Title

Patterns of Diversity in Malagasy Poison Frogs

Author

Klonoski, Karina

Publication Date

2019

Peer reviewed|Thesis/dissertation

Patterns of Diversity in Malagasy Poison Frogs

By

Karina Klonoski

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Environmental Science, Policy, and Management

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Erica Bree Rosenblum, Chair
Professor Rasmus Nielsen
Professor Steven Beissinger

Fall 2019

Patterns of Diversity in Malagasy Poison Frogs

Abstract

Patterns of Diversity in Malagasy Poison Frogs

by

Karina Klonoski

Doctor of Philosophy in Environmental Science, Policy, and Management

University of California, Berkeley

Professor Erica Bree Rosenblum, Chair

The processes that generate and maintain biological diversity in the natural world have long captivated evolutionary biologists. In aposematic organisms, where conspicuous coloration warns predators of toxicity and/or unpalatability, phenotypic variation is particularly compelling because many factors are believed to contribute to color and pattern diversity within and among species. My dissertation focuses on a group of aposematic frogs endemic to Madagascar, commonly called Malagasy poison frogs (genus *Mantella*). The bright coloration observed in many Malagasy poison frogs coincides with the presence of toxic skin compounds derived from arthropod prey. Although several species within this group are endangered, little is known about their natural history and the diversity of color and pattern that occurs is poorly understood. Conservation efforts are further complicated by unresolved genetic relationships among phenotypically divergent species and populations. By characterizing patterns of genetic and phenotypic variation, my dissertation research aims to clarify the evolutionary processes contributing to color diversity and to inform management efforts for this group.

To better understand the factors contributing to phenotypic diversity, I first quantified genetic structure and color and pattern variation across populations of three closely-related species of Malagasy poison frogs, *Mantella aurantiaca*, *Mantella crocea*, and *Mantella milotympanum*. Although my genomic analyses identified three distinct genetic clusters within this complex, they did not correspond to current species designations. In some instances, populations presumed to be "morphs" of one species were actually identified as genetically distinct units. By demonstrating the complexity of distinguishing between species, intraspecific phenotypic variants, and populations, my study highlights the need to re-evaluate how species and morphs are classified, especially in aposematic organisms. In addition to the conservation implications of my study, I also found evidence that a variety of mechanisms, including selection, drift, and hybridization, are contributing to the color diversity observed in these populations.

In my second chapter, I further examined patterns of diversity among *M. aurantiaca*, *M. crocea*, and *M. milotympanum* populations by assessing the relationship between color and pattern variation and alkaloid-based chemical defenses. Unexpectedly, I detected very limited correlations between alkaloid composition and frog color, pattern, or conspicuousness. Additionally, differences in frog conspicuousness were not associated with differences in frog alkaloid diversity or abundance. Collectively, these results suggest that aposematic signals in this group may not be quantitatively honest, where increasing conspicuousness corresponds to increasing toxicity. Geographic distance, however, was highly correlated with alkaloid variation, implicating differences in habitat and associated characteristics as important determinants of variation in chemical defenses. Overall, my results indicate that geography, rather than phenotypic diversity, has a strong role in structuring chemical variation among populations.

In my final chapter, I aimed to clarify the status of the endangered *Mantella cowani* and its relationship to the common *Mantella baroni* by quantifying genetic and phenotypic variation across putative *M. cowani – M. baroni* hybrid and parental populations. Although previous mitochondrial studies have suggested hybridization between *M. cowani* and *M. baroni* in the wild, my genomic analyses did not recover admixture between the two species at this location. In fact, my population genetic analyses did not reveal any substantial genetic structure, even between *M. cowani* and *M. baroni* populations. Despite the lack of genomic differentiation, *M. cowani* was phenotypically distinct from both *M. baroni* and putative *M. cowani – M. baroni* hybrids. These discordant patterns of genomic and phenotypic diversity could possibly be explained by longstanding hybridization, ongoing migration, or selection acting on a small portion of the genome. Despite the lack of genetic variation among populations, I found *M. cowani* to be a morphologically distinct population. Further, the putative *M. cowani – M. baroni* hybrid population demonstrated novel phenotypes not observed in either species. My findings raise important questions as to how different aspects of biological diversity should be prioritized in conservation, especially in aposematic species where phenotypes have clear ecological relevance and may be important for survival.

My dissertation provides integrative studies of phenotypic and genetic diversity for several species of Malagasy poison frogs and has important implications for both evolutionary biology and conservation. Greater understanding of adaptation and diversity in this group will be essential to future conservation efforts and to the preservation of Madagascar's unique biodiversity.

For my father, who gave up everything so that I could have everything

Table of Contents

Acknowledgements

Surprisingly, this has been one of the sections of my dissertation that I've been most resistant to writing. Not because I'm not grateful, but because it is an impossible task to put into words the gratitude that I have for so many people who have seen me at my worst (and my best!) over these past few years. But, I'll try.

Above all, I thank my father, Richard. As a single parent, my dad raised three children while simultaneously balancing a career as a cardiologist. To say he sacrificed a lot is quite the understatement. He worked tirelessly to provide me and my siblings with every opportunity that he possibly could. The greatest gift that my father ever gave me, though, was raising me to believe that I really could be anything that I wanted to be. When I dreamed of being a fashion designer, he proudly posted the (obviously horrendous) outfits that I drew on the refrigerator. When I decided to become a cartoonist, he told me I was great at drawing. When I finally landed on biology, he told me that's what he always wanted to do, and proceeded to come visit me on all my subsequent field excursions. In retrospect, the thing that I'm most grateful for is that my dad always granted me the freedom to figure it out on my own, supported me along the way, and trusted me to make my own decisions. Growing up, it never occurred to me that I couldn't be whoever I wanted to be. That is a mentality that many people are not lucky enough to have, for a variety of reasons. I can never thank my father enough for providing me with a life that seemed wide open to every possibility.

Next, I thank the members of the Rosenblum lab – past and present – for providing the most wonderful, supportive, and inspiring community that I could have ever asked for. I can only hope that I will have the opportunity to work alongside a similarly incredible group of humans again. I especially thank Alex Krohn, who started graduate school with me at the exact same time, and saw me through some of my darkest moments. Even if he was kind of stand-offish when I first met him.

In my fourth year of graduate school, about a month before I was heading to Madagascar for two months to do fieldwork, I met Eric. A month after meeting, Eric came over, put on lab gloves, and helped me wrap and pack my hundreds of field sampling tubes for Madagascar. Despite my being completely out of contact for those two months, he eventually became my partner and a rock for me during some of the most turbulent times of my graduate career. Eric has seen me at my absolute lowest moments, rolled up his sleeves, and picked me up off of the ground. He has been a beacon of positivity and consistently reminded me of the things that truly matter in life. I will be forever grateful for his unwavering support.

I would also like to thank my qualifying exam and dissertation committee members, Steve Beissinger, Rasmus Nielsen, Damian Elias, and Stephanie Carlson, for their valuable feedback and guidance throughout my graduate career. Additionally, a number of scientists have been amazing mentors and were fundamental to my development as a scientist. I'd especially like to thank Katherine Pease, Mini Watsa, Claire Kremen, Betsy Mitchell, and Laura Lammers. For my logistically difficult fieldwork in Madagascar, I am forever indebted to Serge Ndriantsoa for his invaluable assistance over three field seasons. Serge was a tremendous asset in the field and played a large part in bringing these projects to fruition. I am also grateful to Devin Edmonds and Miguel Vences for their support and for sharing their amazing knowledge of *Mantella* and Madagascar with me.

Finally, I thank Bree. When I started this journey seven years ago, I only had an idea that seemed like a dream and not a reality. In manifesting that dream, Bree supported me every step of the way. More importantly, she supported my growth and development as a human being. At times, she saw things in me that I could not see in myself. I've learned many invaluable lessons from Bree, one of which is to embrace the uncertainty. Because of this, I look ahead with excitement rather than fear. I can never thank Bree enough for giving me the opportunity to turn a crazy, far-fetched idea in my head into something real.

1

Phenotypic and genetic diversity in aposematic Malagasy poison frogs (genus *Mantella*)[1]

ABSTRACT

Intraspecific color variation has long fascinated evolutionary biologists. In species with bright warning coloration, phenotypic diversity is particularly compelling because many factors, including natural and sexual selection, contribute to intraspecific variation. To better understand the causes of dramatic phenotypic variation in Malagasy poison frogs, we quantified genetic structure and color and pattern variation across three closely-related species, *Mantella aurantiaca*, *Mantella crocea*, and *Mantella milotympanum*. Although our restriction-site associated DNA (RAD) sequencing approach identified clear genetic clusters, they do not align with current species designations, which has important conservation implications for these imperiled frogs. Moreover, our results suggest that levels of intraspecific color variation within this group have been overestimated, while species diversity has been underestimated. Within major genetic clusters, we observed distinct patterns of variation including: populations that are phenotypically similar yet genetically distinct, populations where phenotypic and genetic breaks coincide, and populations that are genetically similar but have high levels of within-population phenotypic variation. We

[1] This work has previously been published and is reproduced here with permission from co-authors and the Graduate Division at UC Berkeley. See: Klonoski, K., K. Bi, and E. B. Rosenblum. 2019. Phenotypic and genetic diversity in aposematic Malagasy poison frogs (genus *Mantella*). *Ecology and Evolution* 9(5): 2725-2742.

also detected admixture between two of the major genetic clusters. Our study suggests that several mechanisms – including hybridization, selection, and drift – are contributing to phenotypic diversity. Ultimately, our work underscores the need for a re-evaluation of how polymorphic and polytypic populations and species are classified, especially in aposematic organisms.

INTRODUCTION

The origin and maintenance of intraspecific color variation and its consequences for speciation have long captivated evolutionary biologists (e.g., Huxley 1955; Endler 1980; Gray and McKinnon 2007; Corl et al. 2010, Hugall and Stuart-Fox 2012). Examples of color variation in nature are widespread and have been documented across many diverse taxonomic groups, including reptiles, fish, mammals, birds and invertebrates (e.g., Sandoval 1994; Sinervo et al. 2001; Galeotti et al. 2003; Hoekstra et al. 2004; Maan et al. 2008). Previous work has indicated that color variation observed within or between populations may be mediated by natural selection (e.g., apostatic selection, divergent selection in different substrate or light environments as in Sandoval 1994; Hoekstra et al. 2004), sexual selection (e.g., mate choice, variation in male mating strategies as in Sinervo and Lively 1996; Kingston et al. 2003), genetic drift (Hoffman et al. 2006), or some combination of these factors (e.g., Endler 1991; Oxford 2005; Reynolds and Fitzpatrick 2007). Intraspecific color variation can also potentially give rise to new species, especially when natural or sexual selection reduces gene flow between alternative morphs (e.g., Rosenblum and Harmon 2011).

Examples of color variation in aposematic organisms – where conspicuous warning signals advertise toxicity or unpalatability to predators (Poulton 1890) – are particularly compelling. Aposematic colors are often highly contrasting, variable, and potentially exhibit tradeoffs between natural and sexual selection (Summers et al. 1999; Jiggins et al. 2001; Reynolds and Fitzpatrick 2007; Estrada and Jiggins 2008; Nokelainen et al. 2012; Stevens and Ruxton 2012; Crothers and Cummings 2013). Historically, variation in aposematic signals has been considered perplexing from a theoretical perspective because phenotypic diversity is expected to be highly constrained in such systems due to positive frequency-dependent selection via predation (Endler and Greenwood 1988; Mallet and Joron 1999; Briolat et al. 2018). Once a predator has learned to associate toxicity with a particular phenotype, protection should be conferred to those organisms displaying a similar phenotype, encouraging uniformity in warning coloration. According to theoretical predictions of predator avoidance learning, novel phenotypes should be unrecognizable to predators as toxic and thus quickly removed from populations (Müller 1879; Mallet and Barton

1989; Guilford and Dawkins 1993). In recent years, however, studies have identified many biotic and abiotic factors – including intraspecific communication, parasite load, temperature and variability in predator learning and sensory abilities – that contribute to variation in aposematic signals (reviewed in Briolat et al. 2018). Growing awareness of the variety of selective factors influencing aposematic coloration has led scientists to encourage a more holistic approach when investigating diversity in warning coloration and to consider the range of factors that may be at play in maintaining phenotypic variation (Briolat et al. 2018).

Studies of color variation within aposematic species have traditionally focused on systems demonstrating either multiple color morphs within a population (i.e., polymorphism) or geographic color variation among populations (i.e., polytypism; reviewed in Briolat et al. 2018). In systems with high color variability, however, determining whether color variants represent different species, different populations, or different morphs within populations is difficult particularly when genetic structure is not well resolved. Distinguishing between species, populations and morphs can be especially challenging in phenotypically diverse poison frog groups, where high rates of phenotypic variation can confound our understanding of species delimitations (Roland et al. 2017; Tarvin et al. 2017; Posso-Terranova and Andrés 2018).

In many instances, the inability to distinguish whether color variation occurs within populations, between populations, or between species is further compounded by nonexistent or limited genetic datasets, which lack the resolution needed to clarify relationships among populations. Tarvin et al. (2017) recently demonstrated that the level of interspecific mitochondrial divergence among four distinct poison frog species was comparable to the divergence levels observed between populations considered to be a single polymorphic species (Hauswaldt et al. 2011). Noting the limitations of mitochondrial data in resolving species boundaries, the authors explicitly called for genome-level studies, in combination with information on phenotypic diversity and natural history, to understand relationships in these complicated systems (Tarvin et al. 2017). The power of more comprehensive genetic datasets to resolve genetic structure in poison frogs has been demonstrated in recent studies where Neotropical poison frogs considered to be a single species in fact contained multiple genetic lineages that potentially represent new species (Roland et al. 2017; Posso-Terranova and Andrés 2018).

While the relationship between phenotypic and genetic diversity has been extensively studied in the Neotropical poison frogs (e.g., Wang and Summers 2010; Twomey et al. 2013; Roland et al. 2017; Tarvin et al. 2017), there is an entirely separate radiation of poison frogs in which color diversity has never been examined. Endemic to Madagascar, the *Mantella* genus describes sixteen species of toxic, diurnal frogs exhibiting variable coloration and pattern both within and among species (Glaw and Vences 2007). Commonly called Malagasy poison frogs, the bright coloration displayed by many species within this group is presumed to be aposematic. The toxic

skin alkaloids found in *Mantella* species are believed to be derived from arthropod prey (Daly et al. 1996; Daly et al. 1997), similar to the Neotropical poison frogs, and variation in alkaloid composition has been observed among species, populations, and habitats (Daly et al. 2008; Andriamaharavo et al. 2015). The mechanism of chemical defense is hypothesized to have evolved convergently in Neotropical and Malagasy poison frogs (Clark et al. 2005). Despite their high degree of phenotypic variation and apparent similarity to the Neotropical poison frogs (Heying 2001b; Rojas 2016), little is known about the natural history, ecology, and genetic background of Malagasy poison frogs.

Within the *Mantella* genus, one complex of three closely-related species, *Mantella aurantiaca*, *Mantella crocea*, and *Mantella milotympanum*, demonstrates a particularly high degree of variability in conspicuous coloration and pattern. Found in the rainforests of central-eastern Madagascar, the geographic range of all three species is highly restricted and patchy in its distribution (Bora et al. 2008). Among and within populations, there is exceptional phenotypic variation both in dorsal coloration, which ranges from red to green at the extremes, and in patterning elements, which are variable in the degree of ventral spotting and black banding present on the side. Yet, any attempt to understand the phenotypic diversity in this group is hindered by the unresolved taxonomy of these species. Previous genetic work is limited to a handful of mitochondrial DNA and allozyme studies, which have yielded somewhat confusing results (Vences et al. 1998; Schaefer et al. 2002; Chiari et al. 2004; Vences et al. 2004). *M. aurantiaca*, *M. milotympanum*, and *M. crocea* are thought to fall within the *M. madagascariensis* group, one of five clades within the *Mantella* genus, though their position in this group is controversial (Schaefer et al. 2002; Vences et al. 2004). Population genetic studies have detected high degrees of haplotype sharing between *M. milotympanum* and *M. crocea*, resulting in the hypothesis that these two species are conspecific (Chiari et al. 2004; Vences et al. 2004). Additionally, frog populations displaying patterning that is intermediate between that of *M. crocea* and *M. milotympanum* exist in the wild and are referred to as *M. cf. milotympanum* in the literature (Chiari et al. 2004). Evidence of haplotype sharing between *M. aurantiaca* and *M. crocea* (Chiari et al. 2004; Vences et al. 2004) has also prevented taxonomic resolution within this group, and species designations remain controversial.

Given the lack of resolution in previous molecular studies, it is apparent that a high-resolution genetic dataset is needed to both clarify relationships within this group and to determine whether observed color variants represent distinct species or morphs. In this study, we used a restriction-site associated (RAD) sequencing approach, in combination with multiple matrix regression analysis, to compare variation in dorsal coloration and side and ventral patterning with genetic and geographic distance across the entire known range of three species of Malagasy poison frog. Specifically, our objectives were to 1) clarify genetic structure among populations of *M. crocea*, *M. milotympanum* and *M. aurantiaca*, 2) quantify variation in

dorsal coloration and side and ventral patterning both among and within populations, and 3) describe the relationship between genetic diversity, phenotypic diversity, and geographic distance for all major genetic clusters within this three-species complex. This study – the first quantitative and objective exploration of color diversity in the *Mantella* genus – not only provides a foundation for future studies of color evolution in Malagasy poison frogs, but also identifies several critical issues that should be more thoroughly considered in any investigation of aposematic organisms presumed to be polymorphic or polytypic.

MATERIALS AND METHODS

Field Sampling

We sampled three closely-related species, currently named *Mantella aurantiaca*, *Mantella crocea* and *Mantella milotympanum,* throughout their entire known range in central-eastern Madagascar (Fig. 1). We also sampled individuals containing an intermediate phenotype between *M. crocea* and *M. milotympanum*, hereafter referred to as *M. cf. milotympanum*, following previous nomenclature in the literature. Overall, we sampled 88 frogs from 16 populations. Fieldwork was conducted during the rainy breeding season over three years: January – February 2014, January – February 2015, and November 2015 – January 2016. We captured the frogs by hand and transported them back to a field laboratory where all data collection occurred. We collected digital photographs and toe clips from 11 *M. aurantiaca* individuals from 2 populations, 34 *M. crocea* individuals from 6 populations, 19 *M. milotympanum* individuals from 3 populations, and 24 *M. cf. milotympanum* individuals from 5 populations. Several studies have demonstrated that toe-clipping has no significant impact on frog survival, body condition, or growth (reviewed in Perry et al. 2011). Toe clips were preserved in salt-saturated DMSO and stored at room temperature. All data were collected on the same day that frogs were captured. Frogs were held overnight and released to their site of capture the following morning. All animal handling procedures were approved by the Animal Care and Use Committee at the University of California at Berkeley (R347-0314; AUP-2015-01-7083). Collection and exportation of samples were performed under permits issued by the Direction Générale des Forêts, Direction de la Conservation de la Biodiversité et du Système des Aires Protégées, and Ministere de l'Environment, de l'Ecologie et des Forêts in Madagascar (collection permits: 315/13/MEF/SG/DGF/DCB.SAP/SCB, 335/14/MEF/SG/DGF/DCB.SAP/SCB, 336/14/MEF/SG/DGF/DCB.SAP/SCB and 296/15/MEEMF/SG/DGF/DAPT/SCBT; export permits: 051C-EA02/MG14, 048C-EA02/MG15 and 002C-EA01/MG16).

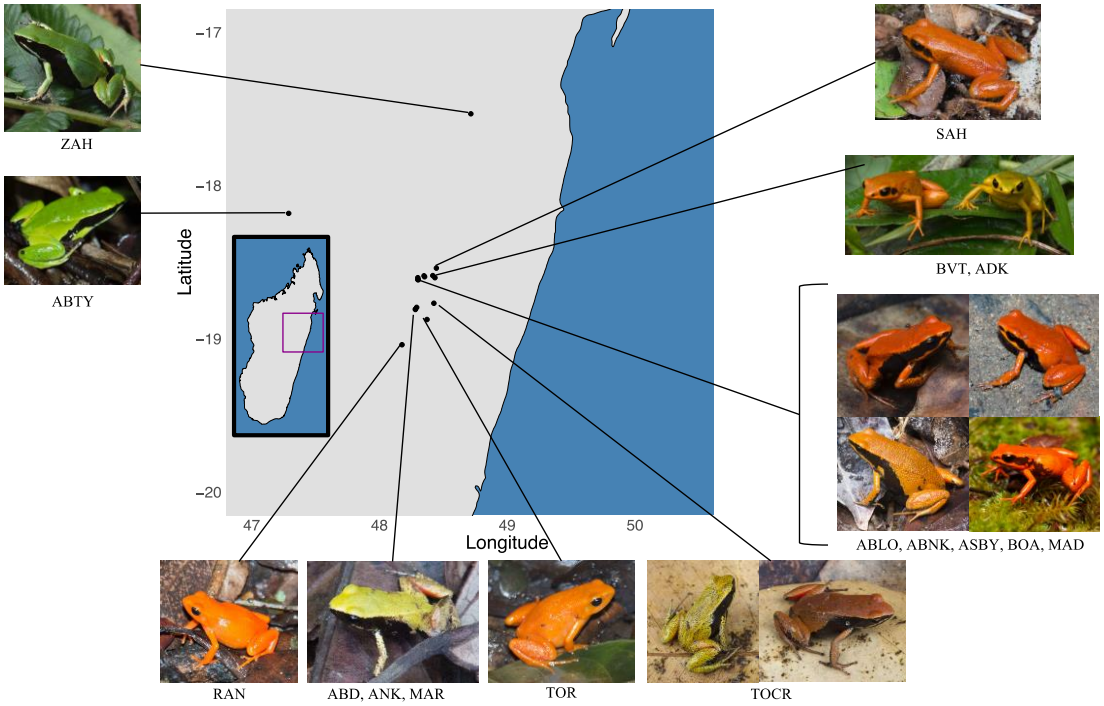

Figure 1: Sampling localities for 16 populations of *Mantella crocea*, *Mantella aurantiaca*, and *Mantella milotympanum* across central-eastern Madagascar. Representative individuals from a population or group of populations are pictured next to population labels. Images shown here do not represent the entire range of observed phenotypic diversity.

Quantification of Phenotypic Variation

To characterize frog coloration and pattern, we photographed the dorsal, ventral and side surfaces of all frogs in a standardized manner following a protocol modified from Stevens et al. (2007). Photographs of all frogs were taken after transportation to the field laboratory, but before any other handling occurred. Frogs were always photographed in natural light during the hours of 1:00-5:00pm using a Pentax K-30 digital single-lens reflex camera fitted with a Pentax 18-135mm lens. All frogs were photographed on a white paper background with a scale bar and a white-gray-black standard present (QPcard 101; gray standard reflectance value of 18%). Two individuals (ABNK09 and RAN11) were excluded from our phenotypic analysis due to unsuitable photographs.

To quantify dorsal coloration, we first used the Image Calibration and Analysis Toolbox (Troscianko and Stevens 2015), utilized within ImageJ v1.51 (Schneider et al. 2012), to generate aligned and normalized images from our RAW photographs, which

enabled us to objectively measure color and pattern. After standardization, we selected two regions of interest on the frog's dorsal surface in which to quantify color: one 3x3 mm square toward the back of the frog dorsum, and one 3x1 mm rectangle behind the frog's right arm. Regions of interest on the frog's dorsum were manually selected in order to avoid any glare present in photographs as well as any injuries on the frog's dorsal surface that resulted in discoloration. After selecting regions of interest, we used the Batch Image Analysis function of the toolbox to extract the red color values, green color values, and blue color values (hereafter referred to as RGB values) for each frog. R values, G values and B values were averaged for all pixels within the regions of interest. Color values were averaged separately for each color channel (R, G, and B).

To transform RGB values into measurements relevant to a vertebrate visual system, we followed the protocol of Krohn and Rosenblum (2016), modified from Endler (2012) and McKay (2013). Briefly, we calculated three axes from our RGB values corresponding to a red-green channel, (R-G)/(R+G), a blue-green channel, (G-B)/(G+B), and a luminance channel, which we defined as untransformed R values. Because luminance is processed separately in vertebrates (Endler and Mielke 2005; Endler 2012), we used our other two axes, (R-G)/(R+G) and (G-B)/(G+B), to plot dorsal coloration as a point in a two-dimensional color space based on a red-green channel and a blue-green channel. In our two-dimensional color space, (R-G)/(R+G) represented the x-axis and (G-B)/(G+B) represented the y-axis. From this color space, we calculated chroma and hue values following Krohn and Rosenblum (2016). Chroma was calculated as the hypotenuse of the triangle formed by the x and y axes, and hue was calculated as the angle between the hypotenuse and the x-axis. We used these values of chroma, hue and luminance to characterize dorsal coloration.

To quantify side and ventral pattern, we again used the Image Calibration and Analysis Toolbox (Troscianko and Stevens 2015) to generate aligned and normalized images from RAW photographs. Next, we manually outlined the ventral and side surfaces of frogs on the standardized images using the polygon and brush selection tools. We selected the entire ventral and side surfaces to obtain a comprehensive measure of overall pattern on each surface. After manually selecting the regions of interest, we used the scale bars present in each photograph to scale all images to the same number of pixels per millimeter (side surfaces = 19 px/mm; ventral surfaces = 18.6 px/mm), which is necessary for pattern analysis. We performed a granularity analysis, which is based on Fast Fourier bandpass filtering, using the Image Calibration and Analysis Toolbox implemented in ImageJ. For our side pattern analysis, we specified Fast Fourier Transform Bandpass filters at 16 levels, starting at 2 pixels and increasing by a multiple of the square root of 2 until 430 pixels. For our ventral pattern analysis, we specified Fast Fourier Transform Bandpass filters at 14 levels, starting at 2 pixels and increasing by a multiple of the square root of 2 until 193 pixels. From our granularity analysis, we generated descriptive summary statistics to estimate pattern contrast, pattern diversity, and luminance contrast for both side and ventral pattern.

Granularity analysis has been increasingly used to measure pattern in a variety of organisms (e.g., Barbosa et al. 2008; Stoddard and Stevens 2010) and draws on basic characteristics of early stage visual processing present across diverse animals (Campbell and Rodson 1968; Godfrey et al. 1987; Stoddard and Stevens 2010; Pérez-Rodríguez et al. 2017). Conversely, color adjacency analysis, a method which has previously been used to quantify pattern in poison frogs, does not represent visual processing of pattern (Pérez-Rodríguez et al. 2017).

Our dataset included both males and females (58 males, 26 females, two juveniles). Preliminary analyses did not reveal an effect of sex on frog coloration or pattern, so sexes were lumped for the analyses presented here. However, our study design did not explicitly aim to quantify sexual dimorphism, and future studies can investigate this question with targeted sampling.

Quantification of Genetic Variation

We extracted DNA from toe clips using Qiagen DNeasy extraction kits (Qiagen, Valencia, CA, USA) generally following the manufacturer's protocol with two modifications: 4 µl of 1 mg/mL RNase A was added to each sample after lysis, and DNA was eluted in 45 µl of 1X LTE buffer to maximize concentration. Prior to library preparation, we checked the quality of extracted DNA using agarose gels and quantified DNA using a Qubit 2.0 Fluorometer (ThermoFisher, Waltham, MA, USA).

We constructed a restriction-site associated DNA (RAD) library following the protocol of Ali et al. (2016), without the targeted bait capture step, otherwise referred to as the "bestRAD" protocol. During preparation of our bestRAD library, we digested 50 ng of DNA from each individual with SbfI-HF (New England Biolabs, Ipswich, MA, USA) restriction enzyme, performed size selection with magnetic beads, and amplified our library using 12 cycles of polymerase chain reaction (PCR). We sequenced our library on two lanes of the Illumina HiSeq 4000 at the U.C. Davis Genome Center with 150 bp paired-end reads.

RADseq Data Processing

To process RADseq data, we used pipelines implemented in a customized PERL workflow that also utilized various external programs (pipelines can be accessed through https://github.com/CGRL-QB3-UCBerkeley/RAD). We first de-multiplexed raw fastq reads using internal barcode sequences and allowing for one mismatch in barcode sequence. We removed de-multiplexed reads that did not include the expected restriction enzyme cut site at the beginning of the read, again allowing for one mismatch in cut site sequence, and also removed exact duplicates using Super Deduper (https://github.com/dstreett/Super-Deduper). To filter reads,

we used Cutadapt (Martin 2011) and Trimmomatic (Bolger et al. 2014) to trim adapter contamination and low quality reads. We removed filtered reads that were shorter than 50bp. After cleaning and filtering reads, we used cd-hit (Li & Godzik 2006; Fu et al. 2012) to cluster forward reads of each individual at 95% similarity, retaining only those clusters with at least two supported reads. For each cluster, we used the sequence identified as representative by cd-hit as our marker. Retained markers were next masked for repetitive elements, low complexities, and short repeats with Ns using RepeatMasker (Smit et al. 2014) with "frog" as a database. Post-masking, we removed markers where more than 30% of nucleotides were Ns. To remove potential paralogs present within each individual, we used Blastn (Altschul et al. 1990) to compare all clustered loci against themselves, and subsequently eliminated any locus that matched a locus other than itself. Next, remaining RAD markers from each individual were combined and clustered for all individuals using cd-hit. We used a similarity threshold of 90% to select for markers containing at least 50 nucleotides and shared by at least 60% of all individuals. This served as our reference genome for all samples, and we subsequently aligned each individual's cleaned sequence reads to this reference using Novoalign (http://www.novocraft.com). We only retained those reads that mapped uniquely to the reference. Using Picard (http://www.picard.sourceforge.net) and GATK (McKenna et al. 2010), we added read groups and performed realignment around indels. To generate quality control information in VCF format, we used SAMtools/BCFtools (Li et al. 2009), after which data were further filtered using a custom method, SNPcleaner (https://github.com/tplinderoth/ngsQC/tree/master/snpCleaner; Bi et al. 2013), that was slightly modified to remove sites around indels and implemented in our pipelines. Additionally, we filtered out markers where more than two alleles were called at any site, and also masked sites that were within 10bp of any indel. Only individual sites falling within the first to ninety-ninth percentile of overall coverage among all samples were retained. We also removed SNPs failing to pass a one-tailed HWE exact test (1e-4) or showing strong base quality bias (1e-100). To avoid bias resulting from excessive missing data, we only used sites present in at least 60% of individuals with at least 3X coverage in our downstream analyses.

