INVESTIGATING THE MICROEVOLUTIONARY PROCESSES THAT PROMOTE AND

MAINTAIN BIODIVERSITY AND TRAIT NOVELTY

by

THOMAS J FIRNENO JR.

Presented to the Faculty of the Graduate School of

The University of Texas at Arlington in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

THE UNIVERSITY OF TEXAS AT ARLINGTON

AUGUST 2021

ACKNOWLEDGEMENTS

my fiancé, Jacob Wadsworth, I know that the past five years have been difficult – being apart from each other has not been ideal or easy – but our love for each other has grown, as have we both as individuals and together. Thank you for allowing me to pursue my passion. No matter what, your love is with me everywhere I go, this I know. I also have to give a special shout out to our cat children, Fitz, Mo, and Milo, who always kept me entertained, warm, and in good company throughout this process. Finally, thank you to my advisor, Matt Fujita. Your support, encouragement, and patience were boundless throughout my time in your lab. I truly could not have asked for a better mentor and friend, and I can only hope to be half the advisor, mentor, and scientist you were to me and my fellow lab members.

<div align="right">July 8th, 2021</div>

DEDICATION

To every science teacher and mentor that instilled in me a passion for scientific inquiry, as well as teaching, mentoring, and passing that passion along to other generations of students. Especially, Diane Jones, Colleen Gatrone, and Eugene Gowisnok.

ABSTRACT

INVESTIGATING THE MICROEVOLUTIONARY PROCESSES THAT PROMOTE AND

MAINTAIN BIODIVERSITY AND TRAIT NOVELTY

Thomas Joseph Firneno Jr.

The University of Texas at Arlington, 2021

Supervising Professor: Matthew K. Fujita, PhD

Gene flow, mutation, selection, and genetic drift influence patterns of speciation, adaptation, and biodiversity. In the age of genomics, we have been increasingly able to understand the interplay between these microevolutionary processes at the genomic level, and their link to the phenotype and the environment. This, in turn, has provided clarity to the patterns of diversity and biological innovation that these processes generate. My dissertation examines how these microevolutionary processes have influenced patterns of diversity and the evolution of a novel trait in true toads (Anura: Bufonidae). First, I explore what processes (gene flow, incomplete lineage sorting, poor taxonomic resolution) have caused patterns of mitonuclear discordance in a closely related species complex of Mesoamerican toads. I then investigate how this pattern of discordance may have hampered the taxonomic resolution of the species in this complex. Finally, I use North American bufonids to explore potential gene and pathway involvement in defensive toxin synthesis.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER 1: INTRODUCTION

The ways in which microevolutionary forces, such as gene flow, mutation, selection, and genetic drift influence patterns of speciation, adaptation, and biodiversity have long interested scientists. These processes and the patterns that they leave behind have been shown and studied within many organismal systems [melanism in peppered moths (Kettlewell, 1973; Majerus, 1988; Coyne, 1998; Cook et al., 2012), adaptation to predation in Trinidadian guppies (Endler, 1980; Reznick et al. 1997), pesticide and antibiotic resistance (Tabashnik, 1994; Baquero & Blazquez, 1997), adaptive radiation in *Anolis* lizards (Losos et al., 2001), and threespine stickleback adaptation (Bell, 2001)]. Moreover, in the age of genomics, we have been increasingly able to understand the interplay between these microevolutionary processes at the genomic level, and their link to the phenotype and the environment. This, in turn, has provided clarity to the patterns of diversity and biological innovation that these processes generate, and continues to provide us with more examples and hypotheses as these processes are further explored in other organismal systems [for example: transmission dynamics of diseases (Kato-Maeda et al., 2013; Roy et al., 2020), genomic adaptation of melanism to variable environments (Nachmann et al., 2003; Mullen and Hoekstra, 2008; Rosenblum et al., 2010), venom evolution and production antivenom therapies (Casewell et al., 2012; Aird et al., 2017; Casewell et al., 2020), adaptation to urbanization/anthropogenic change (Harris & Munshi-South 2017; Theodorou et al., 2018; Homola et al., 2019; Waldvogel et al., 2020), and general patterns of biodiversity (e.g. population structure, gene flow, phylogeography) and conservation management (Mastretta-Yanes et al., 2018; Supple & Shapiro, 2018; Funk et al., 2019; Price et al., 2021)].

My dissertation focuses on true toads (Anura: Bufonidae), animals that are not traditionally considered charismatic and thus causing them to be often overlooked as scientific study systems, but their patterns of diversity have long been believed to be highly influenced by microevolutionary forces, especially gene flow. Bufonid toads are a cosmopolitan group of amphibians distributed across the globe (with the exception of Antarctica), most likely finding their origins in South America during the late Cretaceous to early Paleocene (78-98 Mya), making them a relatively recent radiation of anurans. They subsequently radiated out of South America during the Eocene through the Oligocene, rapidly diverging from each other likely driven by intense climatic fluctuations throughout the Paleogene (Pramuk et al., 2008; Kok et al., 2018). Because of their overall conserved morphology, recent historical divergence, propensity for hybridization, and adaptations to numerous environments, the taxonomy of true toads has been tumultuous, leaving their systematics relatively understudied. However, because of this, they are presented as an excellent system to study the patterns of diversity that have been caused or influenced by microevolutionary forces. In Chapter 2, I explore the potential cause(s) of mitonuclear discordance (e.g. incomplete lineage sorting, gene flow, poor taxonomic resolution) within a species complex of Central American toads. In Chapter 3, I explore how this pattern of mitonuclear discordance has created problems for species delimitation in this complex and provide an integrative hypothesis testing framework to better understand the systematics of these toads.

Mutation, gene flow, and selection can all affect or initiate the biological innovations that we see phenotypically within or across species. Bufonids are one of at least ten amphibian families (six anuran and four salamander) that have evolved the use of a parotoid gland to store suites of toxins used for defense (Hostetler and Cannon, 1974; Cannon and Hostetler, 1976;

AmphibiaWeb 2021). This phenotypic innovation both of the evolution of the parotoid gland and the evolution of the toxins bufonid toads produce has fascinated scientists for decades (Siperstein et al., 1957; Cei et al. 1968; Mebs et al. 2005 Sciani et al. 2013). Most of these studies have focused on histology and chemical ecology of the parotoid gland and toxins within it (reviewed in Rodríguez et al. 2017). However, very few have studied the underlying genetic mechanisms that are involved in toxin production within these organisms. In Chapter 4, I explore the potential genes and pathways that may be implicated in toxin production across North American species of bufonids using transcriptomics.

Ultimately, this dissertation will hopefully serve as an example of the power and insight that studying the patterns and processes of microevolution have on organismal and evolutionary biology. There is still plenty to learn about micro evolutionary forces and the patterns of biodiversity that they produce. Bufonid toads happen to hold great power as an understudied organismal system to study these forces, as well as speciation dynamics, defensive toxin evolution, adaptation, and many other processes both at the micro- and macroevolutionary scales. This system will hopefully become more enriched as advances in genomic and comparative technologies make it more feasible to use them in studies for studying both adaptation and speciation.

LITERATURE CITED

Aird, S. D., Arora, J., Barua A., Qiu, L., Terada, K., & Mikheyev, A. S. (2017). Population Genomic Analysis of a Pitviper Reveals Microevolutionary Forces Underlying Venom Chemistry. *Genome Biology and Evolution*, 9(10): 2640-2649.

AmphibiaWeb. (2021). <https://amphibiaweb.org> University of California, Berkeley, CA, USA.

Baquero F., & Blazquez, J. (1997). Evolution of antibiotic resistance. *Trends in Ecology and Evolution, 12*: 482-487.

Bell, M. (2001). Lateral plate evolution in the threespine stickleback: getting nowhere fast. *Genetica, 112-113*: 445-461.

Cannon, M. S. & Hostetler, J. R. (1976). The anatomy of the parotoid gland in Bufonidae with some histochemical findings. II. *Bufo alvarius*. *Journal of Morphology, 148*(2): 137-160.

Casewell, N. R., Wüster, W., Vonk, F. J., Harrison, R. A., & Fry, B. G. (2013). Complex cocktails: the evolutionary novelty of venoms. *Trends in Ecology and Evolution*, 28(4): 219-229.

Casewell, N. R., Jackson, T. N. W., Laustsen, A., & Sunagar, K. (2020). Causes and Consequences of Snake Venom Variation. *Trends in Pharmacological Sciences*, 41(8): 570-581.

Cei, J. M., Erspamer, V., & Roseghini, M. (1968). Taxonomic and evolutionary significance of biogenic amines and plypeptides in aphibian skin. II. Toads of the genera *Bufo* and *Melanophryniscus*. *Systematic Zoology, 17*(3): 232-245.

Cook, L. M., Grant, B.S., Saccheri, I. J., & Mallet, J. (2012). Selective bird predation on the peppered moth: the last experiment of Michael Majerus. *Biology Letters*, *8*: 609-612.

Coyne, J. A. (1998). Not black and white. *Nature*, 396: 35-36.

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

Funk, W. C., Forester, B. R., Converse, S. J., Darst, C., & Morey, S. (2019). Improving conservation policy with genomics: a guide to integrating adaptive potential into U.S.

Endangered Species Act decisions for conservation practitioners and geneticists. *Conservation Genetics*, 20: 115-134.

Harris, S. E., & Munshi-South, J. (2017). Signatures of positive selection and local adaptation to urbanization in white-footed mice (*Peromyscus leucopus*). *Molecular Ecology*, 26(22): 6336-6350.

Hostetler J. R. & Cannon, M. S. (1974). The anatomy of the parotoid gland in Bufonidae with some histochemical findings. I. *Bufo marinus*. *Journal of Morphology, 142*(2): 225-239.

Homola, J. J., Loftin, C. S., Cammen, K. M., Helbing, C. C., Birol, I., Schultz, T. F., & Kinnison, M. (2019). Replicated Landscape Genomics Identifies Evidence of Local Adaptation to Urbanization in Wood Frogs. *Journal of Heredity*, 110(6): 707-719.

Kato-Maeda, M., Ho, C., Passarelli, B., Banaei, N., Grinsdale, J., Flores, L., Anderson, J., Murray, M., Rose, G, Kawamura, L. M., Pourmand, N., Tariq, M. A., Gagneux, S., & Hopewell, P. C. (2013). Use of Whole Genome Sequencing to Determine the Microevolution of Mycobacterium tuberculosis during an Outbreak. *PLosONE*, 8(3): e58235.

Kettlewell, H. B. D. (1973). *The Evolution of Melanism*. Clarendon Press, Oxford, UK.

Kok, P. J. R., Ratz, S., MacCulloch, R. D., Lathrop, A., Dezfoulian, R., Aubret, F., & Means, D. B. (2018). Historical biogeography of the paleoendemic toad genus *Oreophrynella* (Amphibia: Bufonidae) sheds new light on the origin of the Pantepui endemic terrestrial biota. *Journal of Biogeography, 45*(1): 26-36.

Losos, J. B., Schoener, T. W., Warheit, K. I., & Creer, D. (2001). Experimental studies of adaptive differentiation in Bahamian *Anolis* lizards. *Genetica, 112-113*: 399-415.

Majerus, M. E. N. (1988). *Melanism: Evolution in Action*. Oxford University Press, Oxford, UK.

Mastretta-Yanes, A., Xue, A. T., Moreno-Letelier, A. M., Jorgensen, T. H., Alvarez, N., Piñero, D., & Emerson, B. C. (2018). Long-term in situ persistence of biodiversity in tropical sky islands revealed by landscape genomics. *Molecular Ecology*, 27(2): 432-448.

Mebs, D., Pogoda, W., Maneyro, R., & Kwet, A. (2005). Studies of the poisonous skin secretion of individual red bellied toads, *Melanophryniscus montevidensis* (Anura, Bufonidae), from Uruguay. *Toxicon, 46*(6): 641-650.

Mullen, L.M., & Hoekstra, H.E. (2008). Natural selection along an environmental gradient: A classic cline in mouse pigmentation. *Evolution*, 62:1555-1570.

Nachmann, M.W., Hoekstra, H.E., & D'Agostino, S.L. (2003). The genetic basis of adaptive melanism in pocket mice. *Proceedings of the National Academy of Sciences*, 100:5268-5273.

Pramuk, J. B., Robertson, T., Sites, J. W., & Noonan, B. P. (2008). Around the world in 10 million years: biogeography of the nearly cosmopolitan true toads (Anura: Bufonidae). *Global Ecology and Biogeography, 17*(1): 72-83.

Price, M. R., Hadfield, M. G., Knapp, I. S. S., Toonen, R. J., & Forsman, Z. H. (2021). Evolutionary genomics of endangered Hawaiian tree snails (Achatinellidae: Achatinellinae) for conservation of adaptive capacity. *PeerJ*, 9:e10993.

Reznick, D. N., Shaw, F. H., Rodd, F. H., & Shaw, R. G. (1997). Evaluation of the rate of evolution in natural populations of guppies (*Poecilia reticulata*). *Science, 275*: 1934-1937.

Rodríguez, C., Rollins-Smith, L., Ibáñez, R., Durant-Archibold, A. A., & Gutiérrez, M. (2017). Toxins and pharmacologically active compounds from species of the family Bufonidae (Amphibia, Anura). *Journal of Ethnopharmacology, 198*: 235-254.

Rosenblum, E.B., Rompler, H., Schoneberg, T., & Hoekstra, H.E. (2010). Molecular and

    functional basis of phenotypic convergence in white lizards at White Sands. *Proceedings*

    of the National Academy of Science, 107:2113-2117.

Roy, C., Mandal, S. M., Mondal, S. K., Mukherjee, S., Mapder, T., Ghosh, W., & Chakraborty,

    R. (2020). Trends of mutation accumulation across global SARS-CoV-2 genomes:

    Implications for the evolution of the novel coronavirus. *Genomics*, 112(6): 5331-5342.

Sciani, J. M., Angeli, C. B., Antoniazzi, M. M., Jared, C., & Pimenta, D. C. (2013). Differences

    and similarities among parotoid macrogland secretions in South American toads: a

    preliminary biochemical delineation. *The Scientific World Journal*, *2013*, 1-9.

Siperstein, M. D., Murray, A. W., & Titus, E. (1957). Biosynthesis of cardiotonic sterols from

    cholesterol in the toad, *Bufo marinus*. *Archives of Biochemistry and Biophysics, 67*(1):

    154-160.

Supple, M. A., & Shapiro, B. Conservation of biodiversity in the genomics era. *Genome Biology*,

    19(131): 2018.

Tabashnik, B. E. (1994). Evolution of resistance to *Bacillus thuringiensis*. *Annual Review of*

    Entomology, 39: 47-79.

Theodorou, P., Radzevičiūtė, R., Kahnt, B., Soro, A., Grosse, I., & Paxton, R. J. (2018).

    Genome-wide single nucleotide polymorphism scan suggests adaptation to urbanization

    in an important pollinator, the red-tailed bumblebee (*Bombus lapidarius* L.). *Proceedings*

    of the Royal Society B, 285(1877): 2017806.

Waldvogel, A., Feldmeyer, B., Rolshausen, G., Exposito-Alonso, M., Rellstab, C., Kofler, R.,

    Mock, T., Schmid, K., Schmitt, I., Batallion, T., Savolainen, O., Bergland, A., Flatt, T.,

Guillaume, F., & Pfenninger, M. (2020). Evolutionary genomics can improve prediction of species' responses to climate change. *Evolution Letters*, 4(1): 4-18.

CHAPTER 2:

FINDING COMPLEXITY IN COMPLEXES: ASSESSING THE CAUSE OF MITONUCLEAR DISCORDANCE IN A PROBLEMATIC SPECIES COMPLEX OF MESOAMERICAN TOADS[§]

Thomas J Firneno Jr.[1,2], Justin R. O'Neill[3], Daniel M. Portik[4], Alyson H. Emery[1], Josiah H. Townsend[3,5], and Matthew K. Fujita[1,2]

[1]Department of Biology, University of Texas at Arlington, Arlington, Texas, 76019, USA

[2]Amphibian and Reptile Diversity Research Center, Department of Biology, University of Texas at Arlington, Arlington, Texas, 76019, USA

[3]Department of Biology, Indiana University of Pennsylvania, Indiana, Pennsylvania 15705, USA

[4]California Academy of Sciences, San Francisco, CA, 94118, USA

[5]Centro Zamorano de Biodiversidad, Departamento de Ambiente y Desarrollo, Escuela Agrícola Panamericana Zamorano, Municipalidad de San Antonio de Oriente, Francisco Morazán, Honduras

[§]Used with permission from the publisher

ABSTRACT

Mitonuclear discordance is a frequently encountered pattern in phylogeographic studies and occurs when mitochondrial and nuclear DNA display conflicting signals. Discordance among these genetic markers can be caused by several factors including confounded taxonomies, gene flow, and incomplete lineage sorting. In this study, we present a strong case of mitonuclear discordance in a species complex of toads (Bufonidae: *Incilius coccifer* complex) found in the Chortís Block of Central America. To determine the cause of mitonuclear discordance in this complex, we used spatially explicit genetic data to test species limits and relationships, characterize demographic history, and quantify gene flow. We found extensive mitonuclear discordance among the three recognized species within this group, especially in populations within the Chortís Highlands of Honduras. Our data reveal nuclear introgression within the Chortís Highlands populations that was most likely driven by cyclical range expansions due to climatic fluctuations. Though we determined introgression occurred within the nuclear genome, our data suggest that it is not the key factor in driving mitonuclear discordance in the entire species complex. Rather, due to a lack of discernible geographic pattern between mitochondrial and nuclear DNA, as well as a relatively recent divergence time of this complex, we concluded that mitonuclear discordance has been caused by incomplete lineage sorting. Our study provides a framework to test sources of mitonuclear discordance and highlights the importance of using multiple marker types to test species boundaries in cryptic species.

INTRODUCTION

Understanding the spatial and temporal factors that affect and influence the geographic distribution of genetic variation is a central goal of phylogeography. Many phylogeographic studies utilize an array of genetic markers, including both mitochondrial (mtDNA) and nuclear DNA (nuDNA). While evolutionary patterns from genetic markers often align across geographic landscapes, mtDNA and nuDNA can also have conflicting and discordant signals, often referred to as "mitonuclear discordance" (Avise, 2000; Chan & Levin, 2005; Funk & Omland, 2003; Toews & Brelsford, 2012; Zink & Barrowclough, 2008). A common form of mitonuclear discordance involves differences in nuclear and mitochondrial gene tree topologies and/or branch lengths. These patterns can be caused by several factors, including incomplete lineage sorting (ILS), genetic polymorphism, local adaptation of distinct mtDNA lineages, sex-biased dispersal, unresolved taxonomy, demographic complexity arising from historical fluctuations, and/or gene flow (including hybridization and introgression) (Brandt, Ishida, Georgiadis, & Roca, 2012; Cahill et al., 2013; Chavez, Maher, Arbogast, & Kenagy, 2013; Ivanov, Lee, & Mutanen, 2018; Pavlova et al., 2013; Phuong, Bi, & Moritz, 2016; Ribeiro, Lloyd, & Bowie, 2011; Toews & Brelsford, 2012; Turmelle, Kunz, & Sorenson, 2011). Mitonuclear discordance is indicative of historical and demographic complexities in phylogeographic systems, and provides an opportunity to pursue and test additional evolutionary hypotheses (Ivanov, Lee, & Mutanen, 2018; Singhal & Moritz, 2012; Weigand et al., 2017). Conversely, mitonuclear discordance can interfere with other inferences such as species delimitation, where conflicting signals can prevent establishing conclusive taxonomy (Bonnet, Leblois, Rousset, & Crochet, 2017; Papakostas et al., 2016; Toews & Brelsford, 2012).

Introgression and ILS are two of the most well-described causes of mitonuclear discordance; however, distinguishing between introgression or ILS as the primary cause of discordance has been problematic in many studies (Cahill et al., 2013; Funk & Omland, 2003; McKay & Zink, 2010; Toews & Brelsford, 2012). A common limitation is that many studies lack of appropriate data to test hypotheses of mitonuclear discordance in an integrative framework. For example, many studies utilize genetic data to test demographic fluctuations assuming panmixia (e.g. Tajima's D, Fu's F, mismatch distributions) and integrate spatial data post hoc, rather than simultaneously integrating spatial information (Rowe, Heske, Brown, & Paige, 2004; Perktas, Barrowclough, & Groth, 2011; Orozco-Terwengel, Andreone, Louis, & Vences, 2013). However, newer methods have become available that integrate genome-scale data and spatial data to detect spatial changes in allele frequencies across geography caused by different demographic processes (Bonnet, Leblois, Rousset, & Crochet, 2017; DeBiasse, Nelson, & Hellberg, 2013; Phuong, Bi, Moritz, 2017; Singhal & Moritz, 2012). These methods allow for the testing of hypotheses of diversification and the causes of patterns of mitonuclear discordance in a cohesive framework. For example, Phuong, Bi, & Moritz (2017) used a suite of methods to detect and test mitonuclear discordance in a complex of ground squirrels, and determined that it was due to introgression caused by range fluctuations. We adapt and expand this methodological framework to identify the causes of mitonuclear discordance observed in a species complex of toads located in the Chortís Block of Mesoamerica. Our primary aim toward understanding the evolutionary history of the fauna within the Chortís Block is to identify the causes of this conflicting signal by integrating both genetic and spatially explicit data within an empirical model testing framework.

The Chortís Block in Central America is a remarkable and complex region that harbors high levels of biodiversity (Figure 1; Campbell, 1999; Townsend, 2014). The highlands of the Chortís Block are a mosaic mountainous interior region flanked on both coastal versants by lowland habitats (Carr, 1950; McCranie & Wilson, 2002; Townsend, 2014; Wilson & Meyer, 1985). The geologic and ecological complexity of this region allows for and supports a rich and diverse biota, especially with regard to amphibian and reptile diversity (Wilson & Johnson, 2010; Townsend 2014). The intricate heterogeneity of this region coupled with the rich amphibian diversity offers ample opportunities to investigate patterns and processes of diversification at both deep and recent time scales. The *Incilius coccifer* complex is putatively composed of three distinct, yet closely-related mitochondrial lineages that exhibit a highly conserved morphology: *I. coccifer* in the Pacific lowlands from southern Mexico to northern Costa Rica, *I. ibarrai* in the highlands from western Guatemala to western Honduras, and *I. porteri* in the highlands of central Honduras (Mendelson, Williams, Sheil, & Mulcahy, 2005; Mendelson, Mulcahy, Williams, & Sites, 2011; Firneno & Townsend, 2019). Previous comprehensive sampling and mtDNA sequencing of this species complex revealed numerous sympatric populations of *I. ibarrai* and *I. porteri* within the Chortís Highlands, as well as sympatric individuals of lowland (*I. coccifer*) haplotypes and highland (*I. ibarrai* and/or *I.porteri*) haplotypes (Firneno, Luque-Montes, & Townsend, 2017; Firneno & Townsend, 2019). However, until this study, no nuclear data has been used to evaluate the species limits, assess gene flow among sympatric populations, or investigate the demographic history of this species complex. Upon building a ddRADseq SNP dataset, our preliminary analyses revealed mitonuclear discordance within and among the three species, with the most extensive discordance found in populations within the Chortís Highlands.

Here, we combine mtDNA and ddRADseq data within a spatially explicit model testing framework in order to: (1) test the species boundaries and evolutionary relationships of the *Incilius coccifer* species complex; (2) test for population structure among and within this complex; (3) quantify gene flow among the populations within the Chortís Highlands; (4) identify key mechanisms underlying the diversification of this complex within the Chortís Highlands; and (5) determine the possible causes and consequences of mitonuclear discordance seen in this complex.

MATERIALS AND METHODS

Taxon sampling and DNA extraction

We collected and acquired tissues from 87 individuals across the ranges of the three taxa from the *Incilius coccifer* complex (*I. coccifer*, *I. ibarrai*, and *I. porteri*) with specific focus on highland populations within the Chortís Block (Figure 1). We extracted DNA from muscle or liver tissue stored in SED buffer (250mM EDTA, 20% DMSO, and saturated NaCl; Seutin, White, & Boag, 1991) or 70% ethanol using a standard phenol-chloroform protocol (Sambrook & Russell, 2006). The quality of our DNA extractions was checked using a 1% agarose gel and the DNA concentration was quantified using a QUBIT 2.0 Fluorometer (Life Technologies, Carlsbad, CA).

Molecular data generation, processing, and identifying mitonuclear discordance

We amplified a 658 basepair (bp) segment of the cytochrome oxidase I (COI) gene for 29 individuals using standard primer pairs (LCO-1490, HCO-2198; Folmer et al. 2004) and the PCR profiles used in Firneno & Townsend (2019). PCR products were cleaned using ExoSap-IT and were sequenced in both directions using amplification primers and BigDye v3.1 (Applied Biosystems) on an ABI 3730 capillary sequencer (EuroFins). We edited the resulting sequences

14

in GENEIOUS v6.1 (Biomatters Ltd.) and generated a multiple alignment of 658 bp for 90 samples (included three other species from the *I. coccifer* group as an outgroup; File S2) in MEGA7 (Kumar, Stecher, &Tamura, 2016) with the MUSCLE algorithm (Edgar, 2004) using the default parameters.

We collected ddRADseq data for 84 individuals following the protocol described in Peterson et al. (2012) and following parameters specified in Streicher et al. (2014). Our final library was analyzed on one Illumina HiSeq2500 lane (150 bp single end reads) at the Genomic Sequencing and Analysis Facility (GSAF) at The University of Texas (https://www.wikis.utexas.edu/display/GSAF). The workflow for data processing, filtering, and formatting was automated using scripts available from Portik et al. 2017 (https://github.com/dportik/Stacks_pipeline). In brief, the raw Illumina reads were demultiplexed using STACKS v1.35 (Catchen, Hohenlohe, Bassham, Amores, & Cresko, 2013), the restriction site overhangs were removed using the fastx_trimmer module of the FASTX-TOOLKIT (www.hannonlab.cshl.edu/fastx_toolkit), and the sequencing quality was examined on a per sample basis using FASTQC v0.10.1 (www.bioinformatics.babraham.ac.uk/projects/fastqc). Loci were created, catalogued, and identified using USTACKS, CSTACKS, and SSTACKS, respectively. POPULATIONS was then used to generate alleles for loci present in 70% of all individuals, which resulted in 2,211 loci. Custom filtering removed invariant loci (n=150), non-biallelic loci (n=2), and loci containing at least one individual with more than two alleles (n=854). For loci containing multiple SNP sites (average number of SNPs per locus was 2.13 (± 1.14)), we randomly chose a single SNP to be used for subsequent analyses. Any samples missing data for more than 60% of loci were removed. After completing the above filtering steps, our final SNP dataset consisted of 64 samples and 1,207 loci.

To identify mitonuclear discordance and confirm that it was not an artifact of differently

sized datasets or phylogenetic inference methods we conducted maximum likelihood (ML)

analyses on three datasets: (1) the 658 bp COI dataset of 87 individuals; (2) a 1150 bp

concatenated dataset of COI and 16S of 61 individuals (presented in Firneno & Townsend 2019);

and (3) the 1207 SNP RADseq concatenated data set of 64 individuals (57 of these samples had

corresponding COI data; see Files S1 and S2). ML analyses were carried out in RAxML v8.0

(Stamatakis, 2014) under the following parameters for the three datasets, respectively: (1) 10,000

bootstrap pseudoreplicates under a GTR substitution model; (2) partitioned dataset by gene (16S)

and codon position (COI), 10,000 bootstrap pseudoreplicates under a GTR substitution model

(from Firneno & Townsend 2019); and (3) 10,000 bootstrap pseudoreplicates under a GTR

substitution model.

Population Discovery

We determined the number of discrete populations present across the sampled range of

the *I. coccifer* complex using a combination of Bayesian and likelihood clustering analyses, and

multivariate methods. We used STRUCTURE v2.3.4 (Falush, Stephens, & Pritchard, 2003;

Pritchard, Stephens, & Donnelly, 2000) to examine the number of population clusters and

potential admixture between populations in our dataset using MCMC. Hierarchical analyses were

performed for 10 runs per K, up to a maximum of seven populations, and used the admixture

model with a burn-in of 10,000 steps followed by 100,000 steps. We summarized our results

using STRUCTUREHARVESTER (Earl & vonHoldt, 2012) and evaluated the number of populations

based on inspection of likelihood plots and following Evanno et al (2005). To complement our

STRUCTURE analysis, we used a maximum likelihood approach with ADMIXTURE (Alexander,

Novembre, & Lange, 2009). We performed ten replicate analyses to evaluate up to seven

populations. To assess the best K value, we performed 10-fold cross-validation and determined

the K value with the lowest cross-validation error. We also evaluated the number of discrete

populations using a discriminant analysis of principal components (DAPC) with ADEGENET

v2.0.0 (Jombart, 2008; Jombart & Ahmed, 2011). A maximum of 10 clusters were investigated

using the k-means algorithm. The preferred number of clusters was evaluated using BIC scores.

We explored a range of three to five clusters to describe using DAPC. To minimize overfitting,

an initial DAPC was used to find the a-score for each set of clusters and this value was used to

select the number of principal components to retain in a subsequent reanalysis (Jombart, 2008;

Jombart & Ahmed, 2011). Group membership probabilities were then examined for each cluster.

To independently assess the validity of population differentiation and assignment, we used the

FINERADSTRUCTURE software package (Malinsky, Trucchi, Lawson, & Falush, 2018) to

construct a co-ancestry matrix from our RADseq data. We used a 100,000 burn-in followed by

100,000 MCMC steps sampling every 1,000 steps and the tree was constructed with 10,000 hill-

climbing iterations. The results were visualized using the FINERADSTRUCTUREPLOT.R and

FINESTRUCTURELIBRARY.R scripts (included in the fineRADstructure package file).

Coalescent species delimitation

Patterns of mitonuclear discordance can often arise due to an incorrect taxonomy rather

than actual biological processes. To investigate the species limits of the lineages in this complex

and to determine that incorrect taxonomy is not the cause of mitonuclear discordance, we

conducted Bayes Factor Delimitation implemented in SNAPP v1.3.0 (Bryant, Bouckaert,

Felsenstein, Rosenberg, & RoyChoudhury, 2012) in BEAST2 v2.5.4 (Bouckaert et al., 2014) with

our SNP dataset following Leaché et al. (2014) (e.g. implementing BFD*). The major

advantages of this method for SNP data is that it can accommodate missing data between

individuals/among species, as well as allowing for varying numbers of individuals per species to accommodate unequal sampling among populations/lineages. However, because the coalescent model used in SNAPP cannot accommodate gene flow, we performed our analyses on two different datasets that included: 1) individuals from all five populations (assigned from our hierarchical analyses — two populations from outside the Chortís Block [OCH Lowland and OCH Highland] and three populations from within the Chortís Block [WCH West, WCH Central, and WCH East]); and 2) individuals from the four non-admixed populations (OCH Lowland, OCH Highland, WCH West, and WCH East), removing all admixed individuals with estimated admixture >5% (WCH Central; Grummer et al., 2015). The models that we tested for both datasets are presented in Table 1. Using BFD*, we compared and ranked models to select the best supported species hypothesis. We calculated the Bayes Factor (BF) by subtracting the value of the MLE for the model representing the current taxonomic classification from each alternative model and multiplying the difference by two (BF = 2(model 1 – model 2)). We ranked all of the models and selected the model with the highest BF (Table 1).

We estimated the mutation rates (u and v) from the data (7.339 and 0.536, respectively) within BEAUTI. We assigned a Gamma distribution to our birth rate (λ) of the Yule prior, with and Alpha of 1 and a Beta of 60. Our SNAPP prior was assigned an Alpha of 1, a Beta of 150, and a Lambda of 29. We performed 48 path sampling steps, with 100,000 MCMC generation and a pre-burnin of 1000 generations. The BF was then calculated for each alternative model, the models were ranked, and a best model was chosen (Table 1).

Phylogenetic relationships and divergence dating

We estimated the phylogenetic relationships and associated divergence times of individuals and populations independently for our mtDNA and SNP data sets. We conducted

Bayesian divergence dating analyses with our mtDNA data set using BEAST2 v2.5.4 (Bouckaert et al., 2014). We performed analyses using an HKY model of nucleotide substitution and a constant size growth coalescent tree prior. Because our COI data did not seem clock-like (based on a coefficient of variation > 0.1 when run under a relaxed clock) we used a log-normal relaxed molecular clock calibrated with a 2% per Myr rate of divergence (COI, which has no estimated rate, tends to evolve slower than *cytb* and ND1, which have an estimated rate in bufonids of 1.38% per Myr; Macey & Schulte, 1998; Sequeira et al., 2011). This analysis was run twice with 20,000,000 generations with sampling every 1000 generations, producing a total of 10,000 trees. Runs were assessed using TRACER v1.6 (Rambaut & Drummond, 2009) to examine convergence. A burn-in of 10% was discarded, and a maximum clade credibility (MCC) tree with median heights was created from the remaining 9000 trees.

We estimated the species tree for our SNP dataset using SNAPP v1.3.0 (Bryant, Bouckaert, Felsenstein, Rosenberg, & RoyChoudhury, 2012) implemented in BEAST2. The SNAPP model is based on the coalescent process and can accommodate ILS but assumes a lack of gene flow. We assigned species identities based on the best supported model from out BFD* analyses (see below). To reduce run times, we subsampled each population to include 3–10 representatives, for a total of 32 individuals. We used the same parameters as our BFD* analyses. We performed two independent runs with a chain length of 1,000,000 generations, sampling every 1000 generations. Runs were assessed using TRACER v1.6 (Rambaut & Drummond, 2009) to examine convergence, and tree topologies and node heights were visualized using DENSITREE (Bouckaert, 2010).

