci based on their annotations (see Results section), and visualized genetic variation at these regions across the 86 *Phyllobates* genomes sequenced between Chapters 3 and 4. To gain insight on genetic structure at both broad and narrow genomic resolutions, we built minimum evolution trees (similar to neighbor-joining) [313] of the 50Kb region surrounding each associated locus and plotted alignments of the variable sites within each chosen 5Kb window. Minimum evolution trees were built in FastME [150] using distance matrices generated from genotype likelihoods in ngsDist [151], and alignments were plotted based on posterior mean genotypes calculated from posterior genotype probabilities estimated in Angsd. Only sites with SNP p-values < 0.01 and genotyped for at least 50 individuals were included in alignments. For loci within scaffolds shorter than 50 or 5Kb we analyzed the whole scaffold. Since our selection of candidate loci relied entirely on sequence affinity to genes previously associated with coloration in vertebrates, we also performed the above analyses on non-candidate loci.

4.4 Results

Association and divergence analyses

The genome scan between *P. terribilis* and the southern *P. aurotaenia* produced five candidate regions, one of which formed the "tower-like" pattern characteristic of association in genome-wide studies (Fig. 4.2). Expectedly, this peak was on the largest of the scaffolds with associated regions. Two of the five regions contained annotatable coding sequences, which corresponded to the *dnah9* and *kiaa1217* genes. The San Juan GWA analysis on the other hand, produced three associated regions (Fig. 4.3A), all of which were located in small (12-18.7Kb) scaffolds. Two of them were annotated, one as a transposable element-

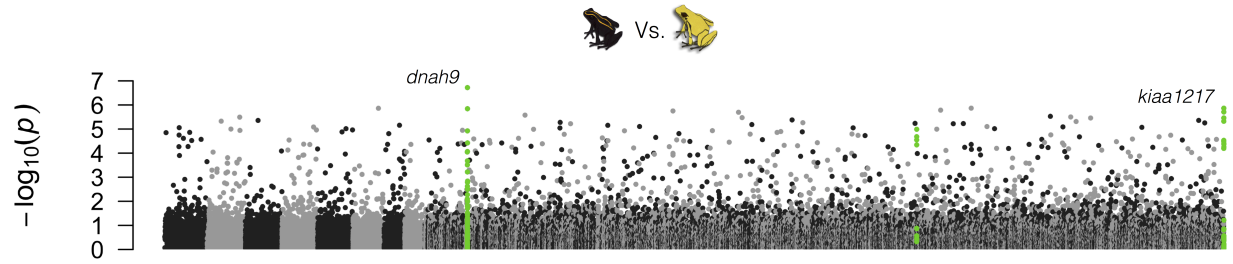

Figure 4.2: Manhattan plot of divergence between *P. terribilis* and the southern *P. aurotaenia*. Points represents SNPs arranged along the X axis according to their position in the reference genome assembly, with scaffolds organized by size. On the Y axis are log-transformed p-values from the Kim et al. [301] test for allele frequency differences. SNPs highlighted in green are within regions identified as associated with color pattern, and the genes near each association peak are indicated next to the peak.

related protein, and the other as a reverse transcriptase, both uncharacterized. Finally, the divergence scan between *P. bicolor* and the narrow-striped populations of *P. aurotaenia* in the San Juan recovered 29 highly diverged 5kb windows, out of which nine were annotated with the *sfpq*, *gpr37l*, *smg5*, *lats2*, and *rdh7* genes, a *Gipsy* retrotransposase, an uncharacterized reverse transcriptase, a *PiggyBac* transposable element-derived protein, and an uncharacterized transposase (Fig. 4.4).

Unexpectedly, we found very little agreement between GWA and divergence analyses (Fig. 4.3B). In light of the highly fragmented state of our reference genome assembly, and the fact that most associated/divergent loci were within scaffolds shorter than 30Kb, we ascribe this result to our inability to detect coincidences of association and divergence at a genomic scales beyond a few kilobases. However, we cannot discard the possibility that either (or both) of our analyses, especially the GWAS, lack statistical power to detect genotype-phenotype associations.

Candidate gene selection

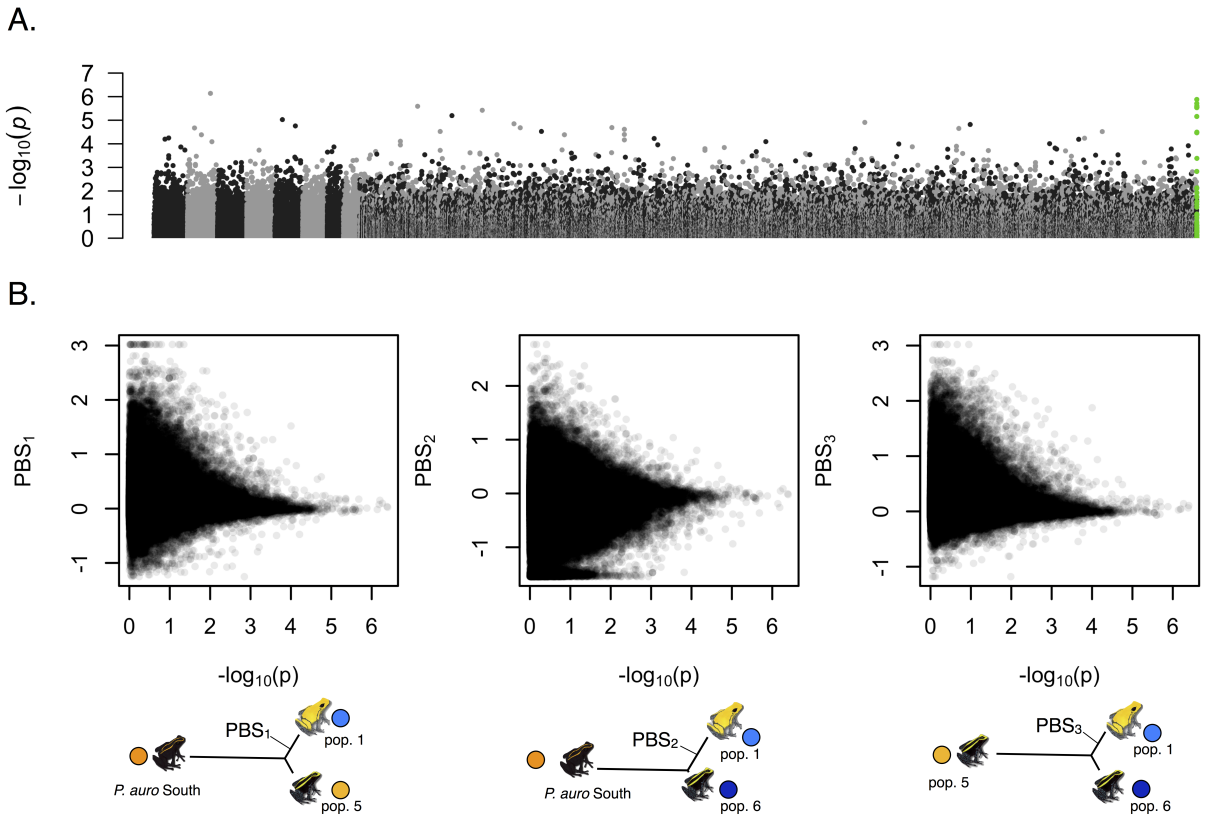

Figure 4.3: Manhattan plot of association with color pattern within populations in the San Juan drainage that display intermediate color patterns (A), and the relationship of association with divergence between *P. bicolor* and several striped populations (B). SNPs in panel A are highlighted as in Fig. 4.2. Tree cartoons in panel B depict the populations used to estimate each population branch statistic.

Out of the 12 annotated protein-coding genes, we selected *gpr37l*, *kiaa1217*, *dnah9*, and *lats2* as candidate color pattern genes, since they potentially play a role in the development or cellular processes of melanophores, the melanin-bearing cells that make up black and dark brown pattern elements in amphibian skin. Although *rdh7* and its human ortholog, *rdh16*, play a role in the vitamin A pathway, which is involved in carotenoid metabolism [314], they act well downstream of carotenoid cleavage, suggesting that variation at *rdh7* is unlikely to affect carotenoid-based coloration. Below we detail our reasoning for considering each gene a viable color pattern candidate in *Phyllobates*:

G-protein coupled receptor 37-like 1 (*gpr37l*). This gene codes for a A G-protein coupled peptide receptor which is involved in the prosaptide-mediated regulation of the MAPK/ERK pathway [315, 316]. Among several other processes, this pathway has been shown to regulate the bidirectional transport of melanosomes (melanin-containing organelles) within melanophores [317, 318], a process involved in vertebrate color change.

Heavy-chain axonemal dynein 9 (*dnah9*). Dyneins are a large gene family composed of cytoskeletal motor proteins, which form large protein complexes that move cargo along microtubules, or move microtubules themselves. Some members of this family are involved in the movement of melanosomes along microtubules [93, 44, 319] in multiple vertebrates, including *Xenopus laevis* [320]. It is worth noting, however, that only cytoplasmic dyneins are known to be involved in melanosome transport [44]. Since *dnah9* codes for an axonemal dynein, more commonly involved in the movement of cilia and flagella [321], we consider this gene a tentative candidate.

Sickle-tail protein homolog (*kiaa1217*). This gene codes for the sickle-tail protein, generally associated with skeletal development [322, 323]. However, a recent study found a SNPs in this gene to be associated with tanning ability in humans [62]. Although the association was weak, and a mechanistic explanation for the involvement of this gene in

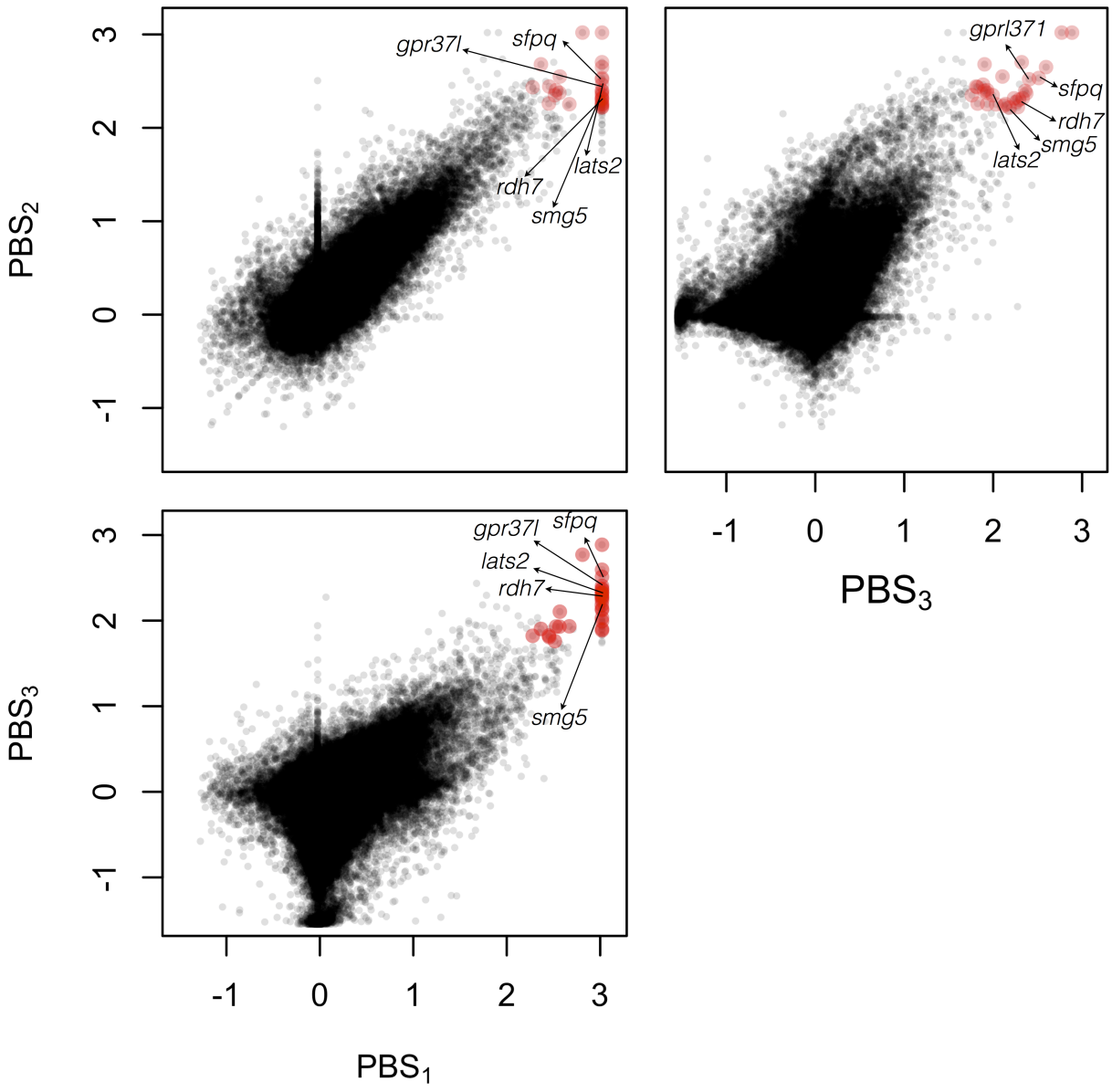

Figure 4.4: Bi-plots of the three population branch statistics (PBS) calculated between *P. bicolor* and different sets of striped taxa for non-overlapping 5Kb windows. The 29 windows that were among the 100 most divergent in the three comparisons are highlighted in red, and the annotated genes near these windows are indicated, except of those affiliated to transposable elements, which were committed.

skin tanning has not been proposed, we retained it as a tentative candidate considering that during suntanning, melanosomes are transported from the center of melanocytes to their neighboring keratinocyte cells [324] along microtubules, in a similar fashion as melanosomes are transported within amphibian melanohpores [325, 326].