Population Genomic Analyses

We used genotype likelihoods instead of genotype calls whenever possible to account for uncertainty in our data. Because our data showed low to medium coverage (1.8-13.9X, with an average of 5.4X), SNP and genotype calls based on allele counts could potentially cause bias or introduce noise in downstream analyses (Johnson & Slatkin 2008; Lynch 2008). Genotype likelihoods were calculated in an empirical Bayesian framework using ANGSD (http://www.popgen.dk/angsd/index.php/ANGSD; Korneliussen et al. 2014), a

software that is specialized for analyzing low to medium coverage next generation sequencing data. The majority of downstream analyses conducted in ANGSD are performed based on likelihood of site allele frequencies, genotype likelihoods, or genotype posterior probabilities. For analyses that required SNP or genotype calling, we only included high confidence variants (identified using a likelihood ratio test to determine variable sites with p-values less than 1e-6 and genotype posterior probabilities greater than 0.95) where at least 80% of individuals showed at least 3X coverage.

To characterize population structure for all samples, we first performed a Principal Components Analysis (PCA) of the covariance matrix of posterior genotype probabilities implemented in ngsTools (http://github.com/mfumagalli/ngsTools; Fumagalli et al. 2014). Next, we used a neighbor-joining network (NeighborNet) analysis based on uncorrelated *p*-distances in Splitstree (Huson 1998; Huson & Bryant 2006) to visualize population structure. We adhered to the stringent thresholds mentioned above to call a set of high quality variants, which were used to compute a genetic distance matrix for Splitstree using the Adegenet package in R (Jombart 2008). We also quantified the population structure of all samples using NGSadmix (Skotte et al. 2013), which relies on genotype likelihoods. Because there were sixteen populations present in our study, we estimated individual admixture proportions with the number of clusters ranging from one to seventeen (K=1 to K=17), with ten replicates per K value. We then used the Evanno method (Evanno et al. 2005) to identify the most likely K value.

To characterize fine-scale population structure, we performed an NGSadmix analysis within each major genetic cluster. We again estimated individual admixture proportions with the number of clusters ranging from one to one more than the total number of populations (K=1 to K=3 for Cluster A; K=1 to K=7 for Cluster B; K=1 to K=9 for Cluster C; K=1 to K=15 for candidate hybrid populations). We ran NGSadmix with ten replicates for each K value and used the Evanno method to determine the most likely K value. We assigned admixed populations to the NGSadmix group from which more than 50% of its admixture was drawn. Finally, we calculated F_{ST} values for all possible pairwise population comparisons using the realSFS function of ANGSD.

Integration of Genomic and Phenotypic Datasets

We used Mantel and partial Mantel tests to investigate the relationship between genetic, geographic and phenotypic distance for each major cluster in our study. Some concerns have been raised over the use of Mantel tests in population genetics, especially in regards to inflated type I error rates for partial Mantel tests when spatial autocorrelation is present (Legendre and Fortin 2010; Diniz-Filho et al. 2013; Guillot and Rousset 2013). Therefore, we rely most heavily on pairwise comparisons, and we

are cautious in our interpretation of partial Mantel results. In addition, we focus on comparisons across major genetic groupings, where any potential biases should be comparable. Although we are conservative throughout about linking pattern to process, our results highlight complexes with interesting spatial patterns of genetic and phenotypic variation.

Our regression analysis was performed on 86 individuals and excluded the two samples indicated above. To generate our geographic distance matrix, we used the Geographic Distance Matrix Generator (Ersts 2018). To generate our genetic distance matrix, we used ngsDist (Vieira et al. 2016) to estimate pairwise genetic distances using genotype posterior probabilities. To quantify phenotypic distances among individuals, we generated three separate distance matrices: a dorsal coloration distance matrix, a side pattern distance matrix, and a ventral pattern distance matrix. To generate our dorsal coloration distance matrix, we first standardized digital photographs, extracted and averaged RGB values from regions of interest, and transformed RGB values to a two-dimensional color space as described above. We then calculated the Euclidean distance between points in this conceptual color space to generate measures of pairwise distance in dorsal coloration. To generate our side and ventral pattern distance matrices, we used differences in luminance (with R as the luminance channel) to characterize the pairwise distances in pattern among individuals. We designated R as the luminance channel because many vertebrates are believed to use long-wavelength sensitive (LWS) cones to detect achromatic signals (Endler and Mielke 2005; Osorio and Vorobyev 2005). After standardizing digital photographs, selecting regions of interest, and scaling all pictures as described above, we calculated the number of pixels that fell into each of 95 luminance bins ranging from 0% to 100% reflectance for each frog's surfaces separately (ventral and side). Next, we used the toolbox's Luminance Distribution Difference Calculator to compare the luminance distribution histograms among each pair of frogs and to generate pairwise measures of difference in luminance distribution, which we used as our measure of variation in ventral and side pattern. This methodology follows the recommendation of the toolbox's user guide because pattern differences in this study system are characterized by discrete patches of high and low luminance values. All Mantel and partial Mantel tests were performed in R version 3.4.3 using the vegan package (R core team 2017), and each test used Pearson's method of correlation and performed 999 permutations.

RESULTS

Our sequencing efforts yielded a high-quality dataset. After filtering reads for intact barcode sequences and restriction enzyme sites, the number of reads per sample

ranged from 0.86 million to 13.8 million, with an average of 4.48 million reads per sample. After self-blasting, we obtained an average of 43,559 RAD loci per individual, with the number of loci ranging from 15,491 to 59,488 per sample. The final pseudo-reference genome, which was generated by clustering loci across all samples and included those loci shared by at least 60% of the samples, contained 35,113 loci. The average coverage per individual was 5.4X and ranged from 1.8X to 13.9X. After raw variant filtering, we retained 2,284,942 sites derived from 21,733 loci for ANGSD analyses. Above 65% of the samples in our dataset have at least 3X coverage at these loci.

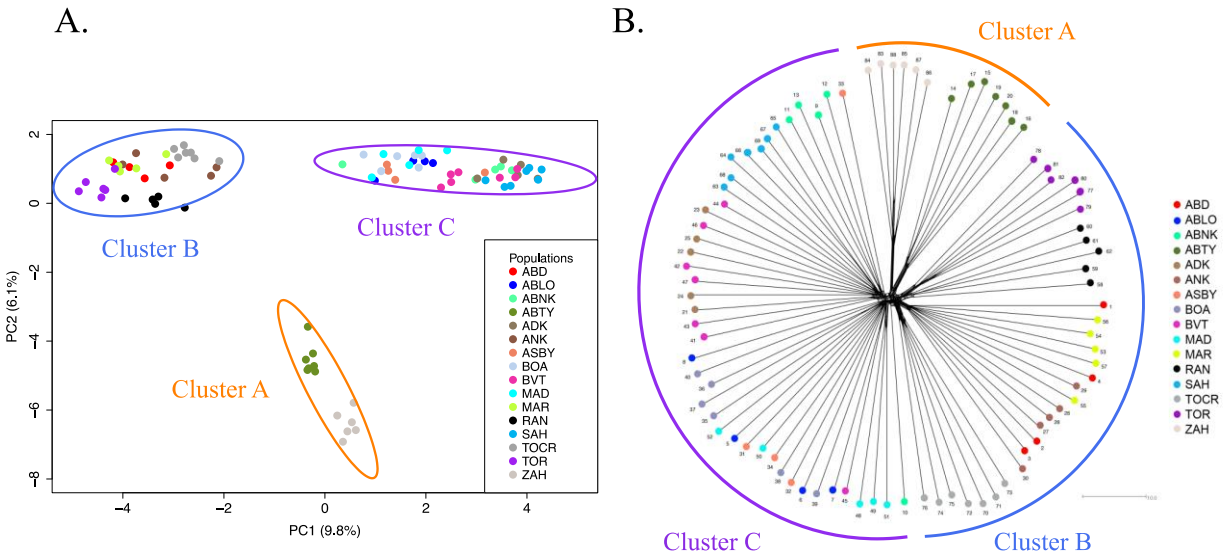

Figure 2: Patterns of genetic variation among individuals from all sampling localities. Plots show PCA (panel *A*) and Splitstree diagram (panel *B*) for all individuals, with colors denoting sampling localities. Ellipses on the PCA plot represent the 95% confidence intervals around the mean of each group identified in our NGSadmix analysis (Clusters A, B, and C). Semi-circles on Splitstree diagram denote groups identified in NGSadmix analysis and again correspond to Clusters A, B and C. NeighborNet analysis in panel *B* was based on *p*-distances in Splitstree.

We found that *M. aurantiaca, M. crocea* and *M. milotympanum* populations were highly structured with three distinct groups emerging. Our PCA, based on 2,284,942 sites, revealed three main clusters: Cluster A contained the two most geographically remote *M. crocea* populations, Cluster B contained the remaining *M. crocea* populations and the two *M. aurantiaca* populations, and Cluster C contained all *M. milotympanum* and all *M. cf. milotympanum* populations (Fig. 2*A*). These three clusters were supported by our NGSadmix analysis, which also identified three groups (*K*=3 based on Evanno et al. 2005 method) corresponding to the same population groupings (Fig. 3). Our splitsTree analysis, based on 14,367 high quality SNPs, further supported this general pattern of genetic variation partitioning (Fig. 2*B*), though there were notable

differences in the degree of genetic admixture present within clusters. The admixture detected in our NGSadmix and splitsTree analysis was concordant with our pairwise population F_{ST} value comparisons, which were lower among admixed populations (Fig. 4). Still, F_{ST} values were relatively high for all pairwise population comparisons, including those that were admixed. It is important to note that geographic and genetic distance were not equivalent across major genetic clusters. While populations within some clusters were both geographically and genetically disparate (as in Cluster A), other populations were geographically more proximate with varying degrees of genetic distance between them (as in Clusters B and C).

Within Cluster A, both our PCA and NGSadmix analysis indicated that each population was a genetically distinct entity. Our NGSadmix analysis identified two groups ($K = 2$ based on Evanno et al. 2005 method) corresponding to each population (Fig. 3). These populations were also clearly differentiated in our splitsTree analysis, with no admixture occurring between them (Fig. 2*B*). In fact, the pairwise F_{ST} value between these two populations (0.43) was among the highest in our dataset (Fig. 4). Within this cluster, Mantel and partial Mantel tests confirmed that genetic distance was most highly correlated with geographic distance ($r = 0.9444$; p-value = 0.001; Table 1). Although dorsal coloration and side pattern were also correlated with genetic variation, neither phenotypic trait remained significantly associated with genetic distance after accounting for the effect of geographic distance in partial Mantel tests (side pattern: $r = -0.1835$; p-value = 0.994; dorsal coloration: $r = 0.1033$, p-value = 0.129). Additionally, the degree of phenotypic variability was much lower in this cluster than in the others, as side pattern was the only phenotypic trait that was significantly correlated with geographic distance when accounting for genetic distance ($r = 0.3006$; p-value = 0.008). Phenotypically, populations within this cluster were relatively uniform, and there was little variation in either dorsal coloration or in side and ventral pattern (Fig. 5).

Within Cluster B, we also found evidence for two distinct groups based on our PCA and NGSadmix analysis. Our NGSadmix analysis identified two groups ($K = 2$ based on Evanno et al. 2005 method), one containing the two *M. aurantiaca* populations and one containing the four remaining *M. crocea* populations (Fig. 3), though there was some admixture between one *M. crocea* and one *M. aurantiaca* population. Our splitsTree analysis supported this general pattern and also showed admixture among the three geographically closest *M. crocea* populations (Fig. 2*B*). Pairwise F_{ST} values among these three admixed *M. crocea* populations ranged from 0.17 - 0.18, which is low in comparison to the range of values in our dataset (Fig. 4). The F_{ST} values between any of these three admixed *M. crocea* populations and the fourth remaining *M. crocea* population ranged from 0.28 – 0.29 (Fig. 4). Comparatively, values between either *M. aurantiaca* population and *M. crocea* populations ranged from

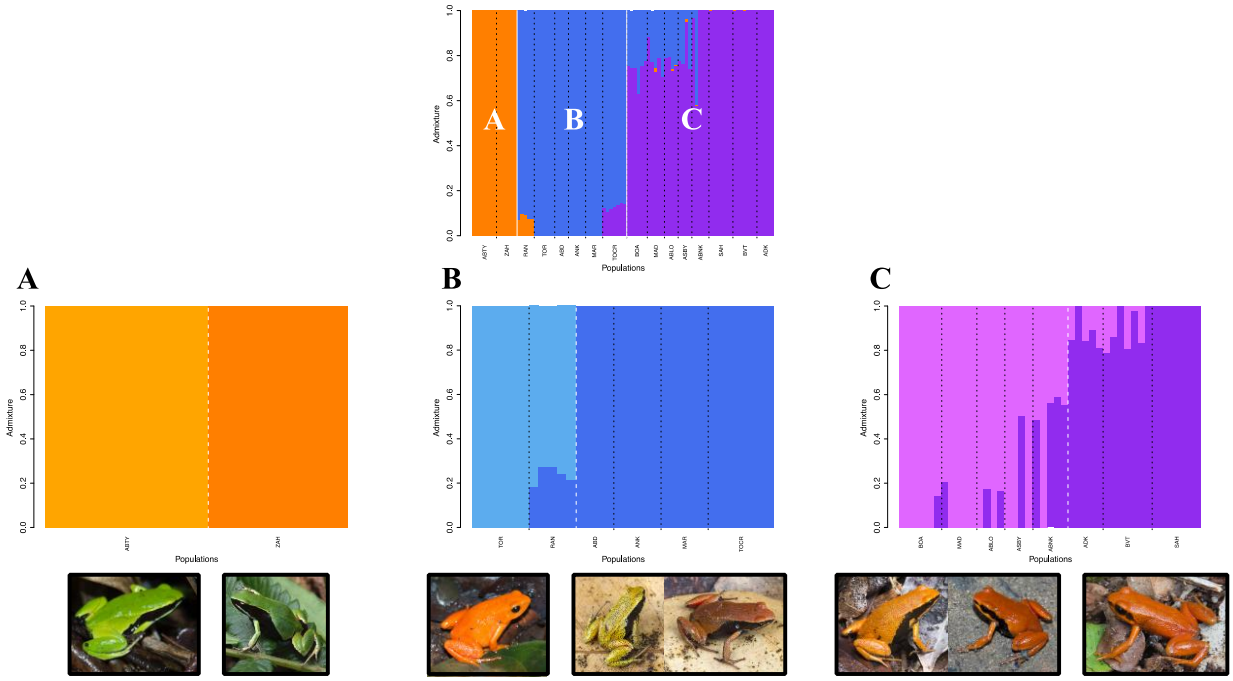

Figure 3: Population structure determined by NGSadmix analysis. The top plot shows NGSadmix clustering for *K*=3 (the most likely *K*) when all samples are included. Dashed black lines separate geographic sampling locations, while solid white lines separate groups identified in NGSadmix analysis (designated here as Clusters A, B, and C). NGSadmix sub-plots show fine-scale population structure within each major genetic cluster (A, B, and C) for *K*=2 (determined to be the most likely *K* for each cluster). Within each sub-plot, black dashed lines separate geographic sampling locations, while dashed white lines separate groups identified in NGSadmix analysis. Below each sub-plot are pictures of representative individuals from the population or populations contained within each NGSadmix group.

0.30 – 0.35 (Fig. 4). Within this group, Mantel and partial Mantel tests confirmed that genetic distance was most highly correlated with variation in dorsal coloration (r = 0.3715; p-value = 0.001; Table 1). Even after accounting for the effects of geographic distance in a partial Mantel test, dorsal coloration remained significantly associated with genetic variation (r = 0.2398; p-value = 0.002). Variation in ventral patterning also remained significantly correlated with genetic distance after controlling for geographic distance (r = 0.2152; p-value = 0.011). All three quantified phenotypic traits were significantly correlated with geography after controlling for genetic distance (dorsal coloration: r = 0.4589; p-value = 0.001; side pattern: r = 0.5174; p-value = 0.001; ventral pattern: r = 0.437; p-value = 0.001). Phenotypic discrepancies between the *M. crocea* and *M. aurantiaca* groups were most evident in dorsal coloration, particularly in dorsal chroma and hue (Fig. 5). Although genetic distance was correlated with variation in side pattern, ventral pattern, and geographic distance, none of these correlations remained significant when accounting for the effects of dorsal coloration, the variable with the highest r value in Mantel tests.

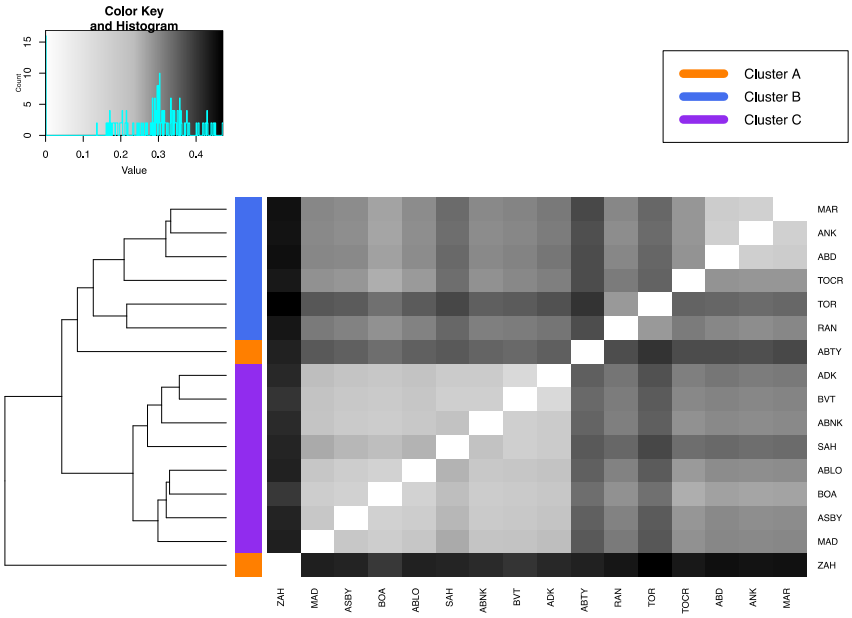

Figure 4: Dendrogram and heat map of pairwise F_{st} values for all sampling localities. Lighter cells denote lower pairwise F_{st} values (less genetic differentiation), while darker cells denote higher pairwise F_{st} values (more genetic differentiation).

Within Cluster C, both our PCA analysis and our NGSadmix analysis indicated two distinct groups. Our NGSadmix analysis revealed two groups ($K = 2$ based on Evanno et al. 2005 method), one containing all *M. milotympanum* populations and one containing all *M. cf. milotympanum* populations (Fig. 3). While our splitsTree analysis supported these two main groupings, there was a high degree of genetic admixture among populations in this cluster (Fig. 2*B*) and populations were less well-defined. The F_{ST} values between populations in this cluster were the lowest out of all pairwise population comparisons in this dataset, ranging from 0.14 – 0.25 (Fig. 4). Despite the high degree of phenotypic variability in populations of this cluster, genetic variation was not correlated with any of the phenotypic traits that we quantified (Table 1). Genetic distance was only significantly correlated with geographic distance ($r = 0.1038$; p-value = 0.018; Table 1). All three phenotypic traits, however, remained significantly correlated with geographic distance when accounting for genetic distance (dorsal coloration: $r = 0.1353$; p-value = 0.009; side pattern: $r = 0.2067$; p-value = 0.002; ventral pattern: $r = 0.387$; p-value = 0.001). Despite this spatial segregation of phenotypes, we observed an exceptionally high degree of variation within populations in this group (Fig. 5). For side and ventral pattern, in particular, within population variation was highest and discrete groupings were not apparent (Fig. 5). Although there was less within population variation in dorsal chroma and hue, there was still not an obvious phenotypic split between *M. milotympanum* and *M. cf. milotympanum*

populations (Fig. 5). Overall, phenotypes in this group demonstrated much higher degrees of intrapopulational variation with less discrete groups emerging.

Group	X Matrix	Mantel *r*	*p*
Cluster A:			
	Dorsal Coloration	0.3089	0.019*
	Geographic Dist.	0.9444	0.001*†
	Side Pattern	0.3066	0.006*
	Ventral Pattern	0.0529	0.258
Cluster B:			
	Dorsal Coloration	0.3715	0.001*†
	Geographic Dist.	0.3414	0.011*
	Side Pattern	0.2880	0.002*
	Ventral Pattern	0.3467	0.001*
Cluster C:			
	Dorsal Coloration	-0.0202	0.537
	Geographic Dist.	0.1038	0.018*†
	Side Pattern	-0.1280	0.897
	Ventral Pattern	-0.0956	0.876
Candidate hybrids:			
	Dorsal Coloration	-0.0012	0.472
	Geographic Dist.	0.1391	0.050*†
	Side Pattern	-0.0386	0.740
	Ventral Pattern	0.0471	0.153

Table 1: Matrix regression results of variables that correlate with genetic distance within major genetic clusters. "Candidate hybrids" refers to the subset of samples that includes admixed individuals and the clusters from which this admixture was drawn (Clusters B and C). Asterisks denote significant *p*-values. † symbols denote the variable with the highest Mantel *r* statistic.

One of the most interesting findings was evidence of possible introgression between *M. crocea/aurantiaca* and *M. milotympanum* populations. To further investigate this phenomenon, we repeated our NGSadmix and Mantel analysis on the subset of samples that included the admixed populations and the two clusters from which this admixture was drawn (Cluster B and Cluster C). Our NGSadmix results yielded a consistent signature of admixture in all *M. cf. milotympanum* populations that we

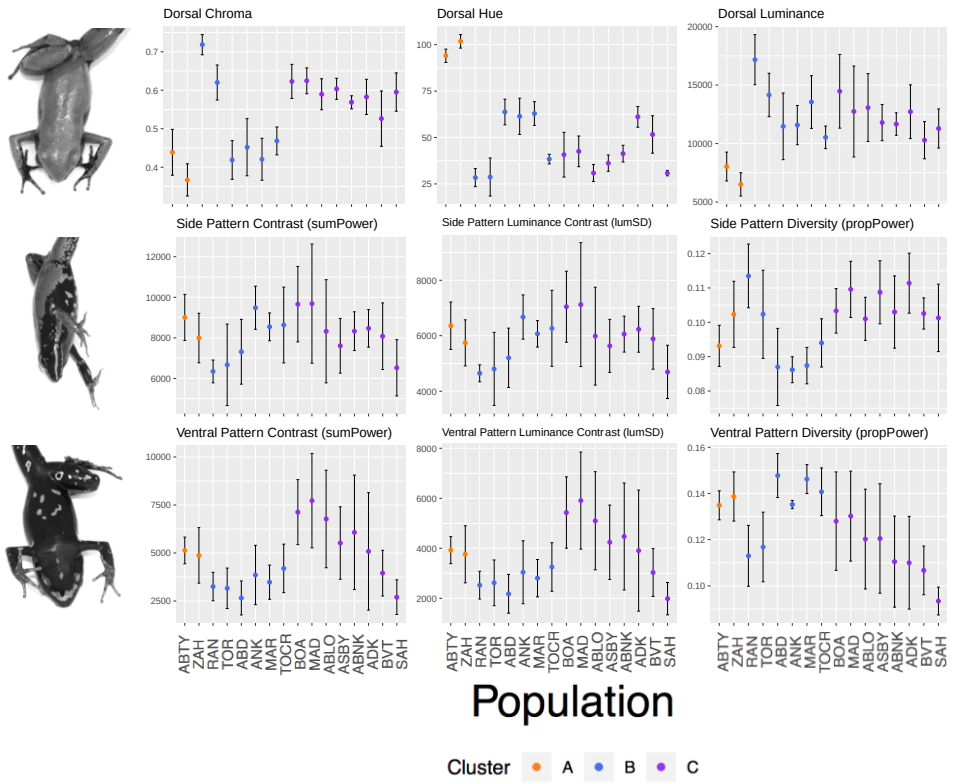

Figure 5: Variation in dorsal coloration (chroma, hue and luminance) and side and ventral pattern (pattern contrast, luminance contrast and pattern diversity) across all sampling localities. Mean values and standard deviation for each population are depicted in the graphs.

sampled (Fig. 6). In the wild, *M. cf. milotympanum* individuals demonstrate a side and ventral pattern that is intermediate between that observed in *M. crocea* and *M. milotympanum* (Fig. 6). Despite the existence of these phenotypically and genotypically intermediate populations, genetic distance was not correlated with any of the phenotypic traits that we quantified, but was significantly correlated with geographic distance ($r = 0.1391$; p-value = 0.050; Table 1). However, variation in dorsal coloration, side pattern, and ventral pattern were all significantly correlated with geographic distance even when controlling for the effects of genetic distance with partial Mantel tests (dorsal coloration: $r = 0.2585$; p-value = 0.001; side pattern: $r = 0.3873$; p-value = 0.001; ventral pattern: $r = 0.3233$; p-value = 0.001). Notably, candidate hybrids, or *M. cf. milotympanum* individuals, demonstrated higher within population variation in patterning characteristics, and also displayed novel pattern phenotypes not present in either "parental" species (Fig. 5, Fig. 6). Within each candidate hybrid population, we observed a spectrum of pattern variants rather than discrete morphs.

DISCUSSION

Our investigation of color diversity and genetic structure in a complex of three closely-related species of Malagasy poison frog revealed that the level of intraspecific color variation in this species complex has likely been overestimated, while the occurrence of distinct species has been underestimated. Although not in alignment with current species designations, we found evidence for several clear genetic clusters, each demonstrating a distinctive pattern of genetic and phenotypic variation, suggesting that a number of mechanisms are contributing to color evolution in this complex. Below we discuss the relevance of our findings for taxonomy, conservation, and the evolutionary processes contributing to phenotypic variation in aposematic organisms. Additionally, we consider the implications of our findings for characterizing morphs in aposematic systems.

Taxonomic Resolution and Implications for Conservation

Populations within this complex were highly structured at relatively small spatial scales, with distinct genetic groups emerging. Unexpectedly, genetic clusters identified in this study do not correspond to current species designations. Specifically, our results indicated that frog populations that have thus far been considered to be a green color variant of the *M. crocea* species form their own genetic cluster (Cluster A) and are in fact distinct from the other *M. crocea* populations included in this study (Fig. 2, Fig. 3). On the basis of this work, we suggest that what has previously been considered to be a green morph of *M. crocea* likely constitutes a new species, and we recommend further investigation of morphology, acoustics and behavior to delineate species boundaries with more certainty. The remaining (non-green) *M. crocea* populations form a distinct cluster with the two *M. aurantiaca* populations (Cluster B; Fig. 2, Fig. 3). Although previous work has found evidence of haplotype sharing between *M. crocea* and *M. aurantiaca* (Chiari et al. 2004; Vences et al. 2004), it is surprising that populations of these two species are identified as a cohesive cluster in our analysis, as they differ greatly in color, pattern and body shape. Our work also demonstrates the validity of the *M. milotympanum* species (Cluster C; Fig. 2, Fig. 3). Previous studies have hypothesized *M. milotympanum* to be a color variant of *M. crocea* (Chiari et al. 2004; Vences et al. 2004), but our results demonstrate that *M. milotympanum* is genetically distinct from the clusters that contain what is currently called *M. crocea* (Clusters A and B).

Finally, we identified candidate hybrid populations (genetically admixed between Clusters B and C) that displayed intermediate genotypes and pattern phenotypes (Fig. 3, Fig. 6). Because earlier studies have hypothesized that *M. milotympanum* and *M. crocea* are conspecific, these intermediates were referred to as *M.*

cf. milotympanum in the literature and assumed to be another phenotypic variant (Chiari et al. 2004; Vences et al. 2004). However, we demonstrate that *M. milotympanum* is its own distinct genetic entity, and likely experienced admixture with *M. crocea* at some point in the past. Although *M. milotympanum* and *M. crocea* do not currently live in sympatry to our knowledge, it is certainly feasible that these species either lived sympatrically in the past or have experienced secondary contact with each other, given the strikingly high degree of deforestation in Madagascar and the highly-fragmented nature of remaining *Mantella* populations.