Analysis of genomic introgression

Introgression, either in the mitochondrial or nuclear genome, is often one of the major causes of mitonuclear discordance in organisms (Toews & Brelsford, 2012). We measured

genomic introgression of WCH parental populations (WCH East and WCH West) genetic

regions in admixed WCH populations (WCH Central). We quantified locus-specific genomic

introgression using the Bayesian genomic cline model and the program *bgc* (Gompert & Buerkle,

2011). We partitioned samples into eastern, western, and admixed populations (Figure 2), and

calculated locus-specific allele frequencies for each population. *bgc* estimates the hybrid index

(*h*) for each individual within an admixed population, which represents the proportion of an

individual's genome that is inherited from one of the parental populations. The hybrid index, the

genomic cline center parameter (α), and the genomic cline rate parameter (β) are used to

estimate the posterior probability of inheritance from one parental population (Φ) at a given

locus within the admixed population. Under this model, if both α and β equal zero, *h* and Φ will

be equivalent and will match the neutral genomic background expectation (Gompert & Buerkle,

2011).

　　　We used GenePop output files from our STACKS pipeline, which were converted into

biallelic input files for *bgc* using the genepopedit R package (Stanley, Jeffery, Wringe, DiBacco,

& Bradbury, 2017). These input files included information for 1,207 loci, though some loci

suffered from missing data for particular individuals. We ran *bgc* using the genotype uncertainty

model (recommended for next-generation sequencing data; Gompert & Buerkle, 2011) on our

parental and admixed population datasets using four chains for 50,000 generations each,

discarding the first 25,000 generations as burnin, and recording parameter estimates for every

fifth MCMC generation. The default settings were used, assuming free recombination between

all loci, and we set the sequence error probability parameter to 0.0001. Parameter estimates for

the four chains were combined after confirming convergence and stationarity of parameters for

each MCMC run.

Joint-demographic history

Understanding the underlying mechanisms of divergence and demographic history can help identify the cause of mitonuclear discordance among organisms (Phuong, Bi, & Moritz, 2016). To investigate hypotheses of alternative demographic histories of populations within the Chortís Highlands, we used the diffusion approximation method of δαδi to analyze the three-dimensional joint site frequency spectrum (3D-JSFS; Gutenkunst, Hernandez, Williamson, & Bustamante, 2009). We used the demographic modeling pipeline (DADI_PIPELINE) of Portik et al. (2017) to conduct all analyses. We created a folded JSFS because we lacked outgroup information. To account for missing data and to maximize the number of segregating sites used in our analyses, we down-projected populations to smaller sample sizes (WCH East: 20 alleles; WCH Central: 32; and WCH West: 14). We aimed to identify the origin of the WCH Central population and identify patterns of isolation and gene flow among the three WCH populations. We constructed three general categories of models, which include 1) a bifurcating branching pattern among populations that matches results from our SNAPP analyses, 2) simultaneous splitting of all three populations, and 3) a hybrid origin of the WCH Central population, resulting from instantaneous fusion of the other populations. Within each category, we created multiple models that varied in migration parameters and in periods of isolation and/or contact. This resulted in 17 models, which are shown visually in File S4.

For all models, we performed consecutive rounds of optimizations following Portik et al. (2017). For each round, we ran multiple replicates and used parameter estimates from the best scoring replicate (highest log-likelihood) to seed searches in the following round. We used the default settings for each round (replicates = 10, 20, 30, 40; maxiter = 3, 5, 10, 15; fold = 3, 2, 2, 1), and optimized parameters using the Nelder-Mead method (*optimize_log_fmin*). Across all

analyses, we used the optimized parameter sets of each replicate to simulate the 3D-JSFS, and the multinomial approach was used to estimate the log-likelihood of the 3D-JSFS given the model. By keeping only a single SNP per RAD locus, we assumed loci are unlinked and consequently the log-likelihood values returned are the true likelihood values rather than composite likelihoods. Models were therefore compared using the Akaike information criterion (AIC), and the replicate with the highest likelihood for each model was used to calculate AIC scores, ΔAIC scores and Akaike weights (ω_i) (Burnham & Anderson, 2002).

Ecological niche modeling

To augment our demographic model testing and generate hypotheses for how glacial cycles may have impacted range dynamics, we modelled the potential distribution of WCH populations across several timescales. We compiled all available museum locality information from VERTNET (Constable, Guralnick, Wieczorek, Spencer, & Peterson, 2010), along with our own sampling, for toads of the *I. coccifer* complex (identified as *I. Ibarrai* or *I. porteri*) within the Chortís Highlands that have been verified morphologically and/or genetically. This resulted in 186 records that were used for ecological niche models (ENM) (File S3). To reduce spatial autocorrelation of our occurrence dataset, we thinned the points by a distance of 5km (Aiello-Lammens, Boria, Radosavljevic, Vilela, & Anderson, 2015; Veloz, 2009), which reduced our dataset to 37 points. We constructed all ecological niche models (ENMs) using the 19 bioclimatic data layers available on Worldclim.org (Hijmans, Cameron, Parra, Jones, & Jarvis, 2005). Because of Maxent's capability to mitigate the contribution of redundant variables (Elith et al, 2011; Feng, Park, Liang, Pandey, & Papes, 2019; Philips & Dudik, 2008; Shcheglovitova & Anderson, 2013), we opted to refrain from reducing the bioclimatic variable set of correlated data layers, so as to not exclude any unique contributions from the omitted data (Graham, 2003).

In addition to data for the current climate, we also used the downscaled paleoclimate layers for three past climate scenarios (also obtained from Worldclim.org): mid-Holocene (~6000 years ago), last glacial maximum (LGM; ~22,000 years ago), and last inter-glacial (LIG; ~120,000 years ago). We masked the bioclimatic layers by a minimum convex polygon buffered by 0.75 degrees around the occurrence points to constrain our background extent to environmental conditions experienced by the species. To build the ENMs, we used MAXENT (Phillips, Anderson, & Schapire, 2006) implemented in the ecological niche modeling application WALLACE (Kass et al., 2018). Because model settings can hold strong influence on model output, determining optimal model complexity is important (Warren & Seifert 2011; Radosavljevic & Anderson 2014). We ran 80 candidate models using a combination of: 1) one of four selected Maxent feature classes (Linear; Linear and Quadratic; Hinge; Linear, Quadratic, and Hinge); and 2) a range of 20 regularization multipliers (0.5–10 in 0.5 increments), allowing our candidate models to range from simple to complex. To test the accuracy of the models, we employed WALLACE'S spatial partition option and split the occurrence data into four spatially independent k-folds for cross-validation (Muscarella et al., 2014; Kass et al., 2018). To choose the top model we looked for the candidate within four ΔAICc that had the lowest testing omission rate and a testing area under the curve (AUC) score above 0.80, ensuring the top model was appropriately complex given our data and sufficiently accurate (Burnham & Anderson 2002; Muscarella et al., 2014; Kass et al., 2018). We projected the models onto the paleoclimate data layers to produce retrojections and summed the binary outputs to produce continuous stability maps (Devitt, Devitt, Hollingsworth, McGuire, & Moritz, 2013; Portik et al., 2017; Yannic et al., 2014).

RESULTS

Confirming the presence and extent of mitonuclear discordance

Both mitochondrial ML trees revealed identical tree topologies (File S1) with well

supported clades. When mitochondrial datasets are compared to the SNP dataset the following

discordances are observed: (1) samples located within the Chortís Highlands that represent

Incilius ibarrai and *I. porteri* are rendered paraphyletic within the mitochondrial data; (2) *I.*

coccifer is no longer nested within the two highland species and is instead basal to them; (3)

other individuals that represent either *I. porteri* (e.g. JHT3302) or *I. coccifer* (e.g. JHT3793)

mitochondrial haplotypes group with their geographically respective nuclear haplotypes within

the SNP tree (File S1).

Population inference

The Bayesian population clustering analysis carried out in STRUCTURE based on 1,207

unlinked SNPs resulted in the initial detection of three main populations (Figure 2a). One

population corresponded to a combination of lowland and highland toads outside of the Chortís

Highlands (OCH). The other two populations correspond to an eastern and western group of

toads within the Chortís Highlands (WCH East and WCH West, respectively) with a division at

Honduran Depression (Figure 1, 2a). Using a hierarchical approach, we detected further

structuring within the OCH group, which delineated these populations into a lowland group

(OCH Lowlands) and a highland group (OCH Highlands). These two lineages correspond to

what is currently recognized as *I. coccifer* and *I. ibarrai*, respectively. No further structuring was

found in the two eastern and western WCH groups. These results were supported by ADMIXTURE

analysis, for which the two population models exhibited the lowest cross-validation error for

both groups and the population assignments of individuals were identical (Figure 2b). There is a

pattern consistent with admixture between the eastern and western WCH populations that

corresponds to an admixed central group (WCH Central), which is associated with other highland areas that are surrounded by two xeric lowland valleys (Figure 1, 2b).

The DAPC supported these results and inferred four or five discrete populations (both population groupings had equal likelihood), with the assignment of individuals matching the results of the maximum-likelihood and Bayesian clustering analyses (Figure 2c). Under a four-population scenario with DAPC, individuals were grouped into the same four populations that were given from our STRUCTURE and ADMIXTURE analyses, with admixed WCH Central individuals being included in either the WCH East or WCH West population clusters. Under a five-populations scenario with DAPC, the same four clusters were given; however, the WCH Central individuals that displayed admixture formed their own cluster (Figure 2c).

The resulting dendrogram and co-ancestry matrix from our FINERADSTRUCTURE analysis (Fig. 2d) confirmed the results of our hierarchical population analyses. Sub-structuring is seen within the WCH Central individuals, which corresponds to differing levels of admixture within this population. Two individuals from the WCH Central population (represented in Fig. 2d by *) also cluster with the WCH East population in our co-ancestry matrix, which is most likely because they have very low levels of admixture within our hierarchical analyses.

Coalescent Species Delimitation, **phylogenetic relationships and divergence times**

Out of the eight delimitation models across our two datasets, the models splitting all of the lineages (5 lineages and 4 lineages, respectively per dataset) were the best supported (Table 1). However, the three, four, and five species models were all weighted very closely (Table 1). These results may indicate the need for a reevaluation of the number of species within this complex, especially with respect to the lineages within the Chortís Highlands.

The phylogenetic analyses of the mtDNA dataset supported three clades that correspond to the three species within the *I. coccifer* complex (Firneno & Townsend, 2019; Mendelson, Mulcahy, Williams, & Sites, 2011). The divergence dating analysis of mtDNA using a calibrated mutation rate supported a mid-Pleistocene divergence (0.83Mya [95% HPD 0.55–1.14 Mya]) for the three species of the *I. coccifer* complex (Figure 3). Given the supported topology, we cannot directly infer divergence dates for the WCH populations detected in our SNP dataset. However, based on the divergence dates of the mitochondrial haplogroups that correspond to the three WCH populations, divergence for this group from the OCH populations may have occurred in the mid- to late-Pleistocene (0.20–0.26 Mya [95% HPD 0.11–0.45 Mya]).

The phylogenetic relationships among populations inferred through Bayesian coalescent analyses of our SNP dataset are discordant to those of the mtDNA (Figure 3). Our phylogenetic analysis of the SNP data renders all samples within the Chortís Highlands as paraphyletic. This analysis also places our OCH Lowlands lineage as sister to an entire highland lineage (includes OCH Highlands and all three WCH lineages), rather than it being nested within the two highland groups (Figure 3). Our SNAPP results have high support at most nodes (≥0.90) for a five-lineage grouping, as indicated by our coalescent species delimitation analysis. The only node that is not well supported in this case is the one between the WCH West and WCH Central lineages (Figure 3). The lack of support at this node is most likely due to a violation of a primary assumption of the SNAPP model, in that gene flow is not present among lineages.

Genomic Introgression

Genomic introgression was apparent in WCH Central populations. The genomic cline parameter α was particularly variable, with a minimum $\alpha = -1.37$ and a maximum $\alpha = 1.47$. The average hybrid index (genomic background of western versus eastern ancestry) of individuals

sampled from the admixed population was 0.4718, which is consistent with a nearly completely

admixed genomic background. Given the hybrid index of an individual, particular loci may

deviate from the background because they are dominated by either western or eastern alleles, as

indicated by the value of α. We detected excess WCH West ancestry (i.e., the lower bound of the

95% CI for α was greater than zero) for 217 loci (17.9% of the loci) and excess WCH East

ancestry (i.e., the upper bound of the 95% CI for α was less than zero) for 231 loci (19.1% of the

loci; Figure 4). Thus, genomic introgression for 37.0% of the sampled loci differed from the

genome average predicted by the hybrid index. Despite a greater number of loci with excess

eastern ancestry than excess western ancestry, this difference was not statistically significant ($p =$

0.24). It is probable that extreme values of α are expected for loci that reside in genetic regions

affected by selection in the WCH Central populations. Genomic cline parameter β was less

variable (min = -0.002; max = 0.002) and the 95% CI for β encompassed zero for all loci.

Joint-demographic history

The best model from our 3D-JSFS analysis occurs in the category of models that have a

bifurcating pattern among populations that matches results from our SNAPP analyses. More

specifically, this model indicates a refugial scenario for the WCH East and the ancestral

population of the WCH Central and WCH West, where an initial divergence in isolation was

followed by secondary contact. This was followed by a subsequent divergence between the

WCH Central and WCH West, which occurred with gene flow (model

"refugia_adj_3_var_sym", ΔAIC = 124.58, ω_I = 1.0; Figure 5, File S5). In the model set that we

examined, we included other refugial models that included the same scenario as our best model,

but with unidirectional gene flow into the WCH Central population, as well as models in which

divergences among all respective populations occurred in isolation and were followed by a single

period of simultaneous gene flow (File S4). These latter models were intended to represent a single forest expansion that facilitated secondary contact concurrently between neighboring populations; however, these models were not favored over our top model.

Ecological niche modeling

The top candidate ENMs that constituted our final models had high AUC scores; models built with 30 arc-second climate data had an AUC statistic of 0.91 and models built with 2.5 arc-minute data had an AUC statistic of 0.90. According to Maxent's analysis of variable contributions, mean temperature of driest quarter (bio9), had the highest permutation importance and had the most information that wasn't present in other variables. Based on the ENMs that we generated for all three WCH populations, we identified concordant areas that are predicted to have remained climatically stable and suitable through time (Figure 6). Our model predicted regions of high habitat stability that correspond to many of the highland regions within the southwest and central Chortís Highlands. Our models also reveal predicted moderate to high habitat stability through the Honduran Depression and the Mejocote-San Juan Depression, primarily during the LGM, which appear to be two of the major biogeographic barriers to highland organisms in this region (Figure 6).

DISCUSSION

Our results revealed extensive mitonuclear discordance within the *Incilius coccifer* complex, causing paraphyly of mitochondrial lineages within the Chortís Highlands of Honduras. Multiple mechanisms can cause these patterns of mitonuclear discordance, including: (1) a confounded taxonomy for this complex; (2) ILS due to rapid divergence among the populations within the species complex; and/or (3) gene flow between lineages, especially with respect to sympatric mitochondrial lineages within the Chortís Highlands. Due to a lack of

discernible geographic pattern between mitochondrial and nuclear DNA, as well as a relatively recent divergence time of this complex, we propose that incomplete lineage sorting explains the patterns of mitonuclear discordance in this complex. Below, we summarize our findings and evidence for our proposed explanation of mitonuclear discordance in the *I. coccifer* complex.

Reevaluating the taxonomy of the *Incilius coccifer* complex

Our study exemplifies the difficulty of species delimitation when mitonuclear discordance exacerbates an already confounded taxonomy. The species boundaries of the *Incilius coccifer* complex have been controversial since Mendelson et al. (2005) provided an initial description of the complex (McCranie & Castañeda, 2007; McCranie & Wilson, 2002; McCranie, 2009; McCranie, 2015; Mendelson, Mulcahy, Williams, & Sites, 2011). While studies have used comprehensive sampling of the region to generate mitochondrial datasets (Firneno & Townsend, 2019), our study is the first to include nuclear data to delimit these species. Based on mitochondrial data, the current taxonomy supports three well supported, yet shallowly diverged lineages, with *I. coccifer* and *I. porteri* forming a clade that is sister to *I. ibarrai* (Firneno & Townsend, 2019; Mendelson, Mulcahy, Williams, & Sites, 2011).

We used nuclear SNP data and coalescent species delimitation methods to test the number of species within the *I. coccifer* complex. Our species delimitation analyses inferred the maximum number of five species in this complex (each representing a distinct population), though the three and four species models had nearly equal weights (Table 1). Given that these coalescent species delimitation methods may be prone to over splitting under certain speciation scenarios (Fujita, Leaché, Burbrink, McGuire, & Moritz, 2012; Grummer, Bryson, & Reeder, 2014; Reid et al., 2013; Sousa & Hey, 2013), we take a conservative approach and propose that there are three species within this complex, including: (1) *I. coccifer*, which comprises all

lowland populations within and outside of the Chortís Block (concordant with Mendelson et al.

(2005)); (2) *I. ibarrai*, which comprises all highland populations west of the Chortís Highlands in

Guatemala, and corresponds to a single sample from Guatemala that was noted by Mendelson et

al. (2011) as being a potentially divergent species/population of *I. ibarrai*; and (3) *I. porteri*,

which comprises all highland populations within the Chortís Highlands in Honduras, and

corresponds to what Mendelson et al. (2005) defined as *I. porteri* and western Honduran

populations of *I. ibarrai*. These three species largely correspond with the original three species

described by Mendelson et al (2005). Our results suggest that WCH populations (*I. porteri*)

require reassessment to determine if there are morphological differences seen between the WCH

West and WCH East mitochondrial lineages, and whether those differences warrant

modifications to our species concepts in the *I. coccifer* complex (e.g. that there are potentially

four species within the complex; Mendelson et al., 2005).

The taxonomy presented by Mendelson et al. (2005) using morphology and mtDNA, and

the taxonomy presented here, both support three species within the *I. coccifer* complex.

However, whichever taxonomy is followed, discordance is always present. Regardless, we

believe that the taxonomic uncertainties surrounding the relationships of these putative taxa is

not the cause, but rather a consequence of mitonuclear discordance in this complex.

Fluctuations in forest refugia drove divergence in the Chortís Highlands

The timing and mechanisms of diversification, as well as the biogeography of a region

can greatly impact patterns of mitonuclear discordance in organisms (Toews & Brelsford, 2012).

Geological and climatic events have greatly influenced the distribution and composition of

Mesoamerican biodiversity, most notably including the formation of the Isthmus of Panama

(~4.5mya) and the fragmentation of tropical forests into refugia during Pleistocene glaciation

(~1.6–0.01mya; Daza, Castoe, & Parkinson, 2010; Hewitt, 2004; Orneales, González, Espinosa de los Monteros, Rodríguez-Gómez, García-Feria, 2014; Raven & Axelrod, 1974). Interglacial periods opened lowland corridors that enabled gene flow between lineages that were initially restricted to refugia during glacial periods (Gutiérrez-Rodríguez, Ornelas, & Rodríguez-Gómez, 2011; Gutiérrez-Garcia & Vásquez-Domínguez, 2012; Ornelas & Rodríguez-Gómez, 2015; Rodríguez-Gómez & Ornelas, 2014; Rovito, Vásquez-Almazán, Papenfuss, Parra-Olea, & Wake, 2015). The distributional and genetic patterns that result from climatic oscillations can aid in the discovery of mitonuclear discordance and help identify its cause, especially in regard to distinguishing between ILS and introgression (Phuong, Bi, & Moritz, 2016; Singhal & Moritz, 2012; Toews & Brelsford, 2012).

The geography of the Chortís Block clearly define the four genetic populations (including the admixed population) that we inferred in the *I. coccifer* complex: (1) OCH Lowlands is separated from OCH Highlands and WCH populations by a highland barrier; (2) OCH Highlands is separated from WCH populations by the Motagua-Polochic fault system at the western border of the Chortís Block; (3) WCH West is separated by the Mejocote-San Juan Depression from WCH Central; and (4) WCH Central is separated by the Honduran Depression from WCH East (Figure 1). These biogeographic barriers are also likely the cause of allopatric divergence in numerous taxa from the region and surrounding areas including several species of birds, mammals, fish, salamanders, and frogs (Consuegra & Vázquez-Domínguez, 2015; Matamoros, Kreiser, & Schaefer, 2012; Rovito & Parra-Olea, 2015; Sánchez-Ramos et al., 2018; Townsend, 2014; Villalobos, 2013).

Our divergence dating analyses revealed diversification in the *I. coccifer* complex began ~830,000 years ago (95% HPD 0.55–1.14 Mya), with subsequent divergences of highland

populations occurring ~437,000 (95% HPD 0.24–0.68 Mya) and ~200,000–260,000 (95% HPD

0.11–0.44 Mya) years ago (Figure 3). These highland divergences coincide with interglacial

periods of warming in which forest refugia most likely receded into the highlands and cut off the

populations from those in the lowlands (Flint, 1947; Hewitt, 2004). Within the Chortís

Highlands, this is supported by the results of our ENMs, in which we see contraction of suitable

habitat in the central and southern *serranía* into the highlands during periods of warming (current

projections and LIG), which likely drove the divergence between our populations. In contrast,

the expansion of suitable habitat during periods of cooling (the LGM; Figure 6) likely caused

subsequent range expansion and gene flow. This concept is also supported by demographic

modeling, in which we found strong support for a refugial model with distinct periods of

isolation and secondary contact (Figure 5).

Ultimately, our results reveal the biogeographic complexity of this region and the *I.*

coccifer complex. We detected patterns of recent divergence among the whole complex, and a

clear biogeographic pattern for populations detected from the SNP data. We did not find the

same biogeographic pattern for the mitochondrial data, with many of the SNP haplotypes being

scattered throughout the mitochondrial tree (e.g. *I. porteri* haplotype on Isla del Tigre (Firneno,

Luque-Montes, & Townsend 2017) and *I. coccifer* in the highlands of western Honduras

(Firneno & Townsend, 2019), group into nuDNA haplotypes that reflect their respective

geographic ranges), especially with respect to haplotypes within the Chortís Highlands (Figure

3). These factors implicate incomplete lineage sorting as the potential cause of mitonuclear

discordance in this complex. However, our demographic analyses suggest that gene flow has

occurred within the WCH populations, which is why we also quantified introgression within

these populations.

Range instability and rapid divergence as drivers of nuclear introgression and mitonuclear discordance

Range instability, caused by climatic fluctuations, has explained patterns of mitonuclear discordance across a diverse array of organisms (Krosby & Rohwler, 2009; Mee & Taylor, 2012; Pereira, Martínez-Solano, & Buckley, 2016; Phuong, Bi, & Moritz, 2016; Singhal & Moritz, 2012). Distributional changes can increase the probability of genetic interactions between closely related lineages through range overlap or secondary contact, which can ultimately lead to introgression between the local and expanding taxa (Currant, Ruedi, Petit, & Excoffier, 2008). Recent range instability (through the mid to late Pleistocene) has also been shown to cause ILS in some taxa, leading to patterns of mitonuclear discordance in some organisms (Fujita, McGuire, Donnellan, & Moritz, 2010). Distinguishing between ILS and introgression can be very difficult; however, it has been noted that discordance arising from ILS is not expected to leave any predictable biogeographic pattern between mitochondrial and nuclear DNA (Funk & Omland, 2003; McKay & Zink, 2010; Toews & Brelsford, 2012).

Our analysis of genomic introgression for our admixed central population (WCH Central) indicate that the genome of this population is almost equal contributions from both parental populations (WCH East and WCH West). Since a hybrid index of 0.5 is indicative of an F1 hybrid, our data (h=0.47) suggests that gene flow amongst these populations may be readily occurring or has occurred in the recent past. Our ENMs indicate that range fluctuations occurred throughout the last ~120,000 years during periods of warming and cooling. Our combined results strongly suggest these climatic events drove divergences and facilitated secondary contact in this species complex. It is assumed that similar patterns of refugial expansion and contraction persisted back in time during other glacial and interglacial periods (Flint, 1947; Gutiérrez-Garcia

& Vásquez-Domínguez, 2012; Moritz, Patton, Schneider, & Smith, 2000). Our demographic modeling results demonstrate that multiple periods of gene flow occurred between the three populations within the Chortís Highlands. In particular, we inferred secondary contact between highland populations, and propose this pattern is driven by refugial expansions. The potentiality of the WCH Central populations being intraspecific hybrids highlights the fact that gene flow can be a major influence on the genetic diversity of these toads, possibly as well as other species of toads, and may allow for mitonuclear discordance to occur more readily.

While range instability often enables gene flow to occur leading to patterns of mitonuclear discordance caused by introgression, on shorter time scales ILS can also be the cause of mitonuclear discordance in the face of range instability (Barrowclough, Groth, Mertz, & Gutiérrez, 2005; Fujita, McGuire, Donnellan, & Moritz, 2010; McLean, Jackson, & Cook, 2016; Wang et al., 2018). Our data indicates introgression within the nuclear genome for our WCH populations caused by secondary contact via cyclical range fluctuations; however, we do not see the same in the mitochondrial lineages. This, along with the recent divergence of this species complex (~790,000 years ago) and that the discordance that we see in this complex does not show a distinct biogeographic pattern between mitochondrial and nuclear DNA, leads us to conclude that the pattern of discordance that we see in this complex has arisen due to ILS.

CONCLUSION

Using a combination of spatially explicit genetic analyses and a model-testing framework, our study identified mechanisms that strongly influenced the diversification of highland taxa in the Chortís Block of Central America, as well as the processes that led to mitonuclear discordance in a highland toad found in the pine-oak and cloud forests of the central and eastern *serranía* of the Chortís Highlands. With the methodological framework offered here,

this study follows an empirical approach for testing sources of mitonuclear discordance, as well as fuels a novel perspective on pursuing phylogenetic and phylogeographic work in this region of Central America. The geological and environmental complexity of the region, along with the high levels of biodiversity and endemism, provide unique resources to continue to investigate phylogeographic questions about the mechanisms of genetic and organismal diversification.

ACKNOWLEDGEMENTS

DATA ACCESSIBILITY

All raw, unprocessed sequences were deposited in NCBI GenBank (MT348741–MT348767; accession numbers are included in File S2) for COI mtDNA data and the Sequence Read Archive (BioProject number PRJNA626342; BioSample accession numbers SAMN14615759–SAMN14615822) for ddRADseq data. We have included a large data package on DRYAD (https://doi.org/10.5061/dryad.q573n5tfw) that includes our final ddRADseq filtered "haplotypes" files and the resulting input files for a number of analysis programs (STRUCTURE, ADMIXTURE, DAPC, FINERADSTRUCTURE, BFD*, BEAST, SNAPP, *bgc*, and δaδI). All newly created 3D demographic models are freely available in an updated version of the δaδI analysis pipeline found at https://github.com/dportik/dadi_pipeline.

LITERATURE CITED

Aiello-Lammens, M. E., Boria, R. A., Radosavljevic, A., Vilela, B., & Anderson, R. P. (2015). spThin: an R package for spatial thinning of species occurrence records for use in ecological niche models. *Ecography*, *38*(5), 541-545. https://doi.org/10.1111/ecog.01132

Alexander, D. H., Novembre, J., & Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, *19*, 1655-1664. https://doi.org/10.1101/gr.094052.109

Avise, J. C. (2000). *Phylogeography: The History and Formation of Species*. Cambridge, MA: Harvard University Press.

Barrowclough, G. F., Groth, J. G., Mertz, L. A., & Gutiérrez, R. J. (2005). Genetic structure, introgression, and a narrow hybrid zone between northern and California spotted owls (*Strix occidentalis*). *Molecular Ecology*, *14*(4), 1109-1120. https://doi.org/10.1111/j.1365-294X.2005.02465.x

Bonnet, T., Leblois, R., Rousset, F., & Crochet, P. (2017). A reassessment of explanations for

    discordant introgressions of mitochondrial and nuclear genomes. *Evolution*, *71*(9), 2140-

    2158. https://doi.org/10.1111/evo.13296

Bouckaert, R. R. (2010). DensiTree: making sense of sets of phylogenetic trees. *Bioinformatics*,

    26(10), 1372-1373. https://doi.org/10.1093/bioinformatics/btq110

Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C. H., Xie, D., … Drummond, A. J.

    (2014). BEAST 2: A Software Platform for Bayesian Evolutionary Analysis. *PLoS*

    Computational Biology, *10*(4), 1-6. https://doi.org/10.1371/journal.pcbi.1003537

Brandt, A. L., Ishida, Y., Georgiadis, N. J., & Roca, A. L. (2012). Forest elephant mitochondrial

    genomes reveal that elephantid diversification in Africa tracked climate transitions.

    Molecular Ecology, *21*(5), 1175-1189. https://doi.org.10.1111/j.1365-294X.2012.05461.x

Bryant, D., Bouckaert, R., Felsenstein, J., Rosenberg, N. A., & RoyChoudhury, A. (2012).

    Inferring species trees directly from biallelic genetic markers: Bypassing gene trees in a

    full coalescent analysis. *Molecular Biology and Evolution*, *29*(8), 1917-1932.

    https://doi.org/10.1093/molbev/mss086

Burnham, K. P., & Anderson, D. R. (2002). *Model selection and multimodel inference: A*

    practical information-theoretical approach. New York, NY: Springer.

Cahill, J. A., Stirling, I., Kistler, L., Salamzade, R., Ersmark, E., Fulton, T.L., … Shapiro, B.

    (2013). Genomic evidence for island population conversion resolves conflicting theories

    of polar bear evolution. *PLoS Genetics*, *9*, e1003345.

    https://doi.org/10.1371/journal.pgen.1003345

Campbell, J. A. (1999). Distribution patterns of amphibians in Middle America. In W. E.

    Duellman (Ed.), *Patterns of Distribution of Amphibians: A Global Perspective*. (pp. 111-

    120). Baltimore, MD: Johns Hopkins University Press.

Carr, A. F. (1950). Outline for a classification of animal habitats in Honduras. *Bulletin of the*

    American Museum of Natural History, *94*, 567-594.

Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A., & Cresko, W. A. (2013). Stacks: an

    analysis tool set for population genomics. *Molecular Ecology*, *22*(11), 3124-3140.

    https://doi.org/10.1111/mec.12354

Chan, K. M. A., & Levin, S. A. (2005). Leaky prezygotic isolation and porous genomes: rapid

    introgression of maternally inherited DNA. *Evolution*, *59*(4), 720-729.

    https://doi.org/10.1111/j.0014-3820.2005.tb01748

Chavez, A. S., Maher, S. P., Arbogast, B. S., & Kenagy, G. J. (2013). Diversification and gene

    flow in nascent lineages of island and mainland North American tree squirrels

    (*Tamiasciurus*). *Evolution*, *68*(4), 1094-1109. https://doi.org/10.1111/evo.12336

Constable, H., Guralnick, R., Wieczorek, J., Spencer, C., & Peterson, A. T. (2010). VertNet: A

    new model for biodiversity data sharing. *PLoS Biology*, *8*(2): e1000309.

    https://doi.org/10.1371/journal.pbio.1000309

Consuegra, S. G., & Vázquez-Domínguez, E. (2015). Mitochondrial diversification of the

    Peromyscus mexicanus species group in Nuclear Central America: biogeographic and

    taxonomic implications. *Journal of Zoological Systematics and Evolutionary Research*,

    53(4), 300-311. https://doi.org/10.1111/jzs.12099

Currant, M., Ruedi, M., Petit, R. J., & Excoffier, L. (2008). The hidden side of invasions:

    Massive introgression by local genes. *Evolution*, *62*, 1908-1920.

    https://doi.org/10.1111/j.15585646.2008.00413.x

Daza, J. M., Castoe, T. A., & Parkinson, C. L. (2010). Using regional comparative

    phylogeographical data from snake lineages to infer historical processes in Middle

    America. *Ecography*, *33*(2), 343-354. https://doi.org/10.1111/j.1600-0587.2010.06281.x

DeBiasse, M. B., Nelson, B. J., & Hellberg, M. E. (2013). Evaluating summary statistics used to

    test for incomplete lineage sorting: mito-nuclear discordance in the reef sponge

    Callyspongia vaginalis. *Molecular Ecology*, *23*(1), 225-238.

    https://doi.org/10.1111/mec.12584

Devitt, T. J., Devitt, S. E. C., Hollingsworth, B. D., McGuire, J. A., & Moritz, C. (2013).

    Montane refugia predict population genetic structure in the large-blotched Ensatina

    salamander. *Molecular Ecology*, *22*(6), 1650-1665. https://doi.org/10.1111/mec.12196

Earl, D. A., & vonHoldt, B. M. (2012). STRUCTUREHARVESTER: a website and program for

    visualizing STRUCTURE output and implementing the Evanno method. *Conservation*

    Genetics Resources, *4*(2), 359-361. https://doi.org/10.1007/s12686-011-9548-7

Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high

    throughput. *Nucleic Acids Research*, *32*(5), 1792-1797.

    https://doi.org/10.1093/nar/gkh340

Elith, J., Phillips, S. J., Hastie, T., Dudík, M., Chee, Y. E., & Yates, C. J. (2011). A statistical

    explanation of MaxEnt for ecologists. *Diversity and Distributions*, *17*(1), 43-57.

    https://doi.org/10.1111/j.1472-4642.2010.00725.x

Evanno, G., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of individuals

using the software STRUCTURE: a simulation study. *Molecular Ecology*, *14*(8), 2611-

2620. https://doi.org/10.1111/j.1365-294X.2005.02553.x

Falush, D., Stephens, M., & Pritchard, J. K. (2003) Inference of population structure using

multilocus genotype data: linked loci and correlated allele frequencies. *Genetics*, *164*(4),

1567-1587.

Feng, X., Park, D. S., Liang, Y., Pandey, R., & Papes, M. (2019). Collinearity in ecological niche

modeling: Confusions and challenges. *Ecology and Evolution*, *9*(18), 10365-10376.

https://doi.org/10.1002/ece3.5555

Firneno, T. J. Jr., Luque-Montes, I., & Townsend, J.H. (2017). An enigmatic record of *Incilius

porteri* (Anura: Bufonidae) from Isla del Tigre, Honduras. *Salamandra*, *53*, 160-162.

Firneno, T.J. Jr., & Townsend, J. H. (2019). Evaluation of species boundaries in sympatric and

parapatric populations of Mesoamerican toads. *Zoologica Scripta*, *48*, 454-465.

https://doi.org/10.1111/zsc.12354

Flint, R. F. (1947). *Glacial Geology and the Pleistocene Epoch*. New York, NY: Wiley & Sons.