Large tumor suppressor kinase 2 (*lats2*). The *lats2* protein is a serine/threonine protein kinase, which is involved in the regulation of *pax3* via the *Hippo* pathway [327]. *Pax3* is a well-known pigmentation gene, which is involved in the development of melanophores/melanocytes [328–330] and xanthophores (yellow and red pigment cells) [330].

Genetic variation at candidate loci

Trees of the 50Kb regions surrounding candidate loci showed little genetic structure in general (Fig. 4.5). In the *dnah9* (Fig. 4.5B) and *kiaa1217* (Fig. 4.5C) trees all *P. terribilis* samples clustered in a single clade, separate from all other samples. The *dnah9* tree also displayed genetic structure consistent with phylogeny, with the clade containing *P. terribilis* and the southern *P. aurotaenia* being reciprocally monophyletic with all other samples (Fig. 4.5B). The *lats2* tree showed two clearly separate clusters, but these did not coincide with geography, phylogeny, or color pattern (Fig. 4.5D), and the *gpr37l* tree displayed no discernible genetic structure. Trees of non-candidate loci showed similar patterns in general, with 11 trees displaying structure consistent with phylogeny, at least to some degree, and four trees where all *P. bicolor* individuals were grouped in a single clade shared with a few *P. aurotaenia* from populations with intermediate color pattern. However, no trees contained clades where *P. bicolor* and *P. terribilis* samples were clustered in the same clade.

Three out of the four alignments of 5Kb candidate regions displayed sites fixed (or nearly fixed) for the minor allele in only one of the solid yellow lineages (*gpr37l* in *P. bicolor*, and

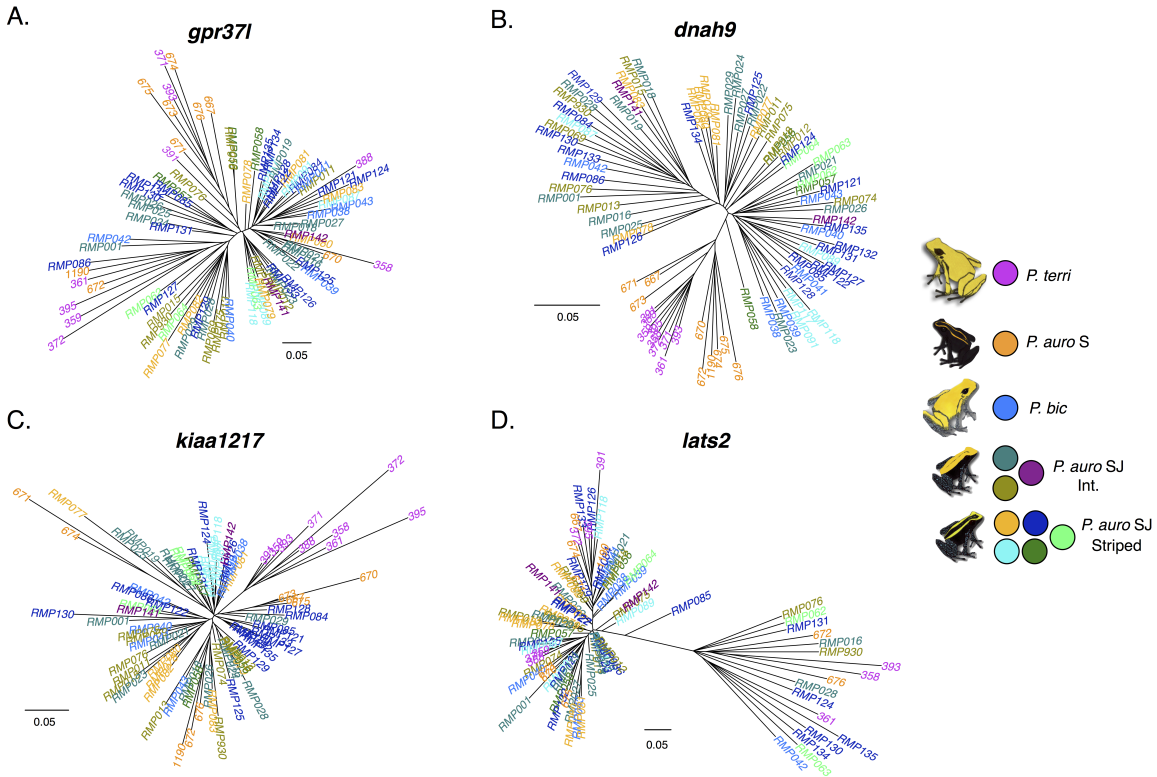

Figure 4.5: Minimum-evolution trees of the 50kb regions surrounding candidate color pattern genes. Sample names are colored following geographic localities as in Fig. 4.1, and the color pattern of each locality is indicated in the legend.

dnah9 and *kiaa1217* in *P. terribilis*; Fig. 4.6A-C). In the fourth alignment (*lats2*), both yellow lineages displayed one site nearly fixed for the minor allele, albeit a different site in each lineage (Fig. 4.6D). It is worth noting, however, that the genotyping rate was low for both lineages in this region ($n = 3$ in both cases). The alignments of non-candidate loci also revealed similar patterns, with no sites being fixed for the minor allele in both *P. terribilis* and *P. bicolor*.

4.5 Discussion

The history of candidate and other associated loci

The main goal of this study was to evaluate the extent to which *de novo* mutation, selection from ancestral variation, and introgression have contributed to the phenotypic similarity between *P. terribilis* and *P. bicolor*. The fact that we did not find evidence of allele sharing or close genetic affinity between these two species at candidate color-pattern loci strongly suggests that introgressive hybridization has not been an important driver of the phylogenetic distribution of solid-yellow patterns in *Phyllobates*. This falls in line with results from previous work [245], which found no evidence of recent admixture between *Phyllobates* species with disjunct, distant distribution ranges, as is the case for *P. terribilis*. The complete absence of overlap between the genomic regions associated with color pattern variation in *P. bicolor* and *P. terribilis* could imply that they acquired their solid-yellow coloration through changes at different genes, further supporting this notion. However, considering that we have not likely identified all loci associated with color pattern, this conjecture must remain speculative.

Determining whether the mutations responsible for color pattern similarity between *P. terribilis* and *P. bicolor* arose *de novo* in each lineage, or were already segregating in their

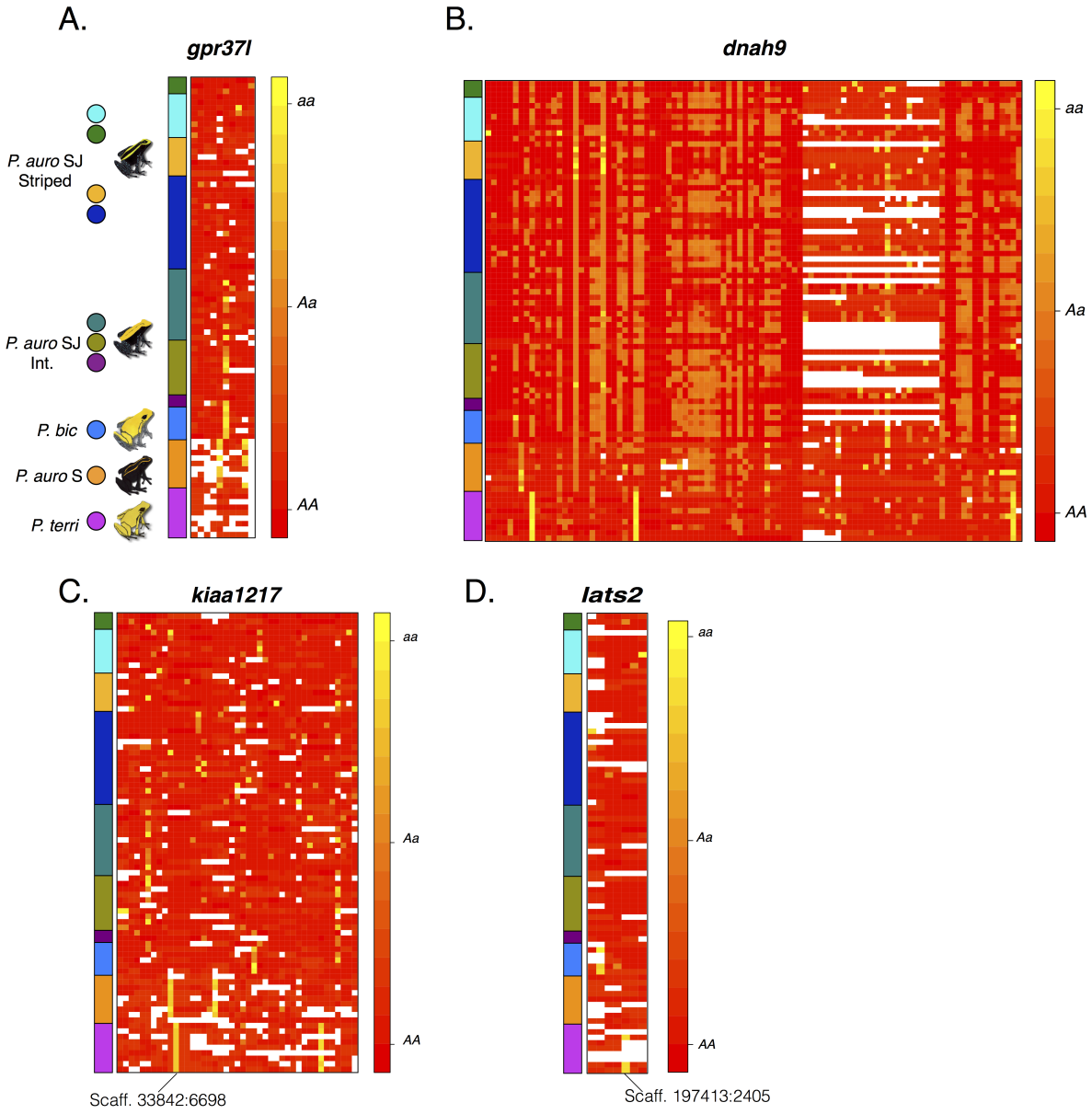

Figure 4.6: Mean posterior genotypes of variable sites in associated 5kb windows. Columns represent sites and rows represent individuals. Rectangles to the left of each alignment denote each individual's geographic locality, and frog cartoons indicate color patterns.

common ancestor is a challenging endeavor. Our results can nonetheless provide initial insight on this question. Some of the minor alleles at candidate loci that are fixed in solid-yellow lineages also occur in some striped lineages at low frequencies, suggesting that they were segregating in the presumably striped ancestral populations of solid-yellow lineages. For example, site 2405 in scaffold 197413 (Fig 4.6D), is nearly fixed in *P. terribilis*, and is also present in one heterozygous *P. aurotaenia* individual from a population in the San Juan drainage, which is more closely related to *P. bicolor*. However, several minor alleles were found exclusively in one of the striped lineages (eg. Scaffold 33842:6698; Fig. 4.6C). Although this is compatible with these alleles having arisen *de novo* in the striped lineage where they occur, we cannot discard the possibility that these alleles are actually segregating at low frequencies in some striped populations, but our low sample sizes impede us from detecting rare alleles. Finally, the fact that we did not find any shared derived alleles between solid-yellow lineages suggests that, even if the alleles responsible for their color pattern similarity were present in their most recent common ancestor, *P. terribilis* and *P. bicolor* appear to have acquired their color patterns through relatively independent evolutionary processes. Further work that more explicitly explores the evolutionary history and selective processes experienced by alleles at these loci, for example whether they rose to high frequencies via hard or soft selective sweeps, will greatly improve our understanding of the evolutionary genetic processes behind color pattern convergence in *Phyllobates*.

Functional mechanism behind color pattern variation in Phyllobates

In most species of poison frogs where development has been observed, coloration develops mainly from the late tadpole to the early post-metamorphic stages, with juvenile coloration largely resembling that of adults [25, 331, 332, R. Márquez *pers. obs.*]. This is the case for the striped *Phyllobates*, where the grayish-brown tadpoles develop a faint dorsolateral stripe

shortly after the emergence of back legs. Within a few days after metamorphosis, the stripe becomes brighter, mostly attaining adult width and pigmentation within a few days. The process starts similarly in the solid-yellow lineages, but after the striped pattern develops, dark pigmentation progressively fades over several months, giving way to yellow until the solid-yellow adult coloration is attained [25, R. Márquez *pers. obs*] (Fig. 1.3A).