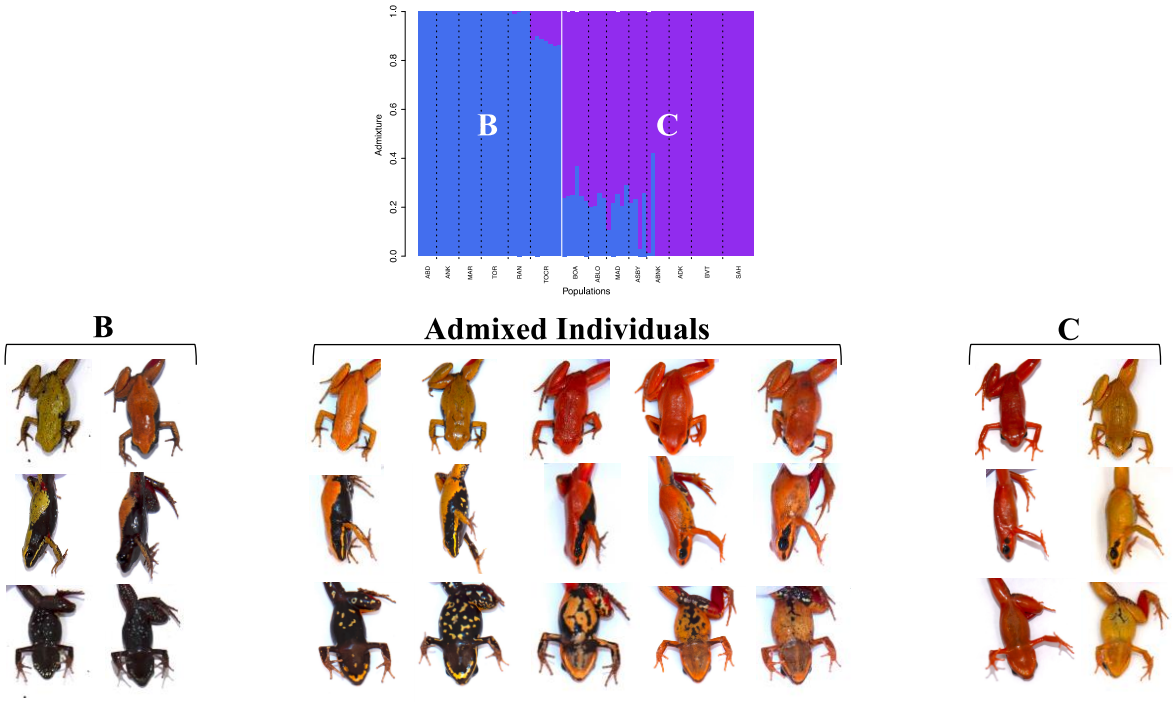

Figure 6: Genetic and phenotypic admixture among candidate hybrid populations. Top plot displays NGSadmix results for the subset of samples from Clusters B and C only. Dashed black lines separate geographic sampling locations, while the solid white line separates groups identified in NGSadmix analysis. Pictures under "B" and "C" headings show representative specimens from populations within each cluster (B or C), while pictures under "Admixed Individuals" show representative variation observed in individuals admixed between Clusters B and C. Individuals that are classified as *Mantella aurantiaca* are not pictured here under Cluster B. All variants displayed under "Admixed Individuals" come from a single population and demonstrate the spectrum of variation that occurs in candidate hybrid populations. Candidate hybrid populations displayed intermediate phenotypes primarily in terms of side and ventral patterning elements.

The taxonomic resolution provided by our study has significant consequences for conservation efforts, given that *M. aurantiaca* and *M. milotympanum* are currently classified as critically endangered, while *M. crocea* is classified as vulnerable (Andreone et al. 2005). Overall, populations were highly genetically structured with substantial phenotypic variation, indicating that there may be a number of distinct units to

consider in management efforts. More importantly, genetic differentiation was not consistently correlated with phenotypic variation, emphasizing the importance of integrating both phenotypic and genetic information in prioritizing units for conservation. Based on our findings, we offer several recommendations for future management efforts. Because *M. milotympanum* represents a distinct species and is not a color morph of *M. crocea*, its status as critically endangered warrants high conservation priority. Future conservation endeavors will need to address the existence of populations admixed between *M. crocea* and *M. milotympanum*, and to determine where these populations fit in a conservation plan. The conservation status of green *M. crocea* populations should be reassessed, as these populations likely constitute a new species rather than a color morph of an already described species. Finally, within each major genetic cluster, our results suggest that there may be two distinct species or subspecies (Fig. 3). We recommend that these sub-clusters be considered as distinct units for Cluster A (the green *M. crocea* cluster, where genetic differentiation is exceptionally high) and Cluster B (the *M. crocea-M. aurantiaca* cluster, where phenotypic differentiation is exceptionally high). Due to the highly isolated and fragmented nature of *Mantella* populations, further intensive survey efforts will be important to identify any additional extant populations.

Overall, while it is premature to delineate new species on the basis of our analyses – especially considering the significant conservation implications in this system – our study confirms that there are at least three genetically distinct groups that do not correspond to current species descriptions. At the least, it seems clear that Cluster A (composed of green "*M. crocea*") should be considered a distinct entity and prioritized given its extremely limited distribution (i.e., known from only two isolated locations). Due to the likely hybridization that we detected between Clusters B and C, and the fine-scale population structure that we observed within each cluster, the status of Clusters B and C is less clear. Rather than revising taxonomy prematurely, we recommend additional studies on gene flow and migration, characteristics of frog calls, and mating behavior before species boundaries can be clarified with any certainty.

Divergent Patterns of Genetic and Phenotypic Diversity

Our findings suggest that a variety of processes at different spatial and genetic scales are likely contributing to differentiation in this Malagasy poison frog complex. Within major genetic clusters, we found evidence of highly regionalized patterns of phenotypic and genetic diversity (Fig. 3). In our regression analysis, while genetic distance was most highly correlated with geographic distance for Clusters A and C, dorsal coloration was most highly correlated with genetic distance within Cluster B (Table 1). We also identified likely *M. crocea-M. milotympanum* hybrid populations (populations genetically admixed between Clusters B and C) that displayed

intermediate genotypes and novel pattern phenotypes (with especially high within population pattern variation; Fig. 3, Fig. 5, Fig. 6). Hybridization has been hypothesized to be an important mechanism of generating novel phenotypes in other poison frog systems (Medina et al. 2013) and may play a similar role here. Based on our findings, patterns of color evolution are likely influenced by variable patterns of drift, selection, and hybridization across major genetic groups in this Malagasy poison frog complex. Regional diversity in patterns of genetic and phenotypic variation has been found in other frog systems, including the red-eyed treefrog *Agalychnis callidryas*, and lends support to the hypothesis that modes of diversification can vary substantially at relatively small spatial scales (Robertson and Zamudio 2009).

While studies of phenotypic variation in Malagasy poison frogs are extremely limited, research on aposematic signal evolution in Neotropical poison frogs suggests that both natural and sexual selection likely contribute to phenotypic diversity (e.g., Reynolds and Fitzpatrick 2007; Noonan and Comeault 2009; Cummings and Crothers 2013; Yang et al. 2016). Previous work in *Oophaga pumilio* has demonstrated that male dorsal coloration is an important component of female mating preferences (Maan and Cummings 2008; Maan and Cummings 2009) and that dorsal brightness is an important signal in male-male competition (Crothers et al. 2011), while also confirming that conspicuous coloration is an honest signal to predators (Maan and Cummings 2011). In fact, studies within this system have even suggested that natural and sexual selection may operate on different aspects of frog coloration, as variation in male brightness, an important component of assortative mating and male-male interactions, is not visible to avian predators (Crothers and Cummings 2013). Additionally, there is growing evidence that pattern traits in poison frogs may also be under selection (Wollenberg et al. 2008; Rojas and Endler 2013; Barnett et al. 2018). Studies of intrapopulation pattern variation in the poison frog *Dendrobates tinctorius* have demonstrated that certain pattern traits are correlated with movement behavior, suggesting that patterning elements may also be important in determining how aposematic organisms are perceived by predators (Rojas et al. 2014a). In fact, recent work in *D. tinctorius* indicates that pattern and color function as an aposematic signal to predators at close range, but are perceived as cryptic when viewed from longer distances (Barnett et al. 2018). Future work on phenotypic diversity in Malagasy poison frogs should draw on this body of literature and consider the relative roles of natural and sexual selection in shaping phenotypes, as well as the relative contributions of color and pattern to predator avoidance.

While Malagasy and Neotropical poison frogs demonstrate interesting parallels with each other, unique characteristics of the *Mantella* system render it particularly interesting from an evolutionary perspective. For example, while studies indicate that birds are particularly important predators of Neotropical poison frogs (Saporito et al. 2007; Maan and Cummings 2011), the only published instances of predation in Malagasy poison frogs were by snakes (*Thamnosophis sp.* in *M. aurantiaca*, and

Acrantophis madagascariensis and *Compsophis laphystius* in *M. laevigata*) and lizards (*Zonosaurus sp.* in *M. aurantiaca* and *Zonosaurus madagascariensis* in *M. laevigata*) (Heying 2001a; Jovanovic et al. 2009; Hutter et al. 2018). Although the extent to which snakes and lizards utilize visual signals – especially in relation to olfactory cues – in predation is not fully understood (Maan and Cummings 2011; Willink et al. 2014), lizards have been found to select for lower conspicuousness in the poison frog *Oophaga granulifera* (Willink et al. 2014). Consequently, there is reason to believe that selective pressures resulting from predation – presumably a major driving force in shaping phenotypic variation in aposematic organisms – may be substantially different for Malagasy poison frogs. Thus, the Malagasy poison frogs represent an important unique and comparative system for developing generality about the selective factors influencing color evolution. To understand the processes generating the interesting and variable patterns of phenotypic diversity, there is a pronounced need for research on the ecology and life history of these frogs. Fundamental research on *Mantella* predation, migration, mating behavior, diet, toxicity, and habitat quantification will be essential in formulating explicit hypotheses regarding color evolution.

Broader Implications for Defining Species and Morphs in Phenotypically Diverse Aposematic Organisms

Our results demonstrate that characterization of morphs based solely on observed phenotypic variation, especially when genetic structure is unresolved, can lead to an overestimation of the degree of polymorphism and/or polytypism that occurs in aposematic systems. Our findings are consistent with other studies in Neotropical poison frogs where sophisticated genetic datasets, utilized either in isolation or paired with morphological and ecological data, have revealed that a single species likely includes several distinct lineages (Roland et al. 2017; Posso-Terranova and Andrés 2018). Thus, overestimation of intraspecific variation and underestimation of species diversity in phenotypically diverse lineages may be more widespread than previously appreciated. In addition to potentially leading to erroneous evolutionary inferences, underestimating species diversity can also have important conservation implications if newly identified lineages are highly restricted in their distribution and/or exist outside of protected areas (Posso-Terranova and Andrés 2018), as is the case for this complex of Malagasy poison frogs.

In this study, although our high-resolution genomic dataset clarified species boundaries at the highest level, the way in which fine-scale population structure is interpreted has significant implications for how hypotheses of color evolution are framed in this system. Within Cluster B, for example, if *M. aurantiaca* and *M. crocea* are considered to be color morphs of the same species, then one interpretation of our results is that differences in dorsal coloration are driving reproductive isolation and

may be a potential mechanism for incipient speciation in this group, as hypothesized in another polymorphic poison frog complex (Wang and Summers 2010). If considered to be separate species, however, then the dramatic phenotypic differences that coincide with genetic variation may serve to reinforce species recognition and to prevent hybridization. Our study demonstrates that even when genetic structure is well understood, distinguishing between species, subspecies and morphs is not always a straightforward process and may require additional information on behavior, acoustics and mating preferences, among other traits.

For aposematic organisms presumed to be polymorphic or polytypic, our results emphasize that more thorough deliberation is necessary not only in the delineation of species, but also in the characterization of morphs. Refining our understanding of what constitutes a morph, and when species or populations are polymorphic and/or polytypic, may be a necessary first step in this regard. Below we highlight three conceptual areas identified on the basis of our findings that merit more careful consideration in the identification and description of phenotypic morphs.

Although phenotypes are multifaceted, morphs are often defined on the basis of one charismatic trait. Consequently, designating variants of one phenotypic axis as morphs is premature and fails to account for variation across multiple traits. In our study, patterns of variation in coloration and pattern traits were not always concordant; interpopulation differences in dorsal coloration were much more discrete than differences in pattern, although both varied among populations (Fig. 5). Within most candidate hybrid populations, dorsal coloration was relatively uniform, but patterning elements varied almost continuously along a spectrum (Fig. 5, Fig. 6). These findings highlight the difficulties of characterizing distinct morphs when there is variation along multiple, potentially correlated phenotypic axes. Similar findings have been reported in other poison frog species, where continuous pattern variation was observed within populations (Rojas & Endler 2013). In such instances, explicitly describing and quantifying variation – whether it is continuous or discrete – in multiple phenotypic traits is essential. We recommend caution in designating discrete morphs when there may be continuous phenotypic variation, especially along multiple phenotypic axes.

In aposematic organisms, designation of morphs is often based on human-observed phenotypic variation. This raises the question: is there really that much diversity in aposematic signals, when considered from the relevant predator and conspecific visual perspectives? In our study, although populations varied in terms of dorsal chroma and hue, dorsal luminance was largely similar across most populations (Fig. 5). Interpreted in light of evidence suggesting that high luminance contrast can serve as an effective warning signal to predators (Prudic et al. 2007), and that dorsal brightness is an important component of conspecific signaling in poison frogs (Maan and Cummings 2009; Crothers et al. 2011), biologically meaningful color diversity in this system may be much less than expected. Although we observed high degrees of

pattern variation on the side and ventral surfaces of frogs in our study (Fig. 5, Fig. 6), the role of this variation is unknown. Recent work in poison frogs linking patterning elements to movement behavior and detectability by predators (Rojas et al. 2014a; Barnett et al. 2018) indicates that pattern variation may be equally, or even more, biologically relevant than color variation. Further, evidence that the detectability of different color pattern variants is influenced by the existing light environment (Rojas et al. 2014b) underscores the importance of incorporating information on both predator sensory abilities and ambient lighting conditions into the characterization of phenotypic variation. Moving forward, it is important to determine which aspects of coloration and/or pattern are perceived by predators and conspecifics. Then, morphs should be defined on the basis on these biologically relevant visual perspectives to accurately capture the degree of diversity in aposematic warning signals.

Finally, our study underscores the need for a reassessment of how morphs are characterized in evolutionary biology. Determining whether color variants represent different species or morphs requires, at the least, sophisticated genomics datasets and an explicit integration of multidimensional phenotypic data. At a broad conceptual scale, our results clearly indicate that levels of intraspecific color variation in this complex of Malagasy poison frogs have been overestimated. At the same time, our results underscore the difficulty in delineating species and morphs at a fine scale in highly complex systems, particularly when genetic and phenotypic breaks are not predictably correlated (Fig. 3). Even with robust genomic data, defining morphs can still be challenging, especially when species boundaries are not straightforward. Polymorphism and polytypism cannot be studied if it is not clear whether discrete morphs exist and whether observed variation is within populations, among populations, or between species.

So, what is a morph? Evolutionary biologists often use the term "morph" without a formal conceptual or operational definition. The research community has thought carefully over decades about how to define and delineate species (e.g., Mayr 1942; Wiley 1978; Mallet 1995; Sites and Marshall 2004; de Queiroz 2007). The same attention has not been paid to the morph concept (but see Teasdale et al. 2013). Not only do we need a more standardized definition of morph, but also operational criteria for delineating morphs when there is phenotypic variation along multiple axes and/or when species boundaries are unclear or dynamic through time.

Our purpose here is not to provide a revised definition of "morph" but rather to highlight the importance of contextualizing phenotypic variation appropriately. When a novel phenotypic variant is discovered, rather than prematurely being described as a new species or morph, it should serve as a launching point for comprehensive studies that integrate phenotypic information with genomic datasets (and, ideally, information on acoustics and behavior). Describing novel phenotypic variants within a species should require – at the least – a high-resolution genetic dataset to confirm intraspecific relationships. In addition, rather than characterizing

human-observed variants as morphs, we recommend that researchers specify the facet of phenotype measured, the sensory perspective used to quantify phenotype, and the level of biological organization where variation is observed. The way we define and delineate morphs is relevant in many systems, but carries particular significance in aposematic organisms, where novel color and pattern variants are regularly found. If we are overestimating intraspecific color variation and underestimating species diversity – as was found in this study – the classification of populations and species as polymorphic or polytypic may need to be re-evaluated. Ultimately, it is time to give as much consideration to the conceptual and operational morph delineation as has been given to species.

ACKNOWLEDGEMENTS

Ke Bi and Erica Bree Rosenblum co-authored this manuscript, which has previously been published in *Ecology and Evolution* (see Klonoski et al. 2019). This work is licensed under a Creative Commons Attribution License (CC BY 4.0) and is reproduced here with slight modifications. We thank Serge Ndriantsoa for his invaluable assistance in the field. We also thank Miguel Vences and Devin Edmonds for sharing their bountiful expertise on Malagasy poison frogs. For logistical support in Madagascar, we thank the Université d'Antananarivo and ICTE/MICET. This work was carried out under a collaboration agreement between MICET, the Département de Biologie Animale of the Université d'Antananarivo, and the Ministère de l'Environnement et des Forêts, Direction du Système des Aires Protégées. We also thank the Direction Générale des Forêts, Ministere de l'Environment et des Forêts, and Madagascar National Parks for granting us research and export permits. We thank the Rosenblum lab for helpful feedback on the manuscript, and Jolyon Troscianko and Alex Krohn for helpful conversations on quantifying color and pattern. Funding was provided by The Mohamed bin Zayed Species Conservation Fund (Project #150511309), the National Geographic Society (Grant Project #9469-14) and the Museum of Vertebrate Zoology at the University of California at Berkeley. All animal work was approved by the Animal Care and Use Committee at the University of California at Berkeley (R347-0314; AUP-2015-01-7083).

LITERATURE CITED

Ali, O. A., S. M. O'Rourke, S. J. Amish, M. H. Meek, G. Luikart, C. Jeffres, and M. R. Miller. 2016. RAD capture (rapture): flexible and efficient sequence-based genotyping. Genetics 202:389-400.

Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. 1990. Basic local alignment search tool. Journal of Molecular Biology 215:403-410.

Andreone, F., J. E. Cadle, N. Cox, F. Glaw, R. A. Nussbaum, C. J. Raxworthy, S. N. Stuart, D. Vallan, and M. Vences. 2005. Species review of amphibian extinction risks in Madagascar: conclusions from the Global Amphibian Assessment. Conservation Biology 19(6):1790-1802.

Andriamaharavo, N. R., H. M. Garraffo, T. F. Spande, L. A. Giddings, D. R. Vieites, M. Vences, and R. A. Saporito. 2015. Individual and Geographic Variation of Skin Alkaloids in Three Swamp-Forest Species of Madagascan Poison Frogs (*Mantella*). Journal of Chemical Ecology 41(9):837-847.

Barbosa, A., L. M. Mäthger, K. C. Buresch, J. Kelly, C. Chubb, C. Chiao, and R. T. Hanlon. 2008. Cuttlefish camouflage: The effects of substrate contrast and size in evoking uniform, mottle or disruptive body patterns. Vision Research 48(10):1242-1253.

Barnett, J. B., C. Michalis, N. E. Scott-Samuel, and I. C. Cuthill. 2018. Distance-dependent defensive coloration in the poison frog *Dendrobates tinctorius*, Dendrobatidae. Proceedings of the National Academy of Sciences 115(25):6416-6421.

Bi, K., T. Linderoth, D. Vanderpool, J. M. Good, R. Nielsen, and C. Moritz. 2013. Unlocking the vault: next generation museum population genomics. Molecular Ecology 22(24):6018-6032.

Bolger, A. M., M. Lohse, and B. Usadel. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30(15):2114-2120.

Bora, P., R. Dolch, R. Jenkins, O. Jovanovic, F. C. Rabemananjara, J. Randrianirina, J. Rafanomezantsoa, L. Raharivololoniaina, O. Ramilijaona, N. Raminosoa, and R. Randrianavelona. 2008. Geographical distribution of three species of Malagasy poison frogs of high conservation priority: *Mantella aurantiaca*, *M. crocea* and *M. milotympanum*. Herpetology Notes 1:39-48.

Briolat, E. S., E. R. Burdfield-Steel, S. C. Paul, K. H. Rönkä, B. M. Seymoure, T. Stankowich, and A. M. M. Stuckert. 2018. Diversity in warning coloration: selective paradox or the norm? Biological Reviews:1-27.

Campbell, F. W., and J. G. Robson. 1968. Application of fourier analysis to the visibility of gratings. The Journal of Physiology 197(3):551-566.

Chiari, Y., M. Vences, D. R. Vieites, F. Rabemananjara, P. Bora, O. Ramilijaona

Ravoahangimalala, and A. Meyer. 2004. New evidence for parallel evolution of colour patterns in Malagasy poison frogs (*Mantella*). Molecular Ecology 13(12):3763-3774.

Clark, V. C., C. J. Raxworthy, V. Rakotomalala, P. Sierwald, and B. L. Fisher. 2005. Convergent evolution of chemical defense in poison frogs and arthropod prey between Madagascar and the Neotropics. Proceedings of the National Academy of Sciences of the United States of America 102(33):11617-11622.

Corl, A., A. R. Davis, S. R. Kuchta, and B. Sinervo. 2010. Selective loss of polymorphic mating types is associated with rapid phenotypic evolution during morphic speciation. Proceedings of the National Academy of Sciences of the USA 107:4254-4259.

Crothers, L. R., and M. E. Cummings. 2013. Warning signal brightness variation: sexual selection may work under the radar of natural selection in populations of a polytypic poison frog. The American Naturalist 181(5):E116-E124.

Crothers, L., E. Gering, and M. Cummings. 2011. Aposematic signal variation predicts male-male interactions in a polymorphic poison frog. Evolution 65(2):599-605.

Cummings, M. E., and L. R. Crothers. 2013. Interacting selection diversifies warning signals in a polytypic frog: an examination with the strawberry poison frog. Evolutionary Ecology 27(4):693-710.

Daly, J. W., N. R. Andriamaharavo, M. Andriantsiferana, and C. W. Myers. 1996. Madagascan poison frogs (*Mantella*) and their skin alkaloids. American Museum Novitates 3177:1-34.

Daly, J. W., H. M. Garraffo, G. S. Hall, and J. F. Cover Jr. 1997. Absence of skin alkaloids in captive-raised Madagascan mantelline frogs (*Mantella*) and sequestration of dietary alkaloids. Toxicon 35(7):1131-1135.

Daly, J. W., H. M. Garraffo, T. F. Spande, L. A. Giddings, R. A. Saporito, D. R. Vieites, and M. Vences. 2008. Individual and geographic variation of skin alkaloids in three species of Madagascan poison frogs (*Mantella*). Journal of Chemical Ecology 34(2):252-279.

de Queiroz, K. 2007. Species concepts and species delimitation. Systematic Biology 56(6):879-886.

Diniz-Filho, J. A. F., T. N. Soares, J. S. Lima, R. Dobrovolski, V. L. Landeiro, M. P. de Campos Telles, T. F. Rangel, and L. M. Bini. 2013. Mantel test in population genetics. Genetics and Molecular Biology 36(4):475-485.

Endler, J. A. 1980. Natural selection on color patterns in *Poecilia reticulata*. Evolution 34(1):76-91.

Endler, J. A. 1991. Variation in the appearance of guppy color patterns to guppies and their predators under different visual conditions. Vision Research 31(3):587-608.

Endler, J. A. 2012. A framework for analysing colour pattern geometry: adjacent colours. Biological Journal of the Linnean Society 107(2):233-253.

Endler, J. A., and J. J. D. Greenwood. 1988. Frequency-dependent predation, crypsis and aposematic coloration. Philosophical Transactions of the Royal Society B 319(1196):505-523.

Endler, J. A., and P. W. Mielke, Jr. 2005. Comparing entire colour patterns as birds see them. Biological Journal of the Linnean Society 86(4):405-431.

Ersts, P. J. 2018. Geographic Distance Matrix Generator(version 1.2.3). American Museum of Natural History, Center for Biodiversity and Conservation. http://biodiversityinformatics.amnh.org/open_source/gdmg.

Estrada, C., and C. D. Jiggins. 2008. Interspecific sexual attraction because of convergence in warning coloration: is there a conflict between natural and sexual selection in mimetic species? Journal of Evolutionary Biology 21(3):749-760.

Evanno, G., S. Regnaut, and J. Goudet. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Molecular Ecology 14(8):2611-2620.

Fu, L., B. Niu, Z. Zhu, S. Wu, and W. Li. 2012. CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics 28(23):3150-3152.

Fumagalli, M., F. G. Vieira, T. Linderoth, and R. Nielsen. 2014. ngsTools: methods for population genetics analyses from next-generation sequencing data. Bioinformatics 30(10):1486-1487.

Galeotti, P., D. Rubolini, P. O. Dunn, and M. Fasola. 2003. Colour polymorphism in birds: causes and functions. Journal of Evolutionary Biology 16(4):635-646.

Glaw, F., and M. Vences. 2007. A field guide to the amphibians and reptiles of Madagascar. 3rd ed. Cologne: Vences and Glaw:496 pages.

Godfrey, D., J. N. Lythgoe, and D. A. Rumball. 1987. Zebra stripes and tiger stripes: the spatial frequency distribution of the pattern compared to that of the background is significant in display and crypsis. Biological Journal of the Linnean Society 32(4):427-433.

Gray, S. M., and J. S. McKinnon. 2007. Linking color polymorphism maintenance and speciation. Trends in Ecology and Evolution 22(2):71-79.

Guilford, T., and M. S. Dawkins. 1993. Receiver psychology and the design of animal signals. Trends in Neurosciences 16(11):430-436.

Guillot, G., and F. Rousset. 2013. Dismantling the Mantel tests. Methods in Ecology and Evolution 4(4):336-344.

Hauswaldt, J. S., A. K. Ludewig, M. Vences, and H. Pröhl. 2011. Widespread co-occurrence of divergent mitochondrial haplotype lineages in a Central American species of poison frog (*Oophaga pumilio*). Journal of Biogeography 38(4):711-726.

Heying, H. E. 2001a. *Mantella laevigata* (climbing Mantella). Aborted predation. Herpetological Review 32:34.

Heying, H. E. 2001b. Social and reproductive behavior in the Madagascan poison frog, *Mantella laevigata*, with comparisons to the dendrobatids. Animal Behaviour 61(3):567-577.
Hoekstra, H. E., K. E. Drumm, and M. W. Nachman. 2004. Ecological genetics of adaptive color polymorphism in pocket mice: geographic variation in selected and neutral genes. Evolution 58(6):1329-1341.
Hoffman, E. A., F. W. Schueler, A. G. Jones, and M. S. Blouin. 2006. An analysis of selection on a colour polymorphism in the northern leopard frog. Molecular Ecology 15(9):2627-2641.
Hugall, A. F., and D. Stuart-Fox. 2012. Accelerated speciation in colour-polymorphic birds. Nature 485(7400):631-634.
Huson, D. H. 1998. SplitsTree: analyzing and visualizing evolutionary data. Bioinformatics 14(1):68-73.
Huson, D. H., and D. Bryant. 2006. Application of phylogenetic networks in evolutionary studies. Molecular Biology and Evolution 23(2):254-267.
Hutter, C. R., Z. F. Andriampenomanana, J. H. Razafindraibe, A. Rakotoarison, and M. D. Scherz. 2018. New dietary data from *Compsophis* and *Alluaudina* species (Squamata: Lamprophiidae: Pseudoxyrhophiinae), and implications for their dietary complexity and evolution. Journal of Natural History 52(39-40):2497-2510.
Huxley, J. 1955. Morphism in birds. In A. Portmann and E. Sutter, ed. Acta XI Congressus Internationalis Ornithologicus. Basel, Switzerland:309-328.
Jiggins, C. D., R. E. Naisbit, R. L. Coe, and J. Mallet. 2001. Reproductive isolation caused by colour pattern mimicry. Nature 411(6835):302-305.
Johnson, P. L., and M. Slatkin. 2008. Accounting for bias from sequencing error in population genetic estimates. Molecular Biology and Evolution 25(1):199-206.
Jombart, T. 2008. adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics 24(11):1403-1405.
Jovanovic, O., M. Vences, G. Safarek, F. C. E. Rabemananjara, and R. Dolch. 2009. Predation upon *Mantella aurantiaca* in the Torotorofotsy wetlands, central-eastern Madagascar. Herpetology Notes 2:95-97.
Kingston, J. J., G. G. Rosenthal, and M. J. Ryan. 2003. The role of sexual selection in maintaining a colour polymorphism in the pygmy swordtail, *Xiphophorus pygmaeus*. Animal Behaviour 65(4):735-743.
Korneliussen, T. S., A. Albrechtsen, and R. Nielsen. 2014. ANGSD: analysis of next generation sequencing data. BMC Bioinformatics 15(1):356.
Krohn, A. R., and E. B. Rosenblum. 2016. Geographic color variation and physiological color change in eastern collared lizards (*Crotaphytus collaris*) from Southern New Mexico, USA. Herpetologica 72(4):318-323.

Legendre, P., and M. J. Fortin. 2010. Comparison of the Mantel test and alternative approaches for detecting complex multivariate relationships in the spatial analysis of genetic data. Molecular Ecology Resources 10(5):831-844.

Li, W., and A. Godzik. 2006. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics 22(13):1658-1659.

Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin, and 1000 Genome Project Data Processing Subgroup. 2009. The sequence alignment/map format and SAMtools. Bioinformatics 25(16):2078-2079.

Lynch, M. 2008. Estimation of nucleotide diversity, disequilibrium coefficients, and mutation rates from high-coverage genome-sequencing projects. Molecular Biology and Evolution 25(11):2409-2419.

Maan, M. E., and M. E. Cummings. 2008. Female preferences for aposematic signal components in a polymorphic poison frog. Evolution 62(9):2334-2345.