Folmer, O., Black, M., Hoeh, W., Lutz, R., & Vrijenhoek, R. (1994). DNA primers for

amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan

invertebrates. *Molecular Marine Biology and Biotechnology 3*(5), 294-299.

https://doi.org/10.1371/journal.pone.0013102

Fujita, M. K., McGuire, J. A., Donnellan, S. C., & Moritz, C. (2010). Diversification and

persistence at the arid-monsoonal interface: Australia-wide biogeography of the Bynoe's

gecko (*Heteronotia binoei*; Gekkonidae). *Evolution*, *64*(8), 2293-2314.

https://doi.org/10.1111/j.1558-5646.2010.00993.x

Fujita, M. K., Leaché, A. D., Burbrink, F. T., McGuire, J. A., & Moritz, C. (2012). Coalescent-based species delimitation in an integrative taxonomy. *Trends in Ecology and Evolution*, *27*(9), 480-488. https://doi.org/10.1016/j.tree.2012.04.012

Funk, D. J., & Omland, K. E. (2003). Species-level paraphyly and polyphyly: frequency, causes, and consequences, with insights from animal mitochondrial DNA. *Annual Reviews*, *34*, 397-423. https://doi.org/10.1146/annurev.ecolsys.34.011802.132421

Gompert, Z., & Buerkle, C. A. (2011). Bayesian estimation of genomic clines. *Molecular Ecology*, *20*(10), 2111-2127. https://doi.org/10.1111/j.1365-294X.2011.05074.x

Graham, M. H. (2003). Confronting multicollinearity in ecological multiple regression. *Ecology*, *84*(11), 2809-2815. https://doi.org/10.1890/02-3114

Grummer, J. A., Bryson, R. W., & Reeder, T. W. (2014). Species delimitation using Bayes factors: simulations and application to the *Sceloporus scalaris* species group (Squamata: Phrynosomatidae). *Systematic Biology*, *63*(2), 119-133. https://doi.org/10.1093/sysbio.syt069

Grummer, J. A., Calderón-Espinosa, M. L., de Oca, A. N., Smith, E. N., Méndez-de la Cruz, F. R., & Leaché, A. D. (2015). Estimating the temporal and spatial extent of gene flow among sympatric lizard populations (genus *Sceloporus*) in the southern Mexican highlands. *Molecular Ecology*, *24*(7), 1523-1542. https://doi.org/10.1111/mec.13122

Gutenkunst, R. N., Hernandez, R. D., Williamson, S. H., & Bustamante, C. D. (2009). Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLoS Genetics*, *5*(10), e1000695. https://doi.org/10.1371/journal.pgen.1000695

Gutiérrez-Rodríguez, C., Ornelas, J. F., & Rodríguez-Gómez, F. (2011). Chloroplast DNA

    phylogeography of a distylous shrub (*Palicourea padifolia*, Rubiaceae) reveals past

    fragmentation and demographic expansion in Mexican cloud forests. *Molecular*

    Phylogenetics and Evolution, *61*(3), 603-615.

    https://doi.org/10.1016/j.ympev.2011.08.023

Gutiérrez-Garcia, T. A., & Vásquez-Domínguez, E. (2012). Biogeographically dynamic genetic

    structure bridging two continents in the monotypic Central American rodent *Ototylomys*

    phyllotis. *Biological Journal of the Linnean Society*, *107*(3), 593-610.

    https://doi.org/10.1111/j.1095-8312.2012.01966.x

Hewitt, G. (2000). The genetic legacy of the Quaternary ice ages. *Nature*, *405*, 907-913.

    https://doi.org/10.1038/35016000

Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G., & Jarvis, A. (2005). Very high

    resolution interpolated climate surfaces for global land area. *International Journal of*

    Climatology, *25*(15), 1965-1978. https://doi.org/10.1002/joc.1276

Ivanov, V., Lee, K. M., & Mutanen, M. (2018). Mitonuclear discordance in wolf spiders:

    Genomic evidence for species integrity and introgression. *Molecular Ecology*, *27*, 1681-

    1695. https://doi.org/10.1111/mec.14564

Jombart, T. (2008). *adegenet*: a R package for the multivariate analysis of genetic markers.

    Bioinformatics, *24*(11), 1403-1405. https://doi.org/10.1093/bioinformatics/btn129

Jombart, T., & Ahmed, I. (2011). *adegene 1.3-1t*: new tools for the analysis of genome-wide

    SNP data. *Bioinformatics*, *27*(21), 3070-3071.

    https://doi.org/10.1093/bioinformatics/btn521

Kass, J. M., Vilela, B., Aiello-Lammens, M. E., Muscarella, R., Merow, C., & Anderson, R. P. (2018). WALLACE: a fexible platform for reproducible modeling of species niches and distributions built for community expansion. *Methods in Ecology and Evolution*, *9*(4), 1151-1156. https://doi.org/10.1111/2041-210X.12945

Krosby, M., & Rohwer, S. (2009). A 2000 km genetic wake yields evidence for northern glacial refugia and hybrid zone movement in a pair of songbirds. *Proceedings of the Royal Society of London, Series B, Biological Sciences*, *276*, 615-621.

Kumar, S., Stecher, G., & Tamura, K. (2016). MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Molecular Biology and Evolution*, *33*(7), 1870–1874. https://doi.org/10.1093/molbev/msw054

Leaché, A. D., Fujita, M. K., Minin, V. N., & Bouckaert, R. R. (2014). Species delimitation using genome-wide SNP data. *Systematic Biology*, *63*(4), 534-542. https://doi.org/10.1093/sysbio/syu018

Macey, J., & Schulte, J. (1998). Phylogenetic relationships of toads in the *Bufo bufo* species group from the Eastern Escarpment of the Tibetan Plateau: A case of vicariance and dispersal. *Molecular Phylogenetics and Evolution*, *9*(1): 80-87.

Malinsky, M., Trucchi, E., Lawson, D. J., & Falush, D. (2018). RADpainter and fineRADstructure: population inference for RADseq data. *Molecular Biology and Evolution*, *35*(5), 1284-1290. https://doi.org/10.1093/molbev/msy023

Matamoros, W. A., Kreiser, B. R., & Schaefer, J. F. (2012). A delineation of Nuclear Middle America biogeographical provinces based on river basin faunistic similarities. *Fish Biology and Fisheries*, *22*(1), 351-365. https://doi.org/10.1007/s11160-011-9232-8

McCranie, J. R. (2009). Amphibians and Reptiles of Honduras. Listas Zoológicas Actualizadas

UCR. *Listas Zoologicas Actualizadas UCR*, Museo de Zoología UCR. San Pedro, Costa

Rica. Retrieved from http://museo.biologia.ucr.ac.cr/Listas/LZAPublicaciones.htm

McCranie, J. (2015). A checklist of the amphibians and reptiles of Honduras, with additions,

comments, on taxonomy, some recent taxonomic decisions, and areas of further study

needed. *Zootaxa*, 3931(3), 352-386.

McCranie, J. R., & Castañeda, F. E. (2007). *Guia de Campo de los Anfibios de Honduras*. Salt

Lake City, UT: Bibliomania!

McCranie, J. R., & Wilson, L. D. (2002). *The Amphibians of Honduras*. Ithaca, NY: Society for

theStudy of Amphibians and Reptiles.

McKay, B. D., & Zink, R. M. (2010) The causes of mitochondrial DNA gene tree paraphyly in

birds. *Molecular Phylogenetics and Evolution*, *54*(2), 647-650.

https://doi.org/10.1016/j.ympev.2009.08.024

McLean, B. S., Jackson, D. J., & Cook, J. A. (2016). Rapid divergence and gene flow at high

latitudes shape the history of Holarctic ground squirrels (*Urocitellus*). *Molecular

Phylogenetics and Evolution*, *102*, 174-188. https://doi.org/10.1016/j.ympev.2016.05.040

Mee, J. A., & Taylor, E. B. (2012). The cybrid invasion: widespread postglacial dispersal by

Phoxinus (Pisces: Cyprinidae) cytoplasmic hybrids. *Canadian Journal of Zoology*, *90*,

577-584.

Mendelson, III, J. R., Williams, B. L., Sheil, C. A., & Mulcahy, D. G. (2005). Systematics of the

Bufo coccifer complex (Anura: Bufonidae) of Mesoamerica. *Scientific Papers: Natural

History Museum The University of Kansas*, *38*, 1-27. https://doi.org/10.1007/s11084-011-

9248-z

Mendelson, III, J. R., Mulcahy, D. G., Williams, T. S., & Sites, J. W. (2011). A phylogeny and
    evolutionary natural history of Mesoamerican toads (Anura: Bufonidae: Incilius) based
    on morphology, life history, and molecular data. *Zootaxa*, *34*(3138), 1-34.
    https://dx.doi.org/10.11646/zootaxa.3138.1.1

Moritz, C. M., Patton, J. L., Schneider, C. J., & Smith, T. B. (2000). Diversification of rainforest
    faunas: an integrated molecular approach. *Annual Reviews*, *31*, 533-563.
    https://doi.org/10.1146/annurev.ecolsys.31.1.553

Muscarella, R., Galante, P. J., Soley-Guardia, M., Boria, R. A., Kass, J. M., Uriarte, M., &
    Anderson, R. A. (2014). ENMeval: An R package for conducting spatially independent
    evaluations and estimating optimal model complextity for MAXENT ecological niche
    models. *Methods in Ecology and Evolution*, *5*(11), 1198-1205.
    https://doi.org/10.1111/2041-210X.12261

Orneales, J. F., González, C., Espinosa de los Monteros, A., Rodríguez-Gómez, F., & García-
    Feria, L. M. (2014). In and out of Mesoamerica: temporal divergence of *Amazilia*
    hummingbirds pre-dates the orthodox account of the completion of the Isthmus of
    Panama. *Journal of Biogeography*, *41*(1), 168-181. https://doi.org/10.1111/jbi.12184

Ornelas, J. F., & Rodríguez-Gómez, F. (2015). Influence of Pleistocene glacial/interglacial
    cycles on the genetic structure of the Mistletoe Cactus *Rhipsalis baccifera* (Cactaceae) in
    Mesoamerica. *Journal of Heredity*, *106*(2), 196-210.
    https://doi.org/10.1093/jhered/esu113

Orozco-terWengel, P., Andreone, F., Louis, E. Jr., & Vences, M. (2013). Mitochondrial
    introgressive hybridization following a demographic expansion in the tomato frogs of

Madagascar, genus *Dyscophus*. *Molecular Ecology*, *22*(24), 6074-6090.

https://doi.org/10.1111/mec.12558

Papakostas, S., Michaloudi, E., Proios, K., Brehm, M., Verhage, L., Rota, J., & Declerck, S. A. (2016). Integrative taonxomy recognizes evolutionary units despite widespread mitonuclear discordance: Evidence from a rotifer cryptic species complex. *Systematic Biology*, *65*, 508-524. https://doi.org/10.1039/sysbio/syw016

Pavlova, A., Amos, J. N., Joseph, L., Loynes, K., Austin, J. J., Keogh, J. S., Stone, G. N., Nicholls, J. A., & Sunnucks, P. (2013). Perched at the mito-nuclear crossroads: divergent mitochondrial lineages correlate with environment in the face of ongoing nuclear gene flow in an Australian bird. *Evolution*, *67*(12), 3412-3428.

https://doi.org/10.1111/evo.12107

Pereira, R. J., Martínez-Solano, I., & Buckley, D. (2016). Hybridization during altitudinal range shifts: nuclear introgression leads to extensive cyto-nuclear discordance in the fire salamander. *Molecular Ecology*, *25*(7), 1551-1565. https://doi.org/10.1111/mec.13575

Perktas, U., Barrowclough, G. F., & Groth, J. G. (2011). Phylogeography and species limits in the green woodpecker complex (Aves: Picidae): multiple Pleistocene refugia and range expansion across Europe and the Near East. *Biological Journal of the Linnean Society*, *104*(3), 710-723. https://doi.org/10.1111/j.1095-8312.2011.01750.x

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

https://doi.org/10.1371/journal.pone.0037135

Phillips, S. J., & Dudík, M. (2008). Modeling of species distributions with Maxent: new

  extensions and a comprehensive evaluation. *Ecography*, *31*(2), 161-175.

  https://doi.org/10.1111/j.0906-7590.2008.5203.x

Phuong, M. A., Bi, K., & Moritz, C. (2016). Range instability leads to cytonuclear discordance

  in a morphologically cryptic ground squirrel species complex. *Molecular Ecology*, *26*,

  4743-4755. https://doi.org/10.1111/mec.14238

Portik, D. M., Leaché, A. D., Rivera, D., Barej, M. F., Burger, M., Hirschfeld, M., … Fujita, M.

  K. (2017). Evaluating mechanisms of diversification in a Guineo-Congolian tropical

  forest frog using demographic model selection. *Molecular Ecology*, *26*(19), 5245-5263.

  https://doi.org/10.1111/mec.14266

Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using

  multilocus genotype data. *Genetics*, *155*(2), 945-959.

  https://doi.org/10.1006%2Ftpbi.2001.1543

Radosavljevic, A., & Anderson, R. P. (2014). Making better MAXENT models of species

  distributions: complexity, overfitting and evaluation. *Journal of Biogeography*, *41*(4),

  629-643. https://doi.org/10.1111/jbi.12227

Rambaut, A., & Drummond, A. (2009). Tracer v1.5.0. Available from: http://beast.bio.ed.ac.uk/

Raven, P. H., & Axelrod, D. I. (1974). Angiosperm biogeography and past continental

  movements. *Annals of the Missouri Botanical Garden*, *61*(3), 539-673.

  https://doi.org/10.2307/2395021

Reid, N. M., Hird, S. M., Brown, J. M., Pelletier, T. A., McVay, J. D., Staler, J. D., & Carstens,

  B. C. (2013). Poor fit to the multispecies coalescent is widely detectable in empirical

  data. *Systematic Biology*, *63*(3), 322-333. https://doi.org/10.1093/sysbio/syt057

Ribeiro, Â. M., Lloyd, P., & Bowie, R. C. K. (2011). A tight balance between natural selection and gene flow in a southern African arid-zone endemic bird. *Evolution*, *65*(12), 3499-3514. https://doi.org/10.1111/j. 1558-5646.2011.01397.x

Rodríguez-Gómez, F., & Ornelas, J. F. (2014). Genetic divergence of the Mesoamerican azure-crowned hummingbird (*Amazilia cyanocephala*, Trochilidae) across the Motagua-Polochic-Jocotán fault system. *Journal of Biological Sysematics and Evolutionary Research*, *52*(2), 142-153. https://doi.org/10.1111/jzs.12047

Rovito, S. M., & Parra-Olea, G. (2015). Two new species of *Chiropterotriton* (Caudata: Plethodontidae) from northern Mexico. *Zootaxa*, *4048*(1), 57-74. https://dx.doi.org/10.11646/zootaxa.4048.1.3

Rovito, S. M., Vásquez-Almazán, C. R., Papenfuss, T. J., Parra-Olea, G., & Wake, D. B. (2015). Biogeography and evolution of Central American cloud forest salamanders (Caudata: Plenthodontidae: *Cryptotriton*), with the description of a new species. *Biological Journal of the Linnean Society*, *175*(1), 150-166. https://doi.org/10.1111/zoj.12268

Rowe, K. C., Heske, E. J., Brown, P. W., & Paige, K. N. (2004). Surviving the ice: Northern refugia and postglacial colonization. *Proceedings of the National Academy of Sciences*, *101*(28), 10355-10359. https://doi.org/10.1073/pnas.0401338101

Sambrook, J., & Russell, D. W. (2006). Purification of nucleic acids by extraction with phenol:chloroform. In J. Sambrook & D. W. Russell (Eds.), *Commonly Used Techniques in Molecular Cloning.* (pp. Appendix 8). Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Sánchez-Ramos, L. E., Gordillo-Martínez, A., Gutíerrez-Arellano, C. R., Kobelkowsky-Vidrio, T., Ríos-Muñoz, C. A., & Navarro-Sigüenza, A. G. (2018). Bird diversity patterns in the

Nuclear Central American highlands: a conservation priority in the northern Neotropics. *Tropical Conservation Science*, *11*, 1-17. https://doi.org/10.1177/1940082918819073

Seutin, G., White, B. N., & Boag, P. T. (1991). Preservation of avian blood and tissue samples for DNA analyses. *Canadian Journal of Zoology*, *69*(1), 82-90. https://doi.org/10.1139/z91-013

Sequeira, F., Sodré, D., Ferrand, N., Bernardi, J. A. R., Sampaio, I., Schneider, H., & Vallinoto, M. (2011). Hybridization and massive mtDNA unidirectional introgression between two closely related Neotropical toads *Rhinella marina* and *R. schneideri* inferred from mtDNA and nuclear markers. *BMC Evolutionary Biology*, *11*(264). https://doi.org/10.1186/1471-2148-11-264

Shcheglovitova, M., & Anderson, R. P. (2013). Estimating optimal complexity for ecological niche models: A jackknife approach for species with small sample sizes. *Ecological Modeling*, *296*(10), 9-17. https://doi.org/10.1016/j.ecolmodel.2013.08.011

Singhal, S., & Moritz, C. (2012). Testing hypotheses for genealogical discordance in a rainforest lizard. *Molecular Ecology*, *21*(20), 5059-5072. https://doi.org/10.1111/j.1365-294X.2012.05747

Sousa, V., & Hey, J. (2013). Understanding the origin of species with genome-scale data: modelling gene flow. *Nature Reviews Genetics*, *14*, 404-414. https://doi.org/10.1038/nrg3446

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis or large phylogenies. *Bioinformatics*, *30*(9), 1312-1313. https://doi,org/10.1093/bioinformatics/btu.003

Stanley, R. R. E., Jeffery, N. W., Wringe, B. F., DiBacco, C., & Bradbury, I. R. (2017). GENEPOPEDIT: a simple and flexible tool for manipulating multilocus molecular data in R. *Molecular Ecology Resources*, *17*(1), 12-18. https://doi.org/10.1111/1755-0998.12569

Streicher, J. W., Devitt, T. J., Golberg, C. S., Malone, J. H., Blackmon, H., & Fujita, M. K. (2014). Diversification and asymmetrical gene flow across time and space: Lineage sorting and hybridization in polytypic barking frogs. *Molecular Ecology*, *23*(13), 3273-3291. https://doi.org/10.1111/mec.12814

Toews, D. L., & Brelsford, A. (2012). The biogeography of mitochondrial and nuclear discordance in animals. *Molecular Ecology*, *21*(16), 3907-3930. https://doi.org/10.1111/j.1365-294X.2012.05664.x

Townsend, J. H. (2014). Characterizing the Chortís Block Biogeographic Province: geological, physiographic, and ecological associations and herpetofaunal diversity. *Mesoamerican Herpetology*, *1*, 204-252.

Turmelle, A. S., Kunz, T. H., & Sorenson, M. D. (2011). A tale of two genomes: contrasting patterns of phylogeographic structure in a widely distributed bat. *Molecular Ecology*, *20*(2), 357-375. https://doi.org/10.1111/j.1365-294X.2010.04947.x

Veloz, S. D. (2009). Spatially autocorrelated sampling falsely inflates measures of accuracy for presence-only niche models. *Journal of Biogeography*, *36*(12), 2290-2299. https://doi.org/10.1111/j.1365-2699.2009.02174.x

Villalobos, F. (2013). Tree squirrels: A key to understand the historic biogeography of Mesoamerica? *Mammalian Biology*, *78*(4), 258-266. https://doi.org/10.1016/j.mambio.2013.02.003

Wang, W., Wang, Y., Lei, F., Liu, Y., Wang, H., & Chen, J. (2018). Incomplete lineage sorting

and introgression in the diversification of Chinese spot-billed ducks and mallards.

Current Zoology, *65*(5), 589-597. https://doi.org/10.1093/cz/zoy074

Warren, D. L., & Seifert, S. N. (2011). Ecological niche modeling in Maxent: the importance of

model complexity and the performance of model selection criteria. *Ecological*

Applications, *21*(2), 335-342. https://doi.org/10.1890/10-1171.1

Weigand, H., Weiss, M., Cai, H., Li, Y., Yu, L., Zhang, C., & Leese, F. (2017). Deciphering the

origin of mito-nuclear discordance in two sibling caddisfly species. *Molecular Ecology*,

26(20), 5075-5715. https://doi.org/10.1111/mec.14292.

Wilson, L. D., & Johnson, J. D. (2010). Distributional patterns of the herpetofauna of

Mesoamerica, a biodiversity hotspot. In L. D. Wilson, J. H. Townsend, & J. D. Johnson

(Eds.), *Conservation of Mesoamerican Amphibians and Reptiles*. (pp. 30-235). Eagle

Mountain, UT: Eagle Mountain Publishing, LC.

Wilson, L. D., & Meyer, J. R. (1985). *The Snakes of Honduras* (2nd Edition). Milwaukee, WI:

Milwaukee Public Museum.

Yannic, G., Pellissier, L., Ortego, J., Lecomte, N., Couturier, S., Cuyler, C., … Dussault, C.

(2014). Genetic diversity in caribou linked to past and future climate change. *Nature*

Climate Change, *4*, 132-137. https://doi.org/10.1038/nclimate2074

Zink, R. M., & Barrowclough, G. F. (2008). Mitochondrial DNA under siege in avian

phylogeography. *Molecular Ecology*, *17*(9), 2107-2121. https://doi.org/10.1111/j.1365-

294X.2008.03737.x

TABLES

Table 1. Bayes Factor Delimitation results for each model analysis. The number of species represents the number of species included in each analysis after lumping or splitting lineages.

Model	Species	MLE	BF	Rank
Analysis 1: With Admixed Individuals				
Current Taxonomy	3	-14016.147	—	4
WCH, OCH H, and OCH L				
Lump Highland	2	-19094.389	5078.242	5
WCH + OCH H and OCH L				
Split All Populations	5	-13402.808	-1226.687	1
WCH W, WCH C, WCH E, OCH H, and OCH L				
Split All, Clump WCH W and WCH C	4	-13550.717	-930.86	3
WCH W + WCH C, WCH E, OCH H, and OCH L				
Split All, Clump WCH E and WCH C	4	-13463.178	-1105.938	2
WCH W, WCH C + WCH E, OCH H, and OCH L				
Analysis 2: Without Admixed Individuals				
Current Taxonomy	3	-8619.359	—	2
WCH, OCH H, and OCH L				
Lump Highland	2	-11789.19	3169.831	3
WCH + OCH H and OCH L				
Split All Populations	4	-8046.2958	-1146.126	1
WCH W, WCH E, OCH H, and OCH L				

FIGURES

Figure 1. Sampling localities and geographic distribution of five populations within the *I. coccifer* complex. Locality colors correspond to population assignments from SNP data. The dotted line indicates the boundaries of the Chortís Block, the solid black line represents the Honduran Depression, and the solid blue line represents the Mejocote-San Juan Depression.
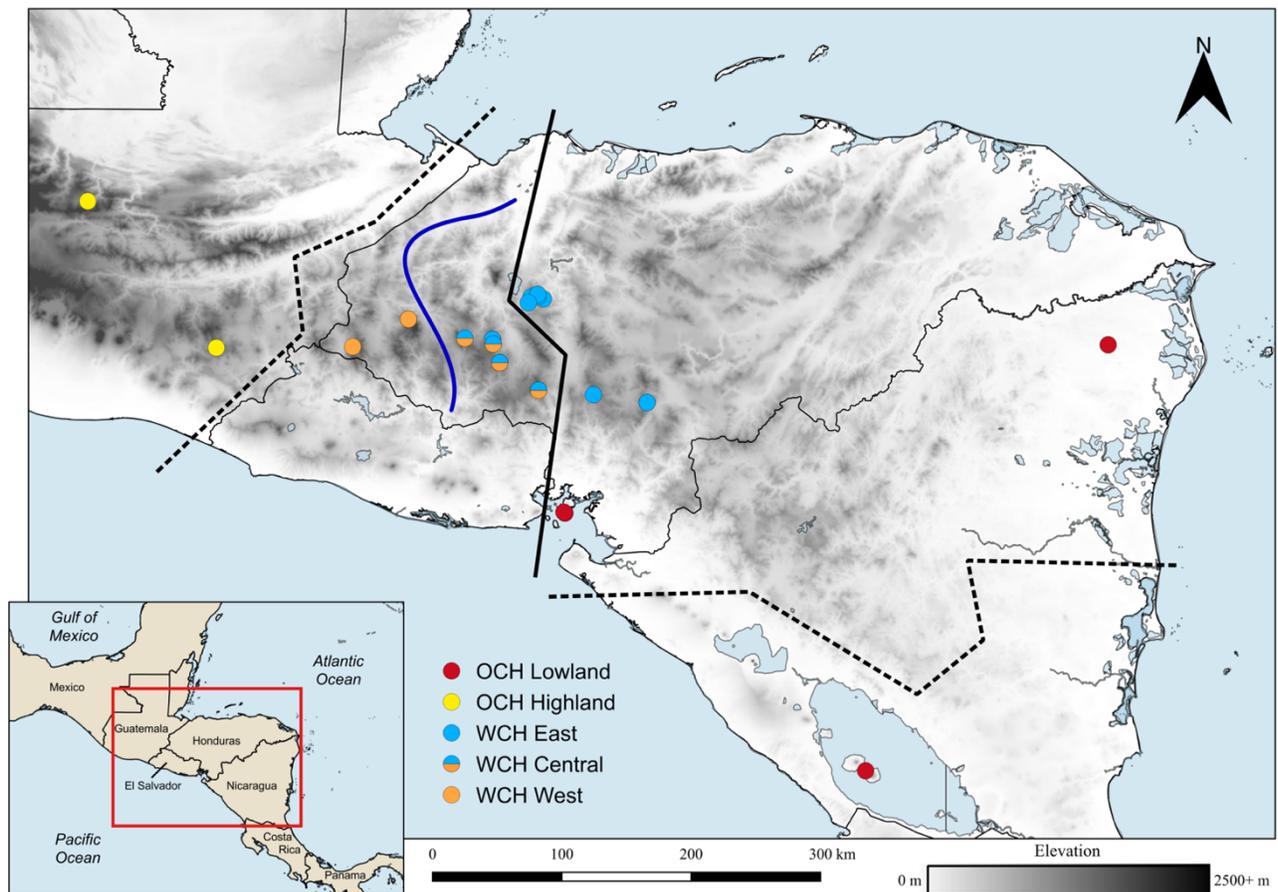
Figure 2. Population assignment for 64 individuals based on (a) hierarchical Bayesian population clustering using STRUCTURE, (b) maximum-likelihood population clustering using ADMIXTURE, (c) discriminant analysis of principal components, and (d) coancestry matrix from FINERADSTRUCTURE (coefficients of coancestry are color coded from low (yellow) to high (black) and the dendrogram depicts a clustering of individual samples based on the pairwise matrix of coancestry coefficients). The spatial distribution of populations is presented with the same color scheme as Figure 1.
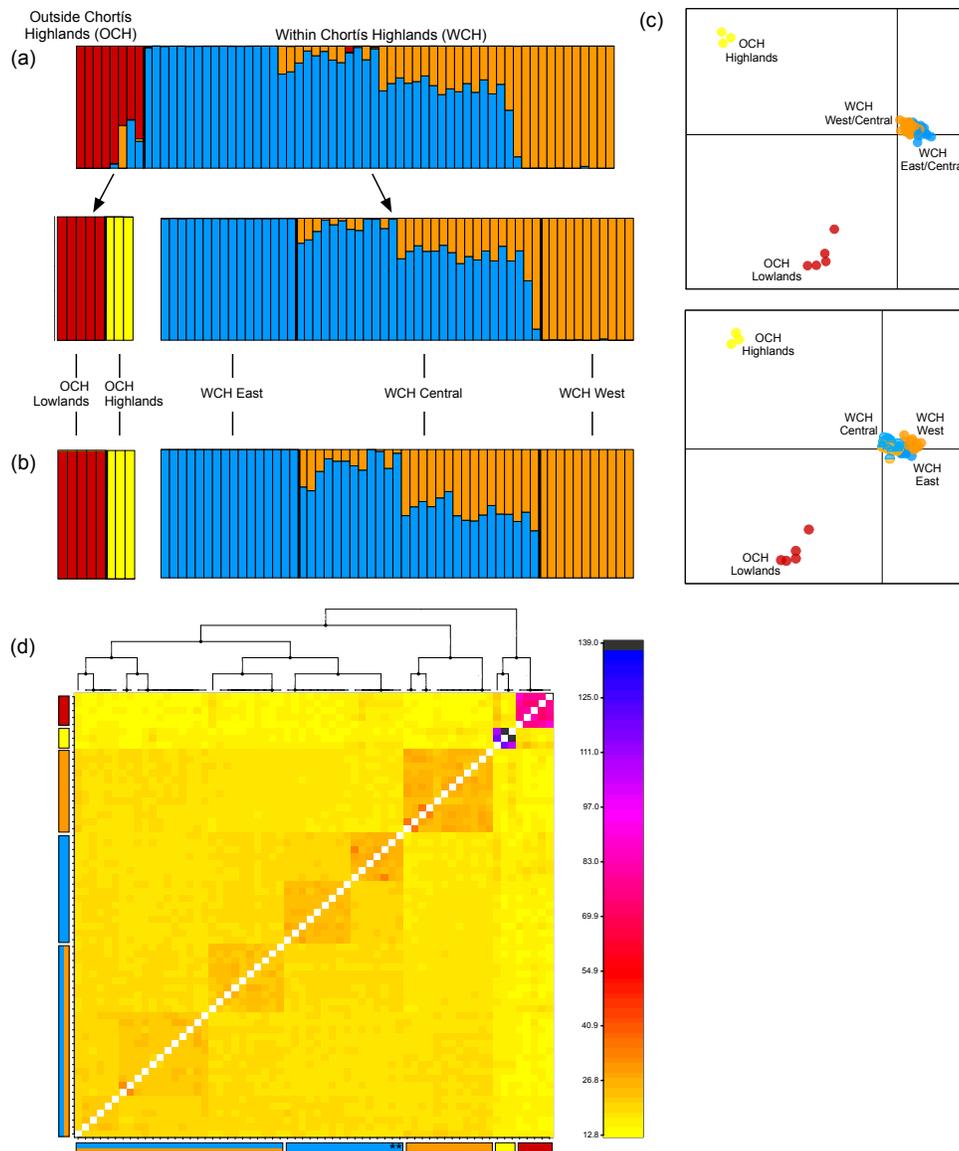
Figure 3. Chronogram of the *Incilius coccifer* complex (left) from a BEAST analysis of mtDNA data calibrated with a mutation rate and a phylogeny of the *I. coccifer* complex (right) from a Bayesian coalescent analysis of SNP data using SNAPP. Nodes with high support (posterior probability >0.9) are indicated by black dots. Median ages on chronogram are provided above nodes, with 95% highest posterior densities (HPD) below and error bars representing the 95% HPD on the nodes. Colored boxes show population assignments based on the results for the mtDNA phylogeny and based on the population clustering analyses of SNP data.
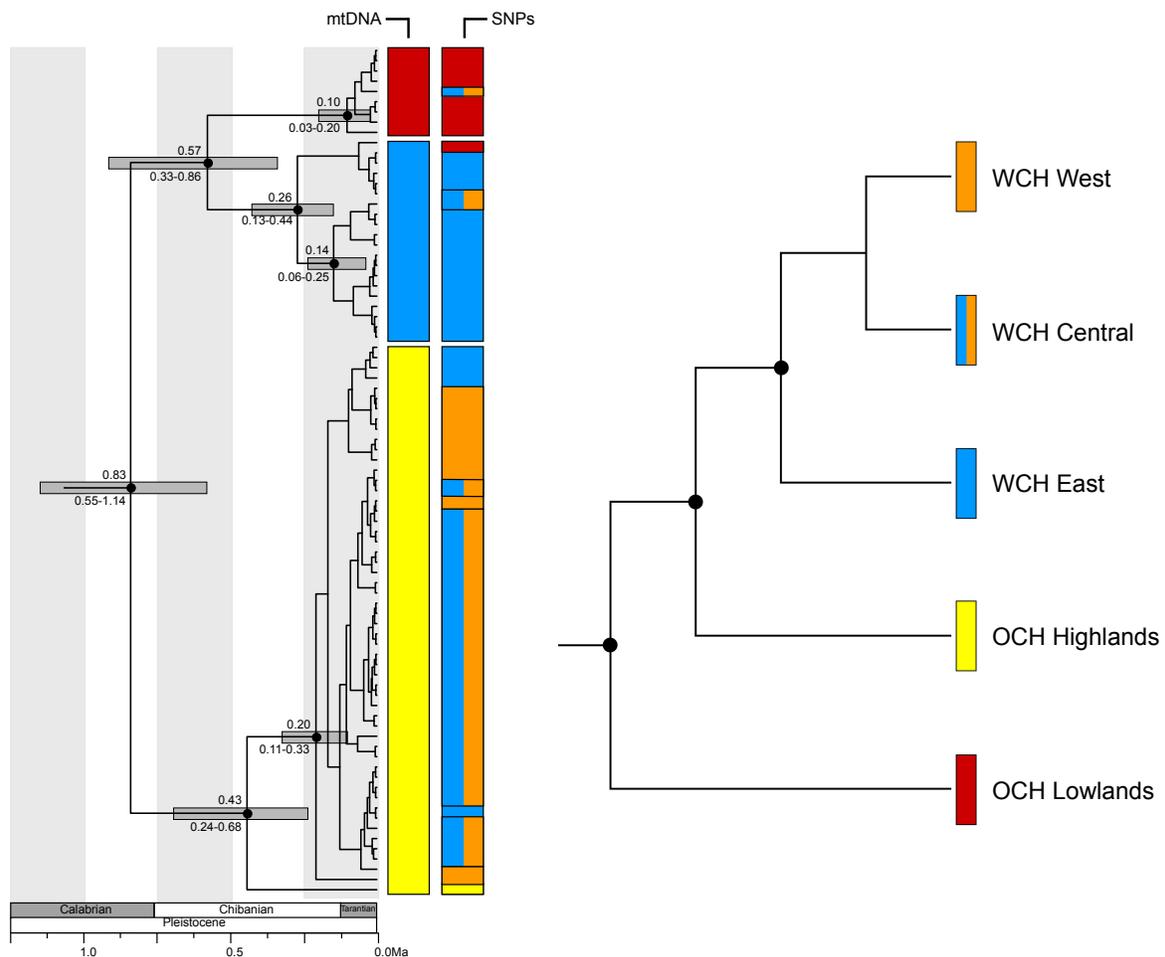
Figure 4. Results of genomic cline analysis depicting the probability of western ancestry given background genomic introgression (e.g. hybrid index). The dashed line represents the perfect linear correlation between the hybrid index and ancestry probability as expected under neutral introgression. Clines shown in blue and orange represent loci with excess ancestry from one parental population (WCH East and WCH West, respectively). The histogram above the cline graph depicts the relative frequencies of individual hybrid indices within the admixed population (WCH Central), and the box is the range of these values on the genomic cline.
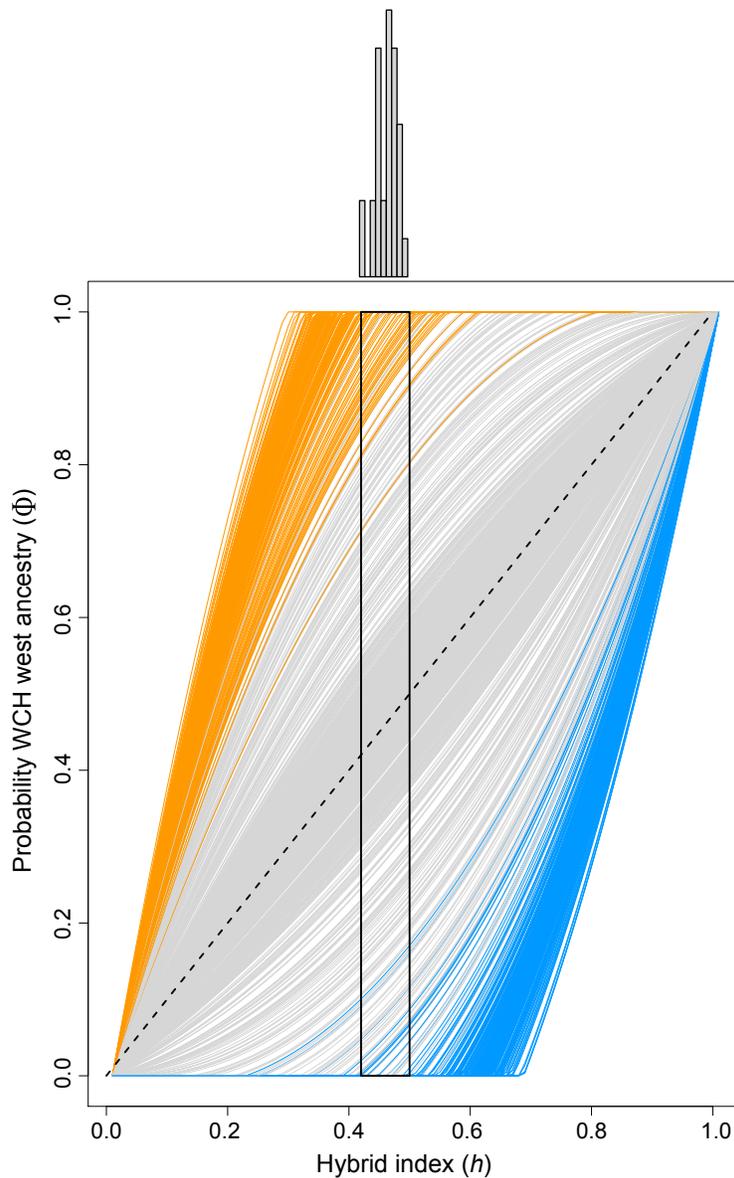
Figure 5. Results of population genetic model comparisons using the 3D site frequency spectrum between the three WCH populations. A simple graphic of the best-fit model is depicted, along with comparisons of the pairwise 2D site frequency spectrum for the data, the model, and the resulting residuals for each population combination. Parameter values for this and additional models are provided in Supporting File 3.
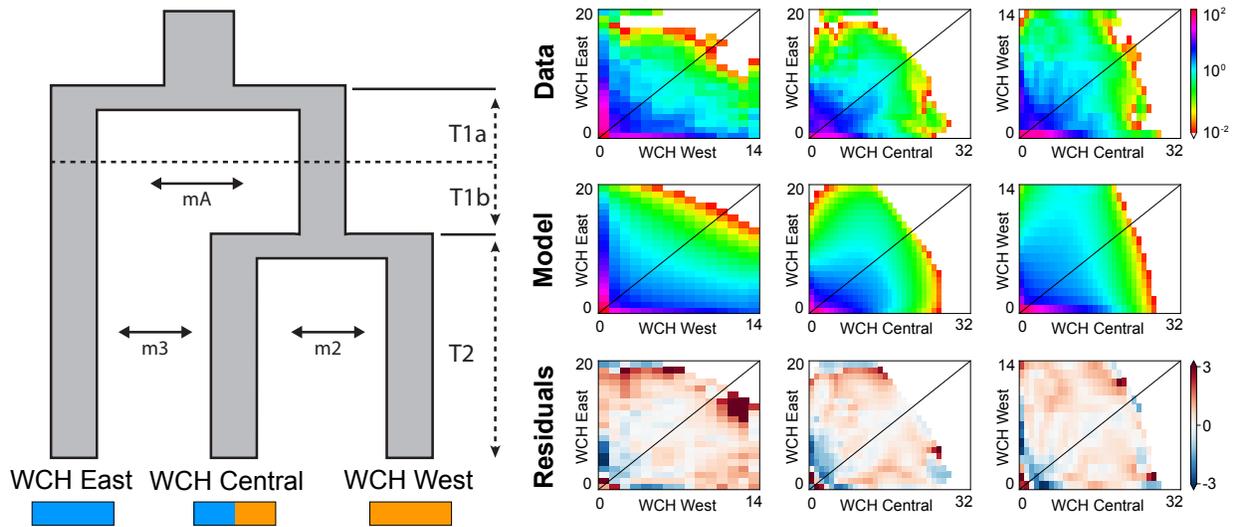
Figure 6. Stability surfaces representing the presence of suitable habitat across the last interglacial period to present day climate within the Chortís Highlands of Honduras. Warmer colors represent areas of higher habitat stability.
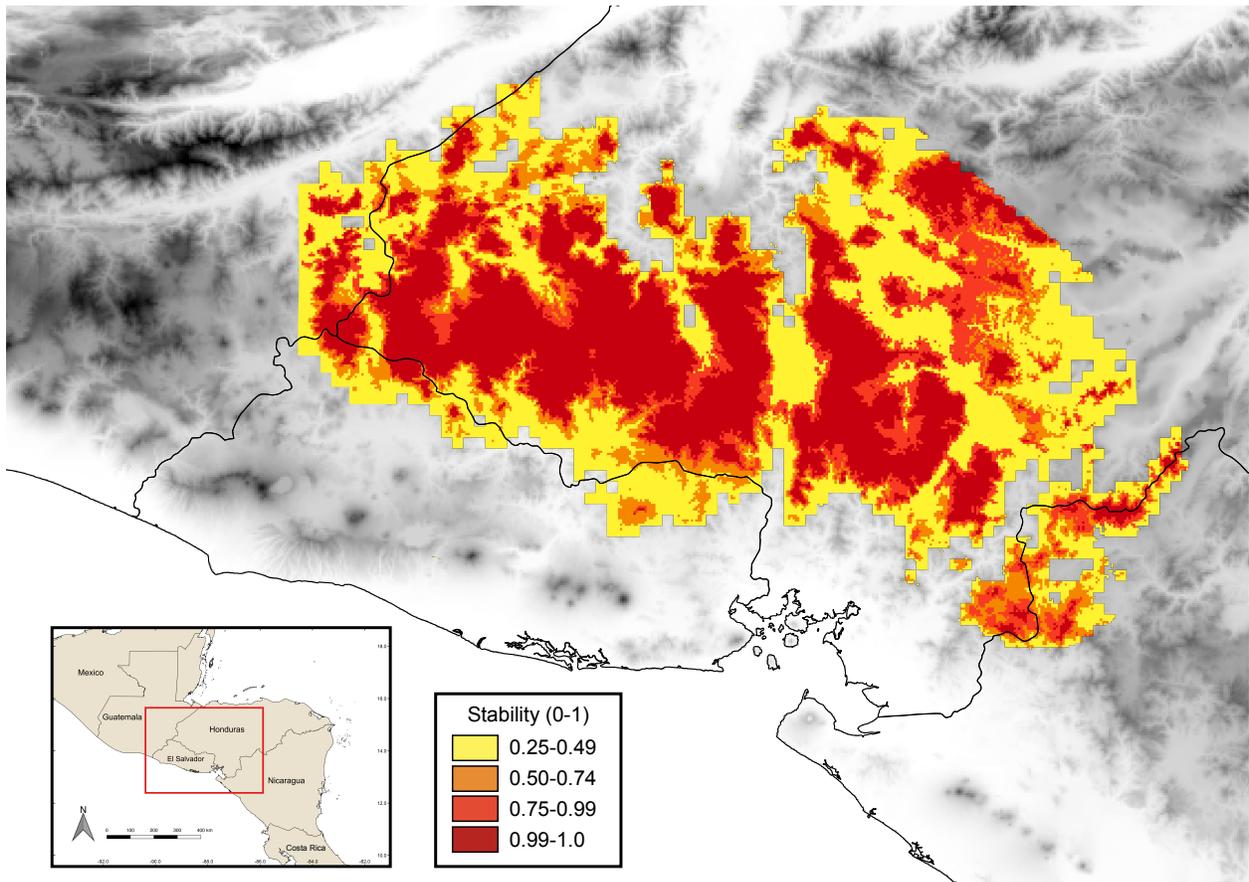
SUPPORTING INFORMATION