Given that three of the four genes with links to melanic coloration (*gpr37l*, *dnah9*, and *kiaa1217*) are potentially involved in melanosome transport, this cellular process could play a role in the ontogenetic color change experienced by solid-yellow *Phyllobates*. Within melanophores, melanosomes can be either aggregated at the center of the cell, giving it a pale appearance, or dispersed throughout the cytosol, making it appear dark [44]. Frog melanophores possess dendritic processes that extend over iridophores, a third kind of pigmented cell, recently shown to be heavily involved in producing orange and yellow coloration in poison frogs, including *P. terribilis* [333]. When melanin is scattered within chromatophores, it can travel through these extensions, occluding the iridophores [47] (see Fig. 1.2). Therefore, the loss of dark pigmentation during the development of solid-yellow coloration could be the result of a gradual retraction of melanosomes to the center of melanophores revealing the underlying iridophores. Future studies characterizing the structure of pigment cells and the expression of candidate loci throughout color pattern development will certainly expand our understanding of the functional mechanisms underlying color pattern evolution in these frogs.

Caveats

Although this study constitutes a substantial advance in our understanding of the evolutionary genetics of convergent evolution in *Phyllobates* poison frogs, it is necessary to point out some caveats which must be considered when interpreting our results. First, although we

undertook a large sampling effort, producing whole-genome resequencing data for 87 frogs from 14 localities, our sample sizes within these localities are small, which has probably reduced our statistical power to detect genotype-phenotype associations. Second the highly fragmented nature of our reference genome assembly impedes us from investigating spatial patterns of genetic variation across the genome, a useful tool to disentangle true associations from sporadic ones, as well as to investigate the history of alleles at associated loci. Considering these two caveats, we would not be surprised if 1) we have not detected some loci involved in color pattern variation, and 2) some of our associated loci represent false positives. As better genomic resources and functional genomic tools become available for poison frogs, we should be able to address these caveats and produce a more robust understanding of the genetic basis of color pattern evolution in *Phyllobates*.

Concluding remarks

Understanding the evolutionary history of alleles involved in convergent evolution can reveal a great deal about the drivers of adaptation in natural populations [334,335]. Overall, our results suggest that the convergent evolution of solid-yellow color patterns in *Phyllobates* has occurred in the absence of introgression between at least two of the three lineages that display such color patterns, indicating that coloration has evolved through relatively independent processes in these two lineages. This study constitutes an important first step towards understanding the genetic and functional mechanisms driving the evolution of aposematic coloration in poison frogs.

4.6 Acknowledgements

We are grateful to Trevor Price, John Novembre, Marko Horb, Nick Van Kuren, and Darli Massardo for insightful comments and discussion, and to Lina M. Arenas for allowing us to use her frog illustrations for the figures. This study was supported by grants DEB-1702014, IOS-1922624, and IOS-1827333 from the National Science Foundation, R35GM131828 from the National Institutes of Health, 9786-15 from the National Geographic Society, 734-2015 from the Colombian Administrative Department of Science, Technology and Innovation (Colciencias) and Empresa de Energía del Pacífico (EPSA), and by a Pew Biomedical Fellowship. Computations were carried out on the University of Chicago's Gardner computer cluster, which is funded by NIH grant TR000430. Tissue and specimen collections were authorized by permits No. 2194, 1177, and 1380 from the Colombian Authority of Environmental Licenses (ANLA), and animal usage protocols approved by the University of Chicago's Institutional Animal Care and Use Committee (ACUP No. 72416).

4.7 Author Contributions

R.M. and M.R.K. conceived and designed the study. R.M. and M.R.K. acquired funding. R.M and M.G. collected samples. R.M. and M.B-C generated sequence data. R.M. analyzed the data and wrote the manuscript.

4.8 Supplementary information

Supplementary Tables

Table 4.1: Samples sequenced in Chapter 4.

Field Number	Species	Locality
MGG350	*P. terribilis*	Timbiqui, Cauca, Colombia
MGG357	*P. terribilis*	Timbiqui, Cauca, Colombia
MGG358	*P. terribilis*	Timbiqui, Cauca, Colombia
MGG388	*P. terribilis*	Timbiqui, Cauca, Colombia
MGG372	*P. terribilis*	Timbiqui, Cauca, Colombia
MGG393	*P. terribilis*	Timbiqui, Cauca, Colombia
MGG361	*P. terribilis*	Timbiqui, Cauca, Colombia
GECOH1190	*P. aurotaenia*	Buenaventura, Valle del Cauca, Colombia
GECOH664	*P. aurotaenia*	Buenaventura, Valle del Cauca, Colombia
GECOH671	*P. aurotaenia*	Buenaventura, Valle del Cauca, Colombia
GECOH674	*P. aurotaenia*	Buenaventura, Valle del Cauca, Colombia
GECOH667	*P. aurotaenia*	Buenaventura, Valle del Cauca, Colombia
GECOH672	*P. aurotaenia*	Buenaventura, Valle del Cauca, Colombia
GECOH675	*P. aurotaenia*	Buenaventura, Valle del Cauca, Colombia
GECOH670	*P. aurotaenia*	Buenaventura, Valle del Cauca, Colombia
GECOH673	*P. aurotaenia*	Buenaventura, Valle del Cauca, Colombia
GECOH676	*P. aurotaenia*	Buenaventura, Valle del Cauca, Colombia

REFERENCES

[1] Cat PT. I Love my White Shoes. New York: Harper Collins; 2008.

[2] Simpson EH. Measurement of diversity. Nature. 1949;163(4148):688–688. doi:10.1038/163688a0.

[3] Faith DP. Conservation evaluation and phylogenetic diversity. Biological Conservation. 1992;61(1):1–10. doi:10.1016/0006-3207(92)91201-3.

[4] Erwin DH. Disparity: Morphological pattern and developmental context. Palaeontology. 2007;50(1):57–73. doi:10.1111/j.1475-4983.2006.00614.x.

[5] Chao A, Chiu CH, Jost L. Unifying species diversity, phylogenetic diversity, functional diversity, and related similarity and differentiation measures through Hill numbers. Annual Review of Ecology, Evolution, and Systematics. 2014;45(1):297–324. doi:10.1146/annurev-ecolsys-120213-091540.

[6] Jablonski D. Approaches to Macroevolution: 1. General concepts and origin of variation. Evolutionary Biology. 2017;44(4):427–450. doi:10.1007/s11692-017-9420-0.

[7] Darwin CR, Wallace AR. On the tendency of species to form varieties; and on the perpetuation of varieties and species by natural means of selection. Journal of the Proceedings of the Linnean Society of London Zoology. 1858;3(9):45–62. doi:10.1111/j.1096-3642.1858.tb02500.x.

[8] Darwin CR. On the origin of species, by means of natural selection; or, The preservation of favoured races in the struggle for life. London: John Murray; 1859.

[9] Huxley JS. Evolution The Modern Synthesis. New York City: Harper and Brothers; 1942.

[10] Gould SJ. Ontogeny and Phylogeny. Belknap Press of Harvard University Press; 1977.

[11] Maynard Smith J, Burian R, Kauffman S, Alberch P, Campbell J, Goodwin B, et al. Developmental constraints and evolution: a perspective from the Mountain Lake conference on development and evolution. The Quarterly Review of Biology. 1985;60(3):265–287. doi:10.1086/414425.

[12] Arthur W. The emerging conceptual framework of evolutionary developmental biology. Nature. 2002;415(6873):757–764. doi:10.1038/415757a.

[13] Gilbert SF. The morphogenesis of evolutionary developmental biology. International Journal of Developmental Biology. 2003;47(7-8):467–477. doi:10.1387/ijdb.14756322.

[14] Lewontin RC. The principle of historicity in evolution. The Wistar Institute symposium monograph. 1967;5:81–94.

[15] Gould SJ, Lewontin RC. The spandrels of San Marco and the Panglossian paradigm: a critique of the adaptationist programme. Proceedings of the Royal Society of London B: Biological Sciences. 1979;205(1161):581–598. doi:10.1098/rspb.1979.0086.

[16] Gould SJ. Wonderful Life: The Burgess Shale and the Nature of History. W. W. Norton; 1990.

[17] Lenski RE, Rose MR, Simpson SC, Tadler SC. Long-term experimental evolution in *Escherichia coli* I. Adaptation and divergence during 2,000 generations. The American Naturalist. 1991;138(6):1315–1341. doi:10.1086/285289.

[18] Gould SJ. The paradox of the first tier: an agenda for paleobiology. Paleobiology. 1985;11(1):2–12. doi:10.1017/S0094837300011350.

[19] Jablonski D. Scale and hierarchy in macroevolution. Palaeontology. 2007;50(1):87–109. doi:10.1111/j.1475-4983.2006.00615.x.

[20] Myers CE, Saupe EE. A macroevolutionary expansion of the modern synthesis and the importance of extrinsic abiotic factors. Palaeontology. 2013;56(6):1179–1198. doi:10.1111/pala.12053.

[21] Mayr E. Systematics and the Origin of Species from, the Viewpoint of a Zoologist. New York: Columbia University Press; 1942.

[22] Wright S. Isolation by distance. Genetics. 1943;28(2):114–38.

[23] Levene H. Genetic equilibrium when more than one ecological niche is available. The American Naturalist. 1953;87(836):331–333. doi:10.1086/281792.

[24] Gavrilets S. Fitness Landscapes and the Origin of Species. Princeton University Press; 2004.

[25] Myers CW, Daly JW, Malkin B. A dangerously toxic new frog (*Phyllobates*) used by Emberá Indians of western Colombia, with discussion of blowgun fabrication and dart poisoning. Bulletin of the American Museum of Natural History. 1978;161:307–366.

[26] Santos JC, Coloma LA, Summers K, Caldwell JP, Ree R, Cannatella DC. Amazonian amphibian diversity is primarily derived from late miocene andean lineages. PLoS Biology. 2009;7(3):e56. doi:10.1371/journal.pbio.1000056.st014.

[27] Grant T, Rada M, Anganoy-Criollo M, Batista A, Dias PH, Jeckel AM, et al. Phylogenetic systematics of Dart-Poison frogs and their relatives revisited (Anura: Dendrobatoidea). South American Journal of Herpetology. 2017;12(s1):S1–S90. doi:10.2994/SAJH-D-17-00017.1.

[28] Remsen JV. High incidence of "leapfrog" pattern of geographic variation in andean birds: implications for the speciation process. Science. 1984;224(4645):171–173. doi:10.1126/science.224.4645.171.

[29] Huxley J. Clines: An auxiliary taxonomic principle. Nature. 1938;142(3587):219–220. doi:10.1038/142219a0.

[30] Wallace AR. The colors of animals and plants. The American Naturalist. 1877;11(11):641–662. doi:10.1086/271979.

[31] Poulton EB. The Colours of Animals. New York,: D. Appleton and company; 1890.

[32] Cuthill IC, Allen WL, Arbuckle K, Caspers B, Chaplin G, Hauber ME, et al. The biology of color. Science. 2017;357(6350):eaan0221. doi:10.1126/science.aan0221.

[33] Endler JA, Mappes J. The current and future state of animal coloration research. Philosophical Transactions of the Royal Society B: Biological Sciences. 2017;372(1724). doi:10.1098/rstb.2016.0352.

[34] Bennett DC, Lamoreux ML. The color loci of mice - A genetic century. Pigment Cell Research. 2003;16(4):333–344. doi:10.1034/j.1600-0749.2003.00067.x.

[35] Sturm RA. Molecular genetics of human pigmentation diversity. Human Molecular Genetics. 2009;18(R1):9–17. doi:10.1093/hmg/ddp003.

[36] Irion U, Singh AP, Nüsslein-Volhard C. The developmental genetics of vertebrate color pattern formation: Lessons from Zebrafish. Current Topics in Developmental Biology. 2016;117:141–169. doi:10.1016/bs.ctdb.2015.12.012.

[37] Patterson LB, Parichy DM. Zebrafish pigment pattern formation: Insights into the development and evolution of adult form. Annual Review of Genetics. 2019;53(1):505–530. doi:10.1146/annurev-genet-112618-043741.

[38] Hofreiter M, Schöneberg T. The genetic and evolutionary basis of colour variation in vertebrates. Cellular and Molecular Life Sciences. 2010;67(15):2591–2603. doi:10.1007/s00018-010-0333-7.

[39] Hubbard JK, Uy JAC, Hauber ME, Hoekstra HE, Safran RJ. Vertebrate pigmentation: From underlying genes to adaptive function. Trends in Genetics. 2010;26(5):231–239. doi:10.1016/j.tig.2010.02.002.

[40] Kronforst MR, Barsh GS, Kopp A, Mallet J, Monteiro A, Mullen SP, et al. Unraveling the thread of nature's tapestry: The genetics of diversity and convergence in animal pigmentation. Pigment Cell & Melanoma Research. 2012;25(4):411–433. doi:10.1111/j.1755-148X.2012.01014.x.

[41] Thibaudeau G, Altig R. Coloration of Anuran tadpoles (Amphibia): Development, dynamics, function, and hypotheses. ISRN Zoology. 2012;2012:1–16. doi:10.5402/2012/725203.

[42] Fujii R. Cytophysiology of fish chromatophores. In: International Review of Cytology. vol. 143; 1993. p. 191–255.

[43] Bagnara JT, Matsumoto J. Comparative Anatomy and Physiology of Pigment Cells in Nonmammalian Tissues. In: Nordlund JJ, Boissy RE, Hearing VJ, King RA, Oetting WS, Ortonne J, editors. The Pigmentary System. Oxford: Blackwell Publishing Ltd; 2006. p. 11–59.

[44] Ligon RA, Mccartney KL. Biochemical regulation of pigment motility in vertebrate chromatophores: A review of physiological color change mechanisms. Current Zoology. 2016;62(3):237–252. doi:10.1093/cz/zow051.