Maan, M. E., and M. E. Cummings. 2009. Sexual dimorphism and directional sexual selection on aposematic signals in a poison frog. Proceedings of the National Academy of Sciences 106(45):19072-19077.

Maan, M. E., and M. E. Cummings. 2011. Poison frog colors are honest signals of toxicity, particularly for bird predators. The American Naturalist 179(1):E1-E14.

Maan, M. E., B. Eshuis, M. P. Haesler, M. V. Schneider, J. J. M. van Alphen, and O. Seehausen. 2008. Color polymorphism and predation in a Lake Victoria cichlid fish. Copeia 2008(3):621-629.

Mallet, J. 1995. A species definition for the modern synthesis. Trends in Ecology and Evolution 10(7):294-299.

Mallet, J. and N. H. Barton. 1989. Strong natural selection in a warning-color hybrid zone. Evolution 43(2):421-431.

Mallet, J., and M. Joron. 1999. Evolution of diversity in warning color and mimicry: polymorphisms, shifting balance, and speciation. Annual Review of Ecology and Systematics 30(1):201-233.

Martin, M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal 17(1):10-12.

Mayr, E. 1942. Systematics and the origin of species. New York: Columbia University Press:334 pages.

McKay, B. D. 2013. The use of digital photography in systematics. Biological Journal of the Linnean Society 110(1):1-13.

McKenna, A., M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, A. Kernytsky, K. Garimella, D. Altshuler, S. Gabriel, M. Daly, and M. A. DePristo. 2010. The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Research 20:1297-1303.

Medina, I., I. J. Wang, C. Salazar, and A. Amézquita. 2013. Hybridization promotes color polymorphism in the aposematic harlequin poison frog, *Oophaga histrionica*. Ecology and Evolution 3(13):4388-4400.
Müller, F. 1879. *Ituna* and *Thyridia*; a remarkable case of mimicry in butterflies. Transactions of the Entomological Society of London 1879:xx-xxix.
Nokelainen, O., R. H. Hegna, J. H. Reudler, C. Lindstedt, and J. Mappes. 2012. Trade-off between warning signal efficacy and mating success in the wood tiger moth. Proceedings of the Royal Society B: Biological Sciences 279(1727):257-265.
Noonan, B. P., and A. A. Comeault. 2009. The role of predator selection on polymorphic aposematic poison frogs. Biology Letters 5(1):51-54.
Osorio, D. and M. Vorobyev. 2005. Photoreceptor spectral sensitivities in terrestrial animals: adaptations for luminance and colour vision. Proceedings of the Royal Society B: Biological Sciences 272(1574):1745-1752.
Oxford, G. S. 2005. Genetic drift within a protected polymorphism: enigmatic variation in color-morph frequencies in the candy-stripe sider, *Enoplognatha ovata*. Evolution 59(10):2170-2184.
Pérez-Rodríguez, L., R. Jovani, and M. Stevens. 2017. Shape matters: animal colour patterns as signals of individual quality. Proceedings of the Royal Society B: Biological Sciences 284(1849):20162446.
Perry, G., M. C. Wallace, D. Perry, H. Curzer, and P. Muhlberger. 2011. Toe clipping of amphibians and reptiles: science, ethics, and the law. Journal of Herpetology 45(4):547-555.
Posso-Terranova, A. and J. A. Andrés. 2018. Multivariate species boundaries and conservation of harlequin poison frogs. Molecular Ecology 27(17):3432-3451.
Poulton, E. B. 1890. The Colours of Animals: Their Meaning and Use, Especially Considered in the Case of Insects. Paul, Trench, Trübner & Co. Ltd., London.
Prudic, K. L., A. K. Skemp, and D. R. Papaj. 2007. Aposematic coloration, luminance contrast, and benefits of conspicuousness. Behavioral Ecology 18(1):41-46.
R Core Team. 2017. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.
Reynolds, R. G., and B. M. Fitzpatrick. 2007. Assortative mating in poison-dart frogs based on an ecologically important trait. Evolution 61(9):2253-2259.
Robertson, J. M., and K. R. Zamudio. 2009. Genetic diversification, vicariance, and selection in a polytypic frog. Journal of Heredity 100(6):715-731.
Rojas, B. 2016. Behavioural, ecological, and evolutionary aspects of diversity in frog colour patterns. Biological Reviews 92(2):1059-1080.
Rojas, B., J. Devillechabrolle, and J. A. Endler. 2014a. Paradox lost: variable colour-pattern geometry is associated with differences in movement in aposematic frogs. Biology Letters 10(6):20140193.

Rojas, B., and J. A. Endler. 2013. Sexual dimorphism and intra-populational colour pattern variation in the aposematic frog *Dendrobates tinctorius*. Evolutionary Ecology 27(4):739-753.

Rojas, B., P. Rautiala, and J. Mappes. 2014b. Differential detectability of polymorphic warning signals under varying light environments. Behavioural Processes 109:164-172.

Roland, A. B., J. C. Santos, B. C. Carriker, S. N. Caty, E. E. Tapia, L. A. Coloma, and L. A. O'Connell. 2017. Radiation of the polymorphic Little Devil poison frog (*Oophaga sylvatica*) in Ecuador. Ecology and Evolution 7(22):9750-9762.

Rosenblum, E. B., and L. J. Harmon. 2011. "Same same but different": replicated ecological speciation at White Sands. Evolution 65(4):946-960.

Sandoval, C. P. 1994. Differential visual predation on morphs of *Timema cristinae* (Phasmatodeae:Timemidae) and its consequences for host range. Biological Journal of the Linnean Society 52(4):341-356.

Saporito, R. A., R. Zuercher, M. Roberts, K. G. Gerow, and M. A. Donnelly. 2007. Experimental evidence for aposematism in the Dendrobatid poison frog *Oophaga pumilio*. Copeia 2007(4):1006-1011.

Schaefer, H. C., M. Vences, and M. Veith. 2002. Molecular phylogeny of Malagasy poison frogs, genus *Mantella* (Anura: Mantellidae): homoplastic evolution of colour pattern in aposematic amphibians. Organisms Diversity & Evolution 2(2):97-105.

Schneider, C. A., W. S. Rasband, and K. W. Eliceiri. 2012. NIH Image to ImageJ: 25 years of image analysis. Nature Methods 9(7):671-675.

Sinervo, B., C. Bleay, and C. Adamopoulou. 2001. Social causes of correlational selection and the resolution of a heritable throat color polymorphism in a lizard. Evolution 55(10):2040-2052.

Sinervo, B., and C. M. Lively. 1996. The rock-paper-scissors game and the evolution of alternative male strategies. Nature 380(6571):240-243.

Sites, J. W., Jr., and J. C. Marshall. 2004. Operational criteria for delimiting species. Annual Review of Ecology, Evolution, and Systematics 35:199-227.

Skotte, L., T. S. Korneliussen, and A. Albrechtsen. 2013. Estimating individual admixture proportions from next generation sequencing data. Genetics 195:693-702.

Smit, A. F. A., R. Hubley, and P. Green. 2014. RepeatMasker Open-4.0. http://www.repeatmasker.org.

Stevens, M., C. A. Párraga, I. C. Cuthill, J. C. Partridge, and T. S. Troscianko. 2007. Using digital photography to study animal coloration. Biological Journal of the Linnean Society 90(2):211-237.

Stevens, M., and G. D. Ruxton. 2012. Linking the evolution and form of warning coloration in nature. Proceedings of the Royal Society B: Biological Sciences 279(1728):417-426.

Stoddard, M. C., and M. Stevens. 2010. Pattern mimicry of host eggs by the common cuckoo, as seen through a bird's eye. Proceedings of the Royal Society B: Biological Sciences 277:1387-1393.

Summers, K., R. Symula, M. Clough, and T. Cronin. 1999. Visual mate choice in poison frogs. Proceedings of the Royal Society B: Biological Sciences 266(1434):2141-2145.

Tarvin, R. D., E. A. Powell, J. C. Santos, S. R. Ron, and D. C. Cannatella. 2017. The birth of aposematism: high phenotypic divergence and low genetic diversity in a young clade of poison frogs. Molecular Phylogenetics and Evolution 109:283-295.

Teasdale, L. C., M. Stevens, and D. Stuart-Fox. 2013. Discrete colour polymorphism in the tawny dragon lizard (*Ctenophorus decresii*) and differences in signal conspicuousness among morphs. Journal of Evolutionary Biology 26(5):1035-1046.

Troscianko, J., and M. Stevens. 2015. Image calibration and analysis toolbox – a free software suite for objectively measuring reflectance, colour and pattern. Methods in Ecology and Evolution 6(11):1320-1331.

Twomey, E., J. Yeager, J. L. Brown, V. Morales, M. Cummings, and K. Summers. 2013. Phenotypic and genetic divergence among poison frog populations in a mimetic radiation. PLoS ONE 8(2):e55443.

Vences, M., Y. Chiari, L. Raharivololoniaina, and A. Meyer. 2004. High mitochondrial diversity within and among populations of Malagasy poison frogs. Molecular Phylogenetics and Evolution 30(2):295-307.

Vences, M., A. Hille, and F. Glaw. 1998. Allozyme differentiation in the genus *Mantella* (Amphibia: Anura: Mantellinae). Folia Zoologica 47(4):261-274.

Vieira, F. G., F. Lassalle, T. S. Korneliussen, and M. Fumagalli. 2016. Improving the estimation of genetic distances from Next-Generation Sequencing data. Biological Journal of the Linnean Society 117(1):139-149.

Wang, I. J., and K. Summers. 2010. Genetic structure is correlated with phenotypic divergence rather than geographic isolation in the highly polymorphic strawberry poison-dart frog. Molecular Ecology 19(3):447-458.

Wiley, E. O. 1978. The evolutionary species concept reconsidered. Systematic Biology 27(1):17-26.

Willink, B., A. García-Rodríguez, F. Bolaños, and H. Pröhl. 2014. The interplay between multiple predators and prey colour divergence. Biological Journal of the Linnean Society 113(2):580-589.

Wollenberg, K. C., S. Lötters, C. Mora-Ferrer, and M. Veith. 2008. Disentangling composite colour patterns in a poison frog species. Biological Journal of the Linnean Society 93(3):433-444.

Yang, Y., C. L. Richards-Zawacki, A. Devar, and M. B. Dugas. 2016. Poison frog color morphs express assortative mate preferences in allopatry but not sympatry. Evolution 70(12):2778-2788.

2

Aposematic coloration does not vary with alkaloid-based chemical defenses in Malagasy poison frogs (genus *Mantella*)

ABSTRACT

Chemical defenses represent a fundamental survival strategy that is found in a wide diversity of taxa. The evolution of chemical defenses is particularly compelling in aposematic organisms, where toxic chemical compounds are presumed to function in tandem with bright warning coloration to deter predation, though in reality this relationship is often poorly understood. To clarify the relationship between phenotypic and chemical components of an aposematic warning signal, we characterized skin alkaloid composition and frog coloration and pattern across populations of a complex of three closely-related species of Malagasy poison frog (currently designated *Mantella aurantiaca*, *Mantella crocea*, and *Mantella milotympanum*). Overall, we detected very few correlations between phenotypic diversity (i.e., frog color, pattern, or conspicuousness) and variation in alkaloid-based chemical defenses. Further, differences in alkaloid diversity or abundance were not correlated with differences in frog conspicuousness. Geographic distance, however, was consistently correlated with alkaloid variation. Our results highlight the importance of geography in structuring chemical variation among populations, while also casting doubt on the quantitative honesty of aposematic signals in this group. Ultimately, our work underscores the need to investigate both chemical and phenotypic diversity from relevant perspectives (i.e., natural predators) in order to better understand the evolutionary ecology of chemical defenses in aposematic organisms. Greater

understanding of phenotypic and functional variation in aposematic traits has important evolutionary and conservation implications for these endangered frogs.

INTRODUCTION

Chemical defenses are widespread across diverse organisms and represent a fundamental strategy for survival. Numerous plant, fungal, insect, and vertebrate species employ chemical defenses to deter predation or herbivory (Speed et al., 2012). In vertebrates, one particularly compelling case of chemical defense is found in brightly colored poison frogs. Poison frogs are characterized by their ability to sequester toxic alkaloid compounds from arthropods consumed as part of their diet and to subsequently store these compounds in glands on their skin (Saporito et al., 2009a; Saporito et al., 2012). The conspicuous coloration observed in many diurnal poison frog species is presumed to function as an aposematic signal that warns potential predators of their toxicity and/or unpalatability (e.g., Santos et al., 2003; Clark et al., 2005; Santos et al., 2016). Currently, poison frogs are known from five families (Dendrobatidae, Bufonidae, Mantellidae, Myobatrachidae, and Eleutherodactylidae) and are found in the disparate geographic locations of Latin America, Madagascar, Australia, and Cuba (Saporito et al., 2012).

Variation in alkaloid composition among species and populations has been widely documented in many poison frog lineages (e.g., Mebs et al., 2005; Saporito et al., 2006; Daly et al., 2007; Saporito et al., 2007a; Daly et al., 2008; Andriamaharavo et al., 2015). Where different species occur in the same location, there are examples of both similarity (Myers et al., 1995; Clark et al., 2005) and dissimilarity (Daly et al., 2008) in alkaloid composition among species. Within species, alkaloid variation among populations is typically correlated with geographic distance (e.g., Saporito et al., 2006; Saporito et al., 2012). Temporal variation in alkaloid composition has also been observed in some poison frog species (Saporito et al., 2006; Daly et al., 2007), though recent work has suggested that alkaloid profiles are relatively robust against seasonal changes in diet (Moskowitz et al., 2018). Given the presumed dietary acquisition of chemical defenses (Daly et al., 1997; reviewed in Saporito et al., 2009a; Saporito et al., 2012), differences in the availability and composition of arthropod prey over space and/or time is considered to be the major driver of variation in alkaloid composition (e.g., Saporito et al., 2006; Saporito et al., 2007a; Daly et al., 2008). Other factors hypothesized to possibly contribute to alkaloid variation include frog age, differences in sequestration mechanisms among frog species, differences in foraging behavior, variation in prey preference, and sex-based differences in diet (Saporito et al., 2007a; Daly et al., 2008; Saporito et al., 2009b; Saporito et al., 2012; Jeckel et al., 2015).

Although the functional consequences of alkaloid variation in poison frogs remain unknown in most instances, such variation in chemical composition may have important implications for predator-prey dynamics. Recent evidence suggests that differences in alkaloid composition are perceived as differences in palatability by arthropods (Bolton et al., 2017). Further, several studies have detected a relationship between toxicity and conspicuousness across poison frog populations or species (Summers & Clough, 2001; Santos & Cannatella, 2011; Wang, 2011; Maan & Cummings, 2012), and there is evidence that frog dorsal conspicuousness serves as an honest signal of toxicity, at least for certain classes of predators (Maan & Cummings, 2012). However, most detailed studies of alkaloid variation have focused on characterizing toxic compounds found in frog skin and lack explicit integration of phenotypic data, limiting our understanding of the relationship between chemical and phenotypic diversity in aposematic frogs.

The Malagasy poison frogs (genus *Mantella*), endemic to Madagascar, provide an exceptional opportunity to investigate whether chemical composition co-varies with frog coloration, pattern, and/or levels of conspicuousness. The *Mantella* genus describes sixteen species of toxic, diurnal frogs that display variation in both coloration and pattern (Glaw & Vences, 2007). Studies of alkaloid composition in this group have detected well over 400 unique compounds across 11 species (e.g., Daly et al., 1984; Garraffo et al., 1993; Daly et al., 1996; Daly et al., 2008; Andriamaharavo et al., 2015). For one complex of closely-related, endangered *Mantella* species found in central-eastern Madagascar, variation in alkaloid composition has been observed among phenotypically variable populations (Andriamaharavo et al., 2015). Although populations within this complex are currently classified as belonging to one of three distinct species, *M. crocea*, *M. milotympanum*, and *M. aurantiaca*, recent work has demonstrated that current species designations do not align with major genetic clusters identified using genomic analysis and suggests that taxonomic revision may be necessary within this group (Klonoski et al., 2019). Regardless, this system provides an opportunity to investigate patterns in phenotypic and chemical components of an aposematic signal across species and populations.

Although alkaloid variation among *Mantella* populations is hypothesized to be due primarily to differences in prey among habitats and geographic locations (Clark et al., 2006; Daly et al., 2008; Andriamaharavo et al., 2015), the relationship between phenotypic diversity and alkaloid variation has not been studied in this group. In this study, we used gas chromatography-mass spectrometry, in combination with phenotypic analysis and multiple matrix regression analysis, to investigate relationships between geography, phenotype, and alkaloid variation across poison frog populations. Our specific objectives were to (a) characterize alkaloid variation across 17 populations within this complex of closely-related species, (b) quantify the dorsal coloration, side and ventral pattern, and dorsal conspicuousness of frogs, (c) describe

the relationship between alkaloid composition, conspicuousness, phenotypic diversity, and geographic distance at several relevant biological scales, and (d) test whether differences in frog alkaloid diversity or abundance were correlated with differences in conspicuousness. By describing patterns of variation in phenotype and chemical composition, we aim to provide a framework that will be helpful in developing hypotheses regarding the evolutionary ecology of chemical defense in putative aposematic frogs.

MATERIAL AND METHODS

Field Sampling

We sampled three closely-related species, currently classified as *Mantella crocea*, *Mantella aurantiaca*, and *Mantella milotympanum*, throughout a large portion of their known range in central-eastern Madagascar (Fig. 1). We also sampled populations displaying a phenotype that is intermediate between that of *M. crocea* and *M. milotympanum*, referred to as *M. cf. milotympanum* in the scientific literature (Chiari et al., 2004). Recent genomic analysis on this group identified three main genetic clusters (designated Clusters A, B, and C) that do not align with current species designations (Klonoski et al., 2019). Cluster A is composed of populations currently classified as *M. crocea* (displaying green dorsal coloration), Cluster B is composed of populations currently classified as *M. aurantiaca* and *M. crocea* (yellow and brown dorsal coloration), and Cluster C is composed of populations classified as *M. milotympanum* and *M. cf. milotympanum*. Because this recent genomic analysis represents the highest resolution genetic analysis available for this group, we hereafter refer to Cluster A, B, and C as the main genetic groupings within this complex of closely-related species, and conduct many of our analyses at the level of major genetic cluster, as identified in Klonoski et al. 2019.

Overall, we sampled 190 individuals from 17 localities (Cluster A: 30 individuals from 2 populations; Cluster B: 65 individuals from 6 populations; Cluster C: 95 individuals from 9 populations). Fieldwork was conducted during the rainy breeding season over two years: January – February 2015, and November 2015 – January 2016. Frogs were captured by hand. At the time of frog capture, we also collected a piece of the substrate on which the frog was first seen. After transportation back a field basecamp, we collected digital photographs of both frogs and their corresponding substrate samples. We also collected skin secretion samples from all frogs. After data collection, frogs were held overnight to recover and released to their site of capture the next morning. All animal handling procedures were approved by the Animal Care and Use Committee at the University of California

at Berkeley (R347-0314; AUP-2015-01-7083). Collection and exportation of samples were performed under permits issued by the Direction Générale des Forêts, Direction de la Conservation de la Biodiversité et du Système des Aires Protégées, and Ministere de l'Environment, de l'Ecologie et des Forêts in Madagascar (collection permits: 335/14/MEF/SG/DGF/DCB.SAP/SCB, 336/14/MEF/SG/DGF/DCB.SAP/SCB and 296/15/MEEMF/SG/DGF/DAPT/SCBT; export permits: 048C-EA02/MG15 and 002C-EA01/MG16).

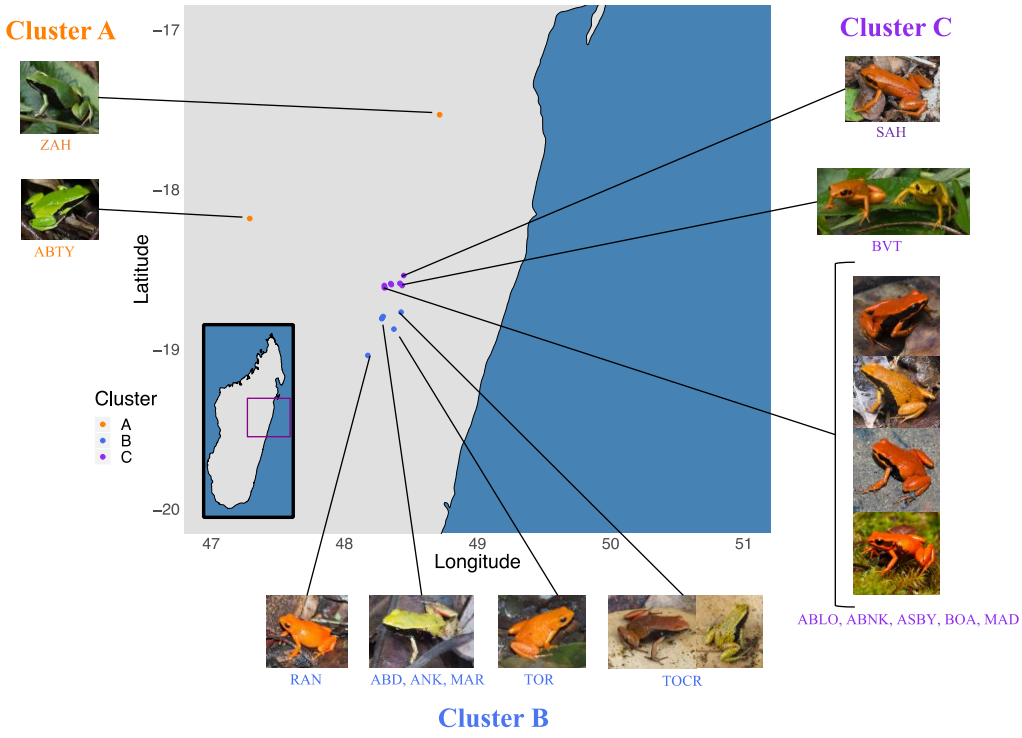

Figure 1: Map of sampling localities in central-eastern Madagascar (figure modified from Klonoski et al. 2019 licensed under CC BY 4.0). Representative individuals from each population or groups of populations are pictured next to labels. Populations are color-coded by major genetic cluster (A, B, or C) as identified in Klonoski et al. 2019.

To characterize frog phenotype and conspicuousness, we photographed substrate samples and frog dorsal, side and ventral surfaces in a standardized manner following a protocol adapted from Stevens et al. (2007). Frogs were photographed after being transported to the field basecamp, but before any other handling occurred. All photographs were taken in natural light between the hours of 1:00 – 5:00pm using a Pentax K-30 digital single-lens reflex camera fitted with a Pentax 18-135 mm lens. All frogs and substrate samples were photographed on a white sheet of paper with the camera flash on and with both a scale bar and a white-gray-black standard present (QPcard 101; gray standard reflectance value of 18%).

To collect skin secretion samples non-lethally, we used a transcutaneous amphibian stimulator (TAS; Grant & Land, 2002), which delivers a mild electrical stimulus to stimulate skin secretions. To obtain skin secretion samples, we first rinsed frogs with water to remove debris. After rinsing, frogs were subjected to two rounds of electrical stimulation via TAS. Each round lasted approximately 20-30 seconds (with on and off electrical stimulation during this period), and voltage was adjusted for each frog until slight muscular contractions were observed. Immediately after each round of electrical stimulation, the entire dorsal, side and ventral surfaces of frogs were wiped with a 1 inch x 1 inch piece of Kimwipe. Frogs were also rinsed with water between rounds of electrical stimulation. Kimwipes were preserved in vials containing 1 mL of 100% methanol. Electrodes of the TAS were also rinsed into vials using methanol to capture any residual skin secretions present on the electrodes. After electrical stimulation was complete, frogs were again rinsed with water and allowed to recover overnight.

Because species within this group are endangered, we used methods appropriate for endangered organisms located in remote field locations (i.e., locations without electricity) by non-lethally collecting alkaloids and using digital photography to quantify coloration and pattern. While it can be challenging to collect data of this nature, we strongly advocate against methods of measuring toxicity or chemical composition that result in frog death, particularly for endangered species.

Quantification of Phenotypic Diversity

To quantify dorsal coloration from digital photographs, we followed the protocol outlined in Klonoski et al. 2019. Briefly, we used the Multispectral Image Calibration and Analysis Toolbox (Troscianko & Stevens, 2015) utilized within ImageJ v1.51 (Schneider et al., 2012) to create aligned and normalized images from our RAW photographs. After standardization, we selected two representative regions on the frog's dorsum in which to quantify coloration: one 3x3 mm square near the back of the frog dorsum, and one 3x1 mm rectangle behind the frog's right arm. Next, we used the Batch Image Analysis function of the toolbox to extract the red, green, and blue color values (RGB values) within selected regions of interest for each frog. R, G, and B values were averaged separately for all pixels within regions of interest. RGB values were subsequently transformed into measurements relevant to vertebrate visual systems using a methodology outlined in Krohn and Rosenblum (2016), modified from Endler (2012) and McKay (2013). Briefly, we calculated three axes from RGB values that corresponded to a red-green channel (R-G)/(R+G), a blue-green channel (G-B)/(G+B), and a luminance channel (untransformed R values). We chose R as the luminance channel because many vertebrates likely use long-wavelength sensitive cones for detection of achromatic cues (Endler & Mielke, 2005; Osorio & Vorobyev, 2005). Because luminance is processed separately in vertebrates

(Endler & Mielke, 2005; Endler, 2012), we used our other two axes to plot dorsal coloration as a point in a two-dimensional color space. To measure the distance in dorsal coloration among individuals, we calculated the Euclidean distances between points in this two-dimensional color space. To measure the distance in dorsal luminance among individuals, we calculated the Euclidean distances between luminance values, which corresponded to the absolute value of the difference between untransformed R values.

To quantify side and ventral pattern, we again used the Multispectral Image Calibration and Analysis Toolbox (Troscianko & Stevens, 2015) in a protocol outlined in detail in Klonoski et al. 2019. After aligning and normalizing RAW images with the toolbox, we manually selected the entire side and ventral surfaces of the frog as regions of interest in order to obtain a comprehensive measure of overall pattern. Next, we used scale bars present in each photograph to scale all images to the same number of pixels per millimeter (side surfaces = 19.2 px/mm; ventral surfaces = 18.6 px/mm). After scaling, we calculated the number of pixels falling into each of 95 luminance bins, ranging from 0% to 100% reflectance, for each frog's side and ventral surfaces separately. Finally, we used the Luminance Distribution Difference Calculator available in the toolbox to generate pairwise distances in luminance distribution by comparing the luminance distribution histograms of each pair of frogs. We used differences in luminance distribution as our measure of variation in side and ventral pattern following the toolbox's recommendation for patterns characterized by discrete patches of high and low luminance values, as is present in the frogs used in this study.

To assess conspicuousness of frogs, we quantified the coloration and luminance of each frog's substrate sample using the methods for quantifying frog dorsal coloration outlined above. For each substrate sample, we manually selected regions of interest with a total area of 25 mm^2 in which to quantify coloration and luminance. Given the diversity of shape and size in substrate samples, we manually selected regions of interest to avoid any areas with glare present in photographs and to compare a standard-sized area across samples. After transforming RGB values of substrate into our three axes as described above, we calculated the Euclidean distance between each frog's dorsal coloration and its substrate's coloration in our two-dimensional color space in order to obtain a measure of each frog's chromatic conspicuousness relative to its background. Similarly, to quantify achromatic (i.e., brightness) conspicuousness, we calculated the Euclidean distance between each frog's luminance axis (i.e., untransformed R values) and its substrate's luminance axis. Pairwise differences in conspicuousness among frogs were generated by calculating the Euclidean distance between chromatic and achromatic conspicuousness values.

Quantification of Chemical Diversity

To characterize alkaloid profiles of frogs, skin secretions collected on Kimwipes were extracted with methanol (4 x 25 μL) and the crude methanolic extracts were used directly for analysis. Extracts were analyzed by GC-MS using a Thermo Trace GC equipped with an AS300 autosampler and interfaced to an iTQ-1100 ion-trap mass spectrometer (Thermo Electron, San Antonio TX). A 5% phenyl-polydimethylsiloxane column (RTX-5MS, Restek, Bellefonte, PA) was used for separation with helium carrier gas at 40 cm/s, with temperature held at 100°C for one minute, followed by a 10°C/minute ramp to 280°C, which was then held for 10 minutes (29 minutes overall run time). Samples (1 μL) were injected splitless at 250°C with a closed time of one minute, followed 50 mL/min split flow and surge pressure of 200 kPa. MS parameters included ionization energy 70 eV with a 250°C source temperature and autotuning with perfluorotributylamine, injection time determined by automatic gain control (max ion time 25 ms), scan rate default 5500 μ/s, 5 microscans averaged to give a total scan time of 0.94 s. A solvent delay of 3 minutes was used and samples were acquired in both electron impact mode at 70 eV and chemical ionization mode with ammonia reagent gas delivered at 1.8 mL/min. This system allowed clean correlation of electron impact (EI) spectra with molecular weight information provided by chemical ionization (CI). Retention times were corrected to those reported (Daly et al., 2005) by injection of known standards and linear regression of reported and observed data. EI and CI spectra were used in concert with corrected retention time to uniquely identify alkaloids. EI provided characteristic fragmentation patterns which were matched against published spectra in the National Institute of Standards and Technology library (NIST08) and the most recent tabulation of amphibian alkaloid mass spectra (Daly et al., 2005) which were manually transcribed into an electronic library within the Xcalibur software. Compounds with correct mass spectra and CI determined mass were considered matches if the retention time was 0.1 minute or less than the reported values (Daly et al., 2005).