File S1. (A) Comparison of maximum likelihood tree topologies using two mitochondrial datasets of differing sizes. The tree on the left was generated using a concatenated dataset of 16S and COI (1150 bp) of 67 individuals. The tree on the right was generated using a dataset of COI (658 bp) of 81 individuals. Colored rectangles around groups represent their currently recognized mitochondrial haplotypes according to Mendelson et al. 2011 and Firneno & Townsend, 2019: *I. coccifer* (red), *I. ibarrai* (yellow), and *I. porteri* (blue). Major nodes with high support (bs ≥ 85) are represented by a black dot. (B) Comparison of maximum likelihood cladograms using 658 bp COI dataset (left) and our 1207 SNP RADseq dataset (right). Colored rectangles over individuals/groups represent the corresponding mitochondrial haplotypes listed above. Individuals that lack any colored rectangles are those that do not have corresponding mitochondrial data associated with them. Rectangles at the end of the branches on both trees represent the broad geographic distributions of the samples: outside of the Chortís Highlands – Lowlands (light gray), within the Chortís Highlands (medium gray), and outside of the Chortís Highlands – Highlands (dark gray). Major nodes with high support (bs ≥ 85) are represented by a black dot.

16S+COI
(1150 bp)

0.02

COI
(658 bp)

0.009

(A)

(B)
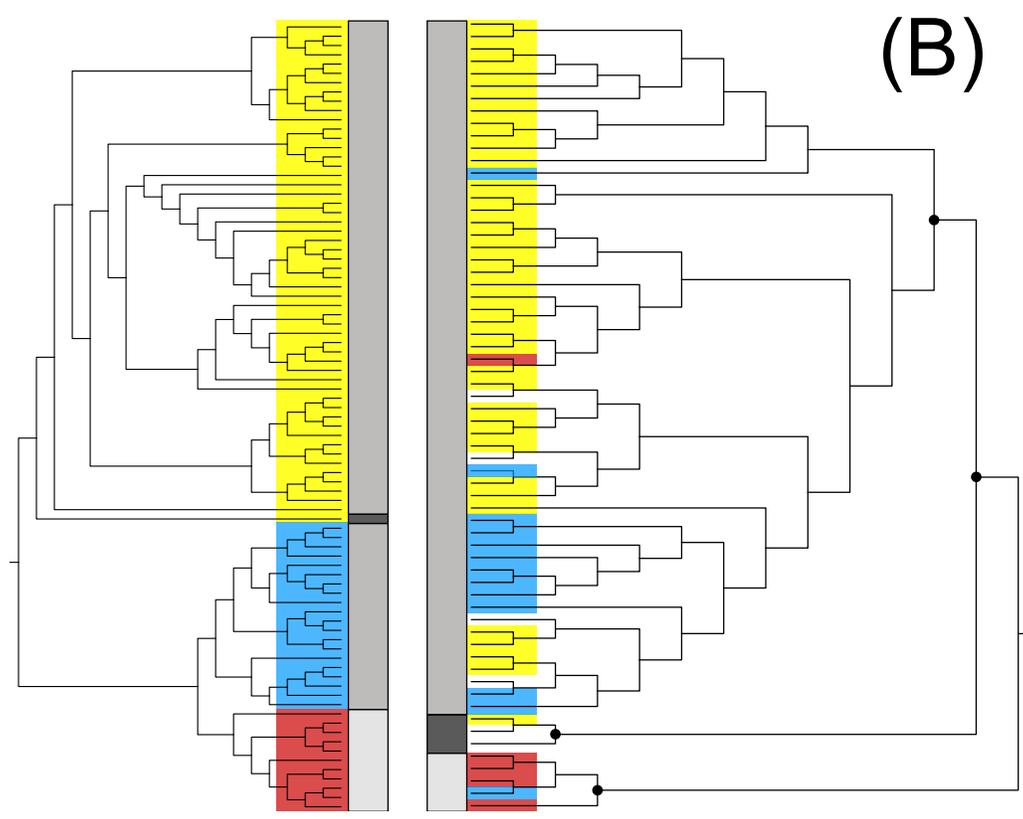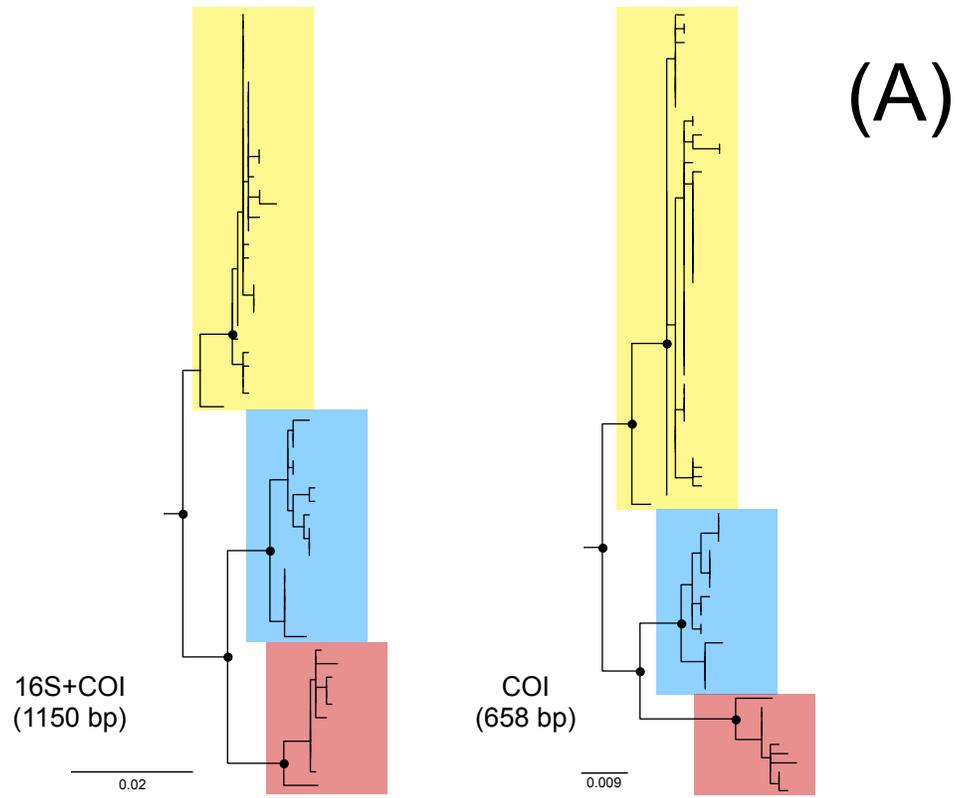
File S2. Samples used for molecular data analyses, with locality data, lineage specificity (both mtDNA and nDNA), voucher numbers, and GenBank Accession numbers; GT, Guatemala; HN, Honduras; MX, Mexico; NI, Nicaragua; PA, Panama.

mtDNA Lineage	Population Assignment	Field Number	Museum Voucher Number	Locality	GenBank: mtDNA	ddRADseq
I. coccifer	OCH Lowland	JS1150	—	NI: Región Autónoma Atlántica Norte	MK720880	X
I. coccifer	OCH Lowland	JS1058	—	NI: Rivas: Ometepe	MK720881	X
I. coccifer	OCH Lowland	JS1016	—	NI: Rivas: Ometepe	MK720883	X
I. coccifer	OCH Lowland	JHT3301	USNM 578693	HN: Valle: Isla el Tigre	KR736042	X
I. porteri	OCH Lowland	JHT3302	USNM 578694	HN: Valle: Isla el Tigre	KR736043	X
I. coccifer	—	JHT3573	CM 163369	HN: Valle: Isla el Tigre	MT348743	—
I. coccifer	—	JHT3546	CM 163367	HN: Valle: Isla el Tigre	MT348741	—
I. coccifer	—	JHT3590	CM 163370	HN: Valle: Isla el Tigre	MT348744	—
I. coccifer	—	JHT3550	CM 163368	HN: Valle: Isla el Tigre	MT348742	—
I. ibarrai	OCH Highland	JAC19612	UTA A-52528	GT: Quiché	JN867971	X
—	OCH Highland	ENS13348	—	GT: Santa Rosa: Carretera Ayarza	—	X
—	OCH Highland	ENS13349	—	GT: Santa Rosa: Carretera Ayarza	—	X
I. porteri	WCH East	CAC044	USNM 578695	HN: Comayagua: Cerro Zarciadero	MK720846	X
I. ibarrai	WCH East	IRL002	—	HN: Comayagua: Cerro Zarciadero	MK720865	X
—	WCH East	IRL003	—	HN: Comayagua: Cerro Zarciadero	—	X
I. ibarrai	WCH East	IRL005	—	HN: Comayagua: Cerro Zarciadero	MK720866	X
—	WCH East	IRL006	—	HN: Comayagua: Cerro Zarciadero	—	X
I. porteri	WCH East	IRL010	—	HN: Comayagua: Cerro Zarciadero	MK720838	X

I. porteri	WCH East	IRL011	—	HN: Comayagua: Cerro Zarciadero	MK720849	X
I. porteri	WCH East	JHT2149	—	HN: Francisco Morazán: Uyuca	MK720847	X
I. ibarrai	WCH East	JHT2205	—	HN: Comayagua: Cerro Zarciadero	MK720863	X
—	WCH East	JHT2206	—	HN: Comayagua: Cerro Zarciadero	—	X
I. porteri	WCH East	JHT2228	—	HN: Francisco Morazán: Finca la Alondra	MK720843	X
I. porteri	WCH East	JHT2246	—	HN: Francisco Morazán: Uyuca	MK720839	X
I. porteri	WCH East	JHT2256	—	HN: Francisco Morazán: Uyuca	MK720844	X
I. porteri	WCH East	JHT2257	—	HN: Francisco Morazán: Uyuca	MK720841	X
I. porteri	WCH East	JHT3947	CM 163380	HN: Francisco Morazán: Uyuca	MT348746	X
I. ibarrai	WCH Central	JHT2604	—	HN: La Paz: Guajiquiro	MK720864	X
I. ibarrai	WCH Central	JHT2605	—	HN: La Paz: Guajiquiro	MK720860	X
I. ibarrai	WCH Central	JHT2608	—	HN: La Paz: Guajiquiro	MK720853	X
I. ibarrai	WCH Central	JHT2609	—	HN: La Paz: Guajiquiro	MK720874	X
I. porteri	WCH Central	JHT2610	—	HN: La Paz: Guajiquiro	MK720837	X
—	WCH Central	JHT2611	—	HN: La Paz: Guajiquiro	—	X
I. ibarrai	WCH Central	JHT2612	—	HN: La Paz: Guajiquiro	MK720871	X
I. ibarrai	WCH Central	JHT2613	—	HN: La Paz: Guajiquiro	MK720862	X
I. ibarrai	WCH Central	JHT2615	—	HN: La Paz: Guajiquiro	MK720855	X
—	WCH Central	JHT2616	—	HN: La Paz: Guajiquiro	—	X
I. ibarrai	WCH Central	JHT2906	—	HN: Intibucá: El Rodeo	MK720852	X
I. ibarrai	WCH Central	JHT3724	CM 163392	HN: Intibucá: El Rodeo	MT348755	X

I. ibarrai	WCH Central	JHT3725	CM 163393	HN: Intibucá: El Rodeo	MT348756	X
I. ibarrai	WCH Central	JHT3726	CM 163394	HN: Intibucá: El Rodeo	MT348757	X
I. ibarrai	WCH Central	JHT3728	CM 163396	HN: Intibucá: El Rodeo	MT348758	X
I. ibarrai	WCH Central	JHT3729	CM 163397	HN: Intibucá: El Rodeo	MT348759	X
I. ibarrai	WCH Central	JHT3730	CM 163398	HN: Intibucá: El Rodeo	MT348760	X
I. ibarrai	WCH Central	JHT3731	CM 163399	HN: Intibucá: El Rodeo	MT348761	X
I. ibarrai	WCH Central	JHT3732	CM 163400	HN: Intibucá: El Rodeo	MT348762	X
I. ibarrai	WCH Central	JHT3733	CM 168168	HN: Intibucá: El Rodeo	MK720875	X
I. ibarrai	WCH Central	JHT3777	CM 168169	HN: Intibucá: El Rodeo	MT348763	X
I. coccifer	WCH Central	JHT3793	CM 168171	HN: Intibucá: San Pedro la Loma	MK720884	X
I. ibarrai	WCH Central	JHT3794	CM 168172	HN: Intibucá: San Pedro la Loma	MK720872	X
I. ibarrai	WCH Central	JHT3795	CM 163306	HN: Intibucá: San Pedro la Loma	MT348764	X
I. ibarrai	WCH Central	JHT3922	CM 168174	HN: Intibucá: El Rodeo	MT348765	X
I. ibarrai	WCH Central	JHT3923	CM 168175	HN: Intibucá: El Rodeo	MT348766	X
I. ibarrai	WCH Central	JHT3924	CM 168176	HN: Intibucá: San Pedro la Loma	MT348767	X
I. porteri	WCH Central	JHT3925	CM 168179	HN: Intibucá: Opalaca	MK720840	X
I. porteri	WCH Central	JHT3932	CM 168186	HN: Intibucá: El Rodeo	MT348745	X
I. ibarrai	WCH West	ENS10270	UTA A-53662	HN: Ocotepeque	JN867970	X
I. ibarrai	WCH West	JHT3692	CM 163384	HN: Lempira: La Ventanas	MT348748	X
I. ibarrai	WCH West	JHT3694	CM 163385	HN: Lempira: La Ventanas	MT348747	X
I. ibarrai	WCH West	JHT3696	CM 163386	HN: Lempira: La Ventanas	MK720873	X
I. ibarrai	WCH West	JHT3697	CM 163387	HN: Lempira: La Ventanas	MT348749	X

I. ibarrai	WCH West	JHT3698	CM 163388	HN: Lempira: La Ventanas	MT348750	X
I. ibarrai	WCH West	JHT3699	CM 163389	HN: Lempira: La Ventanas	MT348751	X
I. ibarrai	WCH West	JHT3700	CM 163390	HN: Lempira: La Ventanas	MT348752	X
I. ibarrai	WCH West	JHT3782	CM 168170	HN: Lempira: La Ventanas	MT348753	X
I. ibarrai	WCH West	JHT3921	CM 168173	HN: Lempira: La Ventanas	MK720869	X
I. ibarrai	WCH West	JHT3945	CM 168177	HN: Lempira: La Ventanas	MK720876	X
I. ibarrai	WCH West	JHT3946	CM 168178	HN: Lempira: La Ventanas	MT348754	X
I. cycladen	—	JRM4607	UTA A-54847	MX: Guerrero: Agua de Obispo	JN867967	—
I. pisinnus	—	JAC26118	UTA A-JAC26118	MX: Michoacan	JN867986	—
I. signifer	—	JRM 4968	UTA A-JRM 4968	PA: Cocle: El Copé	JN867988	—

File S3. Genetically verified (from this and previously published data) and VertNet search results

locality data used for ENM.

Museum Voucher Number	Field Number	Latitude	Longitude
LACM 45248	—	-88.18425	14.30463
LACM 45247	—	-88.18425	14.30463
LACM 47992	—	-88.30748	14.44488
LACM 47993	—	-88.30748	14.44488
LACM 47994	—	-88.30748	14.44488
LACM 47995	—	-88.30748	14.44488
LACM 47996	—	-88.30748	14.44488
LACM 47998	—	-88.32076	14.43145
LACM 47997	—	-88.183	14.3
LACM 48001	—	-88.183	14.3
LACM 48000	—	-88.183	14.3
LACM 48003	—	-88.183	14.3
LACM 48004	—	-88.183	14.3
LACM 48002	—	-88.183	14.3
USNM 523692	—	-88.35177	14.46425
USNM 523691	—	-88.35177	14.46425
USNM 523689	—	-88.35177	14.46425
USNM 523693	—	-88.35177	14.46425
USNM 523690	—	-88.35177	14.46425
USNM 523694	—	-88.30196	14.41568
USNM 523696	—	-88.30196	14.41568
USNM 523695	—	-88.30196	14.41568
KU 194220	—	-88.2167	14.2667
KU 194221	—	-88.2167	14.2667
KU 209250	—	-88.3667	14.35
KU 209251	—	-88.3667	14.35
KU 209240	—	-88.6333	14.5667
LACM 47983	—	-88.99677	14.43333
LACM 47984	—	-88.99677	14.43333
LACM 47985	—	-88.99677	14.43333
LACM 47991	—	-89.08632	14.43333
LACM 47987	—	-89.08632	14.43333
LACM 47988	—	-89.08632	14.43333
LACM 47990	—	-89.08632	14.43333
LACM 47989	—	-89.08632	14.43333
LACM 47986	—	-89.08632	14.43333
USNM 523713	—	-88.8	14.4833
USNM 523720	—	-89.3	14.5
USNM 523721	—	-89.3	14.5
USNM 523722	—	-89.3	14.5
USNM 523723	—	-89.3	14.5

USNM 524697	—	-88.7805	14.91333
USNM 524696	—	-88.7805	14.91333
USNM 514698	—	-88.305531	14.349149
USNM 523999	—	-88.466131	14.233398
USNM 523711	—	-88.466131	14.233398
USNM 514696	—	-88.785696	14.514713
USNM 514697	—	-88.785696	14.514713
USNM 523714	—	-89.122193	14.421867
USNM 523715	—	-89.122193	14.421867
USNM 523716	—	-89.122193	14.421867
USNM 523717	—	-89.122193	14.421867
USNM 523718	—	-89.122193	14.421867
USNM 563911	—	-87.999802	14.6032
USNM 523699	—	-87.912498	14.6297998
USNM 523455	—	-87.912498	14.6297998
KU 97527	—	-87.91806	14.63933
MCZ A-26419	—	-87.838402	14.6501999
KU 209247	—	-87.5	14.5
SDNMH72808	—	-87.248199	13.8696003
KU97524	—	-87.0833	14.0167
LACM 47970	—	-87.09127	14.09343
LACM 47971	—	-87.09127	14.09343
LACM 59428	—	-87.30732	14.17175
LACM 59426	—	-87.30732	14.17175
LACM 59429	—	-87.30732	14.17175
LACM 59430	—	-87.30732	14.17175
LACM 59431	—	-87.30732	14.17175
LACM 59427	—	-87.30732	14.17175
KU 65373	—	-87.44634	14.26667
KU 192294	—	-87.08933	14.18558
KU 194215	—	-87.08933	14.18558
KU 194216	—	-87.08933	14.18558
KU 194218	—	-87.08933	14.18558
KU 194219	—	-87.08933	14.18558
KU 194217	—	-87.08933	14.18558
MCZ A-26441	—	-87.0833	14.0167
KU 97514	—	-87.0833	14.0167
KU 97516	—	-87.0833	14.0167
KU 97517	—	-87.0833	14.0167
KU 97518	—	-87.0833	14.0167
KU 97519	—	-87.0833	14.0167
KU 97521	—	-87.0833	14.0167
KU 97523	—	-87.0833	14.0167
KU 97525	—	-87.0833	14.0167
KU 97526	—	-87.0833	14.0167
KU 97515	—	-87.0833	14.0167

KU 97520	—	-87.0833	14.0167
KU 97522	—	-87.0833	14.0167
KU 103220	—	-87.08148	14.02477
KU 103221	—	-87.08148	14.02477
KU 209249	—	-87.08933	14.18558
KU 194222	—	-88.1167	14.0833
USNM 514695	—	-87.85	14.25
USNM 523709	—	-87.947416	14.664016
USNM 523700	—	-87.947416	14.664016
USNM 523683	—	-87.879798	14.512316
USNM 523684	—	-87.892025	14.505739
USNM 523710	—	-87.452583	14.091147
KU 97524	—	-87.074117	14.029782
SDNHM 72811	—	-86.920867	13.993687
SDNHM 72814	—	-86.920867	13.993687
KU 209257	—	-87.074832	14.031017
KU 209254	—	-87.085206	14.031433
MCZ A-28864	—	-87.09001	14.209852
MCZ A-28865	—	-87.09001	14.209852
MCZ A-28866	—	-87.09001	14.209852
KU 184581	—	-88.051536	14.131103
KU 184582	—	-88.051536	14.131103
KU 184584	—	-88.051536	14.131103
KU 184587	—	-88.051536	14.131103
KU 184591	—	-88.04351	14.127017
KU 184594	—	-88.013455	14.128138
KU 184597	—	-88.013455	14.128138
KU 184598	—	-88.013455	14.128138
KU 184599	—	-88.013455	14.128138
KU 184600	—	-88.013455	14.128138
KU 184602	—	-88.013455	14.128138
KU 184604	—	-88.04351	14.127017
KU 184578	—	-88.04351	14.127017
KU 184579	—	-88.04351	14.127017
KU 184580	—	-88.04351	14.127017
KU 184583	—	-88.04351	14.127017
KU 184585	—	-88.052246	14.138363
KU 184586	—	-88.052246	14.138363
KU 184588	—	-88.052246	14.138363
KU 184589	—	-88.052246	14.138363
KU 184590	—	-88.052246	14.138363
KU 184592	—	-88.052246	14.138363
KU 184593	—	-88.052246	14.138363
KU 184595	—	-88.052246	14.138363
USNM 578695	CAC004	-87.882782	14.767222
—	IRL005	-87.835327	14.786154

—	IRL003	-87.835327	14.786154
—	IRL002	-87.835327	14.786154
—	IRL010	-87.795759	14.755169
—	IRL011	-87.830684	14.772267
—	IRL006	-87.835327	14.786154
—	JHT2149	-87.076059	14.034904
—	JHT2205	-87.9009	14.7283
—	JHT2206	-87.9009	14.7283
—	JHT2228	-87.449045	14.08933
—	JHT2246	-87.076059	14.034904
—	JHT2257	-87.076059	14.034904
—	JHT2256	-87.076059	14.034904
—	JHT2604	-87.82989	14.11992
—	JHT2605	-87.82989	14.11992
—	JHT2608	-87.82989	14.11992
—	JHT2609	-87.82989	14.11992
—	JHT2610	-87.82989	14.11992
—	JHT2612	-87.82989	14.11992
—	JHT2613	-87.82989	14.11992
—	JHT2614	-87.82989	14.11992
—	JHT2615	-87.82989	14.11992
—	JHT2906	-88.145047	14.441386
CM 163385	JHT3694	-88.73347	14.613544
CM 163384	JHT3692	-88.73347	14.613544
CM 163387	JHT3697	-88.73347	14.613544
CM 163388	JHT3698	-88.73347	14.613544
CM 163389	JHT3699	-88.73347	14.613544
CM 163390	JHT3700	-88.73347	14.613544
CM 163392	JHT3724	-88.145047	14.441386
CM 163393	JHT3725	-88.145047	14.441386
CM 163394	JHT3726	-88.145047	14.441386
CM 163396	JHT3728	-88.145047	14.441386
CM 163397	JHT3729	-88.145047	14.441386
CM 163398	JHT3730	-88.145047	14.441386
CM 163399	JHT3731	-88.145047	14.441386
CM 163400	JHT3732	-88.145047	14.441386
CM 168168	JHT3733	-88.15023	14.473329
CM 168169	JHT3777	-88.145047	14.441386
CM 168170	JHT3782	-88.73347	14.613544
CM 168171	JHT3793	-88.099483	14.31196
CM 168172	JHT3794	-88.099501	14.311925
CM 163306	JHT3795	-88.099501	14.311925
CM 168173	JHT3921	-88.15023	14.473329
CM 168174	JHT3922	-88.145047	14.441386
CM 168175	JHT3923	-88.145047	14.441386
CM 168176	JHT3924	-88.099501	14.311925

CM 168179	JHT3925	-88.341786	14.481178
CM 168186	JHT3932	-87.076059	14.034904
CM 168177	JHT3945	-88.733592	14.612588
CM 168178	JHT3946	-88.73347	14.613544
CM 163380	JHT3947	-87.076059	14.034904
—	JHT3949	-87.076059	14.034904
—	TJF107	-86.8855	13.935

File S4. Visual representation of 3D demographic models.