[45] Howard PG. Animal Colour Changes and Their Neurohumours: A Survey of Investigations 1910-1943. Cambridge: Cambridge University Press; 1948.

[46] Bagnara JT. Cytology and cytophysiology of non-melanophore pigment cells. International Review of Cytology. 1966;20(C). doi:10.1016/S0074-7696(08)60801-3.

[47] Bagnara JT, Taylor JD, Hadley ME. The dermal chromatophore unit. The Journal of cell biology. 1968;38(1):67–79. doi:10.1083/jcb.38.1.67.

[48] Weiner L, Han R, Scicchitano BM, Li J, Hasegawa K, Grossi M, et al. Dedicated epithelial recipient cells determine pigmentation patterns. Cell. 2007;130(5):932–942. doi:10.1016/j.cell.2007.07.024.

[49] Mills MG, Patterson LB. Not just black and white: Pigment pattern development and evolution in vertebrates. Seminars in Cell & Developmental Biology. 2009;20(1):72–81. doi:10.1016/j.semcdb.2008.11.012.

[50] D'Ischia M, Wakamatsu K, Cicoira F, Di Mauro E, Garcia-Borron JC, Commo S, et al. Melanins and melanogenesis: From pigment cells to human health and technological applications. Pigment Cell and Melanoma Research. 2015;28(5):520–544. doi:10.1111/pcmr.12393.

[51] Goodwin TW. The Biochemistry of the Carotenoids. vol. II: Animals. Dordrecht: Springer Netherlands; 1984.

[52] Hoekstra HE. Genetics, development and evolution of adaptive pigmentation in vertebrates. Heredity. 2006;97(3):222–234. doi:10.1038/sj.hdy.6800861.

[53] Theron E, Hawkins K, Bermingham E, Ricklefs RE, Mundy NI. 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. 2001;11(8):550–557. doi:10.1016/S0960-9822(01)00158-0.

[54] Uy JAC, Cooper EA, Cutie S, Concannon MR, Poelstra JW, Moyle RG, et al. Mutations in different pigmentation genes are associated with parallel melanism in island flycatchers. Proceedings of the Royal Society B: Biological Sciences. 2016;283(1834):1–9. doi:10.1098/rspb.2016.0731.

[55] Rosenblum EB, Hoekstra HE, Nachman MW. Adaptive reptile color variation and the evolution of the *Mc1R* gene. Evolution. 2004;58(8):1794. doi:10.1554/03-741.

[56] Rosenblum EB, Römpler H, Schöneberg T, Hoekstra HE. Molecular and functional basis of phenotypic convergence in white lizards at White Sands. Proceedings of the National Academy of Sciences of the United States of America. 2010;107(5):2113–2117. doi:10.1073/pnas.0911042107.

[57] Posso-Terranova A, Andrés J. Diversification and convergence of aposematic phenotypes: Truncated receptors and cellular arrangements mediate rapid evolution of coloration in harlequin poison frogs. Evolution. 2017;71(11):2677–2692. doi:10.1111/evo.13335.

[58] Ito S, Wakamatsu K. Chemistry of mixed melanogenesis - Pivotal roles of dopaquinone. Photochemistry and Photobiology. 2008;84(3):582–592. doi:10.1111/j.1751-1097.2007.00238.x.

[59] Miura I, Tagami M, Fujitani T, Ogata M. Spontaneous tyrosinase mutations identified in albinos of three wild frog species. Genes and Genetic Systems. 2017;92(4):189–196. doi:10.1266/ggs.16-00061.

[60] Anello M, Fernández E, Daverio MS, Vidal-Rioja L, Di Rocco F. *TYR* gene in llamas: Polymorphisms and expression study in different color phenotypes. Frontiers in Genetics. 2019;10(JUN):1–9. doi:10.3389/fgene.2019.00568.

[61] Lamason RL. *SLC24A5*, a putative cation exchanger, affects pigmentation in Zebrafish and Humans. Science. 2005;310(5755):1782–1786. doi:10.1126/science.1116238.

[62] Martin AR, Lin M, Granka JM, Myrick JW, Liu X, Sockell A, et al. An unexpectedly complex architecture for skin pigmentation in africans. Cell. 2017;171(6):1340–1353.e14. doi:10.1016/j.cell.2017.11.015.

[63] Bellono NW, Oancea EV. Ion transport in pigmentation. Archives of Biochemistry and Biophysics. 2014;563:35–41. doi:10.1016/j.abb.2014.06.020.

[64] Bagnara JT, Taylor JD, Prota G. Color changes, unusual melanosomes, and a new pigment from leaf frogs. Science. 1973;182(4116):1034–1035. doi:10.1126/science.182.4116.1034.

[65] Bagnara JT. Enigmas of pterorhodin, a red melanosomal pigment of tree frogs. Pigment Cell Research. 2003;16(5):510–516.

[66] Obika M, Bagnara JT. Pteridines as pigments in amphibians. Science. 1964;143(3605):485–487. doi:10.1126/science.143.3605.485.

[67] Stackhouse HL. Some aspects of pteridine biosynthesis in amphibians. Comparative Biochemistry And Physiology. 1966;17(1).

[68] McLean CA, Lutz A, Rankin KJ, Stuart-Fox D, Moussalli A. Revealing the biochemical and genetic basis of color variation in a polymorphic lizard. Molecular Biology and Evolution. 2017;34(8):1924–1935. doi:10.1093/molbev/msx136.

[69] Andrade P, Pinho C, de Lanuza GPi, Afonso S, Brejcha J, Rubin CJ, et al. Regulatory changes in pterin and carotenoid genes underlie balanced color polymorphisms in the wall lizard. Proceedings of the National Academy of Sciences of the United States of America. 2019;116(12):5633–5642. doi:10.1073/pnas.1820320116.

[70] Moran NA, Jarvik T. Lateral transfer of genes from fungi underlies carotenoid production in aphids. Science. 2010;328(5978):624–627. doi:10.1126/science.1187113.

[71] Toews DPL, Hofmeister NR, Taylor SA. The evolution and genetics of carotenoid processing in animals. Trends in Genetics. 2017;33(3):171–182. doi:10.1016/j.tig.2017.01.002.

[72] Eriksson J, Larson G, Gunnarsson U, Bed'hom B, Tixier-Boichard M, Strömstedt L, et al. Identification of the Yellow skin gene reveals a hybrid origin of the domestic chicken. PLoS Genetics. 2008;4(2). doi:10.1371/journal.pgen.1000010.

[73] Mundy NI, Stapley J, Bennison C, Tucker R, Twyman H, Kim KW, et al. Red carotenoid coloration in the Zebra Finch is controlled by a cytochrome P450 gene cluster. Current Biology. 2016;26(11):1435–1440. doi:10.1016/j.cub.2016.04.047.

[74] Toews DPL, Taylor SA, Vallender R, Brelsford A, Butcher BG, Messer PW, et al. Plumage genes and little else distinguish the genomes of hybridizing warblers. Current Biology. 2016;26(17):2313–2318. doi:10.1016/j.cub.2016.06.034.

[75] Hooper DM, Griffith SC, Price TD. Sex chromosome inversions enforce reproductive isolation across an avian hybrid zone. Molecular Ecology. 2019;28(6):1246–1262. doi:10.1111/mec.14874.

[76] O'Quin CT, Drilea AC, Conte MA, Kocher TD. Mapping of pigmentation QTL on an anchored genome assembly of the cichlid fish, *Metriaclima zebra*. BMC Genomics. 2013;14(1). doi:10.1186/1471-2164-14-287.

[77] Riddle MR, Aspiras A, Damen F, Hutchinson JN, Chinnapen D, Tabin CJ. Genetic architecture underlying changes in carotenoid accumulation during the evolution of the Blind Mexican cavefish, *Astyanax mexicanus*. bioRxiv. 2019; p. 788844. doi:10.1101/788844.

[78] Fukuzawa T, Ide H. A ventrally localized inhibitor of melanization in *Xenopus laevis* skin. Developmental Biology. 1988;129(1):25–36. doi:10.1016/0012-1606(88)90158-3.

[79] Fukuzawa T, Samaraweera P, Mangano FT, Law JH, Bagnara JT. Evidence that MIF plays a role in the development of pigmentation patterns in the frog. 1995;167(1):148–158. doi:10.1006/dbio.1995.1013.

[80] Millar SE, Miller MW, Stevens ME, Barsh GS. Expression and transgenic studies of the mouse agouti gene provide insight into the mechanisms by which mammalian coat color patterns are generated. Development. 1995;121(10):3223–3232.

[81] Zuasti A. Melanization stimulating factor (MSF) and melanization inhibiting factor (MIF) in the integument of fish. Microscopy Research and Technique. 2002;58(6):488–495. doi:10.1002/jemt.10167.

[82] Parichy DM. Pigment patterns of larval salamanders (Ambystomatidae, Salamandridae): The role of the lateral line sensory system and the evolution of pattern-forming mechanisms. Developmental Biology. 1996;175(2):265–282. doi:10.1006/dbio.1996.0114.

[83] Parichy DM. When neural crest and placodes collide: Interactions between melanophores and the lateral lines that generate stripes in the salamander *Ambystoma tigrinum tigrinum* (Ambystomatidae). Developmental Biology. 1996;175(2):283–300. doi:10.1006/dbio.1996.0115.

[84] Watanabe M, Iwashita M, Ishii M, Kurachi Y, Kawakami A, Kondo S, et al. Spot pattern of leopard *Danio* is caused by mutation in the zebrafish *connexin41.8* gene. EMBO Reports. 2006;7(9):893–897. doi:10.1038/sj.embor.7400757.

[85] Eom DS, Inoue S, Patterson LB, Gordon TN, Slingwine R, Kondo S, et al. Melanophore migration and survival during zebrafish adult pigment stripe development require the immunoglobulin superfamily adhesion molecule *Igsf11*. PLoS Genetics. 2012;8(8). doi:10.1371/journal.pgen.1002899.

[86] Eom DS, Bain EJ, Patterson LB, Grout ME, Parichy DM. Long-distance communication by specialized cellular projections during pigment pattern development and evolution. eLife. 2015;4(e12401):1–25. doi:10.7554/eLife.12401.

[87] Maderspacher F, Nüsslein-Volhard C. Formation of the adult pigment pattern in zebrafish requires *leopard* and *obelix* dependent cell interactions. Development. 2003;130(15):3447–3457. doi:10.1242/dev.00519.

[88] Frohnhöfer HG, Krauss J, Maischein HM, Nüsslein-Volhard C. Iridophores and their interactions with other chromatophores are required for stripe formation in zebrafish. Development. 2013;140(14):2997–3007. doi:10.1242/dev.096719.

[89] Zimova M, Hackländer K, Good JM, Melo-Ferreira J, Alves PC, Mills LS. Function and underlying mechanisms of seasonal colour moulting in mammals and birds: what keeps them changing in a warming world? Biological Reviews. 2018;93(3):1478–1498. doi:10.1111/brv.12405.

[90] Nilsson Sköld H, Aspengren S, Wallin M. Rapid color change in fish and amphibians - Function, regulation, and emerging applications. Pigment Cell & Melanoma Research. 2012;26(1):29–38. doi:10.1111/pcmr.12040.

[91] Nascimento AA, Roland JT, Gelfand VI. Pigment Cells: A model for the study of organelle transport. Annual Review of Cell and Developmental Biology. 2003;19(1):469–491. doi:10.1146/annurev.cellbio.19.111401.092937.

[92] Aspengren S, Hedberg D, Sköld HN, Wallin M. New insights into melanosome transport in vertebrate pigment cells. International Review of Cell and Molecular Biology. 2008;272(C):245–302. doi:10.1016/S1937-6448(08)01606-7.

[93] Tuma MC, Gelfand VI. Molecular mechanisms of pigment transport in melanophores. Pigment Cell Research. 1999;12(5):283–294. doi:10.1111/j.1600-0749.1999.tb00762.x.

[94] Teyssier J, Saenko SV, Van Der Marel D, Milinkovitch MC. Photonic crystals cause active colour change in chameleons. Nature Communications. 2015;6:1–7. doi:10.1038/ncomms7368.

[95] Cochrane CCS. Journal of a residence and travels in Colombia during the years 1823-1824. In two volumes. London: Henry Colburn; 1825.

[96] Posada Arango A. Mémoire sur le poison de rainette des sauvages du Chocó présentè à la Sociéte Allemande de Paris. Société Médicale Allemande de Paris. 1869;1869:203–213.

[97] Saffray C. Voyage a la Nouvelle-Grenade X: Du Chocó a Panamá. Le Tour du Monde, Nouveau Journal des Voyages. 1873;26(97-112).

[98] Myers CW, Daly JW. Dart-poison frogs. Scientific American. 1983;248(2):120–133. doi:10.1038/scientificamerican0283-120.

[99] Silverstone PA. A revision of the poison-arrow frogs of the genus *Phyllobates* Bibron in Sagra (Family Dendrobatidae). Natural History Museum of Los Angeles County, Science Bulletin. 1976;27:1–53.

[100] Maxson LR, Myers CW. Albumin evolution in tropical poison frogs (Dendrobatidae): A preliminary report. Biotropica. 1985;17(1):50–56. doi:10.2307/2388378.

[101] Chapman FM. Mutation among birds on the genus *Buarremon*. Bulletin of the American Museum of Natural History. 1923;48:243–278.