After obtaining alkaloid profiles for each frog, we created a pairwise distance matrix by calculating the distance between profiles using a Dice coefficient as in Davis et al. (2016). We also calculated alkaloid diversity values for each frog by counting the number of unique alkaloids present in their skin secretion samples. Different isomers of the same compound were counted as unique alkaloids. We calculated the alkaloid abundance of each frog by integration of the extracted ion chromatogram (EIC) for the highest abundance ion. Using the EIC gives greater selectivity for the alkaloid of interest than the total ion chromatogram (TIC), especially given that some alkaloids (and other components) have close retention times and thus overlapping peaks. Given the difficulty in precisely reproducing TAS stimulated secretion between manually restrained frogs of varying sizes, absolute quantitation is not possible and an internal standard was not used with the GC-MS sampling. In addition, ion yields and

fragmentation probabilities for alkaloid species vary widely in mass spectrometry (Cramer et al., 2017; Bergmann et al., 2018). Thus, quantitative values provided herein should only be used comparatively within our dataset.

To visualize differences in alkaloid composition, we performed non-metric multidimensional scaling (nMDS) based on Bray-Curtis dissimilarity matrices. nMDS has been used frequently in studies of poison frog alkaloids to visualize variation in chemical composition (e.g., Saporito et al., 2007a; Daly et al., 2008; Andriamaharavo et al., 2015). We performed nMDS for (a) all samples combined, (b) for the subset of samples from Clusters B, C, and those admixed between B and C, and (c) within each major genetic cluster. To determine significant differences among alkaloid profiles of different populations and/or genetic clusters, we used an analysis of similarity (ANOSIM). In certain instances, we further partitioned data into smaller subsets in order to assess specific pairwise comparisons of alkaloid composition using ANOSIM (i.e., including just those individuals in Clusters A and B to assess whether alkaloid composition varied significantly between Clusters A and B). All analyses were performing using the vegan package implemented in R version 3.3.3 (R Core Team).

Finally, we assigned each alkaloid to its particular alkaloid class based on existing literature (Daly et al., 2005). Compounds whose structure class could not be determined were designated as "Unclassified". Based on this information, we calculated the proportion of alkaloids belonging to each distinctive alkaloid class within each population.

Integration of Phenotypic and Chemical Datasets

We used multiple matrix regression analysis to investigate relationships between geography, phenotype, and chemical composition. We performed multiple matrix regression analysis across all samples, and within each major genetic cluster, to assess correlations between variation in alkaloid composition and (a) geographic distance, (b) dorsal coloration and luminance, (c) side and ventral pattern, and (d) dorsal chromatic and achromatic conspicuousness. Our methodology follows that outlined in Davis et al. (2016) and allows for simultaneous incorporation of multiple distance matrices into our model. In Step 1 of our analysis, we performed pairwise Mantel tests between the y-matrix (alkaloid variation) and each x-matrix (dorsal coloration, dorsal luminance, geographic distance, side pattern, ventral pattern, chromatic conspicuousness, achromatic conspicuousness). All x-matrices that significantly contributed to variation in alkaloid composition were included in Step 2. In Step 2, we performed partial Mantel tests that accounted for geographic distance, which was the variable with the highest r value in Step 1. All x-matrices that remained significantly correlated with alkaloid variation in Step 2 were included in Step 3. In Step 3, we performed a MultiMantel test which allowed us to include multiple variables in our model. All tests were implemented in R version 3.3.3 using the vegan package (R Core Team).

For our multiple matrix regression analysis, we constructed distance matrices for each variable included in our model. To generate our geographic distance matrix, we used the Geographic Distance Matrix Generator (Ersts, 2018). To generate our dorsal coloration and luminance distance matrices, we calculated the Euclidean distances between (a) points in our two-dimensional color space and (b) dorsal luminance values as described above. To generate our side and ventral pattern distance matrices, we quantified pairwise distances between luminance distribution histograms of frogs as described above. To generate our chromatic and achromatic conspicuousness distance matrices, we calculated the Euclidean distance between conspicuousness values of frogs. To create our alkaloid composition distance matrix, we calculated the distance between alkaloid profiles using a Dice coefficient as described in Davis et al. (2016).

Finally, to investigate relationships between frog conspicuousness and alkaloid diversity and abundance, we used simple linear regression and calculated Pearson's correlation coefficient for each comparison. We assessed relationships between chromatic and achromatic conspicuousness and alkaloid diversity and abundance within each major genetic cluster, and across all samples. We also used simple linear regression to characterize relationships between frog size (SVL) and alkaloid diversity, alkaloid abundance, chromatic conspicuousness, and achromatic conspicuousness across all samples. All regression and correlation analyses were performed in R version 3.3.3 (R Core Team).

RESULTS

Detection of Unique Alkaloids

Using gas chromatography-mass spectrometry, we detected 157 unique compounds across all focal *Mantella* populations. Of these, 51 alkaloids have not been previously described. Within Cluster A (N = 30 individuals from 2 sampling localities), we detected 60 distinct alkaloids across sampled populations, with an average of 14.8 alkaloids detected per frog (standard deviation = 7.7). There was variation at the individual level, with the number of alkaloids per frog ranging from 2 to 36. Within Cluster B (N = 65 individuals from 6 sampling localities), we detected 46 distinct compounds across all sampled locations. The number of alkaloids identified per frog ranged from 1 to 19, with an average of 7.5 (standard deviation = 4.3). Within Cluster C (N = 95 individuals from 9 sampling localities), we detected 103 distinct alkaloids across sampling localities, with an average of 13.6 alkaloids identified per individual (standard deviation = 7.0). The number of alkaloids per frog ranged from 1 to 31.

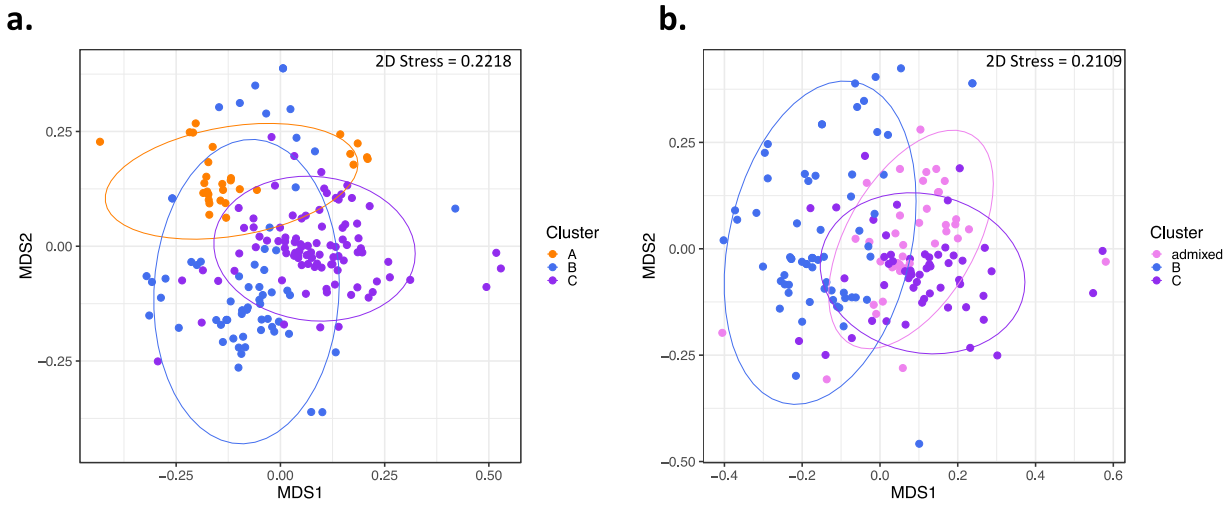

Figure 2: Multidimensional scaling plot of alkaloid composition across all populations of major genetic clusters (A, B, and C) in this complex of closely-related species (panel a) and for the subset of samples including Cluster B, Cluster C, and individuals admixed between Clusters B and C (panel b). Each symbol represents an individual frog, while the distance between points represents the difference in alkaloid composition. Ellipses represent the 95% confidence intervals of each genetic cluster (with admixed populations designated as a separate "cluster" in panel b labeled "admixed"). For panel a, an analysis of similarity (ANOSIM) determined that alkaloid composition is significantly different amongst major genetic clusters (Global $R = 0.4422$; $P = 0.001$). For panel b, ANOSIM determined that alkaloid composition is significantly different amongst Clusters B, C, and admixed populations (Global $R = 0.2734$, $P = 0.001$). Admixed populations correspond to genetically and phenotypically admixed frog populations identified in Klonoski et al. 2019.

Variation in Alkaloid Composition

We detected variation in alkaloid composition among populations at a broad scale (i.e., across all samples; Fig. 2a) and within each major genetic cluster (Fig. 3). When all samples were pooled, alkaloid composition was significantly different amongst the three main genetic clusters according to an analysis of similarity (ANOSIM) (Global $R = 0.4422$, $P = 0.001$; Fig. 2a). ANOSIM performed on subsets of the data revealed that all pairwise combinations of major genetic clusters were significantly different in alkaloid composition from each other (For Clusters A and B: Global $R = 0.4586$, $P = 0.001$; For Cluster A and C: Global $R = 0.5338$, $P = 0.001$; For Clusters B and C: Global $R = 0.3818$, $P = 0.001$).

For the subset of samples that included individuals from Clusters B, C, and those admixed between B and C, we also found that alkaloid composition differed significantly in individuals from Cluster B, those from Cluster C, and those admixed between the two (Global $R = 0.2734$, $P = 0.001$; Fig. 2b). Similarly, all pairwise

combinations of genetic clusters were determined to be significantly different from each other when ANOSIM was performed on subsets of data (For Clusters B and C: Global $R = 0.3818$, $P = 0.001$; For Cluster B and admixed: Global $R = 0.2661$, $P = 0.001$; For Cluster C and admixed: Global $R = 0.1621$, $P = 0.001$). Thus, the variability in chemical composition that we detected here corresponds to genetic and phenotypic admixture that has been previously documented in wild populations (Klonoski et al., 2019).

Within each major genetic cluster, we also detected alkaloid variation among populations, though the degree of overlap in alkaloid composition among populations varied from cluster to cluster (Fig. 3). While Cluster A demonstrated no overlap in alkaloid profiles among populations (Global $R = 0.9595$, $P = 0.001$; Fig. 3), populations within Clusters B and C were less differentiated although their alkaloid composition was still significantly different from each other (For Cluster B: Global $R = 0.6134$, $P = 0.001$; For Cluster C: Global $R = 0.4738$, $P = 0.001$; Fig. 3). Populations also varied in their proportions of alkaloids belonging to distinctive alkaloid classes (Table 1).

Patterns of variation among phenotype, alkaloid composition, and geography

Our matrix regression results consistently identified geography as the variable with the highest explanatory power for variation in alkaloid composition, even at different geographic and biological scales (Table 2). Within each major cluster, our multiple matrix regression analyses indicated that geographic distance was most highly correlated with variation in alkaloid composition (Table 2). When samples were pooled and analyzed across all individuals, geography again remained the variable that best explained variation in chemical composition (slope = 48.1001; p-value = 0.001; Table 2).

In most instances, geographic distance was the only variable significantly correlated with variation in alkaloid composition according to our multiple matrix regression framework. An exception to this result was observed in Cluster B, where side pattern was also significantly correlated with alkaloid variation (slope = 0.1376, p-value = 0.008), although geography was most highly correlated (slope = 0.3252, p-value = 0.001; Table 2). Additionally, at a broader scale – where all samples were pooled – we found that both side and ventral pattern were significantly correlated with alkaloid variation (for side: slope = 11.371, p-value = 0.001; for ventral: slope = 12.473, p-value = 0.001; Table 2). Ultimately, however, geographic distance was most highly correlated with chemical variation (slope = 48.100, p-value = 0.001; Table 2).

We did not detect any significant correlations between alkaloid variation and dorsal coloration, dorsal luminance, dorsal chromatic conspicuousness, and dorsal achromatic conspicuousness at any biological scale (Table 2).

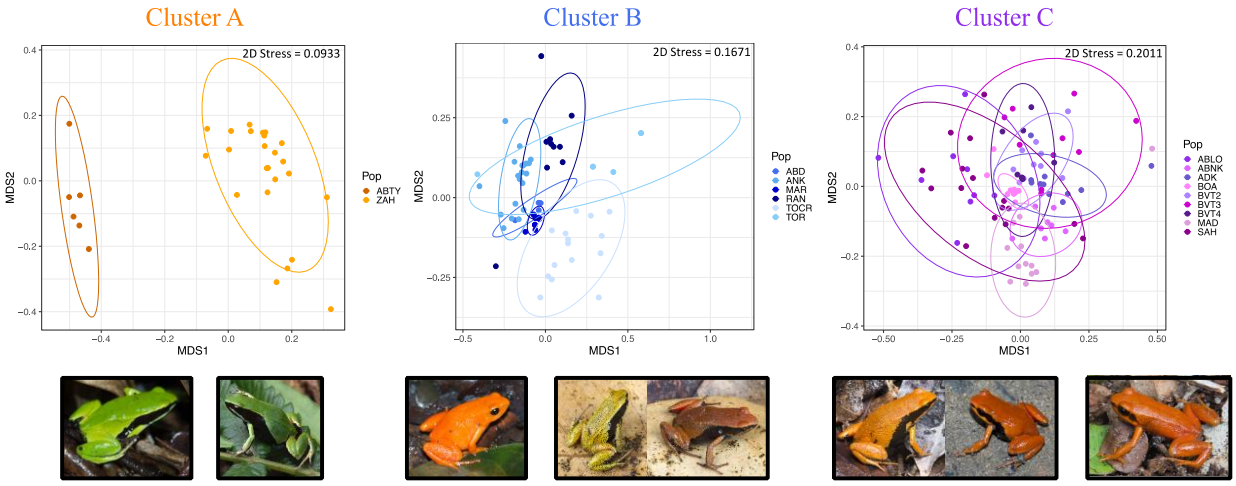

Figure 3: Multidimensional scaling plot of alkaloid composition among populations within each major genetic cluster (A, B, and C). Each symbol represents an individual frog, while the distance between points represents the difference in alkaloid composition. Ellipses represent the 95% confidence intervals of each population. Representative individuals from each genetic cluster are pictured below plots. Based on ANOSIM, alkaloid composition is significantly different amongst populations within each major genetic cluster (For Cluster A: Global $R = 0.9595$, $P = 0.001$; For Cluster B: Global $R = 0.6134$, $P = 0.001$; For Cluster C: Global $R = 0.4738$, $P = 0.001$).

Relationships among frog conspicuousness, alkaloid diversity, alkaloid abundance, and frog size

Both within each major genetic cluster and across all samples, we did not detect any significant relationships between alkaloid diversity (number of alkaloids per frog) and dorsal chromatic or achromatic conspicuousness (Fig. 4). Similarly, we did not detect any significant correlations between alkaloid abundance and dorsal chromatic or achromatic conspicuousness, either across or within genetic clusters (Fig. 5).

Across all samples, frog size was not associated with alkaloid diversity (Fig. 6a), though the two were positively correlated within Cluster B (Adjusted $R^2 = 0.1162$; $p = 0.003$). Although a significant positive relationship was detected between alkaloid abundance and frog size (Adjusted $R^2 = 0.0232$; $p = 0.021$) across all samples, this trend appears to be primarily driven by just a few individuals (Fig. 6b). Within major genetic clusters, alkaloid abundance and frog size were associated only in Cluster A (Adjusted $R^2 = 0.1929$; $p = 0.009$). Finally, we did not detect any significant relationships between frog size and frog conspicuousness across all samples (Fig. 6c, 6d) or within major genetic clusters.

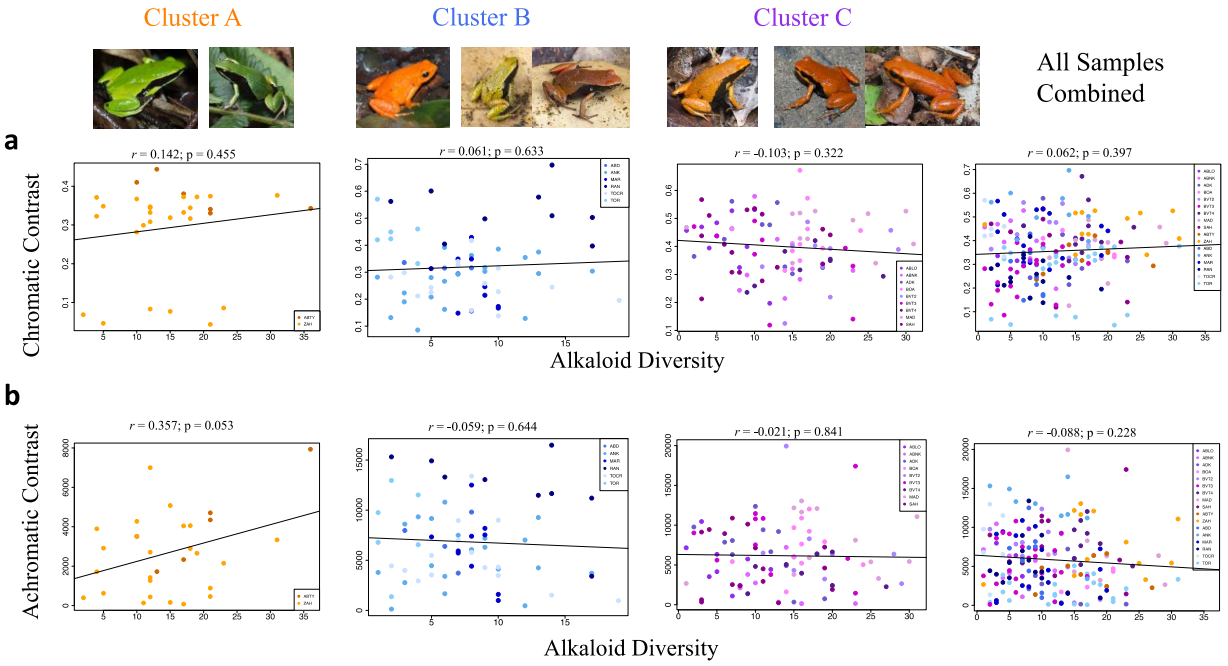

Figure 4: Alkaloid diversity (total number of unique alkaloids per frog) versus chromatic and achromatic conspicuousness of frogs for all samples and within each major genetic cluster. Populations are color coded and representative individuals from each genetic cluster are pictured above plots. Across all samples, and within each genetic cluster, alkaloid diversity was not significantly correlated with chromatic or achromatic contrast, as indicated by Pearson's correlation coefficients.

DISCUSSION

Our investigation of chemical and phenotypic diversity in Malagasy poison frogs revealed the importance of geography in structuring alkaloid variation among populations, even at small spatial scales. Surprisingly, very few aspects of phenotype were correlated with alkaloid composition. Where correlations were observed, they occurred at broader biological scales (i.e., all samples), and were related to side and ventral patterning elements. Further, we detected no significant relationships between frog conspicuousness and alkaloid diversity or abundance, both within and across major genetic clusters, which has important implications for signal honesty in this group. Below, we discuss the relevance of our findings for the evolution of chemical defenses in Malagasy poison frogs, and for future studies of phenotypic and chemical diversity in aposematic poison frogs.

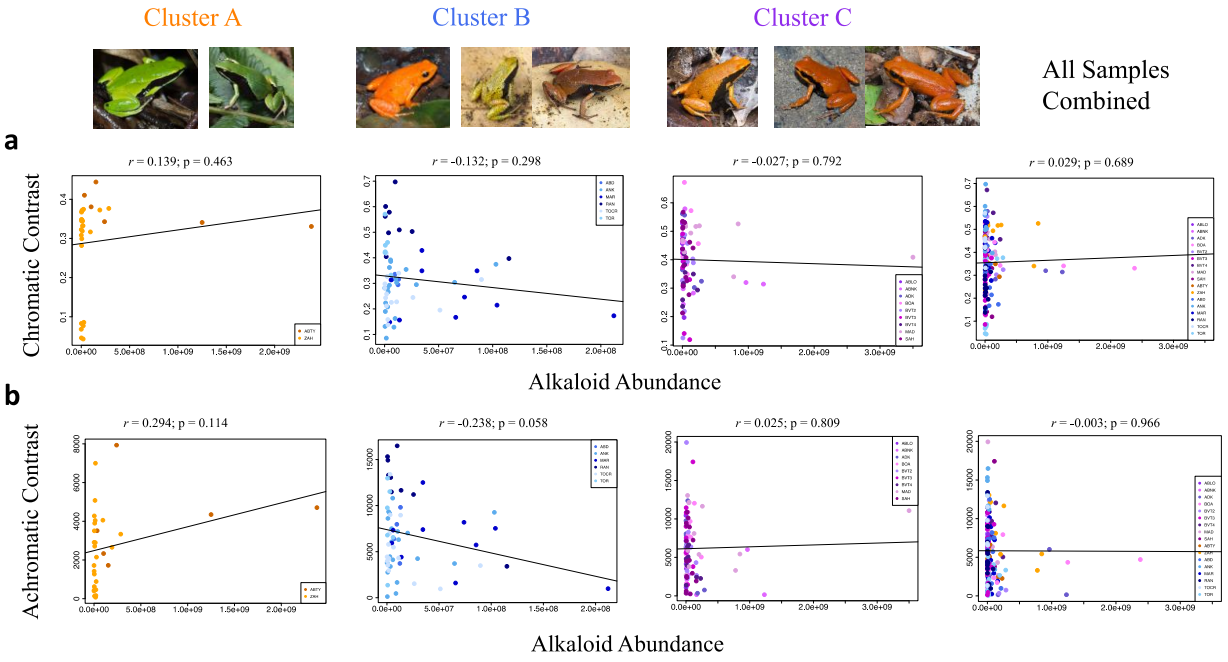

Figure 5: Alkaloid abundance versus chromatic and achromatic conspicuousness of frogs for all samples and within each major genetic cluster. Populations are color coded and representative individuals from each genetic cluster are pictured above plots. Across all samples, and within each genetic cluster, alkaloid abundance was not significantly correlated with chromatic or achromatic contrast, as indicated by Pearson's correlation coefficients.

Variation in alkaloid composition is geographically structured

We detected significant differences in alkaloid composition among all major genetic clusters (A, B, and C), and among populations within each major genetic cluster (Fig. 2a, Fig. 3). We also detected significant differences in alkaloid composition between populations from Cluster B, populations from Cluster C, and populations admixed between Clusters B and C (Fig. 2b). Admixed populations have been previously described (Klonoski et al., 2019) and demonstrate both genetic and phenotypic admixture. The geographic variation in alkaloid composition observed in our study is consistent with a large body of work documenting chemical diversity among poison frog populations of various lineages, including Malagasy poison frogs (e.g., Garraffo et al., 1993; Clark et al., 2006; Saporito et al., 2006; Saporito et al., 2007a; Daly et al., 2008; Andriamaharavo et al., 2015).

Overall, our work substantiates other findings which have implicated differences in geographic location as a major contributor to variation in alkaloid composition among species and populations of poison frogs. Within all major genetic clusters (Clusters A, B, and C), our matrix multiple regression analyses revealed that geographic distance was the variable most highly correlated with variation in alkaloid composition (Table 2). Because populations within each genetic cluster are situated at

disparate geographic scales (167 km apart in Cluster A, 0.37 – 39 km apart in Cluster B, 0.09 – 17 km apart in Cluster C), our findings suggest that geography is an important factor in mediating alkaloid variation at a wide range of spatial scales.

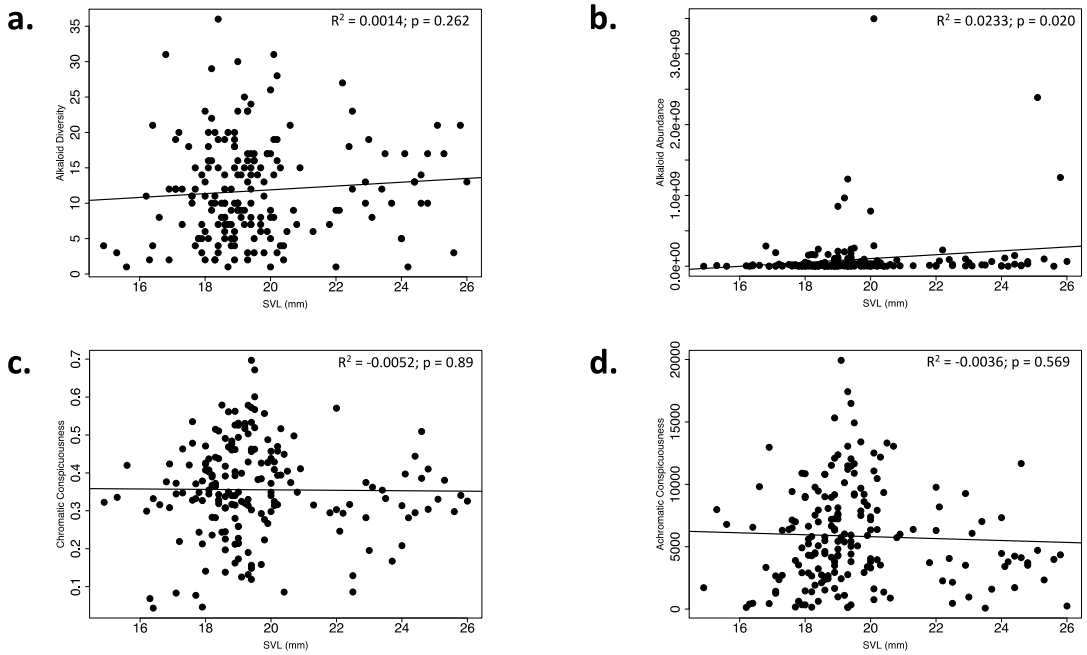

Figure 6: Relationship between frog size (SVL) and alkaloid diversity (number of unique alkaloids per frog; panel a), alkaloid abundance (panel b), chromatic conspicuousness (panel c), and achromatic conspicuousness (panel d) for all samples combined. Based on simple linear regression, no significant relationship was detected between frog size and alkaloid diversity, or either measure of conspicuousness. Alkaloid abundance and frog size, however, were positively related (Adjusted R^2 = 0.0232; p = 0.021).

When assessing correlations between phenotype, geography and alkaloid profiles across all species and populations (i.e., all samples pooled together), geographic distance was again the variable that best explained variation in alkaloid composition (Table 2). Although we detected variation in alkaloid composition among major genetic clusters (Fig. 2a), populations of each major genetic cluster are located in non-overlapping geographic areas. To explicitly test for species differences in alkaloid composition, it would be necessary to characterize chemical profiles in geographic locations where multiple species live in sympatry. To our knowledge, no such locations exist for the focal species examined here. However, given that alkaloid variation has been detected among other sympatric *Mantella* species (Daly et al., 2008), it is possible that some mechanism, such as differences in alkaloid sequestration, maintain species-level differences in alkaloid profiles in this group.

Given the presumed dietary acquisition of alkaloids in poison frogs (referred to as the "dietary acquisition hypothesis" reviewed in Saporito et al., 2009a), many

studies detecting geographic variation in chemical defenses have speculated that differences in prey communities among geographic localities translate into observed alkaloid variation (e.g., Clark et al., 2006, Saporito et al., 2006; Saporito et al., 2007a; Saporito et al., 2009a; Andriamaharavo et al., 2015). Yet, few studies have empirically investigated this hypothesis and relatively little is known about how arthropod communities vary between poison frog sites and habitats, and whether this variation contributes to distinct chemical phenotypes of frogs. In the poison frog *Oophaga sylvatica*, previous work has demonstrated that alkaloid variation among populations reflects the chemical diversity of arthropod prey isolated from frog stomach contents (McGugan et al., 2015), providing compelling evidence that differences in prey availability contribute to variation in chemical defenses.

Recent work in *Mantella laevigata*, however, has suggested that alkaloid profiles are relatively stable even across different seasons with different dietary regimes (Moskowitz et al., 2018). Although frog diet was found to be substantially different between the wet and dry seasons for one population of *M. laevigata*, alkaloid profiles (i.e., presence/absence of alkaloids) were similar across seasons, which may be partially due to long-term retention of alkaloids by poison frogs. Interestingly, the abundance of certain alkaloids was found to vary across seasons, which could potentially arise from seasonal differences in diet or foraging behavior (Moskowitz et al., 2018). Ultimately, additional research on habitat and microhabitat differences, arthropod prey availability and abundance, *Mantella* foraging behavior and diet, and possible interactions among these factors will be necessary in order to disentangle their effects on variation in chemical profiles.

Interpreting variation in chemical defenses

Although we detected variation in alkaloid composition, diversity, and abundance among species, populations, and within populations, little is known regarding the adaptive value of variation in chemical profiles. The function of chemical compounds in relevant biological settings (i.e., ingested by a natural predator) is often unknown in aposematic organisms. In the past, "toxicity" in poison frogs has been quantified primarily through injection of frog skin extracts into mice (as was performed in Wang, 2011; Maan & Cummings, 2012), or by using alkaloid diversity, quantity, and available information on toxicity as proxies for frog "toxicity" (as used in Summers & Clough, 2001). More recently, concerns have been raised over the ability of either of these methods to accurately reflect toxicity of poison frogs in terms of their effect on natural predators (Bolton et al., 2017; Weldon, 2017; Saporito & Grant, 2018). To truly assess the toxicity of frogs, experiments where naturally occurring predators are exposed to frog alkaloids in biologically relevant ways will be required. However, given the conservation status of the species in question, we do not encourage experiments of this nature in this group of poison frogs.