Divergence with no migration
"split_nomig"

Divergence with symmetric migration,
population order variation and migration variation 1
"split_sym_mig_adjacent_var1"

Divergence with symmetric migration,
population order variation and migration variation 2
"split_sym_mig_adjacent_var2"

Adjacent secondary contact, shorter isolation,
population order variation
"refugia_adj_2_var_sym"

Adjacent secondary contact, shortest isolation,
population order variation
"refugia_adj_3_var_sym"

Divergence with asymmetric migration,
population order variation and migration variation 1
"split_uni_mig_adjacent_var1"

Divergence with asymmetric migration,
population order variation and migration variation 2
"split_uni_mig_adjacent_var2"

Adjacent secondary contact, shorter isolation,
population order variation and unidirectional migration
"refugia_adj_2_var_uni"

Adjacent secondary contact, shortest isolation,
population order variation and unidirectional migration
"refugia_adj_3_var_uni"

Simultaneous split, no migration
"sim_split_no_mig"

Simultaneous split, symmetric migration
adjacent populations, population order variation
"sim_split_sym_mig_adjacent_var"

Simultaneous split, unidirectional migration
adjacent populations, population order variation
"sim_split_uni_mig_adjacent_var"

Simultaneous split, symmetric secondary contact
adjacent populations, population order variation
"sim_split_refugia_sym_mig_adjacent_var"

Simultaneous split, unidirectional secondary contact
adjacent populations, population order variation
"sim_split_refugia_uni_mig_adjacent_var"

Admixed origin, no migration
"admix_origin_no_mig"

Admixed origin, symmetric migration
adjacent populations
"admix_origin_sym_mig_adj"

Admixed origin, unidirectional migration
adjacent populations
"admix_origin_uni_mig_adj"

70

File S5. 3D demographic model results.

Model	log-l	AIC	ΔAIC	ωi	theta	nu1	nuA	nu2	nu3	mA	m1	m2	m3	m31	m32	T1a	T1b	T1	T2	f
Adjacent secondary contact with shortest isolation	**-1409.93**	**2839.86**	**0.0**	**1.0**	**147.91**	**0.1733**	**3.1663**	**0.0728**	**0.1895**	**0.7254**	**—**	**8.232**	**10.6627**	**—**	**—**	**0.6063**	**5.759**	**—**	**0.0786**	**—**
Adjacent secondary contact with shorter isolation	-1474.22	2964.44	124.58	<0.01	298.71	0.1444	1.5263	0.1085	0.1656	—	—	6.9306	11.5269	—	—	—	—	0.2113	0.1921	—
Divergence with symmetric migration (variation 2)	-1486.97	2989.94	150.08	<0.01	22.91	1.9415	10.535	1.0707	2.3739	0.2034	—	—	1.2035	—	—	—	—	28.9616	0.3298	—
Divergence with asymmetric migration (variation 1)	-1489.19	2996.38	156.52	<0.01	77.67	1.0181	2.8244	0.4372	0.5062	1.6014	—	—	—	6.4757	2.3047	—	—	5.7741	0.1409	—
Divergence with symmetric migration (variation 1)	-1493.04	3004.08	164.22	<0.01	640.86	0.0804	0.4831	0.048	0.091	0.0526	—	16.605	20.3745	—	—	—	—	2.743	0.5992	—
Adjacent secondary contact with shortest isolation and unidirectional migration	-1499.45	3018.9	179.04	<0.01	36.92	2.6364	3.7255	1.1973	0.8892	1.8491	—	—	—	2.4835	1.601	0.1386	29.3173	—	0.463	—
Admixed origin with symmetric migration between adjacent populations	-1516.51	3049.02	209.16	<0.01	208.16	0.2265	—	0.255	0.3168	—	3.0338	—	8.8053	—	—	—	—	14.3589	3.3403	14.1103
Simultaneous split with no migration	-1523.71	3055.42	215.56	<0.01	223.31	0.1074	—	0.0532	0.6227	—	—	—	—	—	—	—	—	0.0204	—	—
Simultaneous split with unidirectional migration between adjacent populations	-1542.44	3096.88	257.02	<0.01	220.04	0.2413	—	0.0868	0.4557	—	—	—	—	2.307	0.1243	—	—	0.0345	—	—
Simultaneous split with symmetric secondary contact between adjacent populations	-1575.79	3165.58	325.72	<0.01	46.18	0.2804	—	0.2422	0.5965	—	—	2.5098	5.815	—	—	—	—	13.5528	0.5704	—
Divergence with asymmetric migration (variation 2)	-1623.23	3262.46	422.6	<0.01	92.82	0.9536	2.7738	0.5617	0.7866	8.0893	—	—	—	4.6689	—	—	—	2.4534	0.2453	—
Simultaneous split with symmetric migration between adjacent populations	-1658.83	3329.66	489.8	<0.01	185.18	0.2535	—	0.758	0.5697	—	—	0.1541	4.6319	—	—	—	—	0.1506	—	—
Divergence with no migration	-1690.6	3393.2	553.34	<0.01	183.56	0.7463	20.9599	0.1558	0.6416	—	—	—	—	—	—	—	—	0.0295	0.0519	—
Simultaneous split with unidirectional secondary contact between adjacent populations	-1705.81	3425.62	585.76	<0.01	142.86	1.499	—	0.627	0.4033	—	—	—	—	6.3631	3.6504	—	—	0.098	0.2122	—
Adjacent secondary contact with shorter isolation and unidirectional migration	-1729.83	3475.66	635.8	<0.01	147.96	1.3044	1.1561	0.5629	0.4347	—	—	—	—	2.9127	3.6177	—	—	0.0861	0.2081	—
Admixed origin with no migration	-1896.69	3805.38	965.52	<0.01	166.6	0.666	—	2.7431	15.3622	—	—	—	—	—	—	—	—	0.1523	0.0851	1.0247
Admixed origin with unidirectional migration between adjacent populations	-1992.6	4001.2	1161.34	<0.01	84.2	1.5779	—	1.9031	2.0002	—	—	—	—	2.1262	0.0329	—	—	0.4271	0.3499	0.3926

*Abbreviations are as follows: AIC, Akaike Information Criterion; ωI, Akaike weights; theta (4N$_{ref}$μL), the effective mutation rate of the reference population, which here corresponds to the ancestral population; nu1, nuA, nu2, nu3 are various effective population sizes; migration parameters denoted by migration from second label into first label (ex. m31: migration rate from population three to population one); time intervals represented by T and correspond to intervals labeled in Supplemental File 3. Raw parameter values, rather than scaled values, are provided here.

CHAPTER 3:

**DELIMITATION DESPITE DISCORDANCE: EVALUATING THE SPECIES LIMITS
OF A CONFOUNDING SPECIES COMPLEX IN THE FACE OF MITONUCLEAR
DISCORDANCE[§]**

**Thomas J. Firneno, Jr.[1,2,], Justin R. O'Neill[3], Michael W. Itgen[4], Timothy A. Kihneman[1],
Josiah H. Townsend[5,6], & Matthew K. Fujita[1,2]**

[1] *Department of Biology, University of Texas at Arlington, Arlington, Texas, 76019, USA*

[2] *Amphibian and Reptile Diversity Research Center, Department of Biology, University of Texas
at Arlington, Arlington, Texas, 76019, USA*

[3] *Department of Biology, University of Maryland, College Park, MD, 20742, MD, USA*

[4] *Department of Biology, Colorado State University, Fort Collins, CO 80523, USA*

[5] *Department of Biology, Indiana University of Pennsylvania, Indiana, Pennsylvania 15705, USA*

[6] *Centro Zamorano de Biodiversidad, Departamento de Ambiente y Desarrollo, Escuela Agrícola
Panamericana Zamorano, Municipalidad de San Antonio de Oriente, Francisco Morazán,
Honduras*

[§]Used with permission from the publisher

ABSTRACT

The delimitation of species is an essential pursuit of biology, and proper taxonomies are crucial for the assessment and conservation management of organismal diversity. However, delimiting species can be hindered by a number of factors including highly conserved morphologies (e.g. cryptic species), differences in criteria of species concepts, lineages being in the early stages of the speciation or divergence process, and discordance between gene topologies (e.g. mitonuclear discordance). Here we use a taxonomically confounded species complex of toads in Central America that exhibits extensive mitonuclear discordance to test delimitation hypotheses. Our investigation integrates mitochondrial sequences, nuclear SNPs, morphology, and macroecological data to determine which taxonomy best explains the divergence and evolutionary relationships among these toads. We found that a three species taxonomy following the distributions of the nuclear SNP haplotypes offer the best explanation of the species in this complex based off of the integrated data types. Due to the taxonomic instability of this group we also discuss conservation concerns in the face of improper taxonomic delimitation. Our study provides an empirical hypothesis testing framework to assess species delimitation hypotheses in the face of cryptic morphology and mitonuclear discordance and highlights the importance that proper taxonomy has over conservation related actions.

INTRODUCTION

The delimitation of species is an area of great importance across many fields of biology, especially with respect to understanding the depth and breadth of biodiversity, trait evolution, and/or conservation management. While traditional taxonomic practices used accessible phenetic traits (e.g. morphology, developmental traits) to delimit species, the integration of molecular data has exposed cryptic diversity across many organisms (Bickford et al., 2007). Applications of molecular species delimitation have tended to integrate mitochondrial DNA (mtDNA) and/or a limited number of nuclear DNA (nuDNA) markers, though in recent years large genomic datasets that include hundreds to thousands of loci have become much more commonplace (Victor, 2015; Fennessy et al., 2016; Thielsch et al., 2017; Hofmann et al., 2019). This has provided greater taxonomic resolution for many groups of organisms and has given us a true picture of the evolutionary relationships of species (Funk and Omland, 2003; Spinks et al., 2014). While adding more molecular data to taxonomic studies can be useful, it can also sometimes confound the taxonomic and evolutionary relationships of species even further. This is often demonstrated when you have incongruent evolutionary histories, such as between mitochondrial and nuclear loci ("mitonuclear discordance"; Toews and Breslford, 2012). Mitonuclear discordance can arise from a number of factors including introgression, incomplete lineage sorting (ILS), and sex biased dispersal (Funk and Omland, 2003; Toews and Breslford, 2012; Phuong et al., 2016; Ivanov et al., 2018; Firneno et al., 2020). While mtDNA markers have proven useful through molecular barcoding for the detection of cryptic species, there have been a number of cases where mitochondrial-based methodologies conflict with traditional taxonomy and genomic inferences (Toews and Breslford, 2012). The addition of larger genomic datasets, along with a growing prevalence of discordance between genetic datasets has caused for the need

to incorporate more lines of evidence into species delimitation in an integrative taxonomic framework, as well as the reevaluation of cryptic species and/or confounding species complexes (Padial et al. 2010).

Here, we address a controversial issue of species delimitation in a confounding complex of Central American true toads (Anura: Bufonidae: *Incilius coccifer* complex). Since the only taxonomic revision of the species in this complex, which were based on morphology and mitochondrial DNA (mtDNA) by Mendelson et al. (2005), the nominal species have been the subject of continuous debate and reevaluation as a complex (McCranie and Castañeda, 2007; McCranie, 2015; Mendelson et al., 2011; Firneno and Townsend, 2019). The *I. coccifer* complex is currently comprised of three distinct, yet closely-related mitochondrial lineages that exhibit a highly conserved morphology: *I. coccifer* in the Pacific lowlands from southern Mexico to northern Costa Rica, *I. ibarrai* in the highlands from western Guatemala to western Honduras, and *I. porteri* in the highlands of central Honduras (Mendelson et al., 2005; Mendelson et al., 2011; Firneno and Townsend, 2019). The recent addition of genomic SNP data found extensive mitonuclear discordance between the two genetic datasets, which was determined to be the cause of incomplete lineage sorting (ILS) of mitochondrial haplotypes (Firneno et al., 2020). The conflicting genetic evidence has raised a complicated example of species delimitation where a stable and accurate taxonomy is of particular concern, as these organisms have been constantly reassessed as being data deficient, endangered, or least concern by the IUCN Red List of Threatened Species and exist in some of the most disturbed and threatened habitats in Central America (Mendelson et al., 2004; Townsend and Wilson, 2010a; Whitfield et al., 2016; Townsend and Wilson, 2016).

The mitonuclear discordance found in the *I. coccifer* complex involves differences in tree topology and the distribution of populations/species (Mendelson et al., 2005; Mendelson et al., 2011; Firneno and Townsend, 2019; Firneno et al., 2020); therefore, making it unclear exactly how many species there are in the complex and what their geographic range distribution is. Here we integrate and reevaluate previously collected molecular data (mtDNA – Mendelson et al. 2005, Firneno & Townsend, 2019; genome wide SNPs – Firneno et al. 2020), previously and newly collected morphology (Mendelson et al. 2005), and macroecological modeling in a robust statistical and comparative integrative taxonomic framework. We use three hypotheses to test and reevaluate the species limits of the *I. coccifer* complex in the face of mitonuclear discordance: (1) there are three species in the complex that follow the distributions proposed in Mendelson et al. (2005) according to mtDNA and morphological data (currently recognized taxonomy); (2) there are three species in the complex that follow the distributions proposed in Firneno et al. (2020) according to SNP data and biogeographic barriers of the region; or (3) there is some other number of species in the complex that will be revealed by integrating molecular, morphological, and macroecological data that have different range distributions than the currently proposed ones. We test these hypotheses and define species in this paper using the general lineage concept, due to its pluralistic approach and the ability to use numerous criteria to define a species under its definition (Mayden 1997; de Quieroz 1998, 1999; Sangster 2014). Our study highlights the importance of using multiple genetic markers and multiple lines of evidence in an integrative hypothesis-testing framework to resolve complex species delimitation scenarios where different lines of evidence may conflict, as well as the importance of a stable taxonomy to guide proper conservation assessment measures.

METHODS AND MATERIALS

Molecular data generation and processing

We reprocessed and reanalyzed the ddRADseq data of Firneno et al. (2020) (available on GenBank as PRJNA626342: SAMN14615759–SAMN14615822). The workflow for data processing, filtering, and formatting was automated using an updated pipeline available from https://github.com/dportik/Stacks_pipeline (Portik et al., 2017). In brief, the raw Illumina reads were demultiplexed using STACKS v2.53 (Catchen et al., 2013), the restriction site overhangs were removed using the fastx_trimmer module of the FASTX-TOOLKIT (www.hannonlab.cshl.edu/fastx_toolkit), and the sequencing quality was examined on a per sample basis using FASTQC v0.10.1 (www.bioinformatics.babraham.ac.uk/projects/fastqc). Loci were created, catalogued, and identified using USTACKS, CSTACKS, GSTACKS, and SSTACKS, respectively. POPULATIONS was then used to generate alleles for loci present in 80% of all individuals, which resulted in 3,225 loci. Custom filtering removed "blank" loci (n=87), invariant loci (n=1,120), non-biallelic loci (n=0), and loci containing at least one individual with more than two alleles (n=0). For loci containing multiple SNP sites, we randomly chose a single SNP to be used for subsequent analyses. Any samples missing data for more than 50% of loci were removed. After completing the above filtering steps, our final SNP dataset consisted of 64 samples and 2,018 loci.

We also created a complimentary mitochondrial DNA dataset of the cytochrome oxidase I (COI) gene using available sequences on GenBank (Firneno and Townsend, 2019; Firneno et al., 2020). Individuals in this dataset either had complementary RADseq data associated with them or were from roughly the same sampling sites (Table S1; Fig. 1a, b). We generated a multiple alignment of 610 bp for 59 samples (including two outgroups; Table S1) in MEGA7 (Kumar et al., 2016) with the MUSCLE algorithm (Edgar, 2004) using the default parameters.

Population structure

We determined the number of discrete populations present across the sampled range of the *I. coccifer* complex with our RADseq dataset using a combination of Bayesian and likelihood clustering analyses, and multivariate methods. We used STRUCTURE v2.3.4 (Pritchard et al., 2000; Falush et al., 2003) to examine the number of population clusters and potential admixture between populations in our dataset using MCMC. Hierarchical analyses were performed for 10 runs per K, up to a maximum of 6 populations, and used the admixture model with a burn-in of 10,000 steps followed by 100,000 steps. We summarized our results using STRUCTUREHARVESTER (Earl and vonHoldt, 2012) and evaluated the number of populations based on inspection of likelihood plots and following Evanno et al. (2005). To complement our STRUCTURE analysis, we used a maximum likelihood approach with ADMIXTURE (Jombart, 2008). We performed ten replicate analyses to evaluate up to seven populations. To assess the best K value, we performed 10-fold cross-validation and determined the K value with the lowest cross-validation error. We also evaluated the number of discrete populations using a discriminant analysis of principal components (DAPC) with ADEGENET v2.0.0 (Jombart, 2008; Jombart and Ahmed, 2011). A maximum of 10 clusters were investigated using the k-means algorithm. The preferred number of clusters was evaluated using BIC scores. We explored a range of three to five clusters to describe using DAPC. To minimize overfitting, an initial DAPC was used to find the a-score for each set of clusters and this value was used to select the number of principal components to retain in a subsequent reanalysis (Jombart, 2008; Jombart and Ahmed, 2011). Group membership probabilities were then examined for each cluster. To independently assess the validity of population differentiation and assignment, we used the FINERADSTRUCTURE software package (Malinsky et al., 2018) to construct a co-ancestry matrix from our RADseq

data. We used a 100,000 burn-in followed by 100,000 MCMC steps sampling every 1,000 steps and the tree was constructed with 10,000 hill-climbing iterations. The results were visualized using the FINERADSTRUCTUREPLOT.R and FINESTRUCTURELIBRARY.R scripts (included in the FINERADSTRUCTURE package file).

Phylogenetic analyses

We estimated the phylogenetic relationships independently for our mtDNA and SNP data sets. We used a maximum likelihood approach carried out in RAxML v8.0 (Stamatakis, 2014) on both genetic datasets with 1,000 bootstrap replicates under a GTR substitution model. We used a Bayesian approach for our mtDNA dataset by employing BEAST v2.5.1 (Bouckaert et al., 2014) using an HKY model of nucleotide substitution, relaxed clock model, coalescent constant population, and a random starting tree with a Markov Chain Monte Carlo (MCMC) run for 2×10^7 generations, sampling every 1000 generations producing a total of 10,000 trees. We assessed the run using Tracer v1.6 (Rambaut and Drummond, 2009) to examine convergence. A burn-in of 10% was discarded, and a maximum clade credibility (MCC) tree with median heights was created from the remaining 9000 trees. We then estimated the species tree for our SNP dataset using SNAPP v1.3.0 (Bryant et al., 2012) implemented in BEAST2. To reduce run times, we subsampled each population to include 3–6 representatives, for a total of 20 individuals. We estimated the mutation rates (*u* and *v*) from the data (1.06 and 0.94, respectively) within BEAUTI. We assigned a Gamma distribution to our birth rate (λ) of the Yule prior, with and Alpha of 2.0 and a Beta of 2.0. Our SNAPP prior was assigned an Alpha of 11.75, a Beta of 109.73, and a Lambda of 0.01. We performed two independent runs with a chain length of 1,000,000 generations, sampling every 1000 generations. Runs were assessed using TRACER v1.6 (Rambaut

and Drummond, 2009) to examine convergence, and tree topologies and node heights were visualized using DENSITREE (Bouckaert, 2010).

Species delimitation

To estimate the best-fit number of species units based on our two genetic datasets, we used a suite of different species delimitation analyses. We applied three single locus delimitation analyses to our mtDNA dataset including: (1) the General Mixed Yule Coalescent (GMYC) model, which establishes thresholds between the branching patterns of ultrametric gene trees from inter- and intraspecific branches in order to define speciation events (Fujisawa and Barraclough, 2013); (2) the Poisson Tree Process (PTP; Kapli et al., 2016), which, like GMYC, aims to identify the transition between inter- and intraspecific processes, but specifically requires that the phylogenetic gene trees branch lengths are proportional to the number of substitutions, rather than to time as for GMYC; (3) and the Automated Barcode Gap Discovery (ABGD) method (Puillandre et al., 2012), which partitions sequences into groups based on comparisons of pairwise distances and compared to tree-based delimitation methods (GMYC and PTP) that are suspected to over split, ABGD offers a more conservative approach to estimate the number of species given comprehensive sampling (Puillandre et al., 2012). For the GMYC analysis, we applied a single model using the R package *splits* to our ultrametric tree generated in BEAST. Next, we implemented the bPTP analysis using our maximum likelihood tree on the bPTP server (https://species.h-its.org; Zhang et al., 2013). For this analysis, we ran 100,000 MCMC generations, with a thinning of 100 and burnin of 0.1. Finally, we applied the AGBD analysis to our data using the web interface (https://bioinfo.mnhn.fr/abi/public/abgd). For our analysis of each locus dataset, we used the default maximum intraspecific distance values ($Pmax = 0.1$), a

relative barcode gap width of X = 1.5, and a distance correction of Jukes-Cantor (JC69). Default settings were used for all of the remaining parameters.

We then applied four species delimitation methods to our SNP dataset including: (1) Bayes Factor Delimitation (BFD*; Leaché et al., 2014) and (2) Bayesian Phylogenetics and Phylogeography (BPP) v4.2 (Flouri et al., 2018), both of which implement the multispecies coalescent (MSC) model; (3) heuristic species delimitation using the genealogical divergence index (*gdi*) to determine whether putative species boundaries correspond to species-level divergences between populations; and (4) DELINEATE, which implements the protracted speciation model (PSM) to account for the assertion that the MSC delimits populations rather than species (Sukumaran and Knowles, in review). We conducted Bayes Factor Delimitation implemented in SNAPP v1.3.0 (Bryant et al., 2012) in BEAST2 v2.5.4 (Bouckeart et al., 2014) with our SNP dataset following Leaché et al. (2014) (e.g. implementing BFD*). We performed our analyses testing four models, which are presented in Table S2. We compared and ranked models to select the best supported species hypothesis. We calculated the Bayes Factor (BF) by subtracting the value of the log MLE for the model representing the current taxonomic classification from each alternative model and multiplying the difference by two (BF = 2(model 1 – model 2)). We ranked all of the models and selected the model with the highest BF (Table 1). We used the same parameters as our SNAPP analysis (see above). We performed 48 path sampling steps, with 100,000 MCMC generation and a pre-burnin of 1000 generations. The BF was then calculated for each alternative model, the models were ranked, and a best model was chosen (Table S2).

Next, we used Bayes species delimitation as implemented in Bayesian Phylogenetics and Phylogeography (BPP) v4.2 (Flouri et al., 2018; Flouri et al., 2020). We implemented the A10

(species delimitation using a fixed guide tree) analysis under the multispecies coalescent accommodating introgression (MSCi) model using relationships derived from phylogenetic analyses as a guide tree. The mapping of individuals to their identifiers was done in the *imap* file supplied to BPP. Diffuse inverse-gamma priors were designated as follows: θ IG(3, 0.02) with mean 0.02/(3 – 1) = 0.01, and τ IG(3, 0.01) with mean 0.01/(3 – 1) = 0.005. Each analysis was run three times for 100,000 generations with the first 10,000 discarded as burnin. Convergence was assessed by ensuring the stability of parameter estimates.

Following this we used the *gdi* to determine whether putative species boundaries correspond to species-level divergences between populations of the *I. coccifer* complex [25-26]. We calculated *gdi* values in BPP v4.2 (Flouri et al., 2018; Flouri et al., 2020) under the MSCi model. The MSCi approach provides a measure of uncertainty in population estimates, and *gdi* can be calculated using the posterior distributions for θ and τ (Leaché et al., 2019). The posterior probability distributions for θ and τ were estimated in the A00 analysis (Jackson et al., 2017) using a fixed species tree containing our three lineages (*I. coccifer*, *I. ibarrai*, and *I. porteri*, based on SNP haplogroups). We used the same diffuse inverse-gamma priors for θ and τ that we used in our BPP species delimitation analysis (see above). To assess convergence, we compared the posterior distributions from three independent runs (same parameters as BPP above). We then calculated *gdi* for each species by combining all samples from the posterior distributions using the equation: $gdi = 1 - e^{-2\tau/\theta}$ (Leaché et al., 2018; Chan and Grismer, 2019). Lineages were considered distinct species when *gdi* >0.7, the same species when *gdi* <0.2, and/or ambiguous species status if $0.2 \geq gdi \leq 0.7$ (Pinho et al., 2010; Jackson et al., 2017).

Finally, we used DELINEATE (Sukumaran and Knowles, in review) to determine the number of species are within the *I. coccifer* complex under the protracted speciation model

(PSM). We used the species tree from our SNAPP analysis as a guide tree and within our control file we constrained the *I. coccifer* lineage (based on previous descriptions of morphological and ecological distinctiveness) but left the *I. ibarrai* and *I. porteri* (highland) lineages as unconstrained to determine if they are separate species from the *I. coccifer* or each other.

Morphological data acquisition and processing

To determine which genetic species assignment (mtDNA or SNP) and species delimitation model yields more morphological differentiation between species, with special attention paid to discordant highland populations, we collected new measurements from 37 specimens of highland population individuals and combined these data with the measurements of Mendelson et al. (2005) for a total dataset of 386 individuals. The morphological traits measured were snout–vent length (SVL), tibial length (TL), hindfoot length (FL), head length (HL), head width (HW), diameter of tympanum (DT), supratympanic crest length (SC), paratoid length (PL), and paratoid width (PW). All measurements were taken with digital calipers and recorded to the nearest 0.1 mm. A principal components analysis was done using the scaled morphological data. We visualized the clustering of two, three, and four species, as well as the discordant assignments of *I. ibarrai* and *I. porteri* using the mtDNA and SNP assignments based on our molecular delimitation results. PCAs were done in program R using the *princomp* function (R Development Core Team, 2019) and were visualized using the R package *ggplot2* (Wickham, 2016).

We used linear discriminant function analyses (DFA) to determine the extent of morphological distinction among the species. Two analyses were conducted using the classifications of the nuclear and mitochondrial phylogenies due to the mitonuclear discordance seen in highland populations (*I. ibarrai* and *I. porteri*). We log-transformed the dataset and

83

removed highly correlated traits (≥ 0.95). Our final morphological dataset for the DFAs included

FL, HL, HW, DT, SC, PL, and PW. The DFAs were done in R using the *lda* function in package

MASS (Venables and Ripley, 2002).

We further investigated the morphological variation among discordant highland

populations using redundancy analyses (RDA) and variation partitioning (VP). All of the non-

transformed morphological traits were included in these analyses. Taxonomic assignments based

on the mtDNA and nuDNA phylogenies were scored for each individual as dummy variables and

were treated as the explanatory variables. The variation partitioning analysis was used to

calculate the explanatory contributions for each phylogeny. A permutation test with 1,000

permutations was then used to determine the significance of the conditional effects for the

mtDNA and nuDNA species assignments. The RDA and VP analyses were done in program R

v.3.2.2 using the *rda* and *varpart* functions in the package *vegan* (R Development Core Team,

2019; Oksanen et al., 2017).

Species Distribution Modeling

To determine the extent of macroecological differentiation between genetically

discordant populations, we created species distribution models (SDMs) for all three species

based on their distributions according to their respective genetic datasets. We compiled all

available museum locality information from VERTNET (Constable et al., 2010) and our own

sampling for toads of the *I. coccifer* complex. We created two datasets corresponding to our

genetic hypotheses (mtDNA or SNP) and assigned species into their respective genetic datasets

either based on genetic confirmation or being within the distribution of a genetic haplotype. This

resulted in 410 records that were used to create SDMs (Table S3). To reduce spatial

autocorrelation of our occurrence dataset, we thinned points by a distance of 1 km (Veloz, 2009;

Aiello-Lammens et al., 2015). This reduced our datasets to the following for each species:

mtDNA – *I. coccifer* = 43, *I. ibarrai* = 41, *I. porteri* = 30; SNP – *I. coccifer* = 48, *I. ibarrai* = 18,

I. porteri = 50.

To construct our environmental layer dataset, we extracted nineteen bioclimatic variables

from the WorldClim 2.1 database (http://worldclim.org; Fick and Hijmans, 2017) and eighteen

environmental variables from the ENVIREM database (https://envirem.github.io/; Title and

Bemmels, 2018) for present-day conditions (~1970-2020). We also included layers of spatial

homogeneity of global habitat (http://earthenv.org/texture.html) and global percent of tree cover

(https://github.com/globalmaps/gm_ve_v1), and we computed aspect and slope layers within

ARCGIS from the Global 30 Arc-second digital elevation layer (GTOPO30;

https://www.usgs.gov/centers/eros/science/usgs-eros-archive-digital-elevation-global-30-arc-

second-elevation-gtopo30) . Spatial resolution for all environmental layers was 30 arc-seconds

(~1km). To eliminate predictor collinearity among our bioclimatic and ENVIREM variables, we

calculated Pearson's correlation coefficients for all pairs of variables in their respective datasets

using ENMtools (Warren et al., 2019), excluding the variable from a correlated pair ($|r|>0.75$)

that we considered to be less biologically important based on known preferences of *Incilius* toads

(Fig. S3). The resulting dataset contained seven bioclimatic variables (BIO1 = annual mean

temperature; BIO2 = mean diurnal range; BIO3 = isothermality; BIO12 = annual precipitation;

BIO15 = precipitation seasonality; BIO18 = precipitation of warmest quarter; and BIO19 =

precipitation of coldest quarter) and six ENVIREM variables (annual potential evapotranspiration

(PET), thermicity index, climatic moisture index, aridity index, terrain roughness index, and

topographic wetness), bringing our final dataset to a total of seventeen variables. We constrained

our environmental variable layers to the range spanned by all of our occurrence records.

To build SDMs, we used MAXENT (Phillips and Dudík, 2008) implemented in the ecological niche modeling application WALLACE (Kass et al., 2018). Because model settings can hold strong influence on model output, determining optimal model complexity is important (Warren and Seifert, 2011; Radosavljevic and Anderson, 2014). We ran 76 candidate models using a combination of: (a) one of four selected MAXENT feature classes (Linear; Linear and Quadratic; Hinge; Linear, Quadratic, and Hinge); and (b) a range of 19 regularization multipliers (1–10 in 0.5 increments), allowing our candidate models to range from simple to complex. To test the accuracy of the models, we used WALLACE's spatial partition option to split the occurrence data into four spatially independent k-folds for cross validation, whereas, for datasets with under 20 records, we used the jackknife approach because of its accuracy and utility with small datasets (Shcheglovitova and Anderson, 2013; Muscarella et al., 2014; Kass et al., 2018). To choose the top model for each species, we approached it in a hierarchical fashion by: (1) looked for the candidate within four DAICc that had the lowest testing omission rate at the 10th percentile training presence threshold; (2) if there was a tie at the 10th percentile training presence, we used the difference between the training and testing AUC, picking the model with lowest value; (3) if there was a tie at this value, we used the testing AUC score; and (4) if there was a tie at this value, we used the lowest DAICc (Burnham and Anderson, 2002; Muscarella et al., 2014; Kass et al., 2018). Models were reclassified into binary files of suitable and non-suitable habitat based on the 10th percentile training presence threshold, essentially removing the lowest 10% of the prediction values (Kass et al., 2018).

We inferred the amount of niche overlap between species for the mtDNA distribution and the SNP distribution SDMs using by calculating Schoener's *D* and *I* by pairwise comparison using the *nicheOverlap* functions in the R package *dismo* (Hijmans et al., 2011). This metrics

give and output value from 0 to 1, where a value of 0 indicates no overlap between niches/low

niche similarity and a value of 1 indicates that the niches completely overlap/are identical.

RESULTS

Genetic and species delimitation analyses

In agreement with previous studies (Firneno et al., 2020), we see extensive mitonuclear

discordance among our two genetic datasets of individuals within the *Incilius coccifer* complex.

We estimated four populations from our SNP dataset, with a single lowland population, a

highland population that exist west of the Chortís/Guatemala Highlands boundary, and two

populations with admixture east of the Chortís/Guatemala Highlands boundary (Fig. 1d,S1). Our

FINERADSTRUCTURE analysis identified three main populations (as defined above) with some

substructuring among individuals in the population east of the Chortís/Guatemala Highlands

boundary, which may be due to nuclear introgression or isolation-by-distance due to the mosaic

nature of the organisms' highland distributions (Fig. 1c).

The single locus delimitation methods (GMYC, PTP, and ABGD) of our mtDNA dataset

all yielded varied results in terms of the number of species units they estimated. GMYC

estimated five species units, bPTP estimated four species units, and ABGD estimated three

species units (Fig. 1f). Our MSC based methods, BFD* and BPP, both estimated four species

units. Our heuristic method (*gdi*) had ambiguous support ($0.2 \geq gdi \leq 0.7$) for *I. coccifer* and *I.*

ibarrai as distinct species and low support ($0.2 \geq gdi$) for *I. porteri* as a distinct species (Fig. S2),

indicating that there may be fewer than three species in the complex. Finally, our PSM based

method (DELINEATE) estimated one species unit (Fig. 1f, Table S2).

Though ML and Bayesian analyses of both datasets yielded different tree topologies, both

showed high support (bs $\geq$ 90/pp $\geq$ 0.9) for three distinct lineages (Fig. 1e, f). Like previous

studies we see a distinct lowland clade (*I. coccifer*) that is the sister group to the two highland lineages according to our SNP dataset, but is nested within the highland clades according to our mtDNA dataset. We also see two highland lineages according to both datasets (*I. ibarrai* and *I. porteri*); however, the distributions of the respective haplotypes are discordant. Our mtDNA dataset identifies one highland clade solely within central Honduras (*I. porteri*) and another clade that spans from central Honduras to western Guatemala (*I. ibarrai*); whereas our SNP dataset identifies one highland clade east of the Chortís/Guatemalan Highland boundary (*I. porteri*) and the other west of the Chortís/Guatemalan Highland boundary (*I. ibarrai*) (Fig. 1a-b,e). The SNP highland clade that is distributed west of the Chortís/Guatemalan Highland boundary corresponds to a single, divergent sample of *I. ibarrai* that was collected from Guatemala and has previously been identified as possibly being a separate entity from those collected in Honduras (Mendelson et al., 2011; Firneno and Townsend, 2019).

Morphological analyses

The morphological variation across the species complex best supports the two species hypothesis, with an overall correct classification rate of 92.4% for both mtDNA and SNP assignments. The three species and four species hypotheses did have high classification rates but were 10.1–12% lower than the DFAs using a two species classification (Table 1). We found that *I. coccifer* is a morphologically distinct taxon, which had a correct classification rate of 94.9% across all six of the DFAs and showed a high degree of separation in the PCAs (Table 1; Fig. 2a). The decreased rate of correct classification in the three and four species DFAs was associated with delineating the Highland taxon into multiple species (Table 1). These populations, which are currently recognized as *I. porteri* and *I. ibarrai*, show a high degree of morphological similarity. Specifically, it is the Honduran populations of that are very similar.