[102] Norman JA, Christidis L, Joseph L, Slikas B, Alpers D. Unravelling a biogeographical knot: origin of the 'leapfrog' distribution pattern of Australo-Papuan sooty owls (Strigiformes) and logrunners (Passeriformes). Proceedings of the Royal Society B: Biological Sciences. 2002;269(1505):2127–2133. doi:10.1098/rspb.2002.2136.

[103] Cadena CD, Cheviron ZA, Funk WC. Testing the molecular and evolutionary causes of a 'leapfrog' pattern of geographical variation in coloration. Journal of Evolutionary Biology. 2010;24(2):402–414. doi:10.1111/j.1420-9101.2010.02175.x.

[104] Matsumura Si, Yokoyama J, Tateishi Y, Maki M. Intraspecific variation of flower colour and its distribution within a sea lavender, *Limonium wrightii* (Plumbaginaceae), in the northwestern Pacific Islands. Journal of Plant Research. 2006;119(6):625–632. doi:10.1007/s10265-006-0022-7.

[105] Matsumura Si, Yokohama J, Fukuda T, Maki M. Origin of the disjunct distribution of flower colour polymorphism within *Limonium wrightii* (Plumbaginaceae) in the Ryukyu Archipelago. Biological Journal of the Linnean Society. 2009;97(4):709–717. doi:10.1111/j.1095-8312.2009.01253.x.

[106] Emsley MG. The geographical distribution of the color-pattern components of *Heliconius erato* and *Heliconius melpomene* with genetical evidence for the systematic relationship between the two species. Zoologica. 1965;49(15):245–286.

[107] Hovanitz W. Ecological color variation in a butterfly and the problem of "protective coloration". Ecology. 1940;21(3):371–380. doi:10.2307/1930846.

[108] Sheppard PM, Turner JRG, Brown KS, Benson WW, Singer MC. Genetics and the evolution of Muellerian mimicry in *Heliconius* butterflies. Philosophical Transactions of the Royal Society B: Biological Sciences. 1985;308(1137):433–610. doi:10.1098/rstb.1985.0066.

[109] Brower AVZ. Parallel race formation and the evolution of mimicry in *Heliconius* butterflies : A phylogenetic hypothesis from mitochondrial DNA sequences. Evolution. 1996;50(1):195–221. doi:10.2307/2410794.

[110] Quek SP, Counterman BA, Albuquerque de Moura P, Cardoso MZ, Marshall CR, McMillan WO, et al. Dissecting comimetic radiations in *Heliconius* reveals divergent histories of convergent butterflies. Proceedings of the National Academy of Sciences of the United States of America. 2010;107(16):7365–7370. doi:10.1073/pnas.0911572107.

[111] Hines HM, Counterman BA, Papa R, Albuquerque de Moura P, Cardoso MZ, Linares M, et al. Wing patterning gene redefines the mimetic history of *Heliconius* butterflies. Proceedings of the National Academy of Sciences of the United States of America. 2011;doi:10.1073/pnas.1110096108.

[112] James ME, Arenas-Castro H, Groh JS, Engelstaedter J, Ortiz-Barrientos D. Highly replicated evolution of parapatric ecotypes. bioRxiv. 2020; p. 2020.02.05.936401. doi:10.1101/2020.02.05.936401.

[113] Widmer A, Lötters S, Jungfer KH. A molecular phylogenetic analysis of the neotropical dart-poison frog genus *Phyllobates* (Amphibia: Dendrobatidae). Die Naturwissenschaften. 2000;87(12):559–562. doi:10.1007/s001140050.

[114] Grant T, Frost DR, Caldwell JP, Gagliardo R, Haddad CFB, Kok PJR, et al. Phylogenetic systematics of dart-poison frogs and their relatives (Amphibia: Athesphatanura: Dendrobatidae). Bulletin of the American Museum of Natural History. 2006;299:1–262. doi:10.1206/0003-0090(2006)299[1:PSODFA]2.0.CO;2.

[115] Márquez R, Corredor G, Galvis C, Góez D, Amézquita A. Range extension of the critically endangered true poison-dart frog, *Phyllobates terribilis* (Anura: Dendrobatidae), in western Colombia. Acta Herpetologica. 2012;7(2):341–345. doi:10.13128/Acta_Herpetol-11387.

[116] Miller SA, Dykes DD, Polesky HF. A simple salting out procedure for extracting DNA from human nucleated cells. Nucleic Acids Research. 1988;16(3):1215. doi:10.1093/nar/16.3.1215.

[117] Palumbi SR, Martin A, Romano S, McMillan WO, Stice L, Grabowski G. The simple fool's guide to PCR. Honolulu: Department of Zoology, University of Hawaii; 1991.

[118] Che J, Chen HM, Yang JX, Jin JQ, Jiang K, Yuan ZY, et al. Universal COI primers for DNA barcoding amphibians. Molecular Ecology Resources. 2012;12(2):247–258. doi:10.1111/j.1755-0998.2011.03090.x.

[119] Santos JC, Cannatella DC. Phenotypic integration emerges from aposematism and scale in poison frogs. Proceedings of the National Academy of Sciences of the United States of America. 2011;108(15):6175–6180. doi:10.1073/pnas.1010952108.

[120] Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, et al. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics. 2012;28(12):1647–1649. doi:10.1093/bioinformatics/bts199.

[121] Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research. 2004;32(5):1792–1797. doi:10.1093/nar/gkh340.

[122] Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. Systematic Biology. 2010;59(3):307–321. doi:10.1093/sysbio/syq010.

[123] Altekar G, Dwarkadas S, Huelsenbeck JP, Ronquist F. Parallel Metropolis coupled Markov chain Monte Carlo for Bayesian phylogenetic inference. Bioinformatics. 2004;20(3):407–415. doi:10.1093/bioinformatics/btg427.

[124] Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Hohna S, et al. MrBayes 3.2: Efficient Bayesian phylogenetic inference and model choice across a large model space. Systematic Biology. 2012;61(3):539–542. doi:10.1093/sysbio/sys029.

[125] Anisimova M, Gil M, Dufayard JF, Dessimoz C, Gascuel O. Survey of branch support methods demonstrates accuracy, power, and robustness of fast likelihood-based approximation schemes. Systematic Biology. 2011;60(5):685–699. doi:10.1093/sysbio/syr041.

[126] Bouckaert R, Vaughan TG, Barido-Sottani J, Duchêne S, Fourment M, Gavryushkina A, et al. BEAST 2.5: An advanced software platform for

Bayesian evolutionary analysis. PLoS computational biology. 2019;15(4):e1006650. doi:10.1371/journal.pcbi.1006650.

[127] Santos JC, Baquero M, Barrio-Amorós C, Coloma LA, Erdtmann LK, Lima AP, et al. Aposematism increases acoustic diversification and speciation in poison frogs. Proceedings of the Royal Society B: Biological Sciences. 2014;281(1796):20141761. doi:10.1098/rspb.2014.1761.

[128] Rambaut A, Drummond AJ. Tracer v. 1.5. Available from: http://beastbioedacuk/Tracer. 2009;.

[129] Lanfear R, Frandsen PB, Wright AM, Senfeld T, Calcott B. Partitionfinder 2: New methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. Molecular Biology and Evolution. 2017;34(3):772–773. doi:10.1093/molbev/msw260.

[130] Bi K, Vanderpool D, Singhal S, Linderoth T, Moritz C, Good JM. Transcriptome-based exon capture enables highly cost-effective comparative genomic data collection at moderate evolutionary scales. BMC genomics. 2012;13(1):403. doi:10.1186/1471-2164-13-403.

[131] Hodges E, Xuan Z, Balija V, Kramer M, Molla MN, Smith SW, et al. Genome-wide in situ exon capture for selective resequencing. Nature Genetics. 2007;39(12):1522–1527. doi:10.1038/ng.2007.42.

[132] Bolger AM, Lohse M, Usadel B. Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30(15):2114–2120. doi:10.1093/bioinformatics/btu170.

[133] Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson Da, Amit I, et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nature biotechnology. 2011;29(7):644–652. doi:10.1038/nbt.1883.

[134] Fu L, Niu B, Zhu Z, Wu S, Li W. CD-HIT: Accelerated for clustering the next-generation sequencing data. Bioinformatics. 2012;28(23):3150–3152. doi:10.1093/bioinformatics/bts565.

[135] Altschul SF, Madden TL, Schäffer AA. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Research. 1997;25(17):3389–3402. doi:10.1093/nar/25.17.3389.

[136] Slater GSC, Birney E. Automated generation of heuristics for biological sequence comparison. BMC Bioinformatics. 2005;6:31. doi:10.1186/1471-2105-6-31.

[137] Smith AFA, Hubley R, Green P. RepeatMasker Open-4.0; 2013. Available from: http://www.repeatmasker.org.

[138] Meyer M, Kircher M. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. Cold Spring Harbor Protocols. 2010;2010(6):pdb.prot5448–pdb.prot5448. doi:10.1101/pdb.prot5448.

[139] Hodges E, Rooks M, Xuan Z, Bhattacharjee A, Benjamin Gordon D, Brizuela L, et al. Hybrid selection of discrete genomic intervals on custom-designed microarrays for massively parallel sequencing. Nature Protocols. 2009;4(6):960–974. doi:10.1038/nprot.2009.68.

[140] Petersen KR, Streett DA, Gerritsen AT, Hunter SS, Settles ML. Super deduper, fast PCR duplicate detection in fastq files. In: Proceedings of the 6th ACM Conference on Bioinformatics, Computational Biology and Health Informatics - BCB '15. New York, New York, USA: ACM Press; 2015. p. 491–492.

[141] Jiang H, Lei R, Ding SW, Zhu S. Skewer: A fast and accurate adapter trimmer for next-generation sequencing paired-end reads. BMC Bioinformatics. 2014;15(1):1–12. doi:10.1186/1471-2105-15-182.

[142] Magoč T, Salzberg SL. FLASH: Fast length adjustment of short reads to improve genome assemblies. Bioinformatics. 2011;27(21):2957–2963. doi:10.1093/bioinformatics/btr507.

[143] Simpson JT, Wong K, Jackman SD, Schein JE, Jones SJM, Birol I. ABySS: A parallel assembler for short read sequence data. Genome Research. 2009;19(6):1117–1123. doi:10.1101/gr.089532.108.

[144] Huang XQ, Madan A. CAP3: A DNA sequence assembly program. Genome Research. 1999;9(9):868–877. doi:10.1101/gr.9.9.868.

[145] Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nature Methods. 2012;9(4):357–359. doi:10.1038/nmeth.1923.

[146] Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009;25(16):2078–2079. doi:10.1093/bioinformatics/btp352.

[147] McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Research. 2010;20(9):1297–1303. doi:10.1101/gr.107524.110.

[148] Meisner J, Albrechtsen A. Inferring population structure and admixture proportions in low-depth NGS data. Genetics. 2018;210(2):719–731. doi:10.1534/genetics.118.301336.

[149] Skotte L, Korneliussen TS, Albrechtsen A. Estimating individual admixture proportions from next generation sequencing data. Genetics. 2013;195(3):693–702. doi:10.1534/genetics.113.154138.

[150] Lefort V, Desper R, Gascuel O. FastME 2.0: A comprehensive, accurate, and fast distance-based phylogeny inference program. Molecular Biology and Evolution. 2015;32(10):2798–2800. doi:10.1093/molbev/msv150.

[151] Vieira FG, Lassalle F, Korneliussen TS, Fumagalli M. Improving the estimation of genetic distances from next-generation sequencing data. Biological Journal of Linnean Society. 2016;117(1):139–149. doi:10.1111/bij.12511.

[152] Korneliussen TS, Albrechtsen A, Nielsen R. ANGSD: Analysis of next generation sequencing data. BMC Bioinformatics. 2014;15(1):1–13. doi:10.1186/s12859-014-0356-4.

[153] Pickrell JK, Pritchard JK. Inference of population splits and mixtures from genome-wide allele frequency data. PLoS Genetics. 2012;8(11):e1002967. doi:10.1371/journal.pgen.1002967.s016.

[154] Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: A tool set for whole-genome association and population-based linkage analyses. The American Journal of Human Genetics. 2007;81(3):559–575. doi:10.1086/519795.

[155] Bryant D, Bouckaert R, Felsenstein J, Rosenberg NA, RoyChoudhury A. Inferring species trees directly from biallelic genetic markers: Bypassing gene trees in a full coalescent analysis. Molecular Biology and Evolution. 2012;29(8):1917–1932. doi:10.1093/molbev/mss086.

[156] Crawford AJ. Relative rates of nucleotide substitution in frogs. Journal of Molecular Evolution. 2003;57(6):636–641. doi:10.1007/s00239-003-2513-7.

[157] Sun YB, Xiong ZJ, Xiang XY, Liu SP, Zhou WW, Tu XL, et al. Whole-genome sequence of the Tibetan frog *Nanorana parkeri* and the comparative evolution of tetrapod genomes. Proceedings of the National Academy of Sciences of the United States of America. 2015;112(11):E1257–E1262. doi:10.1073/pnas.1501764112.

[158] Fitch WM. Toward defining the course of evolution: Minimum change for a specific tree topology. Systematic Zoology. 1971;20(4):406. doi:10.2307/2412116.