Based on mouse assays, certain classes of alkaloid have been found to be more toxic than others (reviewed in Table 21.2 of Santos et al., 2016) and are presumed to demonstrate similarly high toxicity for poison frog predators. Of the alkaloid classes detected in our dataset, pumiliotoxins and allopumiliotoxins represent two alkaloid types that have been categorized as demonstrating "high" toxicity in mice (reviewed in Table 21.2 of Santos et al., 2016). In our study, proportions of pumiliotoxins (PTX) and allopumiliotoxins (aPTX) did vary among populations, which could reflect differences in toxicity and/or palatability to predators (Table 1). For example, two populations in our study (MAD and TOR) had especially high proportions of allopumiliotoxins present (21.5% and 27.8%; see Table 1), while other populations had particularly high proportions of pumiliotoxins present (ABD = 29.2%, MAR = 32.5%; see Table 1). Further, the diversity of alkaloid classes represented within populations varied, with some populations remarkably conserved (as in MAR, where only four different types of alkaloids were identified; see Table 1). However, given that toxicity assays have been performed in mice and not natural predators, we emphasize that these differences may or may not be biologically relevant. While it is likely that certain compounds are more toxic to poison frog predators and thus contribute more to overall toxicity of frogs, more research in natural systems is needed to identify such compounds.

Given the lack of information on the effects of alkaloid-based defenses in natural systems, ascertaining which aspects of chemical diversity are biologically relevant is nearly impossible at this time. While there is evidence that alkaloid type, quantity and diversity all impact how chemical defenses are perceived by predators (Bolton et al., 2017), more research is needed to determine how natural predators perceive differences in chemical profiles. Ultimately, interpretation of variability in chemical defenses, as detected in this study and many others, will require detailed information on predator communities and sensory capabilities.

Implications for signal honesty in Malagasy poison frogs

Overall, we detected very few correlations between variation in alkaloid composition and variation in frog coloration and pattern. In fact, within major genetic clusters, the only significant correlation detected using our multiple matrix regression analysis was between variation in side pattern and alkaloid composition within Cluster B (Table 2). Despite this association, geography still remained the variable with the highest explanatory power for alkaloid variation (Table 2). When we performed our regression analyses on all samples combined across species and populations, we found that geographic distance again remained the best explainer of alkaloid variation, though variation in side and ventral pattern were also significantly correlated with variation in alkaloid composition (Table 2).

Similarly, we detected no significant relationships between frog conspicuousness and alkaloid composition, diversity, or abundance in our study. Whether looking within major genetic clusters or across all clusters, neither chromatic nor achromatic dorsal conspicuousness was correlated with alkaloid composition (Table 2). Further, our comparisons between frog conspicuousness – both chromatic and achromatic – and alkaloid diversity and abundance did not reveal any significant associations, either within major genetic clusters or across all samples (Fig. 4, Fig. 5).

Research on signal honesty in aposematic organisms has become more popular in recent years (reviewed in Summers et al., 2015), with the Neotropical poison frogs of Latin America serving as a model system for this line of work. Although many studies have confirmed the qualitative honesty of poison frog signals in the sense that conspicuously colored frogs contain alkaloids (e.g., Garraffo et al., 1993; Saporito et al., 2006; Daly et al., 2008), whether or not populations or species display quantitative honesty – where higher levels of defense correspond to stronger visual signals – is less clear (Stuckert et al., 2018). Several studies have detected a relationship between frog conspicuousness and toxicity across populations and/or species (Summers & Clough, 2001; Wang, 2011; Maan & Cummings, 2012). From this body of work, it appears that the direction of the correlation between toxicity and conspicuousness is positive in some lineages (Maan & Cummings, 2012) and negative in others (Wang, 2011). Within populations, there is some evidence that poison frog signals are not quantitatively honest (Stuckert et al., 2018). This somewhat conflicting evidence is difficult to interpret collectively and is likely compounded by other factors, such as the use of dorsal coloration as a conspecific signal in mating and male-male interactions (Maan & Cummings, 2008; Maan & Cummings, 2009; Crothers et al., 2011). Additionally, past studies of signal honesty have utilized either mouse-based toxicity assays (as in Wang, 2011; Maan & Cummings, 2012) or measures of alkaloid diversity, quantity, and/or toxicity information taken from other systems (as in Summers & Clough, 2001; Stuckert et al., 2018) to quantify toxicity. As noted above, these methods may or may not reflect true toxicity of frogs to their natural predators.

While signal honestly has been more extensively investigated in Neotropical poison frogs, our study is the first to address this phenomenon in Malagasy poison frogs. Although we did not explicitly quantify "toxicity" of frogs, our comparison of phenotype to alkaloid composition, diversity, and abundance did not provide compelling evidence for quantitatively honest signals in this group of poison frogs. Ecological differences between Neotropical and Malagasy poison frogs may be important in understanding differences in signal evolution. For example, while only lizards and snakes have been identified as *Mantella* predators (Heying, 2001; Jovanovic et al., 2009; Hutter et al., 2018), birds have been implicated as particularly important predators of Neotropical poison frogs (Saporito et al., 2007b; Maan & Cummings, 2012). Thus, empirical work on signal honesty in Neotropical poison frogs has tended to emphasize an avian visual perspective. Interestingly, using predator visual

models, previous work in *Oophaga pumilio* (formerly *Dendrobates pumilio*) has found that although frog conspicuousness and toxicity were highly correlated for bird-specific models, there was no correlation between the two when assessed with a snake-specific model (Maan & Cummings, 2012). The stark differences in perception between predator taxa highlights the importance of incorporating relevant visual perspectives, and may help explain why we did not detect any correlations between dorsal coloration/conspicuousness and alkaloid composition in this group of Malagasy poison frogs, where predators may differ.

While dorsal coloration and conspicuousness – traits relevant to an avian visual perspective – are associated with toxicity in Neotropical poison frogs, our results suggest that side and ventral pattern are more tightly linked with alkaloid composition in Malagasy poison frogs. A growing body of work in poison frogs has suggested that patterning elements may be more biologically relevant than previously thought. Studies of pattern variation within populations of *Dendrobates tinctorius* have shown that certain elements of frog pattern are correlated with variable movement behavior and may therefore influence perception by predators (Rojas et al., 2014a). More recent work in this same species has demonstrated that although color and pattern are perceived as aposematic signals when viewed at short distances, they are cryptic when viewed by predators from long distances (Barnett et al., 2018). Additionally, ambient light environments also influence the detectability of different color pattern variants (Rojas et al., 2014b). From this work, it is apparent that a more holistic approach encompassing core principles of sensory ecology is needed to examine the functional significance of color pattern variation, especially for aposematic signals.

Conclusions

In this study, our comprehensive sampling across many individuals and populations of closely-related Malagasy poison frog species enabled us to detect overarching patterns in chemical and phenotypic diversity at different biological and geographic scales. Our results confirmed the importance of geography in structuring chemical diversity, which has been implicated in previous studies. Based on these findings, it is clear that more detailed studies on habitat characteristics, arthropod communities, and frog diet are needed to better understand the role of geography in shaping chemical diversity. Unexpectedly, we detected hardly any correlations between frog phenotype and chemical composition, both within and across major genetic clusters. Although we observed substantial variation in alkaloid composition and phenotype across populations, interpreting its significance is difficult when it is unclear which aspects of chemical defense and phenotype are biologically relevant. Ultimately, further research on the evolutionary ecology of chemical defenses in aposematic frogs will need to draw on principles of sensory ecology to investigate how variation in both visual and chemical cues is perceived by natural predators.

Admittedly, this can be a difficult task in endangered species, where lethal methods of alkaloid collection and toxicity assays, such as feeding experiments, are not recommended. Alternatively, we suggest using the following methods to investigate these questions: conducting clay model studies to identify predators and to assess predation on different color and pattern phenotypes, implementing visual models based on principles of sensory ecology to clarify how phenotypic differences are perceived by predators, and collecting alkaloids non-lethally via TAS. Understanding which aspects of defense are most vital for frog survival not only has important evolutionary implications for this group of frogs, but can also contribute meaningfully to conservation of these endangered species.

Table 1: Percentage of total identified alkaloids belonging to each alkaloid class (per population). Note: Population labels correspond to those designated in Figure 1.

Population (Major genetic cluster in parenthesis)	Percentage of total identified alkaloids belonging to each alkaloid class (per population)												
	5,8-I	Unclass/ tricyclic	PTX	3,5-P	5,6,8-I	aPTX	Dehydro-esmethyl PTX	Pip	Dehydro 5,8-I	3,5-I	Hydroxy-Q	Desmethyl hPTX	Deoxy PTX
ABTY (A)	11.9	20.3	0	10.2	0	15.3	4.2	2.5	34.8	0	0	0.9	0
ZAH (A)	28.5	9.5	7.7	3.7	21.8	7.1	3.1	0	2.8	7.4	5.2	3.4	0
ABD (B)	20.8	33.3	29.2	0	0	16.7	0	0	0	0	0	0	0
ANK (B)	36.4	8.0	21.2	16.6	2.6	0	0	2.7	0.7	6.0	6.0	0	0
MAR (B)	36.3	31.3	32.5	0	0	0	0	0	0	0	0	0	0
RAN (B)	17.0	17.0	8.0	16.0	10.0	8.0	16.0	0	0	8.0	0	0	0
TOCR (B)	8.4	16.0	17.6	1.5	9.2	4.6	29.8	0	7.6	0	0	5.3	0
TOR (B)	11.1	38.9	11.1	0	0	27.8	0	0	11.1	0	0	0	0
ABLO (C)	29.6	18.5	16.7	14.8	0	13.0	0	0	1.9	5.6	0	0	0
ABNK (C)	27.8	11.1	1.9	19.4	17.6	11.1	0	7.4	0	0	0	3.7	0
ADK (C)	26.6	6.8	21.4	21.4	8.6	4.3	0	9.4	0	1.7	0	0	0
BOA (C)	27.8	15.3	18.2	17.6	5.7	6.3	3.4	5.7	0	0	0	0	0
BVT2 (C)	21.5	19.4	10.0	12.6	16.8	4.7	0	10.0	4.2	0.5	0	0	0.5
BVT3 (C)	22.8	20.5	5.5	11.0	18.1	4.7	0	7.9	1.6	6.3	0	0.8	0.8
BVT4 (C)	26.9	12.7	11.9	11.9	9.0	4.5	0	11.9	3.0	5.2	0	0	3.0
MAD (C)	14.6	9.9	3.4	12.5	9.9	21.5	20.6	3.0	4.3	0.4	0	0	0
SAH (C)	13.7	17.7	15.7	13.7	6.5	7.8	1.3	2.6	0	11.8	3.3	3.3	2.6

Table 2: Multiple matrix regression results of factors that correlate with alkaloid variation within each cluster (A, B, C) and across all samples.

x-matrix (Determinants of Alkaloid Variation)	**Dorsal Coloration**		**Geographic Distance***		**Side Pattern**		**Ventral Pattern**		**Dorsal Luminance**		**Achromatic Contrast**		**Chromatic Contrast**	
Cluster A	***r***	***p***	***r***	***p***	***r***	***p***	***r***	***p***	***r***	***p***	***r***	***p***	***r***	***p***
Step 1: Mantel	0.2618	**0.012**	0.7922	**0.001**	0.2519	**0.007**	0.0060	0.442	0.2170	**0.023**	0.0690	0.223	0.0430	0.301
Step 2: Partial Mantel	0.0188	0.426	N/A	N/A	-0.0646	0.773	N/A	N/A	0.0490	0.285	N/A	N/A	N/A	N/A
	Slope	***p***	**Slope**	***p***	**Slope**	***p***	**Slope**	***p***	**Slope**	***p***	**Slope**	***p***	**Slope**	***p***
Step 3: MultiMantel	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Cluster B	***r***	***p***	***r***	***p***	***r***	***p***	***r***	***p***	***r***	***p***	***r***	***p***	***r***	***p***
Step 1: Mantel	0.3527	**0.001**	0.4432	**0.001**	0.3830	**0.001**	0.3446	**0.001**	0.1231	**0.028**	0.0782	0.068	0.1516	**0.009**
Step 2: Partial Mantel	0.0709	0.089	N/A	N/A	0.1117	**0.014**	0.0854	0.058	-0.089	0.954	N/A	N/A	-0.0657	0.886
	Slope	***p***	**Slope**	***p***	**Slope**	***p***	**Slope**	***p***	**Slope**	***p***	**Slope**	***p***	**Slope**	***p***
Step 3: MultiMantel	N/A	N/A	0.3252	**0.001**	0.1376	**0.008**	0.0162	0.399	N/A	N/A	N/A	N/A	N/A	N/A
Cluster C	***r***	***p***	***r***	***p***	***r***	***p***	***r***	***p***	***r***	***p***	***r***	***p***	***r***	***p***
Step 1: Mantel	0.0367	0.198	0.2322	**0.001**	0.0403	0.237	0.1239	**0.006**	-0.104	0.980	0.0098	0.394	0.0022	0.456
Step 2: Partial Mantel	N/A	N/A	N/A	N/A	N/A	N/A	0.0247	0.284	N/A	N/A	N/A	N/A	N/A	N/A
	Slope	***p***	**Slope**	***p***	**Slope**	***p***	**Slope**	***p***	**Slope**	***p***	**Slope**	***p***	**Slope**	***p***
Step 3: MultiMantel	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
All Samples	***r***	***p***	***r***	***p***	***r***	***p***	***r***	***p***	***r***	***p***	***r***	***p***	***r***	***p***
Step 1: Mantel	0.2377	**0.001**	0.3496	**0.001**	0.2376	**0.001**	0.1726	**0.001**	0.0971	**0.001**	0.0359	0.135	0.0652	**0.022**
Step 2: Partial Mantel	-0.0019	0.547	N/A	N/A	0.1623	**0.001**	0.166	**0.001**	-0.066	0.984	N/A	N/A	0.0535	0.072
	Slope	***p***	**Slope**	***p***	**Slope**	***p***	**Slope**	***p***	**Slope**	***p***	**Slope**	***p***	**Slope**	***p***
Step 3: MultiMantel	N/A	N/A	48.1001	**0.001**	11.3711	**0.001**	12.4732	**0.001**	N/A	N/A	N/A	N/A	N/A	N/A

Note: For each group, Step 1 shows the results of pairwise Mantel tests between the y-matrix (alkaloid variation) and each x-matrix (dorsal coloration and luminance, geographic distance, side and ventral pattern, chromatic and achromatic conspicuousness). All x-matrices that significantly contributed to alkaloid variation in Step 1 were included in Step 2. In Step 2, we performed partial Mantel tests that accounted for geographic distance, the variable with the highest r value from Step 1 (as denoted by *). Only those variables that remained significant in Step 2 were included in Step 3. If no variables from Step 2 remained significant after controlling for geographic distance, we did not perform a MultiMantel test. In Step 3, we performed a MultiMantel to incorporate multiple variables into our model.

ACKNOWLEDGEMENTS

Richard Fitch, Serge Ndriantsoa, Kimberly Gleason, Ryan Sanders, and Erica Bree Rosenblum co-authored this dissertation chapter. We are grateful to Devin Edmonds and Miguel Vences for their guidance and support. We thank numerous local guides for sharing their knowledge and for providing guidance in the field. We also thank Jacqueline Smith, Kameron Bell, and Nicholas Andreasen for their help in chemical sample preparation, data acquisition and data analysis. We thank Rebecca Tarvin for lending us her TAS. This work was conducted under a collaboration agreement between ICTE/MICET, the Département de Biologie Animale of the Université d'Antananarivo, and the Ministère de l'Environnement et des Forêts, Direction du Système des Aires Protégées. We are grateful to MICET and the Université d'Antananarivo for providing logistical support in Madagascar. We also thank the Direction Générale des Forêts, Ministere de l'Environment et des Forêts, and Madagascar National Parks for granting us research and export permits. We are grateful to Hugo Martin Garraffo and Thomas Spande for their helpful guidance with the alkaloid library, and would like to acknowledge NIH bailment NIDDK – 11-0502 for custody of the John Daly research sample collection. Undergraduate students participated as part of the Indiana State University Summer Undergraduate Research Experiences (SURE) Program. Funding was provided by The Mohamed bin Zayed Species Conservation Fund (Project #150511309), the National Geographic Society (Grant Project #9469-14), the Museum of Vertebrate Zoology at the University of California at Berkeley, and the National Science Foundation (DEB-1556982). All animal work was approved by the Animal Care and Use Committee at the University of California at Berkeley (R347-0314; AUP-2015-01-7083).

LITERATURE CITED

Andriamaharavo, N. R., H. M. Garraffo, T. F. Spande, L. A. Giddings, D. R. Vieites, M. Vences, and R. A. Saporito. 2015. Individual and geographic variation of skin alkaloids in three swamp-forest species of Madagascan poison frogs (Mantella). Journal of Chemical Ecology 41(9):837-847.

Barnett, J. B., C. Michalis, N. E. Scott-Samuel, and I. C. Cuthill. 2018. Distance-dependent defensive coloration in the poison frog Dendrobates tinctorius, Dendrobatidae. Proceedings of the National Academy of Sciences of the United States of America 115(25):6416-6421.

Bergmann, A. J., G. L. Points, R. P. Scott, G. Wilson, and K. A. Anderson. 2018. Development of quantitative screen for 1550 chemicals with GC-MS. Analytical and Bioanalytical Chemistry 410(13):3101-3110.

Bolton, S. K., K. Dickerson, and R. A. Saporito. 2017. Variable Alkaloid Defenses in the Dendrobatid Poison Frog *Oophaga pumilio* are Perceived as Differences in Palatability to Arthropods. Journal of Chemical Ecology 43(3):273-289.

Chiari, Y., M. Vences, D. R. Vieites, F. Rabemananjara, P. Bora, O. Ramilijaona Ravoahangimalala, and A. Meyer. 2004. New evidence for parallel evolution of colour patterns in Malagasy poison frogs (*Mantella*). Molecular Ecology 13(12):3763-3774.

Clark, V. C., V. Rakotomalala, O. Ramilijaona, L. Abrell, and B. L. Fisher. 2006. Individual variation in alkaloid content of poison frogs of Madagascar (Mantella; Mantellidae). Journal of Chemical Ecology 32(10):2219-2233.

Clark, V. C., C. J. Raxworthy, V. Rakotomalala, P. Sierwald, and B. L. Fisher. 2005. Convergent evolution of chemical defense in poison frogs and arthropod prey between Madagascar and the Neotropics. Proceedings of the National Academy of Sciences of the United States of America 102(33):11617-11622.

Cramer, C. J., J. L. Johnson, and A. M. Kamel. 2017. Prediction of mass spectral response factors from predicted chemometric data for druglike molecules. Journal of the American Society for Mass Spectrometry 28(2):278-285.

Crothers, L., E. Gering, and M. Cummings. 2011. Aposematic signal variation predicts male-male interactions in a polymorphic poison frog. Evolution 65(2):599-605.

Daly, J. W., H. M. Garraffo, G. S. E. Hall, and J. F. Cover, Jr. 1997. Absence of skin alkaloids in captive-raised Madagascan mantelline frogs (Mantella) and sequestration of dietary alkaloids. Toxicon 35(7):1131-1135.

Daly, J. W., H. M. Garraffo, T. F. Spande, L. A. Giddings, R. A. Saporito, D. R. Vieites, and M. Vences. 2008. Individual and geographic variation of skin alkaloids in three species of Madagascan poison frogs (*Mantella*). Journal of Chemical Ecology 34(2):252-279.

Daly, J. W., R. J. Highet, and C. W. Myers. 1984. Occurrence of skin alkaloids in non-

dendrobatid frogs from Brazil (Bufonidae), Australia (Myobatrachidae) and Madagascar (Mantellinae). Toxicon 22(6):905-919.
Daly, J. W., N. Rabe Andriamaharavo, M. Andriantsiferana, and C. W. Myers. 1996. Madagascan poison frogs (Mantella) and their skin alkaloids. American Museum Novitates 3177:1-34.
Daly, J. W., T. F. Spande, and H. M. Garraffo. 2005. Alkaloids from amphibian skin: A tabulation of over eight-hundred compounds. Journal of Natural Products 68(10):1556-1575.
Daly, J. W., J. M. Wilham, T. F. Spande, H. M. Garraffo, R. R. Gil, G. L. Silva, and M. Vaira. 2007. Alkaloids in Bufonid toads (*Melanophryniscus*): Temporal and geographic determinants for two Argentinian species. Journal of Chemical Ecology 33(4):871-887.
Davis, L. R., K. Klonoski, H. L. Rutschow, K. J. Van Wijk, Q. Sun, M. M. Haribal, R. A. Saporito, A. Vega, E. B. Rosenblum, K. R. Zamudio, and J. M. Robertson. 2016. Host defense skin peptides vary with color pattern in the highly polymorphic red-eyed treefrog. Frontiers in Ecology and Evolution 4:97-112.
Endler, J. A. 2012. A framework for analysing colour pattern geometry: adjacent colours. Biological Journal of the Linnean Society 107(2):233-253.
Endler, J. A., and P. W. Mielke, Jr. 2005. Comparing entire colour patterns as birds see them. Biological Journal of the Linnean Society 86(4):405-431.
Ersts, P. J. 2018. Geographic Distance Matrix Generator (version 1.2.3). American Museum of Natural History, Center for Biodiversity and Conservation. http://biodiversityinformatics.amnh.org/open_source/gdmg
Garraffo, H. M., J. Caceres, J. W. Daly, and T. F. Spande. 1993. Alkaloids in Madagascan frogs (*Mantella*): pumiliotoxins, indolizidines, quinolizidines, and pyrrolizidines. Journal of Natural Products 56(7):1016-1038.
Glaw, F. and M. Vences. 2007. A field guide to the amphibians and reptiles of Madagascar. (3rd ed.) Cologne, Germany: Vences and Glaw.
Grant, J. B., and B. Land. 2002. Transcutaneous amphibian stimulator (TAS): A device for the collection of amphibian skin secretions. Herpetological Review 33(1):38-41.
Heying, H. E. 2001. *Mantella laevigata* (climbing Mantella). Aborted Predation. Herpetological Review 32(1):34.
Hutter, C. R., Z. F. Andriampenomanana, J. H. Razafindraibe, A. Rakotoarison, and M. D. Scherz. 2018. New dietary data from *Compsophis* and *Alluaudina* species (Squamata: Lamprophiidae: Pseudoxyrhophiinae), and implications for their dietary complexity and evolution. Journal of Natural History 52(39-40):2497-2510.
Jeckel, A. M., R. A. Saporito, and T. Grant. 2015. The relationship between poison

frog chemical defenses and age, body size, and sex. Frontiers in Zoology 12(1):27.
Jovanovic, O., M. Vences, G. Safarek, F. C. E. Rabemananjara, and R. Dolch. 2009. Predation upon *Mantella aurantiaca* in the Torotorofotsy wetlands, central-eastern Madagascar. Herpetology Notes 2:95-97.
Klonoski, K., K. Bi, and E. B. Rosenblum. 2019. Phenotypic and genetic diversity in aposematic Malagasy poison frogs (genus *Mantella*). Ecology and Evolution 9(5):2725-2742.
Krohn, A. R., and E. B. Rosenblum. 2016. Geographic color variation and physiological color change in eastern collared lizards (*Crotaphytus collaris*) from Southern New Mexico, USA. Herpetologica 72(4):318-323.
Maan, M. E. and M. E. Cummings. 2008. Female preferences for aposematic signal components in a polymorphic poison frog. Evolution 62(9):2334-2345.
Maan, M. E. and M. E. Cummings. 2009. Sexual dimorphism and directional sexual selection on aposematic signals in a poison frog. Proceedings of the National Academy of Sciences of the United States of America 106(45):19072-19077.
Maan, M. E., and M. E. Cummings. 2012. Poison frog colors are honest signals of toxicity, particularly for bird predators. The American Naturalist 179(1):E1-E14.
McGugan, J. R., G. D. Byrd, A. B. Roland, S. N. Caty, N. Kabir, E. E. Tapia, S. A. Trauger, L. A. Coloma, and L. A. O'Connell. 2016. Ant and mite diversity drives toxin variation in the little devil poison frog. Journal of Chemical Ecology 42(6):537-551.
McKay, B. D. 2013. The use of digital photography in systematics. Biological Journal of the Linnean Society 110(1):1-13.
Mebs, D., W. Pogoda, R. Maneyro, and A. Kwet. 2005. Studies on the poisonous skin secretion of individual red bellied toads, *Melanophryniscus montevidensis* (Anura, Bufonidae), from Uruguay. Toxicon 46(6):641-650.
Moskowitz, N. A., A. B. Roland, E. K. Fischer, N. Ranaivorazo, C. Vidoudez, M. T. Aguilar, S. M. Caldera, J. Chea, M. G. Cristus, J. P. Crowdis, B. DeMesie, C. R. desJardins-Park, A. H. Effenberger, F. Flores, M. Giles, E. Y. He, N. S. Izmaylov, C. C. Lee, N. A. Pagel, K. K. Phu, L. U. Rosen, D. A. Seda, Y. Shen, S. Vargas, A. W. Murray, E. Abebe, S. A. Trauger, D. A. Donoso, M. Vences, and L. A. O'Connell. 2018. Seasonal changes in diet and chemical defense in the Climbing Mantella frog (*Mantella laevigata*). PLOS ONE 14(6):e021898.
Myers, C. W., J. W. Daly, H. M. Garraffo, and A. Wisnieski. 1995. Discovery of the Costa Rican poison frog *Dendrobates granuliferus* in sympatry with *Dendrobates pumilio*, and comments on taxonomic use of skin alkaloids. American Museum Novitates 3144:1-21.
Osorio, D., and M. Vorobyev. 2005. Photoreceptor spectral sensitivities in terrestrial

animals: adaptations for luminance and colour vision. Proceedings of the Royal Society B: Biological Sciences 272(1574):1745-1752.
R Core Team. 2017. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/
Rojas, B., J. Devillechabrolle, and J. A. Endler. 2014a. Paradox lost: variable colour-pattern geometry is associated with differences in movement in aposematic frogs. Biology Letters 10(6):20140193.
Rojas, B., P. Rautiala, and J. Mappes. 2014b. Differential detectability of polymorphic warning signals under varying light environments. Behavioural Processes 109:164-172.
Santos, J.C., and D. C. Cannatella. 2011. Phenotypic integration emerges from aposematism and scale in poison frogs. Proceedings of the National Academy of Sciences of the United States of America 108(15):6175-6180.
Santos, J. C., L. A. Coloma., and D. C. Cannatella. 2003. Multiple, recurring origins of aposematism and diet specialization in poison frogs. Proceedings of the National Academy of Sciences of the United States of America 100(22):12792-12797.
Santos, J.C., R. D. Tarvin, and L. A. O'Connell. 2016. A review of chemical defense in poison frogs (Dendrobatidae): Ecology, pharmacokinetics, and autoresistance. In: Schulte, B., Goodwin, T., Ferkin, M. (eds) Chemical Signals in Vertebrates 13. Springer, Cham.
Saporito, R. A., M. A. Donnelly, H. M. Garraffo, T. F. Spande, and J. W. Daly. 2006. Geographic and seasonal variation in alkaloid-based chemical defenses of *Dendrobates pumilio* from Bocas del Toro, Panama. Journal of Chemical Ecology 32(4):795-814.
Saporito, R.A., M. A. Donnelly, P. Jain, H. M. Garraffo, T. F. Spande, and J. W. Daly. 2007a. Spatial and temporal patterns of alkaloid variation in the poison frog *Oophaga pumilio* in Costa Rica and Panama over 30 years. Toxicon 50(6):757-778.
Saporito, R. A., M. A. Donnelly, A. A. Madden, H. M. Garraffo, and T. F. Spande. 2009b. Sex-related differences in alkaloid chemical defenses of the Dendrobatid frog *Oophaga pumilio* from Cayo Nancy, Bocas del Toro, Panama. Journal of Natural Products 73(3):317-321.
Saporito, R.A., M. A. Donnelly, T. F. Spande, and H. M. Garraffo. 2012. A review of chemical ecology in poison frogs. Chemoecology 22(3):159-168.
Saporito, R. A., and T. Grant. 2018. Comment on Amézquita et al. (2017) "Conspicuousness, color resemblance, and toxicity in geographically diverging mimicry: The pan-Amazonian frog *Allobates femoralis*". Evolution 72(4):1009-1014.
Saporito, R. A., T. F. Spande, H. M. Garraffo, and M. A. Donnelly. 2009a. Arthropod

alkaloids in poison frogs: a review of the 'dietary hypothesis'. Heterocycles 79(1):277-297.
Saporito, R. A., R. Zuercher, M. Roberts, K. G. Gerow, and M. A. Donnelly. 2007b. Experimental evidence for aposematism in the Dendrobatid poison frog *Oophaga pumilio*. Copeia 2007:1006-1011.
Schneider, C. A., W. S. Rasband, and K. W. Eliceiri. 2012. NIH Image to ImageJ: 25 years of image analysis. Nature Methods 9(7):671-675.
Speed, M. P., G. D. Ruxton, J. Mappes, and T. N. Sherratt. 2012. Why are defensive toxins so variable? An evolutionary perspective. Biological Reviews 87(4):874-884.
Stevens, M., C. A. Párraga, I. C. Cuthill, J. C. Partridge, and T. S. Troscianko. 2007. Using digital photography to study animal coloration. Biological Journal of the Linnean Society 90(2):211-237.
Stuckert, A. M. M., R. A. Saporito, and K. Summers. 2018. An empirical test indicates only qualitatively honest aposematic signaling within a population of vertebrates. Journal of Herpetology 52(2):201-208.
Summers, K., and M. E. Clough. 2001. The evolution of coloration and toxicity in the poison frog family (Dendrobatidae). Proceedings of the National Academy of Sciences of the United States of America 98(11):6227-6232.
Summers, K., M. P. Speed, J. D. Blount, and A. M. M. Stuckert. 2015. Are aposematic signals honest? A review. Journal of Evolutionary Biology 28(9):1583-1599.
Troscianko, J., and M. Stevens. 2015. Image calibration and analysis toolbox – a free software suite for objectively measuring reflectance, colour and pattern. Methods in Ecology and Evolution 6(11):1320-1331.
Wang, I. J. 2011. Inversely related aposematic traits: reduced conspicuousness evolves with increased toxicity in a polymorphic poison-dart frog. Evolution 65(6):1637-1649.
Weldon, P. J. 2017. Poison frogs, defensive alkaloids, and sleepless mice: critique of a toxicity bioassay. Chemoecology 27(4):123-126.