The variation partitioning analysis found that the nuDNA and mtDNA assignments both significantly explained the dataset (Fig. 2b). The nuDNA and mtDNA assignments explained 36% and 32% of the morphological variation, respectively. The variation partitioning analysis revealed that 27% of the explained variation was shared by both assignments (Fig. 2b). The nuDNA and mtDNA assignments significantly explained 9% ($P < 0.001$) and 5% ($P < 0.001$) of the variation as conditional effects, respectively (Fig. 2b).

Ecological niche modeling and niche overlap

The top candidate SDMs that constituted our final models had high AUC scores ranging from 0.81–0.95. Based on the known geographic distributions of the species under either phylogenetic hypothesis, very little geographic overestimation occurred in the models. The only model that we see some overestimation into more lowland regions is the *I. ibarrai* SNP model (Fig. S4), which we attribute to the low number of samples (< 20) that have been used for that model. It should be noted that there is evident estimation of the *I. ibarrai* and *I. porteri* models of both datasets (Fig. S4) outside of the biogeographic barriers that may delineate the true distributions of these species, which is due to the fact that MAXENT does not account for physical biogeographic barriers. Binary maps indicating the modeled occurrence (Fig. 3) revealed very little distributional overlap between *I. coccifer* and *I. ibarrai* or *I. porteri*. We see more overlap between the *I. coccifer* and *I. ibarrai* within the SNP distribution models, which is primarily due to the overestimation of the *I. ibarrai* SNP model into the lowland areas. Our models did reveal broad zones of overlap between *I. ibarrai* and *I. porteri* under both scenarios, though we see less overlap in the SNP models (Fig. 3). The niche overlap and identity tests reflected these results (Fig. 3).

DISCUSSION

Delimiting species can be hindered by a number of factors including highly conserved morphologies (e.g. cryptic species), differences in criteria of species concepts, and discordance between gene tree topologies (e.g. mitonuclear discordance). Here we provided an example of a taxonomically confounding species complex of toads from Central America that exhibits extensive mitonuclear discordance and presented an integrative hypothesis testing framework to assess differentiation of integrated data using the different phylogenetic hypotheses caused by mitonuclear discordance. Below we summarize our findings and evidence for the taxonomic relationships of the species within the *I. coccifer* complex, along with a discussion of the conservation implications for this species complex.

Species delimitation

While strides have been made in species delimitation using genetic data (Fujita et al., 2012; Leaché et al., 2018; Luo et al., 2018), taxonomists are still required to make subjective judgements based on multiple separate lines of evidence to decide on the boundary between populations and species (Sites and Marshall, 2004). Following single lines of evidence, such as a single genetic locus, can lead to discordant results that do not reflect an accurate taxonomy (Jackson et al., 2017). We chose to implement single locus species delimitation methods on our mtDNA dataset, and they suggest that there are between 3–5 species within the complex. Our tree based single locus methods (GMYC and PTP) delineate a species boundary between *I. ibarrai* populations in Guatemala and those in Honduras, which have been previously suggested as two potentially distinct species (Mendelson et al., 2011; Firneno and Townsend, 2019). This is also interesting because the populations of *I. ibarrai* in Honduras correspond to different populations of *I. porteri* according to the SNP dataset. Conversely, ABGD delimits three species

units overall, lumping all *I. ibarrai* populations together, and follows the currently applied taxonomy of this complex.

Single locus species delimitation methods rely on the topologies inferred from a single locus and assume that it represents the current species relationship. However, incorrect topologies due to introgression and/or incomplete lineage sorting (ILS) can mislead these methods (Knowles and Carstens, 2007; Dupuis et al., 2012). Because ILS was found to be the cause of mitonuclear discordance in the *I. coccifer* complex and that nuclear introgression was present within populations of *I. porteri* within the Chortís Highlands (Firneno et al., 2020), we also used species delimitation methods that incorporate multi-locus data in both a multispecies coalescent and protracted speciation model framework. These methods have been shown to take into account ILS and are robust to low levels of introgression (Zhang et al., 2011). Our multi-locus delimitation methods widely varied estimating 1–4 species units. The methods that apply the multispecies coalescent (BFD* and BPP) tended to delimit species along the lines of our population inferences; whereas our other methods (*gdi* and DELINEATE) were more conservative in their estimates, indicating only one or two species units present. An often-cited issue with species delimitation methods that use the MSC framework (BFD* and BPP) is that they tend to delimit species boundaries based on population structure, therefore overestimating the number of species within a focal group (Sukumaran and Knowles, 2017). We see this in our data and is the reason why we used comparative methods that implement other frameworks other than the MSC. It has been remarked that the use of *gdi* can have a number of weaknesses including that the criterion depends on the population divergence time relative to the population size (small populations and recent divergence times may skew results), that ambiguity may arise when two populations being compared are of very different sizes, and that the metric has a wide range of

indecision and reflects a more subjective nature of species delimitation that these methods are trying to avoid (Leaché et al., 2019). Likewise, it has been argued that the PSM implemented by DELINEATE is unrealistic because it models instantaneous speciation in a single generation, which does not seem to reflect how natural species convert from being an incipient species to a true species (Sukumaran and Knowles, 2017; Leaché et al., 2019). These may be reasons why our heuristic and PSM based analyses yielded such conservative results.

The addition of more genetic markers can often aid in the clarity of delimiting species. However, there are cases, as exemplified here, where adding more genetic data can reveal more complexity within and between species than originally assumed or hinder the ability of taxonomists to make objective decisions concerning the delimitation of species (Papakostas et al., 2016; Ivanov et al., 2018; Hinojosa et al., 2019; Pedraza-Marrón et al., 2019). Due to the discordance between our genetic datasets and the wide array of species units estimated by our species delimitation analyses it is difficult to discern the structure and number of species in this complex based on genetic data alone. However, we were able to use these competing phylogenetic and delimitation hypotheses to further integrate and evaluate other lines of evidence (e.g. morphology and macroecology) to determine which hypothesis best describes the variation seen in these other data types and most accurately reflects the proper taxonomic relationships and distributions of the species in this complex.

Taxonomic Implications

Under the general lineage concept (GLC) the only necessary property of a species is its existence as an independently evolving lineage (de Queiroz 2005, 2007). Following our molecular phylogenetic and species delimitation methodologies under the GLC, we cannot clearly and objectively make an inference of the number of species based on our genetic data.

Also, due to mitonuclear discordance it is still not clear which dataset (mtDNA or SNP) gives us a more concrete indication of where the species range limits lie in this complex. Therefore, we integrated morphological and macroecological analyses to further test hypotheses of species number and distribution. These combinations of data in an integrative taxonomic framework have long been shown to clarify problematic species complexes such as the *I. coccifer* complex (Padial et al., 2010). It should be noted that, though macroecological modeling is not often used for species delimitation, when combined with genetic and/or morphological information, species delimitation hypotheses can improve in objectivity and strength (Rissler and Apodaca, 2007; Leaché et al., 2009; Hidalgo-Galiana et al., 2014; Razkin et al., 2016).

Though Firneno et al. (2020) suggest that there may be more than three species in this complex, a hypothesis that was based only on genetic data, we do not find evidence within our genetic, morphological, and macroecological data for this claim. Within our morphological data we see little separation between highland populations within Honduras (Fig. 2a) and a low amount of support for discerning these populations (Table 1) when they are treated as separate species; however, we do see good support for a separate highland population in Guatemala. Instead, we see high support for only two species (a lowland entity and a highland entity) and moderate support for three species (a lowland entity and two highland entities) with relatively high overlap between the two highland entities. Our macroecological analyses clearly show separation of suitable habitat between at least two species (Fig. 3), but with higher niche overlap when splitting highland entities into two species/populations. It is also worth noting that, while we could not test a four species hypothesis within our macroecological analyses (the locality datasets upon spatial rarefication become too small to generate reliable models), we would expect high levels of niche overlap as we see in our current three species models (Fig. 3).

Synthesizing the contrasting datasets that we have collected and analyzed, it is still difficult to concretely say how many species there are in this complex. Though we can safely rule out hypotheses concerning four or more species in this complex, it seems as if there is sufficient evidence for two or three species. We suggest that there are three species in this complex under the GLC based on the following criteria: (1) there are at least three well-supported monophyletic lineages within our phylogenies, though the mtDNA relationship is rendered paraphyletic when compared to the SNP phylogeny due to mitonuclear discordance); (2) there are four distinct populations that reflect the lineages from the SNP dataset, though there is gene flow occurring between two of the population lineages indicating a possible lack of reproductive isolation; (3) there is support for three morphologically distinct entities that reflect our phylogenetic lineages with a high degree of morphological conservation between them; and (4) there is differentiation between three macroecological entities that reflect out phylogenetic lineages, though there is a moderate to high degree of niche overlap between highland lineages.

Along with the species limits, we were also concerned with determining which distribution (mtDNA or SNP) reflected the taxonomy as well. We found support for a more parsimonious distributional hypothesis based on the SNP dataset based on complete monophyly amongst all of our lineages phylogenetically (Fig. 1e), better differentiation among morphological entities under the SNP hypothesis (Fig. 2b; Table 1), and less niche overlap between highland species (Fig. 3). The distributions include: (1) *I. coccifer*, which comprises all lowland populations within and outside of the Chortís Block; (2) *I. ibarrai*, which comprises all highland populations west of the Chortís Highlands in Guatemala; and (3) *I. porteri*, which comprises all highland populations within the Chortís Highlands. These distributions make sense not only based on the genetic, morphological, and macroecological data that we present, but also

94

the biogeographic barriers that divide these species distributions. *I. coccifer* is clearly separate from the two highland taxa by a highland barrier, and *I. ibarrai* and *I. porteri* are separated from each other by the Motagua-Polochic fault system at the western border of the Chortís Block, a well-supported biogeographic boundary for amphibians (Fig. 1b, S5) (Crawford and Smith, 2005; Rovito et al., 2015; Mendoza-Henao et al., 2020). It may also be worth noting that *I. porteri*'s distribution may extend into the highlands of Nicaragua (still comprising the Chortís Highlands), due to the lack of any biogeographic separation. However, no confirmed samples have ever been reported from this region.

Ultimately, we were able to use integrated lines of evidence in a hypothesis testing framework to evaluate the species limits and distributions of the three species found within this highly confounding species complex. A major factor that may also be confounding this species complex is that it is a relatively young complex (divergence was estimated at ~790,000 years ago; Firneno et al. 2020). Young lineages may still be undergoing the process of speciation/divergence, which can prove challenging for the types of delimitation methods that we bring to bear (Padial et al., 2010; Fujita et al., 2012). In the future, more taxon-specific lines of evidence (e.g. bioacoustics, osteology, etc.) may be useful in clarifying and solidifying this complex's taxonomy further, as well as using a process-based delimitation approach that integrates processes of speciation into delimitation practices (Smith and Carstens, 2020). Finally, due to the long-standing taxonomy and distributions that were originally proposed by Mendelson et al. (2005), there have been existing conservation recommendations that may now need to be reevaluated based upon these new findings.

Conservation implications

A stable taxonomy is necessary to guide conservation assessment measures and practices for species. Central American amphibians, such as those within the *I. coccifer* complex, represent some of the world's most vulnerable organisms due to rapid habitat loss and zoonotic disease (e.g., chytridiomycosis) in this region (Lips, 1998; Gallant et al., 2007; Townsend and Wilson, 2010b; James et al., 2015; Lips, 2016; Whitfield et al., 2016); therefore, stable taxonomies are necessary for proper conservation. A recent IUCN reevaluation of this complex assessed the distribution of *I. porteri* on the guidance of Firneno et al. (2020) and changed its conservation status from Data Deficient to Least Concern; currently the IUCN recognizes *I. coccifer* and *I porteri* as Least Concern and *I. ibarrai* as Endangered. However, the distribution and conservation of *I. ibarrai* are awaiting reevaluation in the IUCN Red List, and we propose that the distribution of *I. ibarrai* should be restricted to everything west of the Chortís/Guatemalan Highlands boundary (Fig. 1b,S5) and we suggest that *I. ibarrai*'s conservation status be reevaluated to accurately reflect this restricted range distribution.

While we sought to evaluate the species limits of the *Incilius coccifer* complex, we also wanted to provide an integrative hypothesis testing framework to aid in the delimitation of species complexes in the face of mitonuclear discordance. Our results highlight the importance of using comprehensive and integrated data for addressing complex species delimitation problems. Our findings further emphasize the necessity of an accurate taxonomy to guide conservation measures, especially among organisms that inhabit ecosystems that are highly imperiled. With this integrative taxonomic approach, we hope to fuel a novel perspective on pursuing species delimitation and phylogeographic work using complex and cryptic species and work within this region of Central America where amphibians (and other organisms) are facing grave conservation concerns due to anthropogenic changes to their ecosystems.

ACKNOWLEDGEMENTS

DATA ACCESSIBILITY

We have included a large data package on DRYAD (https://doi.org/10.5061/dryad.j9kd51cb5) that includes our final ddRADseq filtered "haplotypes" files, COI alignment files, resulting input files for a number of analysis programs (STRUCTURE, ADMIXTURE, DAPC, FINERADSTRUCTURE, BFD*, BPP, gdi, DELINEATE, BEAST, and SNAPP), an excel spreadsheet of raw morphological measurements split into the two distributional datasets used, and a .csv file of both locality datasets used for SDMs.

LITERATURE CITED

Aiello-Lammens ME, Boria RA, Radosavljevic A, Vilela B, Anderson RP. 2015 spThin: an R package for spatial thinning of species occurrence records for use in ecological niche models. *Ecography*, **38**, 541-545. (doi: 10.1111/ecog.01132)

Alexander DH, Novembre J, Lange K. 2009. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **19**, 1655-1664. (doi:10.1101/gr.094052.109)

Bickford D, Lohman DJ, Sodhi NS, Ng PKL, Meier R, Winker K, Ingram KK, Das I. 2007 Cryptic species as a window on diversity and conservation. *Trends Ecol. Evol.* **22**, 148-155. (doi:10.1016/j.tree.2006.11.004)

Bouckaert RR. 2010 DensiTree: making sense of sets of phylogenetic trees. *Bioinformatics* **26**, 1372-1373. (doi: 10.1093/bioinformatics/btq110)

Bouckaert R, Heled J, Kühnert D, Vaughan T, Wu CH, Xie D, … Drummond AJ. 2014 BEAST 2: A Software Platform for Bayesian Evolutionary Analysis. *PLoS Comput. Biol.* **10**, 1-6. (doi:10.1371/journal.pcbi.1003537)

Bryant D, Bouckaert R, Felsenstein J, Rosenberg NA, RoyChoudhury A. 2012 Inferring species trees directly from biallelic genetic markers: Bypassing gene trees in a full coalescent analysis. *Mol. Biol. Evol.* **29**, 1917-1932. (doi:10.1093/molbev/mss086)

Burnham KP, Anderson DR. 2002 *Model Selection and Multimodal Inference: A Practical Information-Theoretic Approach*, Second Edition. Springer, Berlin.

Catchen J, Hohenlohe PA, Bassham S, Amores A, Cresko WA. 2013 Stacks: an analysis tool set for population genomics. *Mol. Ecol.* **22**, 3124-3140. (doi: 10.1111/mec.12354)

Chan KO, Grismer LL. 2019 To split or not to split? Multilocus phylogeny and molecular
  species delimitation of southeast Asian toads (family: Bufonidae). *BMC Evol. Biol.* **95**, 1-
  12. (doi: 10.1186/s12862-019-1422-3)

Constable H, Guralnick R, Wieczorek J, Spencer C, Peterson AT. 2010 VertNet: A new model
  for biodiversity data sharing. *PLoS Biol.* **8**: e1000309. (doi:
  10.1371/journal.pbio.1000309)

Corrigan S, Maisano P, Eddy C, Duffy C, Yang L, Li C, Bazinet AL, Mona S, Naylor GJP. 2017
  Historical introgression drives pervasive mitochondrial admixture between two species of
  pelagic sharks. *Mol. Phylogenet. Evol.* **110**, 122-126. (doi:10.1016/ j.ympev.2017.03.011)

Crawford AJ, Smith EN. 2005 Cenozoic biogeography and evolution in direct-developing frogs
  of Central America (Leptodactylidae: Eleutherodactylus) as inferred from a phylogenetic
  analysis of nuclear and mitochondrial genes. *Mol. Phylogenet. Evol.* **35**, 536 – 555.
  (doi:10.1016/j.mpev.2005.03.006)

de Queiroz, K. 1998 The general lineage concept of species, species criteria, and the process of
  speciation: a conceptual unification and terminological recommendations. In *Endless
  Forms: Species and Speciation*, ed. DJ Howard, SH Berlocher, pp. 57–75. Oxford:
  Oxford University Press.

de Queiroz, K. 1999 The general lineage concept of species and the defining properties of the
  species category. In *New interdisciplinary essays*, ed. RA Wilson, pp. 49-89. Cambridge,
  Massachusetts: MIT Press.

de Queiroz, K. 2005 Ernst Mayr and the modern concept of species. *Proc. Natl. Acad. Sci. USA.*
  102, 6600-6607.

de Queiroz, K. 2007 Species Concepts and Species Delimitation. *Syst. Biol.* **56**(6), 879-886.

Dupuis JR, Roe AD, Sperling FAH. 2012 Multi-locus species delimitation in closely related
animals and fungi: one marker is not enough. *Mol. Ecol.* **21**, 4422-4436.
(doi:10.1111/j.1365-294X.2012.05642.x)

Earl DA, vonHoldt BM. 2012 STRUCTUREHARVESTER: a website and program for
visualizing STRUCTURE output and implementing the Evanno method. *Conserv. Genet.
Resour.* **4**, 359-361. (doi: 10.1007/s12686-011-9548-7)

Edgar RC. 2004 MUSCLE: multiple sequence alignment with high accuracy and high
throughput. *Nucleic Acids Res.* **32**, 1792-1797. (doi: 10.1093/nar/gkh340)

Evanno G, Regnaut S, Goudet J. 2005 Detecting the number of clusters of individuals using the
software STRUCTURE: a simulation study. *Mol. Ecol.* **14**, 2611-2620. (doi:
10.1111/j.1365-294X.2005.02553.x)

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

Fennessy J, Bidon T, Reuss F, Kumar V, Elkan P, Nilsson MA, Vamberger M, Fritz U, Janke A.
2016 Multi-locus analyses reveal four giraffe species instead of one. *Curr. Biol.* **26**,
2543-2549. (doi:10. 1016/j.cub.2016.07.036)

Fick SE, Hijmans RJ. 2017 WorldClim2: new 1km spatial resolution climate surfaces for global
land areas. *Int. J. Climatol.* **37**, 4302-4315. (doi: 10.1002/joc.5086)

Firneno TJ Jr, Townsend JH. 2019 Evaluation of species boundaries in sympatric and parapatric
populations of Mesoamerican toads. *Zool. Scr.* **48**, 454-465. (doi:10.1111/zsc.12354)

Firneno TJ Jr, O'Neill JR, Portik DM, Emery AH, Townsend JH, Fujita MK. 2020 Finding
complexity in complexes: assessing the causes of mitonuclear discordance in a

problematic species complex of Mesoamerican toads. *Mol. Ecol.* **29**, 3543-3559.
(doi:10.1111/mec.15496)

Flouri T, Jiao X, Rannala B, Yang Z. 2018 Species tree inference with BPP using genomic
sequences and the multispecies coalescent. *Mol. Biol. Evol.* **35**, 2585-2593. (doi:
10.1093/molbev/msy147)

Flouri T, Jiao X, Rannala B, & Yang Z. 2020 A Bayesian implementation of the multispecies
coalescent model with introgression for phylogenomic analysis. *Mol. Biol. Evol.* **37**,
1211-1223. (doi:10.1093/molbev/msz296)

Fujisawa T, Barraclough TG. 2013 Delimiting species using single-locus data and the
generalized mixed yule coalescent approach: a revised methods and evaluation on
simulated datasets. *Syst. Biol.* **62**, 707-724. (doi:10.1093/sysbio/syt033)

Fujita MK, Leaché AD, Burbrink FT, McGuire JA, Moritz C. 2012 Coalescent-based species
delimitation in an integrative taxonomy. *Trends Ecol. Evol.* **27**, 480-488.
(doi:10.1016/j.tree.2012.04.012)

Funk DJ, Omland KE. 2003 Species-level paraphyly and polyphyly: frequency, causes, and
consequences, with insights from animal mitochondrial DNA. *Annu. Rev. Ecol. Evol.
Syst.* **34**, 397-423. (doi:10.1146/annurev.ecolsys.34.011802. 132421)

Gallant AL, Klaver RW, Casper GS, Lannoo MJ. 2007 Global rates of habitat loss and
implications for amphibian conservation. *Copeia* **2007,** 967-979. (doi:10.1643/0045-
8511(2007)7[967:GROHLA]2.0.CO;2)

Hidalgo-Galiana A, Sánchez-Fernández D, Bilton DT, Cieslak A, Ribera I. 2014 Thermal niche
evolution and geographic range expansion in a species complex of western

Mediterranean diving beetles. *BMC Evol. Biol.* **14**, 1-17. (doi:10.1186/s12862-014-0187-y)

Hijmans RJ, Phillips S, Leathwick J, Elith J. 2011 Package 'dismo'. Available at: http://cran.r-project.org/web/packages/dismo/index.html.

Hinojosa JC, Koubínová D, Szenteczki MA, Pitteloud C, Dincă V, Alvarez N, Vila R. 2019 A mirage of cryptic species: Genomics uncover striking mitonuclear discordance in the butterfly *Thymelicus sylvestris*. *Mol. Ecol.* **28**, 3857-3868. (doi:10.1111/mec.15153)

Hofmann EP, Nicholson KE, Luque-Montes IR, Köhler G, Cerrato-Mendoza CA, Medina-Flores M, Wilson LD, Townsend JH. 2019 Cryptic diversity, but to what extent? Discordance between single-locus species delimitation methods within mainland anoles (Squamata: Dactyloidea) of Northern Central America. *Front. Genet.* **10**:11. (doi:10.3389/fgene.2019.00011)

Ivanov V, Lee KM, Mutanen M. 2018 Mitonuclear discordance in wolf spiders: Genomic evidence for species integrity and introgression. *Mol. Ecol.* **27**, 1681-1695. (doi:10.1111/mec.14564)

Jackson ND, Morales AE, Carstens BC, O'Meara BC. 2017 PHRAPL: Phylogenetic inference using approximate likelihoods. *Syst. Biol.* **66**, 1045-1053. (doi:10.1093/sysbio/syx001)

James TY, Toledo LF, Rödder D, da Silva Leite D, Belasen AM, Betancourt-Román CM, Jenkinson TS, Soto-Azat C, Lambertini C, Longo AV, Ruggeri J, Collins JP, Burrowes PA, Lips KR, Zamudio KR, Longcore JE. 2015 Disentangling host, pathogen, and environmental determinants of a recently emerged wildlife disease: lessons from the first 15 years of amphibian chytridiomycosis research. *Ecol. Evol.* **5**, 4079-4097. (doi:10.1002/ece3.1672)

Jombart T. 2008 *adegenet*: a R package for the multivariate analysis of genetic markers. *Bioinformatics* **24**, 1403-1405. (doi:10.1093/bioinformatics/btn129)

Jombart T, Ahmed I. 2011 *adegene 1.3-1t*: new tools for the analysis of genome-wide SNP data. *Bioinformatics* **27**, 3070-3071. (doi:10.1093/bioinformatics/btn521)

Kapli P, Lutteropp S, Zhang J, Kobert K, Pavlidis P, Stamatakis A, Flouri T. 2016 Multi-rate Poisson tree processes for single-locus species delimitation under maximum likelihood and Markov chain Monte Carlo. *Bioinformatics* **33**, 1630-1638. (doi:10.1093/bioinformatics/btx025)

Kass JM, Vilela B, Aiello-Lammens ME, Muscarella R, Merow C, Anderson RP. 2018 WALLACE: a fexible platform for reproducible modeling of species niches and distributions built for community expansion. *Methods Ecol. Evol.* **9**, 1151-1156. (doi:10.1111/2041-210X.12945)

Knowles LL, Carstens B. 2007 Delimiting species without monophyletic gene trees. *Syst. Biol.* **56**, 887-895. (doi:10.1080/10635150701701091)

Kumar S, Stecher G, Tamura K. 2016 MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol. Biol. Evol.* **33**, 1870–1874. (doi: 10.1093/molbev/msw054)

Leaché AD, Koo MS, Spencer CL, Papenfuss TJ, Fisher RN, McGuire JA. 2009 Quantifying ecological, morphological, and genetic variation to delimit species in the coast horned lizard species complex (*Phrynosoma*). *Proc. Natl. Acad. Sci. USA.* **106,** 12418-12423. (doi:10.1073/pnas.0906380106)

Leaché AD, Fujita MK, Minin VN, Bouckaert RR. 2014 Species delimitation using genome-wide SNP data. *Syst. Biol.* **63**, 534-542. (doi:10.1093/sysbio/syu018)

Leaché AD, McElroy MT, Trinh A. 2018 A genomic evaluation of taxonomic trends through time in coast horned lizards (genus *Phrynosoma*). *Mol. Ecol.* **27**, 2884-2895. (doi:10.111/mec.14715)

Leaché AD, Zhu T, Rannala B, Yang Z. 2019 The spectre of too many species. *Syst. Biol.* **68**, 168-181. (doi:10.1093/sysbio/syy051)

Lips KR. 1998 Decline of a tropical montane amphibian fauna. *Conserv. Biol.* **12,** 106-117. (doi:10.1111/j.1523-1739.1998.96359.x)

Lips KR. 2016 Overview of chytrid emergence and impacts on amphibians. *Philos. T. R. Soc. B.* **371**, 20150465. (doi:10.1098/rstb.2015.0465)

Luo A, Ling C, Ho SYW, Zhu C. 2018 Comparison of methods for molecular species delimitation across a range of speciation scenarios. *Syst. Biol.* **67**, 830-846. (doi:10.1093/sysbio/syy011)

Malinsky M, Trucchi E, Lawson DJ, Falush D. 2018 RADpainter and fineRADstructure: population inference for RADseq data. *Mol. Bio. Evol.* **35**, 1284-1290. (doi:10.1093/molbev/msy023)

Martin SH et al. 2013 Genome-wide evidence for speciation with gene flow in Heliconius butterflies. *Genome Res.* **23**, 1817-1828. (doi:10.1101/gr. 159426.113)

Mayden, RL. 1997 A hierarchy of species concepts: the denoument in the saga of the species problem. In *Species: The units of diversity*, eds. MF Claridge, HA Dawah, and MR Wilson, pp. 381-423. London: Chapman and Hall.

McCranie JR, Castañeda FE. 2007 *Guia de Campo de los Anfibios de Honduras*. Salt Lake City, UT: Bibliomania!

McCranie J. 2015 A checklist of the amphibians and reptiles of Honduras, with additions, comments, on taxonomy, some recent taxonomic decisions, and areas of further study needed. *Zootaxa* **3931**, 352-386.

Mendelson JR III, Brodie ED Jr, Malone JH, Acevedo ME, Baker MA, Smatresk NJ Campbell JA. 2004 Factors associated with the catastrophic decline of a cloudforest frog fauna in Guatemala. *Rev. Biol. Trop.* **52**, 991-1000.

Mendelson JR III, Williams BL, Sheil CA, Mulcahy DG. 2005 Systematics of the *Bufo coccifer* complex (Anura: Bufonidae) of Mesoamerica. *Scientific Papers: Natural History Museum The University of Kansas* **38**, 1-27. (doi:10.1007/s11084-011-9248-z)

Mendelson JR III, Mulcahy DG, Williams TS, Sites JW. 2011 A phylogeny and evolutionary natural history of Mesoamerican toads (Anura: Bufonidae: Incilius) based on morphology, life history, and molecular data. *Zootaxa* **34**, 1-34. (doi:10.11646/zootaxa.3138.1.1)

Mendoza-Henao AM, Arias E, Townsend JH, Parra-Olea G. 2020 Phylogeny-based species delimitation and integrative taxonomic revision of the Hyalinobatrachium fleischmanni species complex, with resurrection of H. viridissimum (Taylor 1942). *Syst. Biodivers.* **18**, 464–484. (doi:10.1080/14772000.2020.1776781)

Muscarella R, Galante PJ, Soley-Guardia M, Boria RA, Kass JM, Uriarte M, Anderson RA. 2014 ENMeval: An R package for conducting spatially independent evaluations and estimating optimal model complextity for MAXENT ecological niche models. *Methods Ecol. Evol.* **5**, 1198-1205. (doi: 10.1111/2041-210X.12261)

Oksanen J, Guillaume-Blanchet F, Kindt R, Legendre P, Minchin PR, O'Hara RB, Simpson GL, Solymos P, Henry M, Stevens H, Wagner H. 2017 vegan: community ecology package R Package Version 2.4-4. https://CRAN.R-project.org/package=vegan

Padial JM, Miralles A, de la Riva I, Vences M. 2010 The integrative future of taxonomy. *Front. Zoo.* **7**(16): 2010. (doi: 10.1186/1742-9994-7-16)

Page TJ, Choy SC, Hughes JM. 2005 The taxonomic feedback loop: symbiosis of morphology and molecules. *Biol. Lett.* **1**, 139-142. (doi:10.1098/rsbl.2005.0298)

Papakostas S, Michaloudi E, Proios K, Brehm M, Verhage L, Rota J, Peña C, Stamou G, Pritchard VL, Fontaneto D, Declerck SAJ. 2016 Integrative taxonomy recognizes evolutionary units despite widespread mitonuclear discordance: evidence from a rotifer cryptic species complex. *Syst. Biol.* **65**, 508-524. (doi:10.1093/sysbio/syw016)

Pedraza-Marrón C, Silva R, Deeds J, Van Belleghem SM, Mastretta-Yanes A, Domínguez-Domínguez O, Rivero-Vega RA, Lutackas L, Murie D, Parkyn D, Bullock LH, Foss K, Ortiz-Zuazaga H, Naváez-Barandica J, Acero A, Gomes G, Betancur-R R. 2019 Genomics overrules mitochondrial DNA, siding with morphology on a controversial case of species delimitation. *P. R. Soc B.* **286**, 20182924. (doi:10.1098/rspb.2018.2924)

Phillips SJ, Dudík M. 2008 Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. *Ecography* **31**, 161-175. (doi:10.1111/j.0906-7590.2008.5203.x)

Phuong MA, Bi K, Moritz C. 2016 Range instability leads to cytonuclear discordance in a morphologically cryptic ground squirrel species complex. *Mol. Ecol.* **26**, 4743-4755. (doi:10.1111/mec.14238)

Pinho C, Hey J. 2010 Divergence with gene flow: models and data. *Annu. Rev. Ecol. Evol. Syst.* **41**, 215-230. (doi: 10.1146/annurev-ecolsys-102209-144644)

Portik DM, Leaché AD, Rivera D, Barej MF, Burger M, Hirschfeld M, Rödel MO, Blackburn DC, Fujita MK. 2017 Evaluating mechanisms of diversification in a Guineo-Congolian tropical forest frog using demographic model selection. *Mol. Ecol.* **26**, 5245-5263. (doi:10.1111/mec14266)

Pritchard JK, Stephens M, Donnelly P. 2000. Inference of population structure using multilocus genotype data. *Genetics* **155**, 945-959. (doi:10.1006%2Ftpbi.2001.1543)

Puillandre N, Lambert A, Brouillet S, Achaz G. 2012 ABGD, automatic barcode gap discovery for primary species delimitation. *Mol. Ecol.* **21**, 1864-1877. (doi:10.1111/j.1365-294X.2011.05239.x)

R Core Team. 2019 R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available at: http://R-project.org/

Radosavljevic A, Anderson RP. 2014 Making better MAXENT models of species distributions: complexity, overfitting and evaluation. *J. Biogeogr.* **41**, 629-643. (doi: 10.1111/jbi.12227)

Rambaut A, Drummond A. 2009 Tracer v1.5.0. Available from: http://beast.bio.ed.ac.uk/

Razkin O, Gómez-Moliner BJ, Vardinoyannis K, Martínez-Ortí A, Madeira MJ. 2016 Species delimitation for cryptic species complexes: case study of *Pyramidula* (Gastropoda, Pulmonata). *Zool. Scr.* **46**, 55-72. (doi:10.1111/zsc.12192)

Rissler LJ, Apodaca JJ. 2007 Adding more ecology into species delimitation: ecological niche models and phylogeography help define cryptic species in the Black Salamander (*Aneides flavipunctatus*). *Syst. Biol.* **56**, 924-942. (doi:10.1080/1063515070170363)

Rovito SM, Parra-Olea G, Recuero E, Wake DB. 2015 Diversification and biogeographical

history of Neotropical plethodontid salamanders. *Zoo. J. Linn. Soc-Lond.* **175**, 167-188.

(doi:10.1111/zoj.12271)

Sangster G. 2014 The application of species criteria in avian taxonomy and its implications for

the debate over species concepts. *Biol. Rev.* **89**(1): 199-214.

Shcheglovitova M, Anderson RP. 2013 Estimating optimal complexity for ecological niche

models: A jackknife approach for species with small sample sizes. *Ecol. Model.* **296**, 9-

17. (doi: 10.1016/j.ecolmodel.2013.08.011)

Sites JW Jr, Marshall JC. 2004 Operational criteria for delimiting species. *Annu. Rev. Ecol. Evol.*

Syst. **35**, 199-227. (doi:10.1146/annurev.ecolsys.35.112202.130128)

Smith ML, Carstens BC. Process-based species delimitation leads to identification of more

biologically relevant species. *Evol.* **74**(2), 216-229.

Spinks PQ, Thomson RC, Shaffer BH. 2014 The advantages of going large: genome-wide SNPs

clarify the complex population history and systematics of the threatened western pond

turtle. *Mol. Ecol.* **23**, 2228-2241. (doi:10.1111/mec. 12736)

Stamatakis A. 2014 RAxML version 8: a tool for phylogenetic analysis and post-analysis or

large phylogenies. *Bioinformatics* **30**, 1312-1313. (doi:10.1093/bioinformatics/btu.003)

Sukumaran J, Knowles LL. 2017 Multispecies coalescent delimits structure, not species. *Proc.*

Natl. Acad. Sci. USA. **114**, 1607-1612. (doi:10.1073/pnas.16-7921114)

Sukumaran J, Holder MT, Knowles LL. In review. Incorporating the speciation process into

species delimitation.