[159] Schliep KP. phangorn: phylogenetic analysis in R. Bioinformatics. 2011;27(4):592–593. doi:10.1093/bioinformatics/btq706.

[160] Pough FH, Taigen TL. Metabolic correlates of the foraging and social behaviour of dart-poison frogs. Animal Behaviour. 1990;39(1):145–155. doi:10.1016/S0003-3472(05)80734-1.

[161] Summers K, Clough ME. 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. 2001;98(11):6227–6232. doi:10.1073/pnas.101134898.

[162] Märki F, Witkop B. The venom of the Colombian arrow poison frog *Phyllobates* bicolor. Experientia. 1963;19(8):329–338. doi:10.1007/bf02152303.

[163] Daly JW, Myers CW, Whittaker N. Further classification of skin alkaloids from neotropical poison frogs (Dendrobatidae), with a general survey of toxic/noxious substances in the Amphibia. Toxicon. 1987;25(10):1023–1095. doi:10.1016/0041-0101(87)90265-0.

[164] Grafen A. The phylogenetic regression. Philosophical Transactions of the Royal Society B: Biological Sciences. 1989;326(1233):119–157. doi:10.1098/rstb.1989.0106.

[165] Martins EP, Hansen TF. Phylogenies and the comparative method: A general approach to incorporating phylogenetic information into the analysis of interspecific data. The American Naturalist. 1997;149(4):646–667. doi:10.1086/286013.

[166] Felsenstein J. Phylogenies and the comparative method. The American Naturalist. 1985;125(1):1–15. doi:10.1086/284325.

[167] Pagel M. Inferring the historical patterns of biological evolution. Nature. 1999;401(6756):877–884. doi:10.1038/44766.

[168] Paradis E, Claude J, Strimmer K. APE: Analyses of phylogenetics and evolution in R language. Bioinformatics. 2004;20(2):289–290. doi:10.1093/bioinformatics/btg412.

[169] Pinheiro J, Bates D, DebRoy S, Sarkar D. nlme: Linear and Nonlinear Mixed Effects models. R package version 31-131. 2017;.

[170] Bradburd GS, Ralph PL, Coop GM. A spatial framework for understanding population structure and admixture. PLoS Genetics. 2016;12(1):1–38. doi:10.1371/journal.pgen.1005703.

[171] Kass RE, Raftery AE. Bayes factors. Journal of the American Statistical Association. 1995;90(430):773–795. doi:10.1080/01621459.1995.10476572.

[172] Dickey JM, Lientz BP. The weighted likelihood ratio, sharp hypotheses about chances, the order of a Markov chain. The Annals of Mathematical Statistics. 1970;41(1):214–226. doi:10.1214/aoms/1177697203.

[173] Petkova D, Novembre J, Stephens M. Visualizing spatial population structure with estimated effective migration surfaces. Nature Genetics. 2015;48(1):94–100. doi:10.1038/ng.3464.

[174] Link V, Kousathanas A, Veeramah K, Sell C, Scheu A, Wegmann D. ATLAS: Analysis Tools for Low-depth and Ancient Samples. bioRxiv. 2017; p. 105346. doi:10.1101/105346.

[175] Wang IJ. Examining the full effects of landscape heterogeneity on spatial genetic variation: A multiple matrix regression approach for quantifying geographic and ecological isolation. Evolution. 2013;67(12):3403–3411. doi:10.1111/evo.12134.

[176] Weir BS, Cockerham CC. Estimating F-statistics for the analysis of population structure. Evolution. 1984;38(6):1358. doi:10.2307/2408641.

[177] Hijmans RJ. Introduction to the geosphere package (Version 1.5-7). 2017; p. 1–19.

[178] Dijkstra EW. A note on two problems in connexion with graphs. Numerische Mathematik. 1959;1(1):269–271. doi:10.1007/BF01386390.

[179] van Etten J. R Package gdistance: Distances and routes on geographical grids. Journal of Statistical Software. 2017;76(13):1–21. doi:10.18637/jss.v076.i13.

[180] Turner JR, Johnson MS, Eanes WF. Contrasted modes of evolution in the same genome: allozymes and adaptive change in *Heliconius*. Proceedings of the National Academy of Sciences of the United States of America. 1979;76(4):1924–1928. doi:10.1073/pnas.76.4.1924.

[181] Turner TL, Hahn MW, Nuzhdin SV. Genomic islands of speciation in *Anopheles gambiae*. PLoS Biology. 2005;3(9):e285. doi:10.1371/journal.pbio.0030285.st001.

[182] Wu CI. The genic view of the process of speciation. Journal of Evolutionary Biology. 2001;14(6):851–865. doi:10.1046/j.1420-9101.2001.00335.x.

[183] Mayr E. Animal Species and Evolution. Cambridge: Harvard University Press; 1963.

[184] Endler JA. Geographic Variation, Speciation, and Clines. Princeton: Princeton University Press; 1977.

[185] Rebelo AG, Siegfried WR. Colour and size of flowers in relation to pollination of *Erica* species. Oecologia. 1985;65(4):584–590. doi:10.1007/BF00379677.

[186] Richmond JQ, Reeder TW. Evidence for parallel ecological speciation in sincid lizards of the *Eumeces skiltonianus* species group (Squamata: Scincidae). Evolution. 2002;56(7):1498–1513. doi:10.1111/j.0014-3820.2002.tb01461.x.

[187] Rios E, Álvarez-Castañeda ST. Environmental responses to altitudinal gradients and subspecific validity in pocket gophers (*Thomomys bottae*) from Baja California Sur, Mexico. Journal of Mammalogy. 2007;88(4):926–934. doi:10.1644/06-mamm-a-226r1.1.

[188] Reguera S, Zamora-Camacho FJ, Moreno-Rueda G. The lizard *Psammodromus algirus* (Squamata: Lacertidae) is darker at high altitudes. Biological Journal of the Linnean Society. 2014;112(1):132–141. doi:10.1111/bij.12250.

[189] Köhler G, Samietz J, Schielzeth H. Morphological and colour morph clines along an altitudinal gradient in the meadow grasshopper *Pseudochorthippus parallelus*. PLoS ONE. 2017;12(12):1–13. doi:10.1371/journal.pone.0189815.

[190] Gamberale G, Tullberg BS. Evidence for a peak-shift in predator generalization among aposematic prey. Proceedings of the Royal Society B: Biological Sciences. 1996;263(1375):1329–1334. doi:10.1098/rspb.1996.0195.

[191] Forsman A, Merilaita S. Fearful symmetry: Pattern size and asymmetry affects aposematic signal efficacy. Evolutionary Ecology. 1999;13(2):131–140. doi:10.1023/A:1006630911975.

[192] Hagman M, Forsman A. Correlated evolution of conspicuous coloration and body size in poison frogs (Dendrobatidae). Evolution. 2003;57(12):2904–2910. doi:10.1111/j.0014-3820.2003.tb01531.x.

[193] Summers K, Symula R, Clough M, Cronin T. Visual mate choice in poison frogs. Proceedings of the Royal Society B: Biological Sciences. 1999;266(1434):2141–2145. doi:10.1098/rspb.1999.0900.

[194] Reynolds RG, Fitzpatrick BM. Assortative mating in poison-dart frogs based on an ecologically important trait. Evolution. 2007;61(9):2253–2259. doi:10.1111/j.1558-5646.2007.00174.x.

[195] Twomey E, Vestergaard JS, Summers K. Reproductive isolation related to mimetic divergence in the poison frog *Ranitomeya imitator*. Nature Communications. 2014;5:4749. doi:10.1038/ncomms5749.

[196] Yang Y, Richards-Zawacki CL, Devar A, Dugas MB. Poison frog color morphs express assortative mate preferences in allopatry but not sympatry. Evolution. 2016;70(12):2778–2788. doi:10.1111/evo.13079.

[197] Olson EC, Miller RL. Morphological integration. Chicago: University of Chicago Press; 1958.

[198] Bacon CD, Silvestro D, Jaramillo C, Smith BT, Chakrabarty P, Antonelli A. Biological evidence supports an early and complex emergence of the Isthmus of Panama. Proceedings of the National Academy of Sciences of the United States of America. 2015;112(19):6110–6115. doi:10.1073/pnas.1423853112.

[199] Duque-Caro JH. The Chocó block in the northwestern corner of South America: Structural, tectonostratigraphic, and paleogeographic implications. Journal of South American Earth Sciences. 1990;3(1):71–84. doi:10.1016/0895-9811(90)90019-W.

[200] Duque-Caro JH. Neogene stratigraphy, paleoceanography and paleobiogeography in northwest South America and the evolution of the Panama seaway. Palaeogeography, Palaeoclimatology, Palaeoecology. 1990;77(3-4):203–234. doi:10.1016/0031-0182(90)90178-A.

[201] Borrero C, Pardo A, Jaramillo CM, Osorio JA, Cardona A, Flores A, et al. Tectonostratigraphy of the Cenozoic Tumaco forearc basin (Colombian Pacific) and its relationship with the northern Andes orogenic build up. Journal of South American Earth Sciences. 2012;39:75–92. doi:10.1016/j.jsames.2012.04.004.

[202] Haffer J. Speciation in Amazonian Forest Birds. Science. 1969;165(3889):131–137. doi:10.1126/science.165.3889.131.

[203] Vanzolini PE, Williams EE. South american anoles: the geographic differentiation and evolution of the *Anolis chrysolepis* species group (Sauria, Iguanidae). Arquivos de Zoologia. 1970;19(1-2):1–258. doi:10.11606/issn.2176-7793.v19i3-4p125-298.

[204] Hooghiemstra H, Van Der Hammen T. Neogene and Quaternary development of the neotropical rain forest: The forest refugia hypothesis, and a literature overview. Earth Science Reviews. 1998;44(3-4):147–183. doi:10.1016/S0012-8252(98)00027-0.

[205] Baker PA, Fritz SC, Battisti DS, Dick CW, Vargas OM, Asner GP, et al. Beyond Refugia: New insights on Quaternary climate variation and the evolution of biotic diversity in tropical South America. In: Rull V, Carnaval AC, editors. Neotropical Diversification: Patterns and Processes. Berlin: Springer; 2020.

[206] Haffer J. Speciation in Colombian forest birds west of the Andes. American Museum Novitates. 1967;2294:1–57.

[207] Gentry AH. Phytogeographic patterns as evidence for a Chocó refuge. In: Prance GT, editor. Biological Diversification in the Tropics. New York: Plenum Press; 1982. p. 112–136.

[208] Behling H, Hooghiemstra H, Negret AJ. Holocene history of the Chocó rain forest from Laguna Piusbi, southern Pacific lowlands of Colombia. Quaternary Research. 1998;50(3):300–308. doi:10.1006/qres.1998.1998.

[209] Berrío JC, Behling H, Hooghiemstra H. Tropical rain-forest history from the Colombian Pacific area: A 4200-year pollen record from Laguna Jotaordó. Holocene. 2000;10(6):749–756. doi:10.1191/09596830094999.

[210] Jaramillo C, Bayona G. Mangrove distribution during the Holocene in Tribugá Gulf, Colombia. Biotropica. 2000;32(1):14. doi:10.1646/0006-3606(2000)032[0014:mddthi]2.0.co;2.

[211] Ramírez LF, Urrego LE. Vegetación holocénica en el delta del río San Juan, Pacífico colombiano. In: Correa ID, Restrepo JD, editors. Geologia y Oceanografía del Dela del Río San Juan. Litoral Pacífico Colombiano.. 1st ed. Medellín: Editorial Universidad EAFIT; 2002. p. 151–169.

[212] González C, Urrego LE, Martínez JI. Late Quaternary vegetation and climate change in the Panama Basin: Palynological evidence from marine cores ODP 677B and

TR 163-38. Palaeogeography, Palaeoclimatology, Palaeoecology. 2006;234(1):62–80. doi:10.1016/j.palaeo.2005.10.019.

[213] Urrego LE, Molina LA, Urrego DH, Ramírez LF. Holocene space–time succession of the Middle Atrato wetlands, Chocó biogeographic region, Colombia. Palaeogeography, Palaeoclimatology, Palaeoecology. 2006;234(1):45–61. doi:10.1016/j.palaeo.2005.10.018.

[214] Myers CW, Daly JW. 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. 1976;157(3):173–262.

[215] Garcia-Moreno J, Fjeldså J. Re-evaluation of species limits in the genus *Atlapetes* based on mtDNA sequence data. Ibis. 1999;141(2):199–207. doi:10.1111/j.1474-919X.1999.tb07542.x.

[216] Toon A, Austin JJ, Dolman G, Pedler L, Joseph L. Evolution of arid zone birds in Australia: Leapfrog distribution patterns and mesic-arid connections in quail-thrush (*Cinclosoma*, Cinclosomatidae). Molecular Phylogenetics and Evolution. 2012;62(1):286–295. doi:10.1016/j.ympev.2011.09.026.

[217] Avise JC, Robinson TJ. Hemiplasy: A new term in the lexicon of phylogenetics. Systematic Biology. 2008;57(3):503–507. doi:10.1080/10635150802164587.

[218] Hahn MW, Nakhleh L. Irrational exuberance for resolved species trees. Evolution. 2015;70(1):7–17. doi:10.1111/evo.12832.

[219] Jeffreys H. Theory of Probability. Oxford: Clarendon Press; 1938.

[220] Marin JM, Robert CP. Importance sampling methods for Bayesian discrimination between embedded models. arXiv. 2009; p. 0910.2325.