3

Discordant patterns of phenotypic and genetic variation in an endangered Malagasy poison frog, *Mantella cowani*, the common *Mantella baroni*, and putative *Mantella cowani* – *Mantella baroni* hybrids

ABSTRACT

In the face of ongoing habitat loss and forest fragmentation, understanding population genetic structure and levels of phenotypic variation is crucial for management of endangered species and populations. To clarify the status of an endangered Malagasy poison frog, *Mantella cowani*, and its relationship to *Mantella baroni*, we quantified genetic and phenotypic diversity among *M. cowani*, *M. baroni*, and putative *M. cowani* – *M. baroni* hybrid populations. Our restriction site-associated (RAD) sequencing approach did not reveal substantial genetic structure among populations. In addition to a lack of genetic differentiation between *M. cowani* and *M. baroni*, we did not recover any genomic signatures of admixture in putative *M. cowani* – *M. baroni* hybrids. Despite the lack of genetic structure, we did detect phenotypic variation among populations, particularly in dorsal patterning traits. Our results have important implications for the evolutionary processes contributing to patterns of phenotypic divergence despite genetic homogeneity that we observed in these species. Additionally, our study indicates that genetic data alone is not sufficient for making conservation decisions for this group of Malagasy poison frogs. Ultimately, our work highlights the complex nature of designating management units where discordant patterns of phenotypic and genetic variation are observed.

INTRODUCTION

Globally, amphibians are experiencing a variety of threats that are leading to dramatic and severe population declines (e.g., Houlahan et al. 2000; Blaustein and Kiesecker 2002; Hof et al. 2011; IUCN 2019). Habitat loss and fragmentation constitute major threats to amphibian biodiversity in many regions (e.g., Alford and Richards 1999; Hof et al. 2011; Ficetola et al. 2014; IUCN 2019). In Madagascar, habitat destruction and consequential forest fragmentation are occurring on a staggering scale (Achard et al. 2002; Harper et al. 2007; Grinand et al. 2013; Vieilledent et al. 2013; Vieilledent et al. 2018), often resulting in increasingly isolated amphibian populations. In such instances, a major conservation challenge comes in determining whether populations and/or species are distinct and subsequently designating management units for conservation efforts.

Among the amphibians threatened by widespread habitat loss are the unique Malagasy poison frogs (genus *Mantella*), a group of sixteen toxic, diurnal frog species displaying variable coloration and pattern (Glaw and Vences 2007). The bright coloration found in many species in this group is presumed to be aposematic and to function as a warning signal to potential predators (e.g., Schaefer et al. 2002; Chiari et al. 2004). Several species within this group are threatened with extinction, and one in particular, *Mantella cowani*, is especially vulnerable due to its severely restricted range (Vences et al. 1999; Andreone et al. 2005). Historically, *M. cowani* was overharvested for the pet trade, which dramatically impacted populations and led to implementation of a moratorium on its export in 2003 (Andreone et al. 2005; Rabemananjara et al. 2008). Currently, *M. cowani* is known from only a handful of isolated and fragmented locations, none of which occur in protected areas, and therefore merits a high conservation priority (Vences et al. 1999; Andreone et al. 2005; Chiari et al. 2005).

Previous mitochondrial studies indicate that *M. cowani* forms a monophyletic group with three other *Mantella* species: *M. baroni*, *M. nigricans*, and *M. haraldmeieri* (Schaefer et al. 2002). However, nuclear data has not confirmed these relationships and several phenotypically intermediate forms exist among species. Further, there is mitochondrial evidence of hybridization between *M. cowani* and *M. baroni* at one site in the wild where phenotypically intermediate frogs are found (Chiari et al. 2005; Rabemananjara et al. 2007). Mitochondrial analyses conducted on the entire group, including putative hybrids, has suggested that there are three distinct haplotype clades: 1) *M. baroni*, *M. aff. baroni*, *M. nigricans*, and putative hybrids of *M. cowani – M. baroni*, 2) *M. cowani* and putative hybrids of *M. cowani – M. baroni*, and 3) *M. haraldmeieri* (Rabemananjara et al. 2007). Additional work focused solely on populations of *M. cowani* and *M. baroni* has also recovered the genetic differentiation between the two species that has been detected in other studies (Chiari et al. 2005). Despite the

possible hybridization between *M. cowani* and *M. baroni*, mitochondrial studies thus far agree that the two species represent distinct units for conservation.

Genetic data indicating a distinction between *M. cowani* and *M. baroni* is reinforced by obvious morphological differences between the two species. While *M. cowani* contains smaller reddish-orange flank blotches, *M. baroni* is characterized by larger yellow-greenish flank blotches in addition to the presence of a supraocular stripe that is notably absent in *M. cowani*. Additionally, *M. baroni* displays orange and black patterned tibiae, while *M. cowani* exhibits orange-red bands (Chiari et al. 2005; Glaw and Vences 2007). However, little is known about the life history of either species and the adaptive value – if any – of observed phenotypic variation remains unknown.

Given the imminent threats facing *M. cowani* populations, and evidence for hybridization in the wild, it is essential to test mitochondrial relationships with nuclear data in order to clarify the status of *M. cowani* and its relationship to *M. baroni*. In this study, we used a restriction-site associated (RAD) sequencing approach, in combination with phenotypic analysis, to further investigate patterns of genetic and phenotypic diversity among and within putative *M. cowani – M. baroni* hybrid and relevant parental populations. Using a targeted sampling approach, our objectives were to (a) characterize the degree of phenotypic and genetic diversity among and within *M. cowani*, *M. baroni*, and *M. cowani – M. baroni* populations and to (b) determine whether there were consistent genomic and phenotypic signatures of admixture in putative hybrids. Our work has important implications for future conservation efforts for *M. cowani*, as well as broader implications regarding conservation of morphological and genetic biodiversity in endangered species.

MATERIALS AND METHODS

Field Sampling

We sampled three focal populations of *Mantella cowani*, *Mantella baroni*, and putative *Mantella cowani – Mantella baroni* hybrids from eastern Madagascar (Figure 1). To document patterns of phenotypic and genetic diversity within the putative hybrid population, we performed more in-depth sampling across this population (n = 78) relative to *M. cowani* and *M. baroni* populations. Fieldwork was conducted during the rainy season over two years: January – February 2014 and 2015. Frogs were captured by hand and transported back to a field basecamp where all subsequent processing occurred. We collected digital photographs and toe clips from 8 *M. cowani* individuals, 8 *M. baroni* individuals, and 78 putative *M. cowani – M. baroni* hybrid individuals. All sample collection occurred on the same day that frogs were captured. Toe clips were

stored in salt-saturated DMSO at room temperature. After data collection, frogs were held overnight and released to their site of capture the next morning. The Animal Care and Use Committee at the University of California at Berkeley approved all animal handling procedures (R347-0314; AUP-2015-01-7083). Collection and exportation of samples were conducted under permits issued by the Direction de la Conservation de la Biodiversité et du Système des Aires Protégées , Direction Générale des Forêts, and Ministere de l'Environment, de l'Ecologie et des Forêts in Madagascar (collection permits: 315/13/MEF/SG/DGF/DCB.SAP/SCB, 335/14/MEF/SG/DGF/DCB.SAP/SCB, and 336/14/MEF/SG/DGF/DCB.SAP/SCB; export permits: 051C-EA02/MG14 and 048C-EA02/MG15).

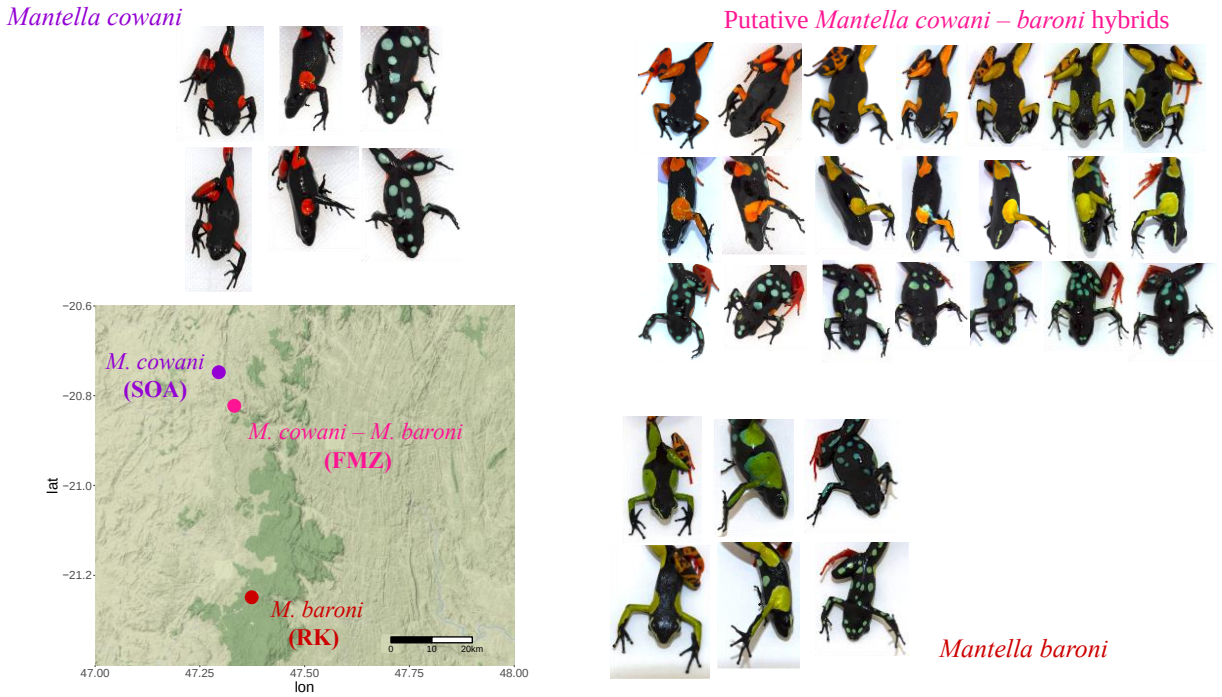

Figure 1: Sampling localities for *Mantella cowani*, *Mantella baroni*, and putative *Mantella cowani – baroni* hybrid populations. Representative individuals from each population are pictured next to species name. Map was provided courtesy of Google Maps and was created using the "ggmap" R package (Kahle and Wickham 2013).

Quantification of Phenotypic Variation

To characterize variation in dorsal and ventral pattern, we photographed the dorsal and ventral surfaces of frogs using a standardized technique based on a protocol modified from Stevens et al. (2007). Frogs were photographed after transport to our field basecamp, but before any other processing occurred. All

photographs were taken with a Pentax K-30 digital single-lens reflex camera fitted with a Pentax 18-135 mm lens between the hours of 1:00-5:00pm in natural light. Frogs were always photographed on a white background with a scale bar and white-gray-black standard (QPcard 101; gray standard reflectance value of 18%) present to account for ambient lighting conditions.

To quantify dorsal and ventral pattern, we used the Multispectral Image Calibration and Analysis Toolbox (MICA, Troscianko and Stevens 2015) implemented within ImageJ v1.49 (Schneider et al. 2012) in a protocol outlined in Klonoski et al. (2019). We first created aligned and normalized images from RAW photographs using the MICA toolbox. After standardization of photographs, we manually outlined the entire dorsal and ventral surfaces of frogs as regions of interest. Entire surfaces were selected in order to obtain comprehensive measures of overall frog pattern. Next, we scaled images to the same number of pixels per millimeter using the scale bar present in all photographs (for dorsal and ventral surfaces = 15.6 px/mm). After scaling, we performed a granularity analysis using the MICA toolbox. For dorsal pattern analysis, we specified Fast Fourier Transform Bandpass filters at 15 levels, starting at two pixels and increasing by a multiple of the square root of two until 290 pixels. For ventral pattern analysis, we specified filters at 14 levels, starting at two pixels and increasing by a multiple of the square root of two until 255 pixels. We then generated descriptive summary statistics from our granularity analysis. For both dorsal and ventral pattern, we estimated pattern contrast, pattern diversity, luminance contrast, and luminance mean (see Troscianko and Stevens 2015). After quantifying pattern characteristics, we performed an ANOVA for each trait to determine if populations were significantly different in dorsal or ventral pattern. Where we detected significant results with our ANOVA, we performed post hoc Tukey's tests to determine which populations differed in each patterning element.

Finally, we also approximated the phenotypic distance between frogs for both dorsal and ventral pattern using the Luminance Distribution Difference Calculator function of the MICA toolbox. We used differences in luminance to quantify variation in dorsal and ventral pattern because patterning elements of focal populations are characterized by discrete patches of high and low luminance values (see Figure 1). To quantify the phenotypic distance between each pair of frogs, we first calculated the number of pixels in each of 95 luminance bins (ranging from 0% to 100% reflectance) for dorsal and ventral surfaces separately. We then used the Luminance Distribution Difference Calculator to compare luminance distribution histograms and generate pairwise comparisons of differences among frogs.

Generation of RAD Libraries

To extract DNA from toe clips, we used Qiagen DNeasy extraction kits (Qiagen, Valencia, CA, USA). We followed the manufacturer's protocol with two

notable exceptions: 4 µl of RNase A was added to samples after lysis, and DNA was eluted in 40 µl of 1X LTE buffer in order to maximize the concentration. Before library preparation, we quantified DNA concentration of samples using a Qubit 3.0 Fluorometer (ThermoFisher, Waltham, MA, USA). We created a restriction-site associated (RAD) library following a protocol outlined in Ali et al. (2016). During library preparation, we digested 50 ng of DNA from each individual using SbfI – HF restriction enzyme (New England Biolabs, Ipswich, MA, USA). Our library was sequenced on one half of a lane of the Illumina HiSeq 4000 at the UC Davis Genome Center with 150 bp paired-end reads.

RADseq Data Processing

To process our RADseq data, we followed a protocol outlined in Klonoski et al. (2019) with limited modifications. Generally, we used pipelines in a customized PERL workflow to process our sequencing data (pipelines accessible at https://github.com/CGRL-QB3-UCBerkeley/RAD). First, we used internal barcode sequences to demultiplex raw fastq reads while allowing for one mismatch in barcode sequence. Demultiplexed reads that did not include the expected restriction enzyme cut site, again allowing for one mismatch in cut site sequence, were removed. We also removed exact duplicates using Super Deduper (https://github.com/dstreett/Super-Deduper). Next, we filtered reads by using Cutadapt (Martin 2011) and Trimmomatic (Bolger et al. 2014) to trim low quality reads and adaptor contamination. We also removed filtered reads that were less than 50 bp. After cleaning and filtering, we clustered forward reads of each individual at 95% similarity using cd-hit (Li and Godzik 2006; Fu et al 2012). We retained only those clusters with at least two supported reads. For each cluster, the sequence that was identified as representative by cd-hit was used as our marker. All retained markers were masked for low complexities, repetitive elements, and short repeats with Ns using RepeatMasker (Smit et al. 2014) and "frog" as a database. We eliminated markers where more than 30% of nucleotides were Ns. We also used Blastn (Altschul et al. 1990) to remove potential paralogs present within each individual by comparing clustered loci against themselves and removing any locus that matched a locus other than itself. All RAD markers that remained from each individual were then combined and clustered across all individuals using cd-hit, and we retained markers shared by at least 40% of all individuals to create our reference genome. After creation of our reference genome, we aligned each individual's cleaned reads to our reference using Novoalign (http://ww.novocraft.com), retaining only those reads that mapped uniquely to the reference. Next, we added read groups and performed realignment around indels using Picard (http://www.picard.sourceforge.net) and GATK (McKenna et al. 2010). We generated quality control information in VCF format using SAMtools/BCFtools

(Li et al. 2009) and then further filtered data with a custom method, SNPcleaner (https://github.com/tplinderoth/ngsQC/tree/master/snpCleaner; Bi et al. 2013) that was slightly modified to eliminate sites around indels. To ensure data quality, we also masked sites within 10 bp of any indel and removed markers where more than two alleles were called at any site. SNPs that failed to pass a one-tailed HWE exact test (1e-4) or that showed strong base quality bias (1e-100) were also removed. For downstream analyses, we only used sites present in at least 70% of individuals with at least 3x coverage in order to avoid bias from excessive missing data.

Quantification of Genetic Variation

To account for uncertainty in our data, we used genotype likelihoods instead of genotype calls whenever possible. We calculated genotype likelihoods using ANGSD (http://www.popgen.dk/angsd/index.php/ANGSD; Korneliussen et al. 2014), a software specifically designed to analyze low to medium coverage next-generation sequencing data. Although many downstream analyses performed in ANGSD are based on genotype likelihoods, genotype posterior probabilities, or likelihood of site allele frequencies, some analyses require SNP calling. When analysis required that SNPs be called, we only used high-confidence variants which were identified by performing a likelihood ratio test to determine variable sites with p-values <1e-3 and genotype posterior probabilities >0.95.

To visualize population genomic structure among populations, we first performed a principal components analysis (PCA) of the covariance matrix of posterior genotype probabilities using ngsTools (http://github.com/mfumagalli/ngsTools; Fumagalli et al. 2014). We performed PCA on (a) all individuals, but only at variable sites (number of sites = 14,311), and (b) only on higher-coverage individuals (i.e., individuals with >7x coverage) with no SNP calling. Next, we quantified population structure using NGSadmix (Skotte et al. 2013), which utilizes genotype likelihoods. Because we sampled three populations in our study, we estimated individual admixture proportions with the number of clusters ranging from one to four ($K = 1 – 4$), with ten replicates per K value. We also calculated F_{ST} values for all pairwise combinations of populations using the realSFS function of ANGSD. We then used the thetaStat function of ANGSD to calculate genetic diversity statistics (Watterson's Theta and Tajima's Estimator) based on folded site frequency spectra generated for each population. To further characterize genetic diversity of populations, we also calculated the relatedness coefficient of each population using ngsRelate (Korneliussen and Moltke 2015). We calculated the relatedness coefficient of populations following the protocol outlined in Krohn et al. (2017).

Finally, we used ngsDist (Vieira et al. 2016) to estimate pairwise genetic distance among individuals using genotype posterior probabilities. After generating a

genetic distance matrix based on ngsDist pairwise comparisons, we also generated distance matrices for geographic distance and for phenotypic distance (both dorsal and ventral pattern). We generated our phenotypic distance matrices using the Luminance Distribution Difference Calculator function of the MICA toolbox as outlined above. After creation of distance matrices, we performed Mantel tests to assess relationships between geography, genetics, and phenotype. We performed matrix regression analyses using geographic, genetic, and phenotypic distance matrices across all samples. We also performed matrix regression analyses solely within the putative hybrid population (FMZ) using genetic and phenotypic distance matrices. All Mantel tests were performed in R version 3.6.1 using the vegan package (R Core Team, 2019). To visualize these comparisons, we plotted pairwise geographic distance versus pairwise phenotypic and geographic distance and fitted a regression line to our plots. We also plotted pairwise genetic distance against phenotypic distance across all samples, and exclusively within the putative hybrid population.

RESULTS

Our sequencing efforts generated 350,677,797 reads that passed quality filtering. After additional filtering, we recovered an average of 4,816,748 high-quality sites on 36,372 contigs. Our final reference genome, which was created by clustering loci across all samples and retaining those markers shared by at least 40% of individuals, contained 39,265 loci. The average coverage per sample was 12.2x and ranged from 1.9x to 36.8x. For ANGSD analysis, we used only those sites present in at least 70% of individuals with at least 3x coverage.

Our PCA and NGSadmix analysis revealed no population genetic structure among putative parental (*M. cowani* and *M. baroni*) and hybrid (*M. cowani-M. baroni*) populations (Figure 2; Figure 3). Despite the lack of genetic structure, we did detect phenotypic differences among populations, particularly for dorsal patterning elements. For dorsal pattern, "pure" *M. cowani* individuals were significantly differentiated from both "pure" *M. baroni* and putative hybrid individuals in pattern contrast, luminance contrast, and mean luminance (Figure 4). Putative hybrids were not significantly different from "pure" *M. baroni* individuals in any of the pattern traits that we quantified, though hybrid individuals did typically demonstrate greater variability in dorsal patterning elements than either of the "pure" *M. cowani* or *M. baroni* populations (Figure 4). Populations were less phenotypically differentiated in ventral patterning traits, with similar trait values for pattern diversity and mean luminance. *M. cowani* and *M. baroni* were significantly different from each other in pattern and luminance contrast, but putative hybrids were not significantly different from either "pure" population (Figure 5).

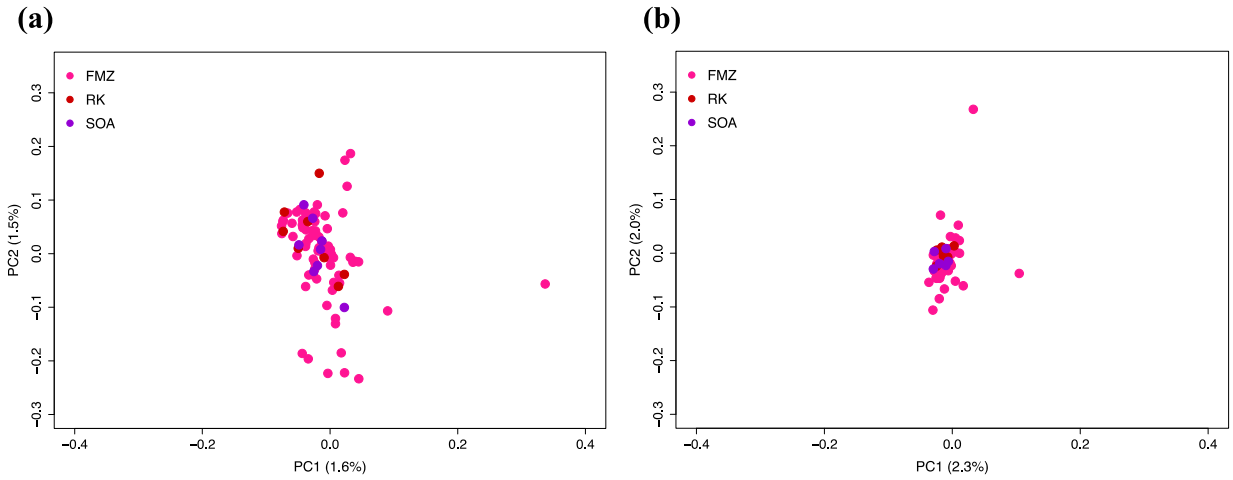

Figure 2: Principal components analysis (PCA) of genetic variation among individuals from all sampling localities. In panel a, PCA was performed on all individuals at 14,311 variable sites. In panel b, PCA was performed only on those individuals with greater than 7x coverage and with no SNP calling. Patterns of genetic variation were consistent across PCA analyses with no genetic structure observed among populations of *M. cowani* (SOA), *M. baroni* (RK), and putative *M. cowani* – *M. baroni* hybrids (FMZ).

Within this complex of populations, our ngsDist and Mantel analysis did not reveal a pattern of isolation by distance, as genetic distance was not correlated with geographic distance (r = -0.013; p = 0.430; see Table 1 and Figure 6). Geographic distance was, however, positively correlated with variation in dorsal pattern (r = 0.086; p = 0.049; see Table 1 and Figure 6), though this was not a particularly strong association. Variation in ventral pattern – which was less pronounced according to our phenotypic analysis – was not correlated with geographic distance (Table 1; Figure 6). Neither dorsal nor ventral pattern variation was associated with genetic distance (Table 1; Figure 7). Within the putative *M. cowani-M. baroni* hybrid population, where we had more in-depth sampling, we did not detect any associations between phenotypic diversity (either dorsal or ventral pattern) and genetic distance (Figure 7).

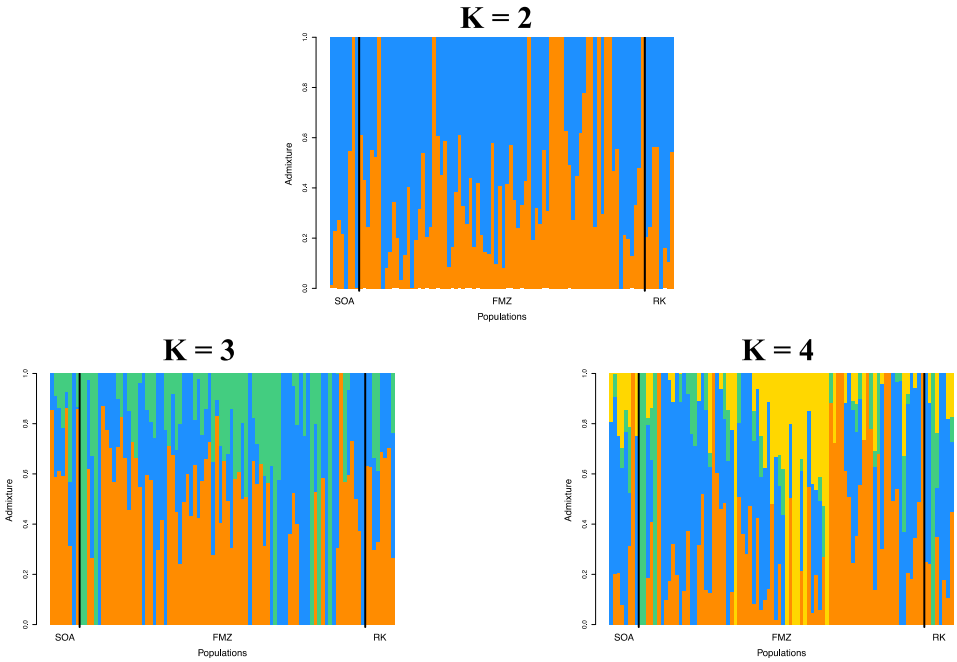

Figure 3: NGSadmix analysis performed across all individuals from all sampling localities (SOA: *M. cowani*, FMZ: putative *M. cowani* – *M. baroni* hybrids, RK: *M. baroni*). NGSadmix results are pictured for $K = 2$, $K = 3$, and $K = 4$. Solid black lines separate populations. NGSadmix analysis did not recover population structure for any value of K that we tested.

Mantel test	***r***	***p***
Geographic dist. x genetic dist.	-0.013	0.430
Geographic dist. x phenotypic dist. (dorsal pattern)	**0.086**	**0.049**
Geographic dist. x phenotypic dist. (ventral pattern)	0.056	0.18
Genetic dist. x phenotypic dist. (dorsal pattern)	-0.037	0.708
Genetic dist. x phenotypic dist. (ventral pattern)	-0.037	0.688

Table 1: Matrix regression results of comparisons among geographic, phenotypic, and genetic distance across all populations. Bolded r and p values indicate significant associations. The only significant association we detected was between geographic distance and variation in dorsal pattern ($r = 0.086$; $p = 0.049$).

Finally, F_{ST} values among populations were generally low (Table 2), although the F_{ST} value was highest between the "pure" *M. baroni* (RK) and *M. cowani* (SOA) populations. For all populations, values of genetic diversity were similarly low (Table 3). Additionally, measures of genetic relatedness within populations was exceptionally high for all focal populations (Table 3).

	FMZ (*M. cowani – M. baroni*)	**SOA (*M. cowani*)**
SOA (*M. cowani*)	0.033	
RK (*M. baroni*)	0.036	0.067

Table 2: F_{ST} values for all populations.

	θ_W	π	**Relatedness**	**n**
FMZ	0.149	0.129	0.420	78
RK	0.122	0.118	0.379	8
SOA	0.125	0.120	0.398	8

Table 3: Values of genetic diversity (Watterson's Theta, θ_W, and Tajima's Estimator, π), the relatedness coefficient, and sample sizes (n) for populations of *M. cowani* (SOA), *M. baroni* (RK), and putative *M. cowani – M. baroni* hybrids (FMZ).

DISCUSSION

Our investigation of phenotypic and genetic diversity among *M. cowani*, *M. baroni*, and putative *M. cowani – M. baroni* hybrids revealed a lack of genetic structure among populations, despite phenotypic differences. Our genomic dataset did not recover the differentiation between *M. cowani* and *M. baroni* that has been detected in previous mitochondrial studies, nor did we detect signatures of genomic admixture in putative hybrid individuals. However, we did find that *M. cowani* was phenotypically distinct from *M. baroni* and putative hybrids, particularly in dorsal patterning elements. Below, we discuss the relevance of our findings for conservation and consider the evolutionary processes that may be contributing to the discordant patterns of genetic and phenotypic variation observed here.