Title PO, Bemmels JB. 2018 ENVIREM: an expanded set of bioclimatic and topographic variables increases flexibility and improves performance of ecological niche modeling. *Ecography* **41**, 291-307. (doi: 10.1111/ecog.02880)

Thielsch A, Knell A, Mohammadyari A, Petrusek A, Schwenk K. 2017 Divergent clades or cryptic species? Mito-nuclear discordance in a Daphnia species complex. *BMC Evol. Biol.* **17**, 1-9. (doi:10. 1186/s12862-017-1070-4)

Toews DPL, Brelsford A. 2012 The biogeography of mitochondrial and nuclear discordance in animals. *Mol. Ecol.* **21**, 3907-3930. (doi:10.1111/j.1365- 294X.2012.05664.x)

Townsend JH, Wilson LD. 2010a Conservation status of the Honduran herpetofauna: issues and imperatives. In *Conservation of Mesoamerican Amphibians and Reptiles* (eds. LD Wilson, JH Townsend, JD Johnson), pp. 488-509. Eagle Mountain, Utah, USA: Eagle Mountain Publishing.

Townsend JH, Wilson LD, Johnson JD. 2010b *Conservation of Mesoamerican Amphibians and Reptiles*. Eagle Mountain, Utah, USA: Eagle Mountain Publishing.

Townsend JH, Wilson LD. 2016 Amphibians of the Cordillera Nombre de Dios, Honduras: COI barcoding suggests underestimated taxonomic richness in a threatened endemic fauna. *Mesoamerican Herpetology* **3**, 910-927.

Venables WN, Ripley BD. 2002 *Modern Applied Statistics with S*, Fourth edition. Springer, New York.

Veloz SD. 2009 Spatially autocorrelated sampling falsely inflates measures of accuracy for presence-only niche models. *J. Biogeogr.* **36**, 2290-2299. (doi: 10.1111/j.1365-2699.2009.02174.x)

Victor BC. 2015 How many coral reef fish species are there? Cryptic diversity and the new
molecular taxonomy. In *Ecology of fishes on coral reefs: the functioning of an ecosystem
in a changing world* (ed. C Mora), pp. 76-87. Cambridge, UK: Cambridge University
Press.

Warren DL, Glor RE, Turelli M. 2010 ENMTools: a toolbox for comparative studies of
environmental niche models. *Ecography* **33**, 607-611. (doi:10.1111/j.600-
0587.2009.06142.x)

Warren DL, Seifert SN. 2011. Ecological niche modeling in Maxent: the importance of model
complexity and the performance of model selection criteria. *Ecol. Appl.* **21**, 335-342.
(doi: 10.1890/10-1171.1)

Warren DL, Matzke N, Cardillo M, Baumgartner J, Beaumont L, Huron N, Simões M, Iglesias
TL, Dinnage R. 2019 ENMTools (software package). Available at:
https://github.com/danlwarren/ENMTools. (doi: 10.5281/zenodo.3268814)

Weigand H, Weiss M, Cai H, Li Y, Yu L, Zhang C, Leese F. 2017 Deciphering the origin of
mitonuclear discordance in two sibling caddisfly species. *Mol. Ecol.* **26**, 5705-5715.
(doi:10.1111/mec. 14292)

Whitfield SM, Lips KR, Donnelly MA. 2016 Amphibian decline and conservation in Central
America. *Copeia* **104**, 351-379. (doi:10.1643/CH-15-300)

Wickham H. 2016 *ggplot2: Elegant graphics for data analysis*. Springer-Verlag, New York.

Zhang C, Zhang DX, Zhu T, Yang Z. 2011 Evaluation of a Bayesian coalescent model of species
delimitation. *Syst. Biol.* **60**, 747-761. (doi:10.1093/sysbio/syr071)

Zhang J, Kapli P, PavlidisP, Stamatakis A. 2013 A general species delimitation method with applications to phylogenetic placements. *Bioinformatics* **29**, 2896-2876. (doi: 10.1093/bioinformatics/btt499)

TABLES

Table 1. Classification rates of the linear discriminant functions analysis.

Taxonomic ID	Delimitation Hypothesis – Two Species				
	Nuclear SNP Assignment				
	Lowland	Highland	-	-	Total
Lowland	**130 (94.9%)**	7	-	-	137
Highland	22	**225 (91.1%)**	-	-	247
					384
	mtDNA Assignment				
	Lowland	Highland	-	-	Total
Lowland	**130 (94.9%)**	7	-	-	137
Highland	22	**225 (91.1%)**	-	-	247
					384
	Delimitation Hypothesis – Three Species				
	Nuclear SNP Assignment				
	I. coccifer	*I. ibarrai*	*I. porteri*	-	Total
I. coccifer	**130 (94.9%)**	4	4	-	138
I. ibarrai	0	**92 (89.3%)**	11	-	103
I. porteri	22	27	**94 (65.7%)**	-	143
					384
	mtDNA Assignment				
	I. coccifer	*I. ibarrai*	*I. porteri*	-	Total
I. coccifer	**130 (94.9%)**	3	4	-	137
I. ibarrai	7	**122 (81.9%)**	20	-	149
I. porteri	15	25	**58 (58.2%)**	-	98
					384
	Delimitation Hypothesis – Four Species				
	Nuclear SNP Assignment				
	I. coccifer	*I. ibarrai*	*I. porteri* C/E	*I. porteri* W	Total
I. coccifer	**131 (94.9%)**	1	4	2	138
I. ibarrai	0	**85 (92.4%)**	4	3	7
I. porteri C/E	19	18	**83 (61.5%)**	15	52
I. porteri W	2	0	7	**10 (52.6%)**	9
					384
	mtDNA Assignment				
	I. coccifer	*I. ibarrai* GT	*I. ibarrai* HN	*I. porteri*	Total
I. coccifer	**131 (94.9%)**	2	1	4	138
I. ibarrai GT	0	**85 (91.4%)**	5	3	93
I. ibarrai HN	10	3	**31 (60.8%)**	7	51
I. porteri	11	14	9	**68 (66.7%)**	102
					384

FIGURES

Figure 1. Distribution of haplotypes in the *I. coccifer* complex (*I. coccifer* in red, *I. ibarrai* in yellow, and *I. porteri* in blue) and sampling for our (a) mitochondrial and (b) SNP datasets. The Chortís/Guatemala Highland boundary is indicated by the dashed line. (c) Coancestry matrix from FINERADSTRUCTURE based on our SNP dataset (coefficients of coancestry are color coded from low (yellow) to high (black) and the dendrogram depicts a clustering of individual sampled based on the pairwise matrix of coancestry coefficients) indicating main populations (bolded squares) with substructuring identified within the sampling east of the Chortís/Guatemala Highlands boundary (dashed squares). (d) Discriminant analysis of principal components of SNP dataset. (e) Maximum likelihood cladograms of mtDNA (left) and SNP (right) datasets exhibiting extent of mitonuclear discordance. Nodes with high support (bootstrap $\geq$ 90) are indicated by black dots and asterisks (*) indicate samples with haplotypes that do not follow the exact pattern of mitonuclear discordance shown by dashed lines. (f) Summary of species delimitation analyses, with collapsed cladogram inferred from BEAST analysis of mtDNA (left) and SNAPP species tree of SNP dataset (right). Nodes with high support (posterior probability $\geq$ 0.9) are indicated by black dots. Blocks represent species units that have been estimated for each analysis. Gray areas connecting blocks indicate that the support for the split between those taxa is low to moderate (*gdi* between 0.1 and 0.7).

114
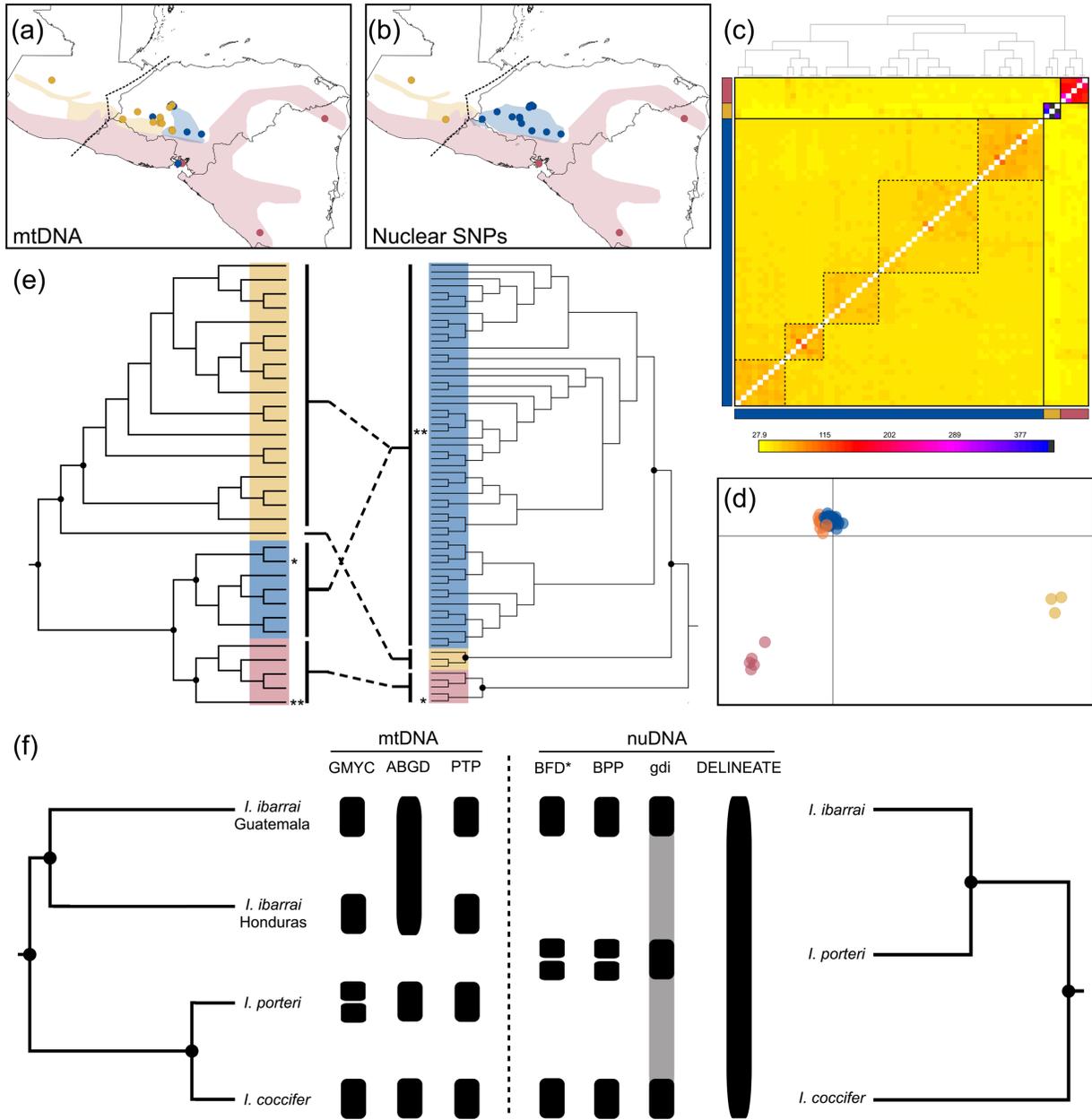
Figure 2. Analyses of the association between species delimitation and morphometry. (a) Visualization of the first two PCA loadings of the morphological variation. Convex polygons were produced using the mtDNA and SNP species assignments for across our molecular species delimitation hypotheses to show the degree of morphological variation and overlap between the species. (b) Venn diagram of the results of the variation partitioning analysis that depicts the adjusted R^2 values and significance levels of the mtDNA- and SNP-based delimitations for discordant highland populations (*I. ibarrai* and *I. porteri*) Values outside of the shaded areas represent marginal effects (R^2 adj; e.g. the amount of variation explained when testing each delimitation separately). Values in the intersection of the shaded area represent variation explained in common by both delimitations. Values outside of the intersection of the shaded area represent the conditional effects (variation uniquely explained by each delimitation). *$P < 0.001$.

(a)

mtDNA

SNP

- Highland - Lowland
- *I. coccifer* - *I. ibarrai*
- *I. porteri*
- *I. coccifer* - *I. ibarrai* GT
- *I. ibarrai* HN - *I. porteri*

- Highland - Lowland
- *I. coccifer* - *I. ibarrai*
- *I. porteri*
- *I. coccifer* - *I. ibarrai*
- *I. porteri* C/E - *I. porteri* W

(b)

mtDNA
(COI)
R^2adj = 0.32*

nuDNA
(SNPs)
R^2adj = 0.36*

0.05*   0.27   0.09*

Residuals = 0.58
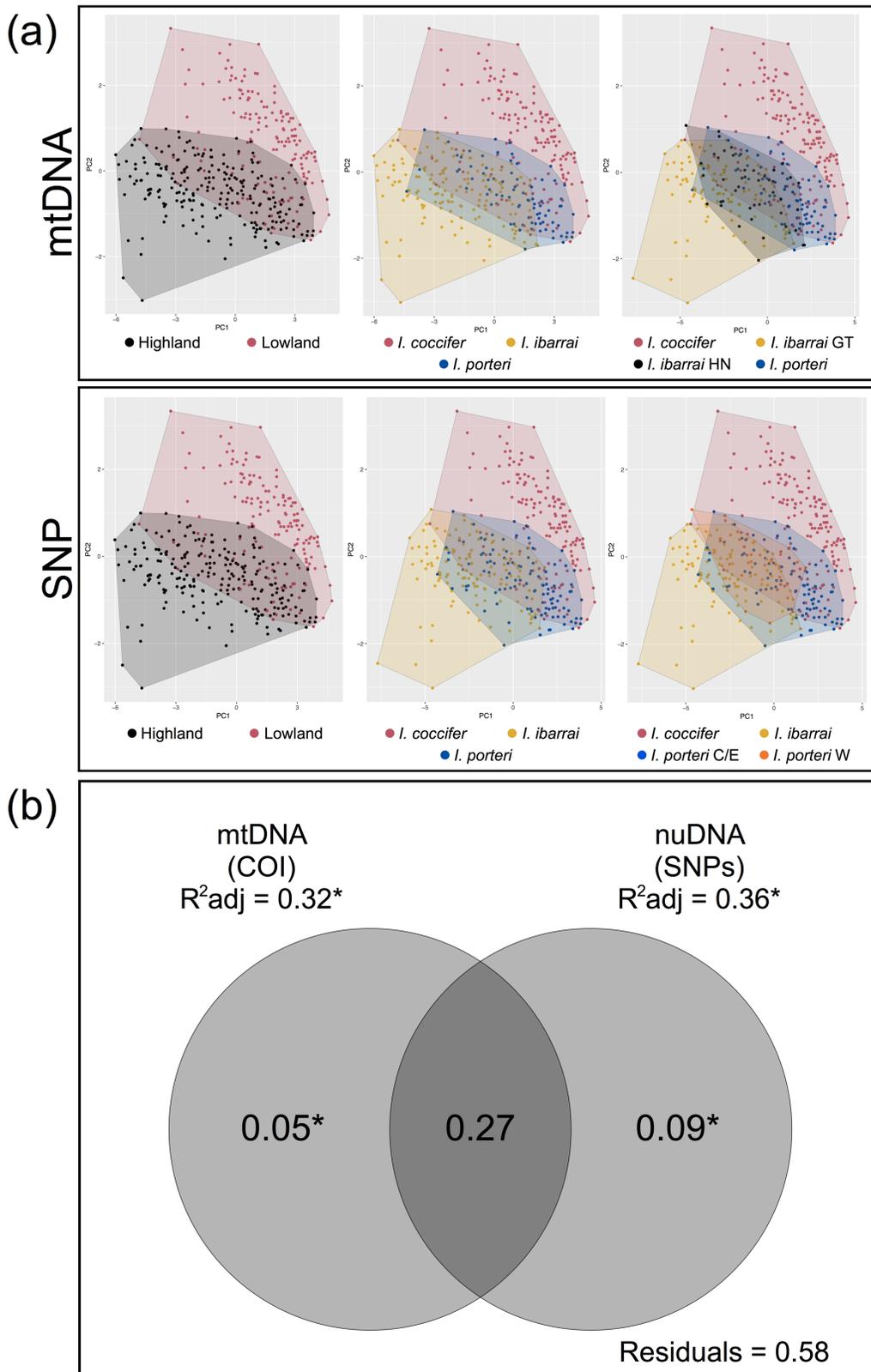
Figure 3. Combined binary maps for the modeled occurrence of the focal taxa based on genetic species assignments. Colors correspond as follows: *I. coccifer* only = red; *I. ibarrai* only = yellow; *I. porteri* only = blue; *I. coccifer* + *I. ibarrai* = orange; *I. coccifer* + *I. porteri* = purple; *I. ibarrai* + *I. porteri* = teal; all species present = black; none present = white. Dotted lines represent the boundaries of the Chortís Block. Intersections of focal taxa in the Venn diagrams include the results of niche overlap (Schoener's *D*: top number) and niche identity (*I*: bottom number) tests.
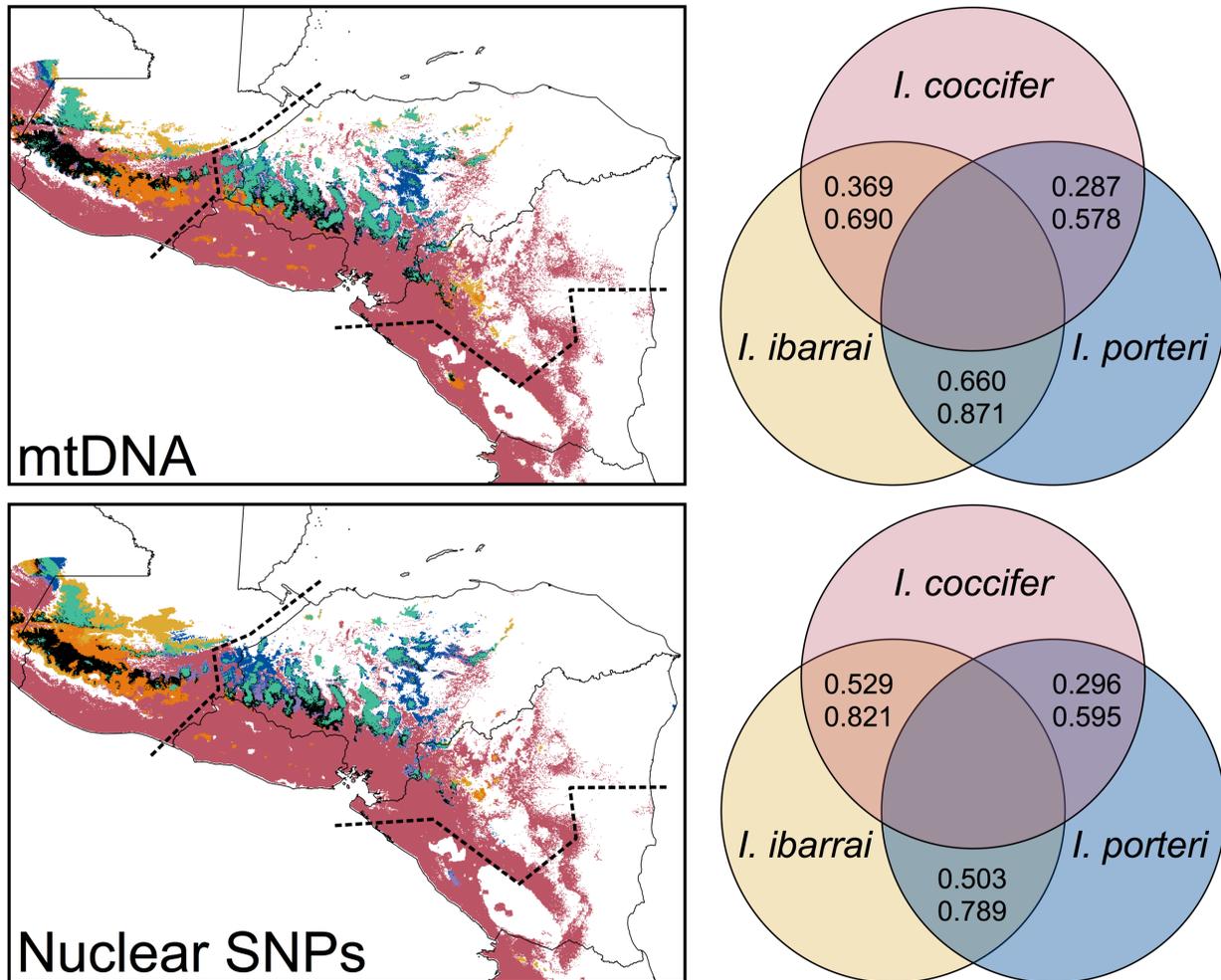
SUPPORTING INFORMATION

Table S1. Samples used for molecular data analyses, with locality data, lineage specificity (both mtDNA and nDNA), voucher

numbers, and GenBank Accession numbers; GT, Guatemala; HN, Honduras; MX, Mexico; NI, Nicaragua; PA, Panama.

mtDNA Lineage	SNP Lineage	Field Number	Museum Voucher Number	Locality	GenBank: mtDNA
I. coccifer	*I. coccifer*	JS1150	—	NI: Región Autónoma Atlántica Norte	MK720880
I. coccifer	*I. coccifer*	JS1058	—	NI: Rivas: Ometepe	MK720881
I. coccifer	*I. coccifer*	JS1016	—	NI: Rivas: Ometepe	MK720883
I. coccifer	*I. coccifer*	JHT3301	USNM 578693	HN: Valle: Isla el Tigre	KR736042
I. porteri	*I. coccifer*	JHT3302	USNM 578694	HN: Valle: Isla el Tigre	KR736043
I. ibarrai	*I. ibarrai*	JAC19612	UTA A-52528	GT: Quiché	JN867971
—	*I. ibarrai*	ENS13348	—	GT: Santa Rosa: Carretera Ayarza	—
—	*I. ibarrai*	ENS13349	—	GT: Santa Rosa: Carretera Ayarza	—
I. porteri	*I. porteri*	CAC044	USNM 578695	HN: Comayagua: Cerro Zarciadero	MK720846
I. ibarrai	*I. porteri*	IRL002	—	HN: Comayagua: Cerro Zarciadero	MK720865
—	*I. porteri*	IRL003	—	HN: Comayagua: Cerro Zarciadero	—
I. ibarrai	*I. porteri*	IRL005	—	HN: Comayagua: Cerro Zarciadero	MK720866
—	*I. porteri*	IRL006	—	HN: Comayagua: Cerro Zarciadero	—
I. porteri	*I. porteri*	IRL010	—	HN: Comayagua: Cerro Zarciadero	MK720838
I. porteri	*I. porteri*	IRL011	—	HN: Comayagua: Cerro Zarciadero	MK720849
I. porteri	*I. porteri*	JHT2149	—	HN: Francisco Morazán: Uyuca	MK720847
I. ibarrai	*I. porteri*	JHT2205	—	HN: Comayagua: Cerro Zarciadero	MK720863

—	*I. porteri*	JHT2206	—	HN: Comayagua: Cerro Zarciadero	—
I. porteri	*I. porteri*	JHT2228	—	HN: Francisco Morazán: Finca la Alondra	MK720843
I. porteri	*I. porteri*	JHT2246	—	HN: Francisco Morazán: Uyuca	MK720839
I. porteri	*I. porteri*	JHT2256	—	HN: Francisco Morazán: Uyuca	MK720844
I. porteri	*I. porteri*	JHT2257	—	HN: Francisco Morazán: Uyuca	MK720841
I. porteri	*I. porteri*	JHT3947	CM 163380	HN: Francisco Morazán: Uyuca	MT348746
I. ibarrai	*I. porteri*	JHT2604	—	HN: La Paz: Guajiquiro	MK720864
I. ibarrai	*I. porteri*	JHT2605	—	HN: La Paz: Guajiquiro	MK720860
I. ibarrai	*I. porteri*	JHT2608	—	HN: La Paz: Guajiquiro	MK720853
I. ibarrai	*I. porteri*	JHT2609	—	HN: La Paz: Guajiquiro	MK720874
I. porteri	*I. porteri*	JHT2610	—	HN: La Paz: Guajiquiro	MK720837
—	*I. porteri*	JHT2611	—	HN: La Paz: Guajiquiro	—
I. ibarrai	*I. porteri*	JHT2612	—	HN: La Paz: Guajiquiro	MK720871
I. ibarrai	*I. porteri*	JHT2613	—	HN: La Paz: Guajiquiro	MK720862
I. ibarrai	*I. porteri*	JHT2615	—	HN: La Paz: Guajiquiro	MK720855
—	*I. porteri*	JHT2616	—	HN: La Paz: Guajiquiro	—
I. ibarrai	*I. porteri*	JHT2906	—	HN: Intibucá: El Rodeo	MK720852
I. ibarrai	*I. porteri*	JHT3724	CM 163392	HN: Intibucá: El Rodeo	MT348755
I. ibarrai	*I. porteri*	JHT3725	CM 163393	HN: Intibucá: El Rodeo	MT348756
I. ibarrai	*I. porteri*	JHT3726	CM 163394	HN: Intibucá: El Rodeo	MT348757
I. ibarrai	*I. porteri*	JHT3728	CM 163396	HN: Intibucá: El Rodeo	MT348758

I. ibarrai	*I. porteri*	JHT3729	CM 163397	HN: Intibucá: El Rodeo	MT348759
I. ibarrai	*I. porteri*	JHT3730	CM 163398	HN: Intibucá: El Rodeo	MT348760
I. ibarrai	*I. porteri*	JHT3731	CM 163399	HN: Intibucá: El Rodeo	MT348761
I. ibarrai	*I. porteri*	JHT3732	CM 163400	HN: Intibucá: El Rodeo	MT348762
I. ibarrai	*I. porteri*	JHT3733	CM 168168	HN: Intibucá: El Rodeo	MK720875
I. ibarrai	*I. porteri*	JHT3777	CM 168169	HN: Intibucá: El Rodeo	MT348763
I. coccifer	*I. porteri*	JHT3793	CM 168171	HN: Intibucá: San Pedro la Loma	MK720884
I. ibarrai	*I. porteri*	JHT3794	CM 168172	HN: Intibucá: San Pedro la Loma	MK720872
I. ibarrai	*I. porteri*	JHT3795	CM 163306	HN: Intibucá: San Pedro la Loma	MT348764
I. ibarrai	*I. porteri*	JHT3922	CM 168174	HN: Intibucá: El Rodeo	MT348765
I. ibarrai	*I. porteri*	JHT3923	CM 168175	HN: Intibucá: El Rodeo	MT348766
I. ibarrai	*I. porteri*	JHT3924	CM 168176	HN: Intibucá: San Pedro la Loma	MT348767
I. porteri	*I. porteri*	JHT3925	CM 168179	HN: Intibucá: Opalaca	MK720840
I. porteri	*I. porteri*	JHT3932	CM 168186	HN: Intibucá: El Rodeo	MT348745
I. ibarrai	*I. porteri*	ENS10270	UTA A-53662	HN: Ocotepeque	JN867970
I. ibarrai	*I. porteri*	JHT3692	CM 163384	HN: Lempira: La Ventanas	MT348748
I. ibarrai	*I. porteri*	JHT3694	CM 163385	HN: Lempira: La Ventanas	MT348747
I. ibarrai	*I. porteri*	JHT3696	CM 163386	HN: Lempira: La Ventanas	MK720873
I. ibarrai	*I. porteri*	JHT3697	CM 163387	HN: Lempira: La Ventanas	MT348749
I. ibarrai	*I. porteri*	JHT3698	CM 163388	HN: Lempira: La Ventanas	MT348750
I. ibarrai	*I. porteri*	JHT3699	CM 163389	HN: Lempira: La Ventanas	MT348751
I. ibarrai	*I. porteri*	JHT3700	CM 163390	HN: Lempira: La Ventanas	MT348752

I. ibarrai	*I. porteri*	JHT3782	CM 168170	HN: Lempira: La Ventanas	MT348753
I. ibarrai	*I. porteri*	JHT3921	CM 168173	HN: Lempira: La Ventanas	MK720869
I. ibarrai	*I. porteri*	JHT3945	CM 168177	HN: Lempira: La Ventanas	MK720876
I. ibarrai	*I. porteri*	JHT3946	CM 168178	HN: Lempira: La Ventanas	MT348754
I. cycladen	—	JRM4607	UTA A-54847	MX: Guerrero: Agua de Obispo	JN867967
I. signifer	—	JRM 4968	UTA A-JRM 4968	PA: Cocle: El Copé	JN867988

Table S2. Bayes Factor Delimitation results for each model analysis. The number of species represents the number of species included in each analysis after lumping or splitting lineages.

Model	Species	MLE	BF	Rank
Current Taxonomy	3	-16377.806	—	2
I. coccifer, *I. ibarrai*, and *I. porteri*				
Lump Highland	2	-19958.855	7162.098	3
I. coccifer, *I. ibarrai* + *I. porteri*				
Split All Populations	4	-15893.671	-484.135	1
I. coccifer, *I. ibarrai*, *I. porteri* West, and *I. porteri* East				
Split All, Clump Some Highlands	3	-19253.439	5751.266	4
I. coccifer, *I. ibarrai* + *I. porteri* West, and *I. porteri* East				

Table S3. Genetically verified (from this and previously published data) and VertNet search results locality data used for species distribution models and the species assignments of occurrence records based on discordant genetic datasets.