[221] Oaks JR, Cobb KA, Minin VN, Leaché AD. Marginal likelihoods in phylogenetics: A review of methods and applications. Systematic Biology. 2019;68(5):681–697. doi:10.1093/sysbio/syz003.

[222] Stone CJ, Hansen MH, Kooperberg C, Truong YK. Polynomial Splines and their Tensor Products in Extended Linear Modeling. The Annals of Statistics. 1997;25(4):1371–1425.

[223] Huxley JS. Clines: An auxiliary method in taxonomy. Bijdragen tot de Dierkunde. 1939;27:491–520.

[224] Haldane JBS. The theory of a cline. Journal of Genetics. 1948;48(3):277–284. doi:10.1007/BF02986626.

[225] Szymura JM, Barton NH. Genetic analysis of a hybrid zone between the Fire-Bellied toads, *Bombina bombina* and *B. variegata*, near cracow in southern Poland. Evolution. 1986;40(6):1141–1159. doi:10.2307/2408943.

[226] Mullen LM, Hoekstra HE. Natural selection along an environmental gradient: A classic cline in mouse pigmentation. Evolution. 2008;62(7):1555–1570. doi:10.1111/j.1558-5646.2008.00425.x.

[227] Curran E, Stankowski S, Pardo-Diaz C, Salazar CA, Lineares M, Nadeau NJ. Parallel clines in a quantitative trait in butterfly co-mimics species despite different levels of genomic divergence and selection. bioRxiv. 2019;doi:10.1101/842708.

[228] Baldassarre DT, White TA, Karubian J, Webster MS. Genomic and morphological analysis of a semipermeable avian hybrid zone suggests asymetrical introgression of a sexual signal. Evolution. 2014;68(9):2644–2657. doi:10.1111/evo.12457.

[229] Cooper BS, Sedghifar A, Nash WT, Comeault AA, Matute DR. A maladaptive combination of traits contributes to the maintenance of a *Drosophila* hybrid zone. Current Biology. 2018;28(18):2940–2947.e6. doi:10.1016/j.cub.2018.07.005.

[230] Payseur BA, Krenz JG, Nachman MW. Differential patterns of introgression across the X chromosome in a hybrid zone between house mice. Evolution. 2004;58(9):2064–2078. doi:10.1111/j.0014-3820.2004.tb00490.x.

[231] Turner LM, Harr B. Genome-wide mapping in a house mouse hybrid zone reveals hybrid sterility loci and Dobzhansky-Muller interactions. eLife. 2014;3(2002):1–25. doi:10.7554/eLife.02504.

[232] Nadeau NJ, Whibley A, Jones RT, Davey JW, Dasmahapatra KK, Baxter SW, et al. Genomic islands of divergence in hybridizing *Heliconius* butterflies identified by large-scale targeted sequencing. Philosophical Transactions of the Royal Society B: Biological Sciences. 2012;367(1587):343–353. doi:10.1098/rstb.2011.0198.

[233] Crawford JE, Nielsen R. Detecting adaptive trait loci in nonmodel systems: Divergence or admixture mapping? Molecular Ecology. 2013;22(24):6131–6148. doi:10.1111/mec.12562.

[234] Vestergaard JS, Twomey E, Larsen R, Summers K, Nielsen R. Number of genes controlling a quantitative trait in a hybrid zone of the aposematic frog *Ranitomeya imitator*. Proceedings of the Royal Society B: Biological Sciences. 2015;282(1807):20141950. doi:10.1111/j.1365-294X.2012.05644.x.

[235] Durrett R, Buttel L, Harrison R. Spatial models for hybrid zones. Heredity. 2000;84(1):9–19. doi:10.1046/j.1365-2540.2000.00566.x.

[236] Barton NH, Hewitt GM. Analysis of hybrid zones. Annual Review of Ecology and Systematics. 1985;16:113–148. doi:10.1146/annurev.es.16.110185.000553.

[237] Cruzan MB, Templeton AR. Paleoecology and coalescence: Phylogeographic analysis of hypotheses from the fossil record. Trends in Ecology & Evolution. 2000;15(12):491–496. doi:10.1016/S0169-5347(00)01998-4.

[238] Morales-Rozo A, Tenorio EA, Carling MD, Cadena CD. Origin and cross-century dynamics of an avian hybrid zone. BMC Evolutionary Biology. 2017;17(1):257. doi:10.1186/s12862-017-1096-7.

[239] Excoffier L, Dupanloup I, Huerta-Sánchez E, Sousa VC, Foll M. Robust demographic inference from genomic and SNP data. PLoS Genetics. 2013;9(10). doi:10.1371/journal.pgen.1003905.

[240] Steinrücken M, Kamm JA, Song YS. Inference of complex population histories using whole-genome sequences from multiple populations. bioRxiv. 2015; p. 026591. doi:10.1101/026591.

[241] Kamm JA, Terhorst J, Durbin R, Song YS. Efficiently inferring the demographic history of many populations with allele count data. bioRxiv. 2018; p. 1–29. doi:10.1101/287268.

[242] Al-Asadi H, Petkova D, Stephens M, Novembre J. Estimating recent migration and population-size surfaces. PLoS Genetics. 2019;15(1):1–21. doi:10.1371/journal.pgen.1007908.

[243] Bradburd GS, Coop GM, Ralph PL. Inferring continuous and discrete population genetic structure across space. Genetics. 2018;210(1):33–52. doi:10.1534/genetics.118.301333.

[244] Bradburd GS, Ralph PL. Spatial population genetics: It's about time. Annual Review of Ecology, Evolution, and Systematics. 2019;50(1):427–449. doi:10.1146/annurev-ecolsys-110316-022659.

[245] Márquez R, Linderoth TP, Mejía-Vargas D, Nielsen R, Kronforst MR, Amézquita A. Divergence, gene flow and the origin of leapfrog geographic distributions: The history of color pattern variation in *Phyllobates* poison-dart frogs. bioRxiv. 2020; p. 2020.02.21.960005. doi:10.1101/2020.02.21.960005.

[246] Zimin AV, Marçais G, Puiu D, Roberts M, Salzberg SL, Yorke JA. The MaSuRCA genome assembler. Bioinformatics. 2013;29(21):2669–2677. doi:10.1093/bioinformatics/btt476.

[247] Zimin AV, Puiu D, Luo MC, Zhu T, Koren S, Marçais G, et al. Hybrid assembly of the large and highly repetitive genome of *Aegilops tauschii* hybrid, a progenitor of bread wheat, with the MaSuRCA mega-reads algorithm. Genome Research. 2017;27(5):787–792. doi:10.1101/gr.213405.116.

[248] Ghurye J, Pop M, Koren S, Bickhart D, Chin CS. Scaffolding of long read assemblies using long range contact information. BMC Genomics. 2017;18(1):1–11. doi:10.1186/s12864-017-3879-z.

[249] Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv. 2013; p. 1303.3997.

[250] Linderoth TP. Identifying Population Histories, Adaptive Genes, and Genetic Duplication from Population- Scale Next Generation Sequencing. University of California Berkeley; 2018. Available from: https://escholarship.org/uc/item/5kp4q40k.

[251] Van Belleghem SM, Papa R, Ortiz-Zuazaga H, Hendrickx F, Jiggins CD, Owen McMillan W, et al. patternize: An R package for quantifying colour pattern variation. Methods in Ecology and Evolution. 2018;9(2):390–398. doi:10.1111/2041-210X.12853.

[252] Abràmoff MD, Magalhães PJ. Image processing with ImageJ. Biophotonics International. 2004;11(7):36–42.

[253] Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. Genetics. 2000;155(2):945–959.

[254] Peter BM, Slatkin M. Detecting range expansions from genetic data. Evolution. 2013;67(11):3274–3289. doi:10.1111/evo.12202.

[255] Slatkin M, Excoffier L. Serial founder effects during range expansion: A spatial analog of genetic drift. Genetics. 2012;191(1):171–181. doi:10.1534/genetics.112.139022.

[256] Peter BM, Slatkin M. The effective founder effect in a spatially expanding population. Evolution. 2015;69(3):721–734. doi:10.1111/evo.12609.

[257] Szymura JM, Barton NH. The genetic structure of the hybrid zone between the firebellied toads *Bombina bombina* and *B. variegata*: comparisons between transects and between loci. Evolution. 1991;45(2):237–261. doi:10.1111/j.1558-5646.1991.tb04400.x.

[258] Elzhov TV, Mullen KM, Spiess AN, Bolker B. minpack.lm: R interface to the Levenberg-Marquardt nonlinear least-squares algorithm found in MINPACK, plus support for bounds. R package version 12-1. 2016;.

[259] Flaxman SM, Wacholder AC, Feder JL, Nosil P. Theoretical models of the influence of genomic architecture on the dynamics of speciation. Molecular Ecology. 2014;23(16):4074–4088. doi:10.1111/mec.12750.

[260] Gompert Z, Mandeville EG, Buerkle CA. Analysis of population genomic data from hybrid zones. Annual Review of Ecology, Evolution, and Systematics. 2017;48(August):207–29. doi:10.1146/annurev-ecolsys-110316.

[261] Mallet J. The genetics of warning colour in Peruvian hybrid zones of *Heliconius* erato and H. melpomene. Proceedings of the Royal Society B: Biological Sciences. 1989;236(1283):163–185. doi:10.1098/rspb.1989.0019.

[262] Nosil P. Reproductive isolation caused by visual predation on migrants between divergent environments. Proceedings of the Royal Society B: Biological Sciences. 2004;271(1547):1521–1528. doi:10.1098/rspb.2004.2751.

[263] Nokelainen O, Valkonen J, Lindstedt C, Mappes J. Changes in predator community structure shifts the efficacy of two warning signals in Arctiid moths. Journal of Animal Ecology. 2014;83(3):598–605. doi:10.1111/1365-2656.12169.

[264] Barrett RDH, Laurent S, Mallarino R, Pfeifer SP, Xu CCY, Foll M, et al. Linking a mutation to survival in wild mice. Science. 2019;363(6426):499–504. doi:10.1126/science.aav3824.

[265] Noonan BP, Comeault AA. The role of predator selection on polymorphic aposematic poison frogs. Biology Letters. 2008;5(1):51–54. doi:10.1098/rsbl.2008.0586.

[266] Brown JL, Maan ME, Cummings ME, Summers K. Evidence for selection on coloration in a Panamanian poison frog: a coalescent-based approach. Journal of Biogeography. 2010;37(5):891–901. doi:10.1111/j.1365-2699.2009.02260.x.

[267] Chouteau M, Angers B. The role of predators in maintaining the geographic organization of aposematic signals. The American Naturalist. 2011;178(6).

[268] Sumner FB. The analysis of a concrete case of intergradation between two subpsecies. Proceedings of the National Academy of Sciences. 1929;15(2):110–120. doi:10.1073/pnas.15.2.110.

[269] Saccheri IJ, Rousset F, Watts PC, Brakefield PM, Cook LM. Selection and gene flow on a diminishing cline of melanic peppered moths. Proceedings of the National Academy of Sciences of the United States of America. 2008;105(42):16212–7. doi:10.1073/pnas.0803785105.

[270] Amézquita A, Castro L, Arias M, González M, Esquivel C. Field but not lab paradigms support generalisation by predators of aposematic polymorphic prey: the *Oophaga histrionica* complex. Evolutionary Ecology. 2013;27(4):769–782. doi:10.1007/s10682-013-9635-1.

[271] Casas-Cardona S, Márquez R, Vargas-Salinas F. Different colour morphs of the poison frog *Andinobates bombetes* (Dendrobatidae) are similarly effective visual predator deterrents. Ethology. 2018;124(4):245–255. doi:10.1111/eth.12729.

[272] Rößler DC, Pröhl H, Lötters S. The future of clay model studies. BMC Zoology. 2018;3(1):1–5. doi:10.1186/s40850-018-0033-6.

[273] Protti-Sánchez F. Aposematismo en la rana *Phyllobates vittatus* (Anura: Dendrobatidae): Integrando defensas, señales y riesgo de depredación [Masters]. Universidad de Costa Rica; 2019.

[274] Wright S. Evolution in Mendelian populations. Genetics. 1931;16(2):97–159. doi:10.1016/S0092-8240(05)80011-4.

[275] Buggs RJA. Empirical study of hybrid zone movement. Heredity. 2007;99(3):301–312. doi:10.1038/sj.hdy.6800997.

[276] Hegna RH, Saporito RA, Donnelly MA. Not all colors are equal: predation and color polytypism in the aposematic poison frog *Oophaga pumilio*. Evolutionary Ecology. 2013;27(5):831–845. doi:10.1007/s10682-012-9605-z.

[277] Medina I, Wang IJ, Salazar C, Amézquita A. Hybridization promotes color polymorphism in the aposematic harlequin poison frog, *Oophaga histrionica*. Ecology and Evolution. 2013;3(13):4388–4400. doi:10.1002/ece3.794.

[278] Frantz AC, Cellina S, Krier A, Schley L, Burke T. Using spatial Bayesian methods to determine the genetic structure of a continuously distributed population: Clusters or isolation by distance? Journal of Applied Ecology. 2009;46(2):493–505. doi:10.1111/j.1365-2664.2008.01606.x.

[279] Meirmans PG. The trouble with isolation by distance. Molecular Ecology. 2012;21(12):2839–2846. doi:10.1111/j.1365-294X.2012.05578.x.