Patterns of Phenotypic and Genetic Diversity

Our phenotypic analyses revealed differences among populations, especially in dorsal pattern. *M. cowani* populations were consistently differentiated from both *M. baroni* and *M. cowani – M. baroni* populations in terms of dorsal pattern contrast, mean luminance and luminance contrast (Figure 4). Although *M. cowani – M. baroni* putative

hybrids were not significantly different from *M. baroni* individuals for any aspects of dorsal pattern that we quantified, putative hybrids did typically demonstrate a wider range of variability in patterning traits. Variation in ventral pattern was less pronounced among populations, though *M. cowani* and *M. baroni* were significantly differentiated in luminance and pattern contrast (Figure 5). This is consistent with our Mantel analysis results, which indicated a marginally significant association between variation in dorsal pattern – but not ventral pattern – and geographic distance (Table 1; Figure 6). We did not characterize some additional phenotypic traits due to limitations of our photographic analysis methodology (e.g., leg pattern, arm coloration, and presence/absence of a supraocular stripe). Characterization of these phenotypic elements may yield additional signatures of differentiation among populations.

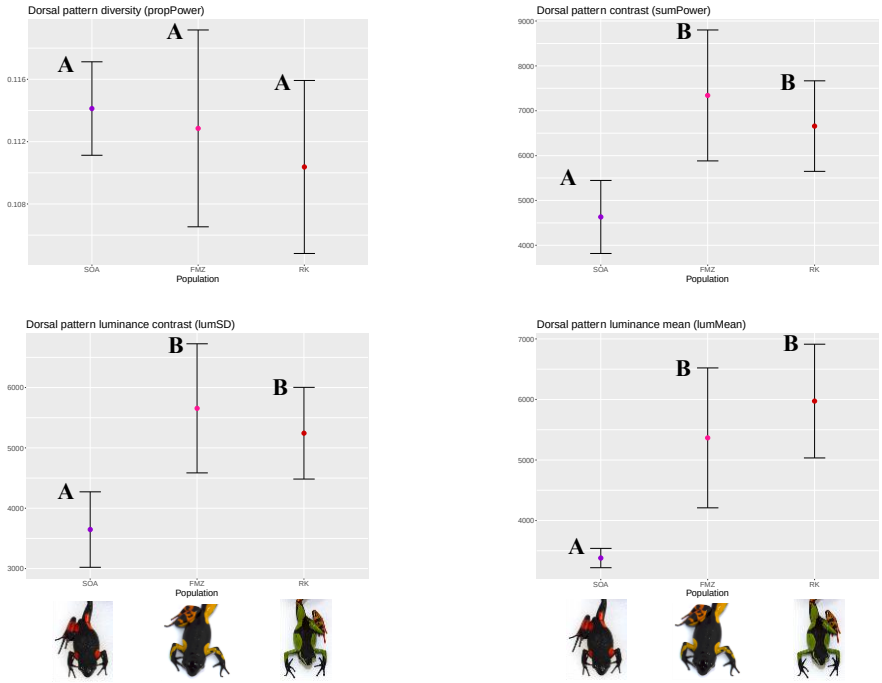

Figure 4: Variation in dorsal pattern (pattern diversity, pattern contrast, luminance contrast, and luminance mean) across populations. For each pattern trait, mean values and standard deviation are displayed for each population. Different letters next to error bars designate significantly different groups based on post hoc Tukey's tests. In most dorsal pattern traits quantified here, SOA was significantly different from both FMZ and RK.

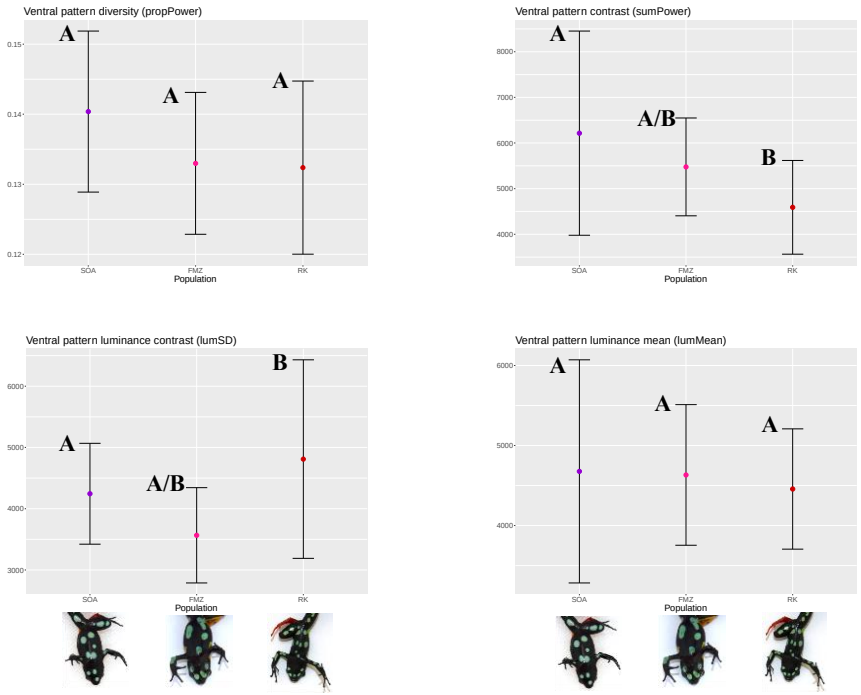

Figure 5: Variation in ventral pattern (pattern diversity, pattern contrast, luminance contrast, and luminance mean) across populations. For each pattern trait, mean values and standard deviation are displayed for each population. Different letters next to error bars designate significantly different groups based on post hoc Tukey's tests. Less population-level variation was observed in ventral pattern traits than in dorsal pattern, though SOA and RK were significantly different from each other in both ventral pattern and luminance contrast.

Despite the phenotypic variability present among populations, we did not detect genome-wide differentiation. Our genomic dataset indicates a lack of genetic structure among *M. cowani*, *M. baroni*, and putative *M. cowani-M. baroni* hybrids and contradicts earlier mitochondrial studies on this complex (Figure 2; Figure 3). Unsurprisingly, given the lack of genetic structure, we did not observe any significant associations between variation in either dorsal or ventral pattern and genetic distance, either across all samples or exclusively within the putative hybrid population (Table 1; Figure 7).

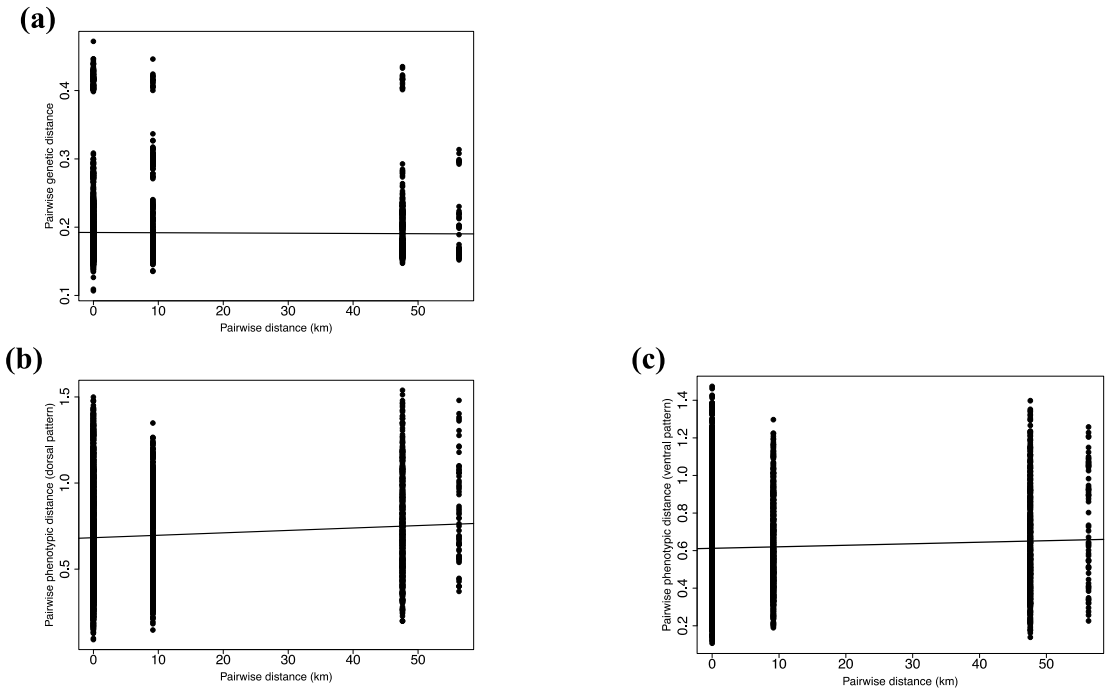

Figure 6: Graphs depicting the relationship between geographic distance and genetic distance (panel a), variation in dorsal pattern (panel b), and variation in ventral pattern (panel c) across all populations. Each point represents a pairwise comparison of two individuals. Regression lines are indicated on the plots. Mantel tests confirmed that geographic distance was not correlated with genetic distance or ventral pattern variation, though it was correlated with variation in dorsal pattern (panel b; $r = 0.086$; $p = 0.049$; see Table 1).

Our findings are consistent with a growing body of evidence that phenotypic diversity is not always accompanied by genome-wide divergence (e.g., Toews et al. 2016; Krohn et al. 2019). One explanation for this phenomenon is ongoing migration between populations. There are many examples of phenotypic differentiation despite ongoing gene flow (e.g., Saint-Laurent et al. 2003; Fitzpatrick et al. 2014; Toews et al. 2016) and persistence of divergent phenotypes in such instances may indicate an adaptive value. The low F_{ST} values that we detected among focal populations (Table 2), combined with the absence of a pattern of genetic isolation by distance (Table 1; Figure 6), are consistent with a scenario of gene flow among populations. However, we did not explicitly quantify migration among populations and cannot distinguish between historical or contemporary gene flow. Although *M. baroni* and *M. cowani* are typically currently found in fragmented and non-overlapping habitat (Chiari et al. 2005), it is important to note that Madagascar's heavy anthropogenic influence has resulted in a complex history of land use, with large changes occurring over rapid timescales (Scales 2011; Vieilledent et al. 2018). When considered in light of the fact that many *Mantella* populations, including those of *M. cowani*, have not been historically or contemporarily well surveyed, ascertaining the impact of past and current landscape changes on gene flow and the geographic distributions of these

species is difficult. Demographic modeling will be essential in elucidating historical and contemporary patterns of migration and gene flow in this group.

Alternatively, the lack of genome-wide differentiation among phenotypically divergent populations could also be explained if color and pattern differences are dictated by relatively few genes. The genetic basis of coloration and pattern has been demonstrated in several taxonomic groups (e.g., Hoekstra et al. 2006; Rosenblum et al. 2010; Wittkopp et al. 2003). In many of these instances, a few genes control coloration and thus a small number of genomic regions are subject to divergent selection (e.g., Steiner et al. 2007; Papa et al. 2008) when variable phenotypes are functionally significant. While the adaptive value of phenotypic variation remains unknown for Malagasy poison frogs, previous work suggests that different color and pattern variants may be important signals to local predators (Noonan and Comeault 2008). Given the maintenance of distinct phenotypes despite genomic similarity, scans for outlier loci could be useful in identifying candidate genomic regions associated with color and pattern differences in this group.

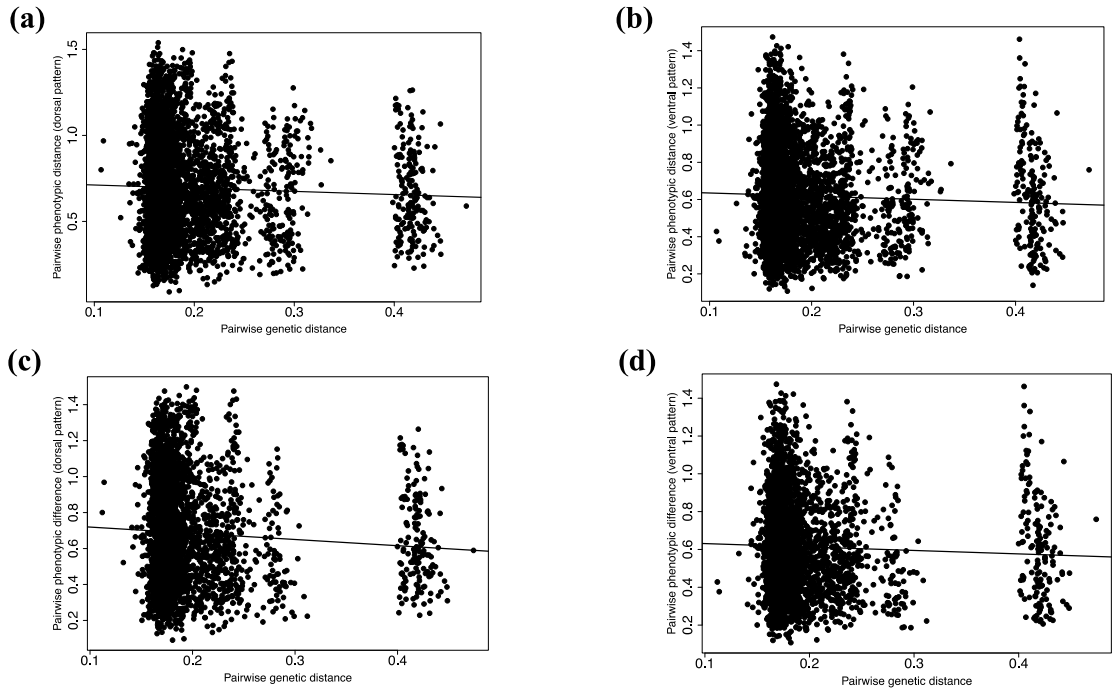

Figure 7: Graphs depicting the relationship between genetic distance and variation in dorsal and ventral pattern across all populations (panels a and b), and within a putative hybrid population (FMZ; panels c and d). Each point represents a pairwise comparison between two individuals. Regression lines are indicated on the plots. Mantel tests confirmed that there was no significant relationship between genetic distance and phenotypic distance, either across all populations or within FMZ (see Table 1).

Conservation Implications

Each of our focal populations demonstrated similarly low levels of genetic diversity and remarkably high levels of relatedness within populations (Table 3), which is concerning interpreted in light of continuing forest fragmentation and habitat loss. Given the unique phenotypic characteristics and extremely limited occurrences of populations known as *M. cowani*, conservation decisions regarding this species should not be made lightly. Thus, we believe that detailed genomic studies across more populations within the *M. cowani* group are necessary before any management decisions are made. To that end, we emphasize that the dataset presented here is preliminary in its ability to contextualize genetic relationships within the *M. cowani* group overall.

Although our sampling is limited to one population of *M. cowani*, one population of *M. baroni*, and one population of putative *M. cowani – M. baroni* hybrids, the genetic similarity that we detected across populations is interesting from a conservation perspective. Despite earlier genetic studies, our genomic data did not find signatures of admixture between "pure" *M. cowani* and *M. baroni* populations in phenotypically intermediate *M. cowani – M. baroni* frogs (Figure 3). One possible explanation is that hybridization between *M. baroni* and *M. cowani* has been longstanding and stable, leading to genetic homogenization across populations, as in Toews et al. (2016). Considering our genomic data, it is difficult to determine whether phenotypically intermediate frogs should be considered as hybrids or as morphological variants of *M. cowani* and/or *M. baroni*.

Ultimately, our work here raises the intriguing question of how different aspects of biological diversity should be prioritized in designating management units for conservation. Although populations appear to be genomically similar, *M. cowani* is clearly phenotypically distinct. Additionally, individuals from the putative *M. cowani-M. baroni* hybrid population demonstrate novel phenotypes not exhibited by either *M. cowani* or *M. baroni*. In the presence of discordant patterns of genotypic and phenotypic diversity, how should morphological and genetic variants be parsed in terms of conservation priority?

While this question applies to many endangered species, it is especially significant in aposematic organisms such as Malagasy poison frogs, where phenotypes have clear ecological relevance. Although the adaptive value of phenotypic variation in aposematic frogs is largely unknown, there is evidence that different variants may be specifically adapted to local predator communities (Noonan and Comeault 2008; Chouteau and Angers 2011). In cases such as these, conservation of morphological diversity may be as important as maintaining genetic variation if distinct phenotypes are essential in ensuring protection against predators. Even though the adaptive value of divergent phenotypes is unclear for our focal populations, *M. cowani* clearly represents a unique morphological variant and thus warrants protection. Similarly,

individuals from the putative *M. cowani – M. baroni* hybrid population demonstrated novel phenotypes and high intrapopulation variability in pattern. For these reasons, we recommend that *M. cowani – M. baroni* individuals also be considered as a morphologically unique population worthy of protection, even despite their genetic similarity to *M. cowani* and *M. baroni* individuals.

Within the *Mantella* genus in particular, previous studies have consistently documented discordant relationships between phenotypic and genetic diversity, and have indicated that relationships between these two aspects of biodiversity are not straightforward in this group (Crottini et al. 2019; Klonoski et al. 2019). Even in the *M. cowani* group itself, mitochondrial data suggests that *M. baroni*, *M. nigricans*, and *M. aff. baroni* constitute a single species despite dramatic phenotypic variation and the existence of multiple intermediate forms (Rabemananjara et al. 2007). Our findings here add to this growing body of work, and highlight the complicated nature of designating conservation units in systems with exceptionally high phenotypic diversity among genetically similar populations. Consequently, we recommend that detailed information on both morphology and genetics should be used to inform conservation decisions for this group and that genetic data alone is not sufficient to designate management units. Given our limited understanding of the extreme phenotypic variation that is often observed among and within *Mantella* species, basic life history information and additional studies on the functional significance of color and pattern differences will also be critical in deciding which aspects of phenotypic variation are most important to conserve.

ACKNOWLEDGEMENTS

Ke Bi and Erica Bree Rosenblum co-authored this dissertation chapter. We thank Serge Ndriantsoa for his tremendous help in the field. We are also grateful to Devin Edmonds, Molly Bletz, and Miguel Vences for guidance about working in Madagascar and for support in the field. We thank numerous local guides for sharing their incredible knowledge and for their assistance in the field. We also thank Sean O'Rourke for his assistance in creating RAD libraries for this project. This work was conducted under a collaboration agreement between the Zoological Institute of the Technical University of Braunschweig, the Département de Biologie Animale of the Université d'Antananarivo, and the Ministère de l'Environnement et des Forêts, Direction du Système des Aires Protégées. We are grateful to the Direction Générale des Forêts, Ministere de l'Environment et des Forêts, and Madagascar National Parks for providing collection and export permits. We also thank MICET and the Université d'Antananarivo for logistical support in Madagascar. Funding was provided by the Society for the Study of Evolution (the Rosemary Grant Graduate

Student Research Award), the National Geographic Society, the Museum of Vertebrate Zoology at the University of California, Berkeley, and the Department of Environmental Science, Policy, and Management at the University of California, Berkeley. All animal work performed as part of this study was approved by the Animal Care and Use Committee at the University of California at Berkeley (R347-0314; AUP-2015-01-7083).

LITERATURE CITED

Achard, F., H. D. Eva, H-J. Stibig, P. Mayaux, J. Gallego, T. Richards, and J-P. Malingreau. 2002. Determination of deforestation rates of the world's humid tropical forests. Science 297(5583):999-1002.

Alford, R. A., and S. J. Richards. 1999. Global amphibian declines: a problem in applied ecology. Annual Review of Ecology and Systematics 30(1):133-165.

Ali, O. A., S. M. O'Rourke, S. J. Amish, M. H. Meek, G. Luikart, C. Jeffres, and M. R. Miller. 2016. RAD capture (rapture): flexible and efficient sequence-based genotyping. Genetics 202(2):389-400.

Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. 1990. Basic local alignment search tool. Journal of Molecular Biology 215(3):403-410.

Andreone, F., J. E. Cadle, N. Cox, F. Glaw, R. A. Nussbaum, C. J. Raxworthy, S. N. Stuart, D. Vallan, and M. Vences. 2005. Species review of amphibian extinction risks in Madgascar: conclusions from the Global Amphibian Assessment. Conservation Biology 19(6):1790-1802.

Bi, K., T. Linderoth, D. Vanderpool, J. M. Good, R. Nielsen, and C. Moritz. 2013. Unlocking the vault: next generation museum population genomics. Molecular Ecology 22(24):6018-6032.

Blaustein, A. R., and J. M. Kiesecker. 2002. Complexity in conservation: lessons from the global decline of amphibian populations. Ecology Letters 5(4):597-608.

Bolger, A. M., M. Lohse, and B. Usadel. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30(15):2114-2120.

Chiari, Y., F. Andreone, M. Vences, and A. Meyer. 2005. Genetic variation of an endangered Malagasy frog, *Mantella cowani*, and its phylogeographic relationship to the widespread *M. baroni*. Conservation Genetics 6(6):1041-1047.

Chiari, Y., M. Vences, D. R. Vieites, F. Rabemananjara, P. Bora, O. Ramilijaona Ravoahangimalala, and A. Meyer. 2004. New evidence for parallel evolution of colour patterns in Malagsay poison frogs (*Mantella*). Molecular Ecology 13(12):3763-3774.

Chouteau, M., and B. Angers. 2011. The role of predators in maintaining the geographic organization of aposematic signals. The American Naturalist 178(6):810-817.

Crottini, A., P. Orozco-terWengel, F. C. E. Rabemananjara, J. S. Hauswaldt, and M. Vences. 2019. Mitochondrial introgression, color pattern variation, and severe demographic bottlenecks in three species of Malagasy poison frogs, genus *Mantella*. Genes 10(4):317.

Fitzpatrick, S. W., J. C. Gerberich, J. A. Kronenberger, L. M. Angeloni, and W. C. Funk. 2014. Locally adapted traits maintained in the face of high gene flow. Ecology Letters 18(1):37-47.

Fu, L., B. Niu, Z. Zhu, S. Wu, and W. Li. 2012. CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics 28(23):3150-3152.
Fumagalli, M., F. G. Vieira, T. Linderoth, and R. Nielsen. 2014. ngsTools: methods for population genetics analyses from next-generation sequencing data. Bioinformatics 30(10):1486-1487.
Glaw, F., and M. Vences. 2007. A field guide to the amphibians and reptiles of Madagascar. 3rd ed. Cologne: Vences and Glaw.
Grinand, C., F. Rakotomalala, V. Gond, R. Vaudry, M. Bernoux, and G. Vieilledent. 2013. Estimating deforestation in tropical humid and dry forests in Madagascar from 2000 to 2010 using multi-date Landsat satellite images and the random forests classifier. Remote Sensing of Environment 139:68-80.
Harper, G. J., M. K. Steininger, C. J. Tucker, D. Juhn, and F. Hawkins. 2007. Fifty years of deforestation and forest fragmentation in Madagascar. Environmental Conservation 34(4):1-9.
Hof, C, M. B. Araujo, W. Jetz, and C. Rahbek. 2011. Additive threats from pathogens, climate and land-use change for global amphibian diversity. Nature 480(7378):516-519.
Hoekstra, H. E., R. J. Hirschmann, R. A. Bundey, P. A. Insel, and J. P. Crossland. 2006. A single amino acid mutation contributes to adaptive beach mouse color pattern. Science 313(5783):101-104.
Houlahan, J. E., C. S. Findlay, B. R. Schmidt, A. H. Meyer, and S. L. Kuzmin. 2000. Quantitative evidence for global amphibian population declines. Nature 404(6779):752-755.
IUCN. 2019. The IUCN Red List of Threatened Species. Version 2019-2. http://www.iucnredlist.org
Kahle, D., and H. Wickham. 2013. ggmap: A package for spatial visualization with Google Maps and OpenStreetMap. R package version 2.3. R Foundation for Statistical Computing, Vienna, Austria.
Klonoski, K., K. Bi, and E. B. Rosenblum. 2019. Phenotypic and genetic diversity in aposematic Malagasy poison frogs (genus *Mantella*). Ecology and Evolution 9(5):2725-2742.
Korneliussen, T. S., A. Albrechtsen, and R. Nielsen. 2014. ANGSD: analysis of next generation sequencing data. BMC Bioinformatics 15(1):356.
Korneliussen, T. S., and I. Moltke. 2015. NgsRelate: a software tool for estimating pairwise relatedness from next-generation sequencing data. Bioinformatics 31(24):4009-4011.
Krohn, A. R., C. J. Conroy, R. Pesapane, K. Bi, J. E. Foley, and E. B. Rosenblum. 2017. Conservation genomics of desert dwelling California voles (*Microtus californicus*) and implications for management of endangered Amargosa voles (*Microtus californicus scirpensis*). Conservation Genetics 19(2):383-395.
Krohn, A. R., E. T. Diepeveen, K. Bi, and E. B. Rosenblum. 2019. Local adaptation

does not lead to genome-wide differentiation in lava flow lizards. Ecology and Evolution 9(12):6810-6820.
Li, W., and A. Godzik. 2006. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics 22(13):1658-1659.
Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin, and 1000 Genome Project Data Processing Subgroup. 2009. The sequence alignment/map format and SAMtools. Bioinformatics 25(16):2078-2079.
Martin, M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal 17(1):10-12.
McKenna, A., M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, A. Kernytsky, K. Garimella, D. Altshuler, S. Gabriel, M. Daly, and M. A. DePristo. 2010. The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Research 20(9):1297-1303.
Noonan, B. P., and A. A. Comeault. 2009. The role of predator selection on polymorphic aposematic poison frogs. Biology Letters 5(1):51-54.
Papa, R., A. Martin, and R. D. Reed. 2008. Genomic hotspots of adaptation in butterfly wing pattern evolution. Current Opinion in Genetics & Development 18(6):559-564.
R Core Team. 2019. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.
Rabemananjara, F. C. E., Y. Chiari, O. Ravoahangimalala Ramilijaona, and M. Vences. 2007. Evidence for recent gene flow between north-eastern and south-eastern Madagascan poison frogs from a phylogeography of the *Mantella cowani* group. Frontiers in Zoology 4:1.
Rabemananjara, F. C. E., N. Rasoamampionona Raminosoa, O. Ravoahangimalala Ramilijaona, D. Rakotondravony, F. Andreone, P. Bora, A. I. Carpenter, F. Glaw, T. Raxafindrabe, D. Vallan, D. R. Vieites, and M. Vences. 2008. Malagasy poison frogs in the pet trade: a survey of levels of exploitation of species in the genus *Mantella*. Amphibian and Reptile Conservation 5:3-16.
Rosenblum, E. B., H. Rompler, T. Schoneberg, and H. E. Hoekstra. 2010. Molecular and functional basis of phenotypic convergence in white lizards at White Sands. Proceedings of the National Academy of Sciences 107(5):2113-2117.
Saint-Laurent, R., M. Legault, and L. Bernatchez. 2003. Divergent selection maintains adaptive differentiation despite high gene flow between sympatric rainbow smelt ecotypes (*Osmerus mordax* Mitchill). Molecular Ecology 12(2):315-330.
Scales, I. R. 2011. Farming at the forest frontier: Land use and landscape change in Western Madagascar, 1896-2005. Environment and history 17(4):499-524.
Schaefer, H. C., M. Vences, and M. Veith. 2002. Molecular phylogeny of Malagasy

poison frogs, genus *Mantella* (Anura: Mantellidae): homoplastic evolution of colour pattern in aposematic amphibians. Organisms Diversity & Evolution 2(2):97-105.
Schneider, C. A., W. S. Rasband, and K. W. Eliceiri. 2012. NIH Image to ImageJ: 25 years of image analysis. Nature Methods 9(7):671-675.
Skotte, L., T. S. Korneliussen, and A. Albrechtsen. 2013. Estimating individual admixture proportions from next generation sequencing data. Genetics 195(3):693-702.
Smit, A. F. A., R. Hubley, and P. Green. 2014. RepeatMasker Open-4.0. http://www.repeatmasker.org.
Steiner, C. C., J. N. Weber, and H. E. Hoekstra. 2007. Adaptive variation in beach mice produced by two interacting pigmentation genes. PLoS Biology 5(9):e219.
Stevens, M., C. A. Párraga, I. C. Cuthill, J. C. Partridge, and T. S. Troscianko. 2007. Using digital photography to study animal coloration. Biological Journal of the Linnean Society 90(2):211-237.
Toews, D. P. L., S. A. Taylor, R. Vallender, A. Brelsford, B. G. Butcher, P. W. Messer, and I. J. Lovette. 2016. Plumage genes and little else distinguish the genomes of hybridizing warblers. Current Biology 26(17):2313-2318.
Troscianko, J., and M. Stevens. 2015. Image calibration and analysis toolbox – a free software suite for objectively measuring reflectance, colour and pattern. Methods in Ecology and Evolution 6(11):1320-1331.
Vences, M., F. Glaw, and W. Bohme. 1999. A review of the genus *Mantella* (Anura, Ranidae, Mantellinae): taxonomy, distribution and conservation of Malagasy poison frogs. Alytes 17:3-72.
Vieilledent, G., C. Grinand, F. A. Rakotomalala, R. Ranaivosoa, J-R. Rakotoarijaona, T. F. Allnutt, and F. Achard. 2018. Combining global tree cover loss data with historical national forest cover maps to look at six decades of deforestation and forest fragmentation in Madagascar. Biological Conservation 222:189-197.
Vieilledent, G., C. Grinand, and R. Vaudry. 2013. Forecasting deforestation and carbon emissions in tropical developing countries facing demographic expansion: a case study in Madagascar. Ecology and Evolution 3(6):1702-1716.
Vieira, F. G., F. Lassalle, T. S. Korneliussen, and M. Fumagalli. 2016. Improving the estimation of genetic distances from Next-Generation Sequencing data. Biological Journal of the Linnean Society 117(1):139-149.
Wittkopp, P. J., S. B. Carroll, and A. Kopp. 2003. Evolution in black and white: genetic control of pigment patterns in *Drosophila*. TRENDS in Genetics 19(9):495-504.