[280] Price GR. Selection and covariance. Nature. 1970;227(5257):520–521. doi:10.1038/227520a0.

[281] Colosimo PF, Hosemann KE, Balabhadra S, Villarreal G, Dickson M, Grimwood J, et al. Widespread parallel evolution in sticklebacks by repeated fixation of ectodysplasin alleles. Science. 2005;307(5717):1928–1933. doi:10.1126/science.1107239.

[282] Feldman CR, Brodie ED, Brodie ED, Pfrender ME. The evolutionary origins of beneficial alleles during the repeated adaptation of garter snakes to deadly prey. Proceedings of the National Academy of Sciences. 2009;106(32):13415–13420. doi:10.1073/pnas.0901224106.

[283] The Heliconius Genome Consortium, Dasmahapatra KK, Walters JR, Briscoe AD, Davey JW, Whibley A, et al. Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. Nature. 2012;487(7405):94–98. doi:10.1038/nature11041.

[284] Zhen Y, Aardema ML, Medina EM, Schumer M, Andolfatto P. Parallel molecular evolution in an herbivore community. Science. 2012;337(6102):1634–1637. doi:10.1126/science.1226630.

[285] Huerta-Sánchez E, Jin X, Asan, Bianba Z, Peter BM, Vinckenbosch N, et al. Altitude adaptation in Tibetans caused by introgression of Denisovan-like DNA. Nature. 2014;512(7513):194–197. doi:10.1038/nature13408.

[286] Pease JB, Haak DC, Hahn MW, Moyle LC. Phylogenomics reveals three sources of adaptive variation during a rapid radiation. PLoS Biology. 2016;14(2):e1002379. doi:10.1371/journal.pbio.1002379.

[287] Arnold ML. Natural hybridization and evolution. New York, NY: Oxford Univ. Press; 1997.

[288] Orr HA, Betancourt AJ. Haldane's sieve and adaptation from the standing genetic variation. Genetics. 2001;157(2):875–884.

[289] Barrett R, Schluter D. Adaptation from standing genetic variation. Trends in Ecology & Evolution. 2008;23(1):38–44. doi:10.1016/j.tree.2007.09.008.

[290] Abbott R, Albach D, Ansell S, Arntzen JW, Baird SJE, Bierne N, et al. Hybridization and speciation. Journal of Evolutionary Biology. 2013;26(2):229–246. doi:10.1111/j.1420-9101.2012.02599.x.

[291] Arnold ML, Martin NH. Hybrid fitness across time and habitats. Trends in Ecology & Evolution. 2010;25(9):530–536. doi:10.1016/j.tree.2010.06.005.

[292] Santos JC, Coloma LA, Cannatella DC. Multiple, recurring origins of aposematism and diet specialization in poison frogs. Proceedings of the National Academy of Sciences of the United States of America. 2003;100(22):12792–12797. doi:10.1073/pnas.2133521100.

[293] Vences M, Kosuch J, Boistel R, Haddad CFB, La Marca E, Lötters S, et al. Convergent evolution of aposematic coloration in Neotropical poison frogs: a molecular phylogenetic perspective. Organisms Diversity & Evolution. 2003;3(3):215–226.

[294] Rojas B. Behavioural, ecological, and evolutionary aspects of diversity in frog colour patterns. Biological Reviews. 2017;92(2):1059–1080. doi:10.1111/brv.12269.

[295] Saporito RA, Zuercher R, Roberts M, Gerow KG, Donnelly MA. Experimental evidence for aposematism in the dendrobatid poison frog *Oophaga pumilio*. Copeia. 2007; p. 1006–1011. doi:10.1643/0045-8511(2007)7[1006:EEFAIT]2.0.CO;2.

[296] Comeault AA, Noonan BP. Spatial variation in the fitness of divergent aposematic phenotypes of the poison frog, *Dendrobates tinctorius*. Journal of Evolutionary Biology. 2011;24(6):1374–1379. doi:10.1111/j.1420-9101.2011.02258.x.

[297] Rojas B, Devillechabrolle J, Endler JA. Paradox lost: variable colour-pattern geometry is associated with differences in movement in aposematic frogs. Biology Letters. 2014;10(6):20140193–20140193. doi:10.1098/rsbl.2014.0193.

[298] Maan ME, Cummings ME. Female preferences for aposematic signal components in a polymorphic poison frog. Evolution. 2008;62(9):2334–2345. doi:10.1111/j.1558-5646.2008.00454.x.

[299] Yang Y, Servedio MR, Richards-Zawacki CL. Imprinting sets the stage for speciation. Nature. 2019;574(7776):99–102. doi:10.1038/s41586-019-1599-z.

[300] Pröhl H. Territorial behavior in dendrobatid frogs. Journal of Herpetology. 2005;39(3):354–365. doi:10.1670/162-04A.1.

[301] Kim SY, Lohmueller KE, Albrechtsen A, Li Y, Korneliussen T, Tian G, et al. Estimation of allele frequency and association mapping using next-generation sequencing data. BMC Bioinformatics. 2011;12. doi:10.1186/1471-2105-12-231.

[302] Devlin B, Roeder K. Genomic control for association studies. Biometrics. 1999;55(4):997–1004. doi:10.1111/j.0006-341X.1999.00997.x.

[303] Aulchenko YS, Ripke S, Isaacs A, van Duijn CM. GenABEL: An R library for genome-wide association analysis. Bioinformatics. 2007;23(10):1294–1296. doi:10.1093/bioinformatics/btm108.

[304] Benjamini Y, Hochberg Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing. Journal of the Royal Statistical Society: Series B (Methodological). 1995;57(1):289–300. doi:10.1111/j.2517-6161.1995.tb02031.x.

[305] Yi X, Liang Y, Huerta-Sanchez E, Jin X, Cuo ZXP, Pool JE, et al. Sequencing of 50 human exomes reveals adaptation to high altitude. Science. 2010;329(5987):75–78. doi:10.1126/science.1190371.

[306] Hudson RR, Slatkin M, Maddison WP. Estimation of levels of gene flow from DNA sequence data. Genetics. 1992;132(2):583–589.

[307] Bhatia G, Patterson N, Sankararaman S, Price AL. Estimating and interpreting F_{ST}: The impact of rare variants. Genome Research. 2013;23(9):1514–1521. doi:10.1101/gr.154831.113.

[308] Zhou X, Stephens M. Genome-wide efficient mixed-model analysis for association studies. Nature Genetics. 2012;44(7):821–824. doi:10.1038/ng.2310.

[309] Browning BL, Browning SR. A unified approach to genotype imputation and haplotype-phase inference for large data sets of trios and unrelated individuals. American Journal of Human Genetics. 2008;84(2):210–223. doi:10.1016/j.ajhg.2009.01.005.

[310] Stanke M, Morgenstern B. AUGUSTUS: A web server for gene prediction in eukaryotes that allows user-defined constraints. Nucleic Acids Research. 2005;33(SUPPL. 2):465–467. doi:10.1093/nar/gki458.

[311] Stanke M, Schöffmann O, Morgenstern B, Waack S. Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources. BMC Bioinformatics. 2006;7:1–11. doi:10.1186/1471-2105-7-62.

[312] Mi H, Muruganujan A, Ebert D, Huang X, Thomas PD. PANTHER version 14: More genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. Nucleic Acids Research. 2019;47(D1):D419–D426. doi:10.1093/nar/gky1038.

[313] Gascuel O, Steel M. Neighbor-joining revealed. Molecular Biology and Evolution. 2006;23(11):1997–2000. doi:10.1093/molbev/msl072.

[314] Su J, Chai X, Kahn B, Napoli JL. cDNA cloning, tissue distribution, and substrate characteristics of a cis-retinol/3α-hydroxysterol short-chain dehydrogenase isozyme. Journal of Biological Chemistry. 1998;273(28):17910–17916. doi:10.1074/jbc.273.28.17910.

[315] Campana WM, Hiraiwa M, O'Brien JS. Prosaptide activates the MAPK pathway by a G-protein-dependent mechanism essential for enhanced sulfatide synthesis by Schwann cells. The FASEB Journal. 1998;12(3):307–314. doi:10.1096/fasebj.12.3.307.

[316] Meyer RC, Giddens MM, Schaefer SA, Hall RA. GPR37 and GPR37L1 are receptors for the neuroprotective and glioprotective factors prosaptide and prosaposin. Proceedings of the National Academy of Sciences of the United States of America. 2013;110(23):9529–9534. doi:10.1073/pnas.1219004110.

[317] Andersson TPM, Svensson SPS, Karlsson AM. Regulation of melanosome movement by map kinase. Pigment Cell Research. 2003;16(3):215–221. doi:10.1034/j.1600-0749.2003.00048.x.

[318] Deacon SW, Nascimento A, Serpinskaya AS, Gelfand VI. Regulation of bidirectional melanosome transport by organelle bound MAP kinase. Current Biology. 2005;15(5):459–463. doi:10.1016/j.cub.2004.12.074.

[319] Nilsson H, Wallin M. Evidence for several roles of dynein in pigment transport in melanophores. Cell Motility and the Cytoskeleton. 1997;38(4):397–409. doi:10.1002/(SICI)1097-0169(1997)38:4¡397::AID-CM9¿3.0.CO;2-0.

[320] Tuma MC, Zill A, Le Bot N, Vernos I, Gelfand V. Heterotrimeric kinesin II is the microtubule motor protein responsible for pigment dispersion in Xenopus melanophores. Journal of Cell Biology. 1998;143(6):1547–1558. doi:10.1083/jcb.143.6.1547.

[321] King SM. Composition and assembly of axonemal dyneins. In: King SM, editor. Dyneins: The Biology of Dynein Motors: Second Edition. second edi ed. Elsevier Inc.; 2017. p. 163–201. Available from: https://doi.org/10.1016/B978-0-12-809471-6.00005-X.

[322] Semba K, Araki K, Li Z, Matsumoto KI, Suzuki M, Nakagata N, et al. A novel murine gene, Sickle tail, linked to the Danforth's short tail locus, is required for normal development of the intervertebral disc. Genetics. 2006;172(1):445–456. doi:10.1534/genetics.105.048934.

[323] Karasugi T, Semba K, Hirose Y, Kelempisioti A, Nakajima M, Miyake A, et al. Association of the tag SNPs in the human SKT gene (KIAA1217) with lumbar disc herniation. Journal of Bone and Mineral Research. 2009;24(9):1537–1543. doi:10.1359/jbmr.090314.

[324] Lavker RM, Kaidbey KH. Redistribution of melanosomal complexes within keratinocytes following UV-A irradiation: A possible mechanism for cutaneous darkening in man. Archives of Dermatological Research. 1982;272(3-4):215–228. doi:10.1007/BF00509049.

[325] Wu X, Bowers B, Rao K, Wei Q, Hammer JA. Visualization of melanosome dynamics within wild-type and dilute melanocytes suggests a paradigm for myosin v function in vivo. Journal of Cell Biology. 1998;143(7):1899–1918. doi:10.1083/jcb.143.7.1899.

[326] Rogers SL, Gelfand VI. Membrane trafficking, organelle transport, and the cytoskeleton. Current Opinion in Cell Biology. 2000;12(1):57–62. doi:10.1016/S0955-0674(99)00057-5.

[327] Manderfield LJ, Engleka KA, Aghajanian H, Gupta M, Yang S, Li L, et al. Pax3 and hippo signaling coordinate melanocyte gene expression in neural crest. Cell Reports. 2014;9(5):1885–1895. doi:10.1016/j.celrep.2014.10.061.

[328] Galibert MD, Yavuzer U, Dexter TJ, Goding CR. Pax3 and regulation of the melanocyte-specific tyrosinase-related protein-1 promoter. Journal of Biological Chemistry. 1999;274(38):26894–26900. doi:10.1074/jbc.274.38.26894.

[329] Lang D, Lu MM, Huang L, Engleka KA, Zhang M, Chu EY, et al. Pax3 functions at a nodal point in melanocyte stem cell differentiation. Nature. 2005;433(7028):884–887. doi:10.1038/nature03292.

[330] Minchin JEN, Hughes SM. Sequential actions of Pax3 and Pax7 drive xanthophore development in zebrafish neural crest. Developmental Biology. 2008;317(2):508–522. doi:10.1016/j.ydbio.2008.02.058.

[331] Silverstone P. Two new species of Colostethus (Amphibia, Anura, Dendrobatidae) from Colombia. Contributions In Science Natural History Museum of Los Angeles County. 1975;268:1–10.

[332] Myers CW, Daly JW. A name for the poison frog of Cordillera Azul, Eastern Perú, with notes on its biology and skin toxins. Natural History. 1979;(2674):1–28.

[333] Twomey E, Kain M, Claeys M, Summers K, Castroviejo-Fisher S, Van Bocxlaer I. Mechanisms for color convergence in a mimetic radiation of poison frogs. The American Naturalist. 2020;195(5):E132–E149. doi:10.1086/708157.

[334] Stern DL. The genetic causes of convergent evolution. Nature Reviews Genetics. 2013;14(11):751–764. doi:10.1038/nrg3483.

[335] Rosenblum EB, Parent CE, Brandt EE. The Molecular Basis of Phenotypic Convergence. Annual Review of Ecology, Evolution, and Systematics. 2014;45(1):203–226. doi:10.1146/annurev-ecolsys-120213-091851.