Locality		Genetic Assignment	
Latitude	Longitude	mtDNA	SNP
9.9167	-84.1333	*I. coccifer*	*I. coccifer*
9.952109	-84.373238	*I. coccifer*	*I. coccifer*
9.981572	-84.746237	*I. coccifer*	*I. coccifer*
9.927956	-84.067083	*I. coccifer*	*I. coccifer*
9.909131	-84.135021	*I. coccifer*	*I. coccifer*
10.026993	-84.723326	*I. coccifer*	*I. coccifer*
14.0904	-89.07446	*I. coccifer*	*I. coccifer*
14.0904	-89.07446	*I. coccifer*	*I. coccifer*
14.0904	-89.07446	*I. coccifer*	*I. coccifer*
14.0904	-89.07446	*I. coccifer*	*I. coccifer*
14.36667	-89.1	*I. coccifer*	*I. coccifer*
14.36667	-89.1	*I. coccifer*	*I. coccifer*
14.36667	-89.1	*I. coccifer*	*I. coccifer*
14.36667	-89.1	*I. coccifer*	*I. coccifer*
14.36667	-89.1	*I. coccifer*	*I. coccifer*
14.36667	-89.1	*I. coccifer*	*I. coccifer*
14.36667	-89.1	*I. coccifer*	*I. coccifer*
13.71667	-89.00729	*I. coccifer*	*I. coccifer*
13.71667	-89.00729	*I. coccifer*	*I. coccifer*
13.71667	-89.00729	*I. coccifer*	*I. coccifer*
13.71667	-89.00729	*I. coccifer*	*I. coccifer*

13.71667	-89.00729	*I. coccifer*	*I. coccifer*
13.71667	-89.00729	*I. coccifer*	*I. coccifer*
13.71667	-88.93333	*I. coccifer*	*I. coccifer*
13.71667	-88.93333	*I. coccifer*	*I. coccifer*
13.71667	-88.93333	*I. coccifer*	*I. coccifer*
13.71667	-88.93333	*I. coccifer*	*I. coccifer*
13.71667	-88.93333	*I. coccifer*	*I. coccifer*
13.71152	-88.92807	*I. coccifer*	*I. coccifer*
13.71152	-88.92807	*I. coccifer*	*I. coccifer*
13.71667	-88.91484	*I. coccifer*	*I. coccifer*
13.78619	-89.38853	*I. coccifer*	*I. coccifer*
13.78619	-89.38853	*I. coccifer*	*I. coccifer*
13.78619	-89.38853	*I. coccifer*	*I. coccifer*
13.78619	-89.38853	*I. coccifer*	*I. coccifer*
13.78619	-89.38853	*I. coccifer*	*I. coccifer*
13.78619	-89.38853	*I. coccifer*	*I. coccifer*
13.78619	-89.38853	*I. coccifer*	*I. coccifer*
13.78619	-89.38853	*I. coccifer*	*I. coccifer*
13.78619	-89.38853	*I. coccifer*	*I. coccifer*
13.78619	-89.38853	*I. coccifer*	*I. coccifer*
13.78619	-89.38853	*I. coccifer*	*I. coccifer*
13.78619	-89.38853	*I. coccifer*	*I. coccifer*
13.78619	-89.38853	*I. coccifer*	*I. coccifer*
13.78619	-89.38853	*I. coccifer*	*I. coccifer*
13.78619	-89.38853	*I. coccifer*	*I. coccifer*
13.78619	-89.38853	*I. coccifer*	*I. coccifer*
13.78619	-89.38853	*I. coccifer*	*I. coccifer*
13.78619	-89.38853	*I. coccifer*	*I. coccifer*
13.78619	-89.38853	*I. coccifer*	*I. coccifer*
13.7	-89.40772	*I. coccifer*	*I. coccifer*
13.7	-89.40772	*I. coccifer*	*I. coccifer*
13.7	-89.40772	*I. coccifer*	*I. coccifer*
13.7	-89.40772	*I. coccifer*	*I. coccifer*
13.7	-89.40772	*I. coccifer*	*I. coccifer*
13.7	-89.40772	*I. coccifer*	*I. coccifer*
13.7	-89.40772	*I. coccifer*	*I. coccifer*
13.7	-89.40772	*I. coccifer*	*I. coccifer*
13.7	-89.40772	*I. coccifer*	*I. coccifer*
13.7	-89.40772	*I. coccifer*	*I. coccifer*
13.7	-89.40772	*I. coccifer*	*I. coccifer*
13.7	-89.40772	*I. coccifer*	*I. coccifer*
13.7	-89.40772	*I. coccifer*	*I. coccifer*
13.7	-89.40772	*I. coccifer*	*I. coccifer*
13.95	-88.16667	*I. coccifer*	*I. coccifer*
13.71029	-89.21052	*I. coccifer*	*I. coccifer*
13.71029	-89.21052	*I. coccifer*	*I. coccifer*

13.71029	-89.21052	*I. coccifer*	*I. coccifer*
13.71029	-89.21052	*I. coccifer*	*I. coccifer*
13.71029	-89.21052	*I. coccifer*	*I. coccifer*
13.71029	-89.21052	*I. coccifer*	*I. coccifer*
13.71029	-89.21052	*I. coccifer*	*I. coccifer*
13.71029	-89.21052	*I. coccifer*	*I. coccifer*
13.71029	-89.21052	*I. coccifer*	*I. coccifer*
13.71029	-89.21052	*I. coccifer*	*I. coccifer*
13.71029	-89.21052	*I. coccifer*	*I. coccifer*
13.71029	-89.21052	*I. coccifer*	*I. coccifer*
13.7	-89.2	*I. coccifer*	*I. coccifer*
13.7	-89.2	*I. coccifer*	*I. coccifer*
13.7	-89.2	*I. coccifer*	*I. coccifer*
13.7	-89.2	*I. coccifer*	*I. coccifer*
13.7	-89.05209	*I. coccifer*	*I. coccifer*
13.66667	-89.1	*I. coccifer*	*I. coccifer*
13.66667	-89.1	*I. coccifer*	*I. coccifer*
13.66667	-89.1	*I. coccifer*	*I. coccifer*
13.66667	-89.1	*I. coccifer*	*I. coccifer*
13.66667	-89.1	*I. coccifer*	*I. coccifer*
13.66667	-89.1	*I. coccifer*	*I. coccifer*
13.66667	-89.1	*I. coccifer*	*I. coccifer*
13.66667	-89.1	*I. coccifer*	*I. coccifer*
13.48333	-89.22431	*I. coccifer*	*I. coccifer*
13.48333	-89.22431	*I. coccifer*	*I. coccifer*
13.48333	-89.22431	*I. coccifer*	*I. coccifer*
13.48333	-89.22431	*I. coccifer*	*I. coccifer*
13.48333	-89.22431	*I. coccifer*	*I. coccifer*
13.48333	-89.22431	*I. coccifer*	*I. coccifer*
13.48333	-89.22431	*I. coccifer*	*I. coccifer*
13.48333	-89.22431	*I. coccifer*	*I. coccifer*
13.48333	-89.22431	*I. coccifer*	*I. coccifer*
13.48333	-89.22431	*I. coccifer*	*I. coccifer*
13.48333	-89.22431	*I. coccifer*	*I. coccifer*
13.48333	-89.22431	*I. coccifer*	*I. coccifer*
13.48333	-89.22431	*I. coccifer*	*I. coccifer*
13.48333	-89.22431	*I. coccifer*	*I. coccifer*
13.48333	-89.22431	*I. coccifer*	*I. coccifer*
13.48333	-89.22431	*I. coccifer*	*I. coccifer*
13.48333	-89.22431	*I. coccifer*	*I. coccifer*
13.48333	-89.22431	*I. coccifer*	*I. coccifer*
13.48333	-89.22431	*I. coccifer*	*I. coccifer*
13.48333	-89.22431	*I. coccifer*	*I. coccifer*
13.48333	-89.22431	*I. coccifer*	*I. coccifer*

13.48333	-89.22431	*I. coccifer*	*I. coccifer*
13.48333	-89.22431	*I. coccifer*	*I. coccifer*
13.48333	-89.22431	*I. coccifer*	*I. coccifer*
13.48333	-89.22431	*I. coccifer*	*I. coccifer*
13.48333	-89.22431	*I. coccifer*	*I. coccifer*
13.48333	-89.22431	*I. coccifer*	*I. coccifer*
13.48333	-89.22431	*I. coccifer*	*I. coccifer*
13.48333	-89.22431	*I. coccifer*	*I. coccifer*
13.48333	-89.22431	*I. coccifer*	*I. coccifer*
13.48333	-89.22431	*I. coccifer*	*I. coccifer*
13.48333	-89.22431	*I. coccifer*	*I. coccifer*
13.48333	-89.22431	*I. coccifer*	*I. coccifer*
13.48333	-89.22431	*I. coccifer*	*I. coccifer*
13.48333	-89.22431	*I. coccifer*	*I. coccifer*
13.48333	-89.22431	*I. coccifer*	*I. coccifer*
13.48333	-89.22431	*I. coccifer*	*I. coccifer*
13.48333	-89.22431	*I. coccifer*	*I. coccifer*
13.48333	-89.22431	*I. coccifer*	*I. coccifer*
13.48333	-89.22431	*I. coccifer*	*I. coccifer*
13.48333	-89.22431	*I. coccifer*	*I. coccifer*
13.48333	-89.22431	*I. coccifer*	*I. coccifer*
13.48333	-89.22431	*I. coccifer*	*I. coccifer*
13.48333	-89.22431	*I. coccifer*	*I. coccifer*
13.48333	-89.22431	*I. coccifer*	*I. coccifer*
13.7	-89.2	*I. coccifer*	*I. coccifer*
13.7	-89.2	*I. coccifer*	*I. coccifer*
13.86667	-89.51667	*I. coccifer*	*I. coccifer*
13.66667	-88.85742	*I. coccifer*	*I. coccifer*
13.66667	-88.80659	*I. coccifer*	*I. coccifer*
13.835036	-88.991881	*I. coccifer*	*I. coccifer*
13.72114	-89.00632	*I. coccifer*	*I. coccifer*
13.637212	-89.104758	*I. coccifer*	*I. coccifer*
13.327846	-88.044912	*I. coccifer*	*I. coccifer*
13.5391998	-87.346001	*I. coccifer*	*I. coccifer*
13.30028	-87.19083	*I. coccifer*	*I. coccifer*
13.31465	-87.19769	*I. coccifer*	*I. coccifer*
13.31465	-87.19769	*I. coccifer*	*I. coccifer*
14.80639	-85.90657	*I. coccifer*	*I. coccifer*
14.80639	-85.90657	*I. coccifer*	*I. coccifer*
14.80639	-85.90657	*I. coccifer*	*I. coccifer*
14.80639	-85.90657	*I. coccifer*	*I. coccifer*
14.80639	-85.90657	*I. coccifer*	*I. coccifer*
14.80639	-85.90657	*I. coccifer*	*I. coccifer*
14.80639	-85.90657	*I. coccifer*	*I. coccifer*
14.80639	-85.90657	*I. coccifer*	*I. coccifer*
14.8	-85.9	*I. coccifer*	*I. coccifer*

15.1	-85.53333	*I. coccifer*	*I. coccifer*
13.58333	-87.57206	*I. coccifer*	*I. coccifer*
13.58333	-87.57206	*I. coccifer*	*I. coccifer*
13.58333	-87.57206	*I. coccifer*	*I. coccifer*
13.58333	-87.57206	*I. coccifer*	*I. coccifer*
13.58333	-87.57206	*I. coccifer*	*I. coccifer*
14.486911	-89.632368	*I. coccifer*	*I. coccifer*
14.596426	-89.455693	*I. coccifer*	*I. coccifer*
14.596426	-89.455693	*I. coccifer*	*I. coccifer*
14.320935	-91.91243	*I. coccifer*	*I. coccifer*
14.320935	-91.91243	*I. coccifer*	*I. coccifer*
14.320935	-91.91243	*I. coccifer*	*I. coccifer*
14.320935	-91.91243	*I. coccifer*	*I. coccifer*
14.320935	-91.91243	*I. coccifer*	*I. coccifer*
14.320935	-91.91243	*I. coccifer*	*I. coccifer*
14.320935	-91.91243	*I. coccifer*	*I. coccifer*
14.320935	-91.91243	*I. coccifer*	*I. coccifer*
14.320935	-91.91243	*I. coccifer*	*I. coccifer*
14.320935	-91.91243	*I. coccifer*	*I. coccifer*
14.325942	-91.913725	*I. coccifer*	*I. coccifer*
14.325942	-91.913725	*I. coccifer*	*I. coccifer*
14.325942	-91.913725	*I. coccifer*	*I. coccifer*
14.325942	-91.913725	*I. coccifer*	*I. coccifer*
14.325942	-91.913725	*I. coccifer*	*I. coccifer*
14.223417	-90.147145	*I. coccifer*	*I. coccifer*
11.483676	-85.559265	*I. coccifer*	*I. coccifer*
11.906393	-85.915398	*I. coccifer*	*I. coccifer*
12.163883	-86.142838	*I. coccifer*	*I. coccifer*
12.163883	-86.142838	*I. coccifer*	*I. coccifer*
12.163883	-86.142838	*I. coccifer*	*I. coccifer*
12.163883	-86.142838	*I. coccifer*	*I. coccifer*
12.163883	-86.142838	*I. coccifer*	*I. coccifer*
12.206155	-86.103053	*I. coccifer*	*I. coccifer*
11.541592	-85.945336	*I. coccifer*	*I. coccifer*
14.547125	-83.99829	*I. coccifer*	*I. coccifer*
14.609592	-83.944591	*I. coccifer*	*I. coccifer*
13.267276	-87.644555	*I. coccifer*	*I. coccifer*
14.31196	-88.099483	*I. porteri*	*I. coccifer*
13.267276	-87.644555	*I. porteri*	*I. coccifer*
14.6032	-87.999802	*I. porteri*	*I. porteri*
14.6297998	-87.912498	*I. porteri*	*I. porteri*
14.6297998	-87.912498	*I. porteri*	*I. porteri*
14.63933	-87.91806	*I. porteri*	*I. porteri*
14.6501999	-87.838402	*I. porteri*	*I. porteri*
14.5	-87.5	*I. porteri*	*I. porteri*
13.8696003	-87.248199	*I. porteri*	*I. porteri*

14.0167	-87.0833	*I. porteri*	*I. porteri*
14.09343	-87.09127	*I. porteri*	*I. porteri*
14.09343	-87.09127	*I. porteri*	*I. porteri*
14.17175	-87.30732	*I. porteri*	*I. porteri*
14.17175	-87.30732	*I. porteri*	*I. porteri*
14.17175	-87.30732	*I. porteri*	*I. porteri*
14.17175	-87.30732	*I. porteri*	*I. porteri*
14.17175	-87.30732	*I. porteri*	*I. porteri*
14.17175	-87.30732	*I. porteri*	*I. porteri*
14.26667	-87.44634	*I. porteri*	*I. porteri*
14.18558	-87.08933	*I. porteri*	*I. porteri*
14.18558	-87.08933	*I. porteri*	*I. porteri*
14.18558	-87.08933	*I. porteri*	*I. porteri*
14.18558	-87.08933	*I. porteri*	*I. porteri*
14.18558	-87.08933	*I. porteri*	*I. porteri*
14.18558	-87.08933	*I. porteri*	*I. porteri*
14.0167	-87.0833	*I. porteri*	*I. porteri*
14.0167	-87.0833	*I. porteri*	*I. porteri*
14.0167	-87.0833	*I. porteri*	*I. porteri*
14.0167	-87.0833	*I. porteri*	*I. porteri*
14.0167	-87.0833	*I. porteri*	*I. porteri*
14.0167	-87.0833	*I. porteri*	*I. porteri*
14.0167	-87.0833	*I. porteri*	*I. porteri*
14.0167	-87.0833	*I. porteri*	*I. porteri*
14.0167	-87.0833	*I. porteri*	*I. porteri*
14.0167	-87.0833	*I. porteri*	*I. porteri*
14.0167	-87.0833	*I. porteri*	*I. porteri*
14.0167	-87.0833	*I. porteri*	*I. porteri*
14.02477	-87.08148	*I. porteri*	*I. porteri*
14.02477	-87.08148	*I. porteri*	*I. porteri*
14.18558	-87.08933	*I. porteri*	*I. porteri*
14.0833	-88.1167	*I. porteri*	*I. porteri*
14.25	-87.85	*I. porteri*	*I. porteri*
14.664016	-87.947416	*I. porteri*	*I. porteri*
14.664016	-87.947416	*I. porteri*	*I. porteri*
14.512316	-87.879798	*I. porteri*	*I. porteri*
14.505739	-87.892025	*I. porteri*	*I. porteri*
14.091147	-87.452583	*I. porteri*	*I. porteri*
14.029782	-87.074117	*I. porteri*	*I. porteri*
13.993687	-86.920867	*I. porteri*	*I. porteri*
13.993687	-86.920867	*I. porteri*	*I. porteri*
14.031017	-87.074832	*I. porteri*	*I. porteri*
14.031433	-87.085206	*I. porteri*	*I. porteri*
14.034735	-87.065131	*I. porteri*	*I. porteri*
14.209852	-87.09001	*I. porteri*	*I. porteri*

14.209852	-87.09001	*I. porteri*	*I. porteri*
14.209852	-87.09001	*I. porteri*	*I. porteri*
14.131103	-88.051536	*I. porteri*	*I. porteri*
14.131103	-88.051536	*I. porteri*	*I. porteri*
14.131103	-88.051536	*I. porteri*	*I. porteri*
14.131103	-88.051536	*I. porteri*	*I. porteri*
14.127017	-88.04351	*I. porteri*	*I. porteri*
14.128138	-88.013455	*I. porteri*	*I. porteri*
14.128138	-88.013455	*I. porteri*	*I. porteri*
14.128138	-88.013455	*I. porteri*	*I. porteri*
14.128138	-88.013455	*I. porteri*	*I. porteri*
14.128138	-88.013455	*I. porteri*	*I. porteri*
14.128138	-88.013455	*I. porteri*	*I. porteri*
14.127017	-88.04351	*I. porteri*	*I. porteri*
14.127017	-88.04351	*I. porteri*	*I. porteri*
14.127017	-88.04351	*I. porteri*	*I. porteri*
14.127017	-88.04351	*I. porteri*	*I. porteri*
14.127017	-88.04351	*I. porteri*	*I. porteri*
14.138363	-88.052246	*I. porteri*	*I. porteri*
14.138363	-88.052246	*I. porteri*	*I. porteri*
14.138363	-88.052246	*I. porteri*	*I. porteri*
14.138363	-88.052246	*I. porteri*	*I. porteri*
14.138363	-88.052246	*I. porteri*	*I. porteri*
14.138363	-88.052246	*I. porteri*	*I. porteri*
14.138363	-88.052246	*I. porteri*	*I. porteri*
14.138363	-88.052246	*I. porteri*	*I. porteri*
14.7991	-87.84105	*I. porteri*	*I. porteri*
14.7742666	-87.84676	*I. porteri*	*I. porteri*
14.780083	-87.8462	*I. porteri*	*I. porteri*
14.771187	-87.88838	*I. porteri*	*I. porteri*
14.034904	-87.076059	*I. porteri*	*I. porteri*
14.034904	-87.076059	*I. porteri*	*I. porteri*
14.034904	-87.076059	*I. porteri*	*I. porteri*
14.034904	-87.076059	*I. porteri*	*I. porteri*
14.034904	-87.076059	*I. porteri*	*I. porteri*
14.034904	-87.076059	*I. porteri*	*I. porteri*
14.034904	-87.076059	*I. porteri*	*I. porteri*
14.11922	-87.82989	*I. porteri*	*I. porteri*
14.14654	-87.84424	*I. porteri*	*I. porteri*
14.30463	-88.18425	*I. ibarrai*	*I. porteri*
14.30463	-88.18425	*I. ibarrai*	*I. porteri*
14.44488	-88.30748	*I. ibarrai*	*I. porteri*
14.44488	-88.30748	*I. ibarrai*	*I. porteri*
14.44488	-88.30748	*I. ibarrai*	*I. porteri*
14.44488	-88.30748	*I. ibarrai*	*I. porteri*
14.44488	-88.30748	*I. ibarrai*	*I. porteri*

14.43145	-88.32076	*I. ibarrai*	*I. porteri*
14.3	-88.183	*I. ibarrai*	*I. porteri*
14.3	-88.183	*I. ibarrai*	*I. porteri*
14.3	-88.183	*I. ibarrai*	*I. porteri*
14.3	-88.183	*I. ibarrai*	*I. porteri*
14.3	-88.183	*I. ibarrai*	*I. porteri*
14.3	-88.183	*I. ibarrai*	*I. porteri*
14.46425	-88.35177	*I. ibarrai*	*I. porteri*
14.46425	-88.35177	*I. ibarrai*	*I. porteri*
14.46425	-88.35177	*I. ibarrai*	*I. porteri*
14.46425	-88.35177	*I. ibarrai*	*I. porteri*
14.46425	-88.35177	*I. ibarrai*	*I. porteri*
14.41568	-88.30196	*I. ibarrai*	*I. porteri*
14.41568	-88.30196	*I. ibarrai*	*I. porteri*
14.41568	-88.30196	*I. ibarrai*	*I. porteri*
14.2667	-88.2167	*I. ibarrai*	*I. porteri*
14.2667	-88.2167	*I. ibarrai*	*I. porteri*
14.35	-88.3667	*I. ibarrai*	*I. porteri*
14.35	-88.3667	*I. ibarrai*	*I. porteri*
14.5667	-88.6333	*I. ibarrai*	*I. porteri*
14.43333	-88.99677	*I. ibarrai*	*I. porteri*
14.43333	-88.99677	*I. ibarrai*	*I. porteri*
14.43333	-88.99677	*I. ibarrai*	*I. porteri*
14.43333	-89.08632	*I. ibarrai*	*I. porteri*
14.43333	-89.08632	*I. ibarrai*	*I. porteri*
14.43333	-89.08632	*I. ibarrai*	*I. porteri*
14.43333	-89.08632	*I. ibarrai*	*I. porteri*
14.43333	-89.08632	*I. ibarrai*	*I. porteri*
14.43333	-89.08632	*I. ibarrai*	*I. porteri*
14.4833	-88.8	*I. ibarrai*	*I. porteri*
14.5	-89.3	*I. ibarrai*	*I. porteri*
14.5	-89.3	*I. ibarrai*	*I. porteri*
14.5	-89.3	*I. ibarrai*	*I. porteri*
14.5	-89.3	*I. ibarrai*	*I. porteri*
14.91333	-88.7805	*I. ibarrai*	*I. porteri*
14.91333	-88.7805	*I. ibarrai*	*I. porteri*
14.349149	-88.305531	*I. ibarrai*	*I. porteri*
14.233398	-88.466131	*I. ibarrai*	*I. porteri*
14.233398	-88.466131	*I. ibarrai*	*I. porteri*
14.514713	-88.785696	*I. ibarrai*	*I. porteri*
14.514713	-88.785696	*I. ibarrai*	*I. porteri*
14.421867	-89.122193	*I. ibarrai*	*I. porteri*
14.421867	-89.122193	*I. ibarrai*	*I. porteri*
14.421867	-89.122193	*I. ibarrai*	*I. porteri*
14.421867	-89.122193	*I. ibarrai*	*I. porteri*
14.421867	-89.122193	*I. ibarrai*	*I. porteri*

14.7991	-87.84105	*I. ibarrai*	*I. porteri*
14.7991	-87.84105	*I. ibarrai*	*I. porteri*
14.7991	-87.84105	*I. ibarrai*	*I. porteri*
14.7284	-87.9009	*I. ibarrai*	*I. porteri*
14.11992	-87.82989	*I. ibarrai*	*I. porteri*
14.11992	-87.82989	*I. ibarrai*	*I. porteri*
14.11992	-87.82989	*I. ibarrai*	*I. porteri*
14.11992	-87.82989	*I. ibarrai*	*I. porteri*
14.11992	-87.82989	*I. ibarrai*	*I. porteri*
14.11992	-87.82989	*I. ibarrai*	*I. porteri*
14.11992	-87.82989	*I. ibarrai*	*I. porteri*
14.11992	-87.82989	*I. ibarrai*	*I. porteri*
14.11992	-87.82989	*I. ibarrai*	*I. porteri*
14.11992	-87.82989	*I. ibarrai*	*I. porteri*
14.14654	-87.84424	*I. ibarrai*	*I. porteri*
14.14654	-87.84424	*I. ibarrai*	*I. porteri*
14.31311	-88.103352	*I. ibarrai*	*I. porteri*
14.441386	-88.145047	*I. ibarrai*	*I. porteri*
14.441386	-88.145047	*I. ibarrai*	*I. porteri*
15.07148	-90.1	*I. ibarrai*	*I. ibarrai*
15.089	-90.03124	*I. ibarrai*	*I. ibarrai*
15.089474	-90.119552	*I. ibarrai*	*I. ibarrai*
15.089474	-90.119552	*I. ibarrai*	*I. ibarrai*
15.089474	-90.119552	*I. ibarrai*	*I. ibarrai*
15.18819	-90.198416	*I. ibarrai*	*I. ibarrai*
15.23774	-90.272305	*I. ibarrai*	*I. ibarrai*
15.23774	-90.272305	*I. ibarrai*	*I. ibarrai*
15.23774	-90.272305	*I. ibarrai*	*I. ibarrai*
15.23774	-90.272305	*I. ibarrai*	*I. ibarrai*
15.23774	-90.272305	*I. ibarrai*	*I. ibarrai*
15.23774	-90.272305	*I. ibarrai*	*I. ibarrai*
15.256577	-90.259178	*I. ibarrai*	*I. ibarrai*
15.256577	-90.259178	*I. ibarrai*	*I. ibarrai*
14.99896	-90.807487	*I. ibarrai*	*I. ibarrai*
14.475295	-90.663773	*I. ibarrai*	*I. ibarrai*
14.527159	-90.60902	*I. ibarrai*	*I. ibarrai*
14.527159	-90.60902	*I. ibarrai*	*I. ibarrai*
14.527159	-90.60902	*I. ibarrai*	*I. ibarrai*
14.527159	-90.60902	*I. ibarrai*	*I. ibarrai*
14.527159	-90.60902	*I. ibarrai*	*I. ibarrai*
14.527159	-90.60902	*I. ibarrai*	*I. ibarrai*
14.450443	-90.584001	*I. ibarrai*	*I. ibarrai*
14.450443	-90.584001	*I. ibarrai*	*I. ibarrai*
14.450443	-90.584001	*I. ibarrai*	*I. ibarrai*
14.450443	-90.584001	*I. ibarrai*	*I. ibarrai*

15.352505	-91.313679	*I. ibarrai*	*I. ibarrai*
15.129168	-91.534854	*I. ibarrai*	*I. ibarrai*
15.316266	-91.542503	*I. ibarrai*	*I. ibarrai*
15.316266	-91.542503	*I. ibarrai*	*I. ibarrai*
15.316266	-91.542503	*I. ibarrai*	*I. ibarrai*
15.316266	-91.542503	I. ibarrai	I. ibarrai
15.316266	-91.542503	I. ibarrai	I. ibarrai
15.316266	-91.542503	I. ibarrai	I. ibarrai
15.514315	-91.870315	I. ibarrai	I. ibarrai
15.514315	-91.870315	I. ibarrai	I. ibarrai
15.514315	-91.870315	I. ibarrai	I. ibarrai
15.514315	-91.870315	I. ibarrai	I. ibarrai
15.514315	-91.870315	I. ibarrai	I. ibarrai
14.69325	-90.003305	I. ibarrai	I. ibarrai
14.61389	-89.934537	I. ibarrai	I. ibarrai
14.61389	-89.934537	I. ibarrai	I. ibarrai
14.61389	-89.934537	I. ibarrai	I. ibarrai
14.61389	-89.934537	I. ibarrai	I. ibarrai
14.531054	-90.828745	I. ibarrai	I. ibarrai
14.516881	-90.826863	I. ibarrai	I. ibarrai
14.516881	-90.826863	I. ibarrai	I. ibarrai
14.516881	-90.826863	I. ibarrai	I. ibarrai

Figure S1. Population assignment for 64 individuals based on *(a)* hierarchical Bayesian population clustering using STRUCTURE and *(b)* maximum-likelihood population clustering using ADMIXTURE. The spatial distribution of populations is presented with the same color scheme as figure 1b in the main manuscript.
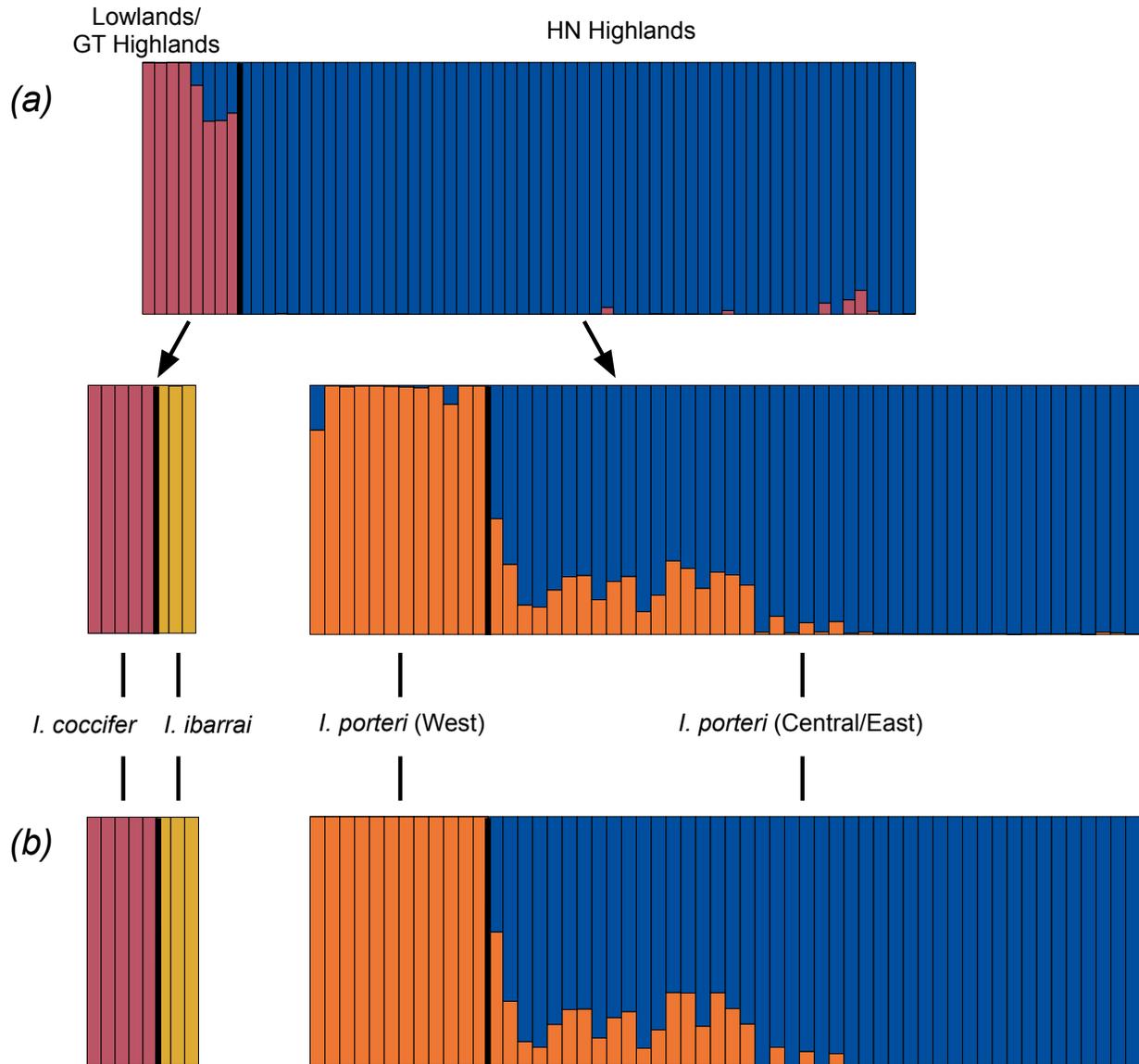
Figure S2. Posterior distribution of genealogical divergence indexes (*gdi*) estimated from the

BPP parameters.
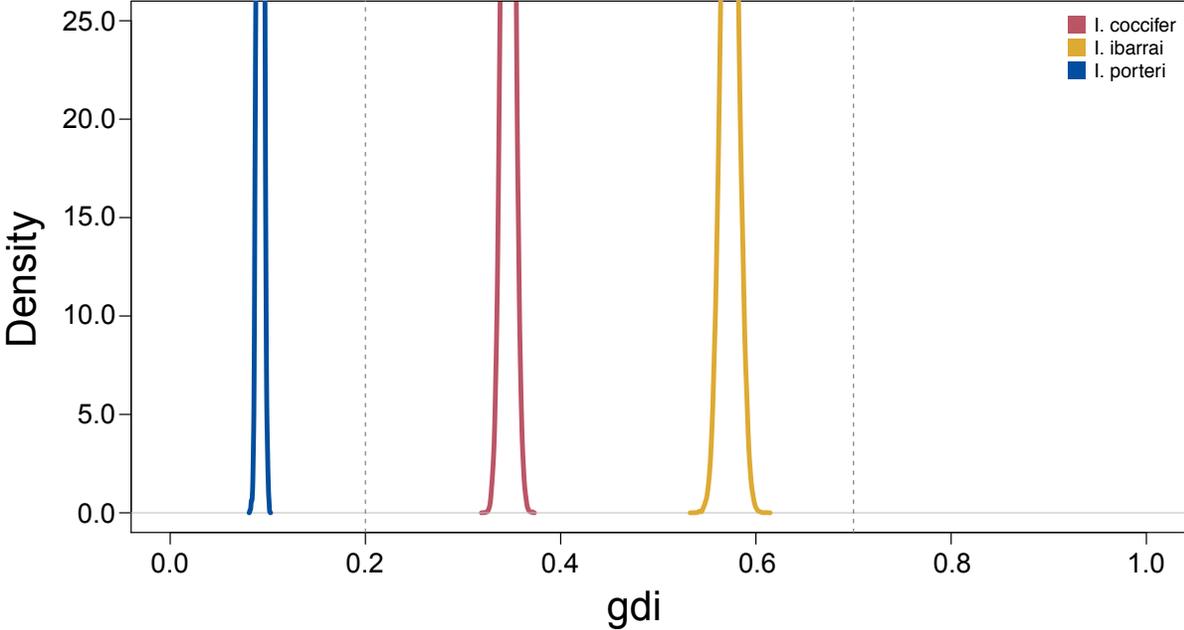
Figure S3. Heatmaps of (*a*) bioclim variable and (*b*) ENVIREM variable correlation.
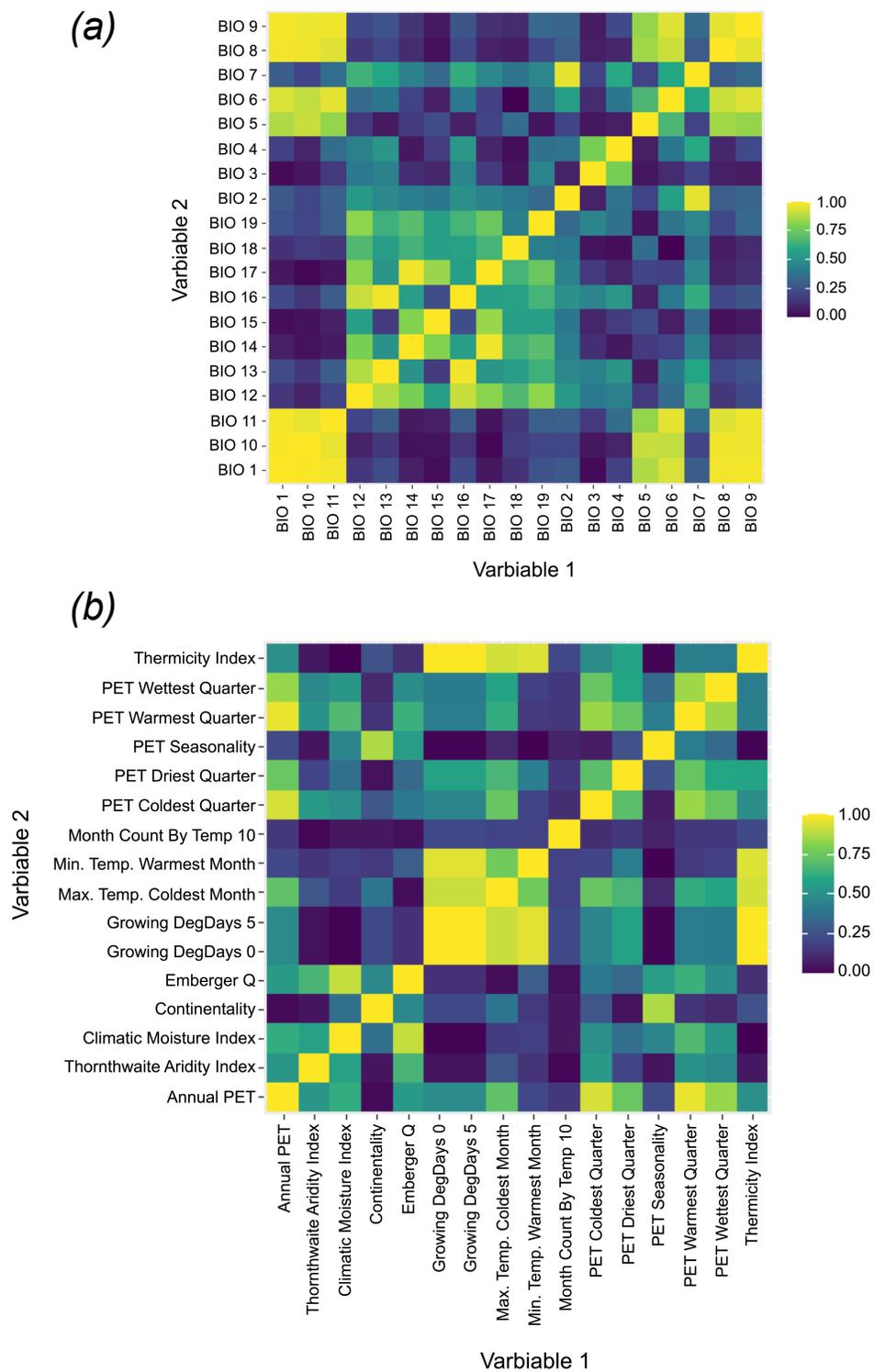
Figure S4. Continuous outputs from the species distribution models of *I. coccifer*, *I. ibarrai*, and *I. porteri* based on their respective genetic datasets. Warmer colors indicate areas of high probability of occurrence, whereas cooler colors indicate a lower probability of occurrence.
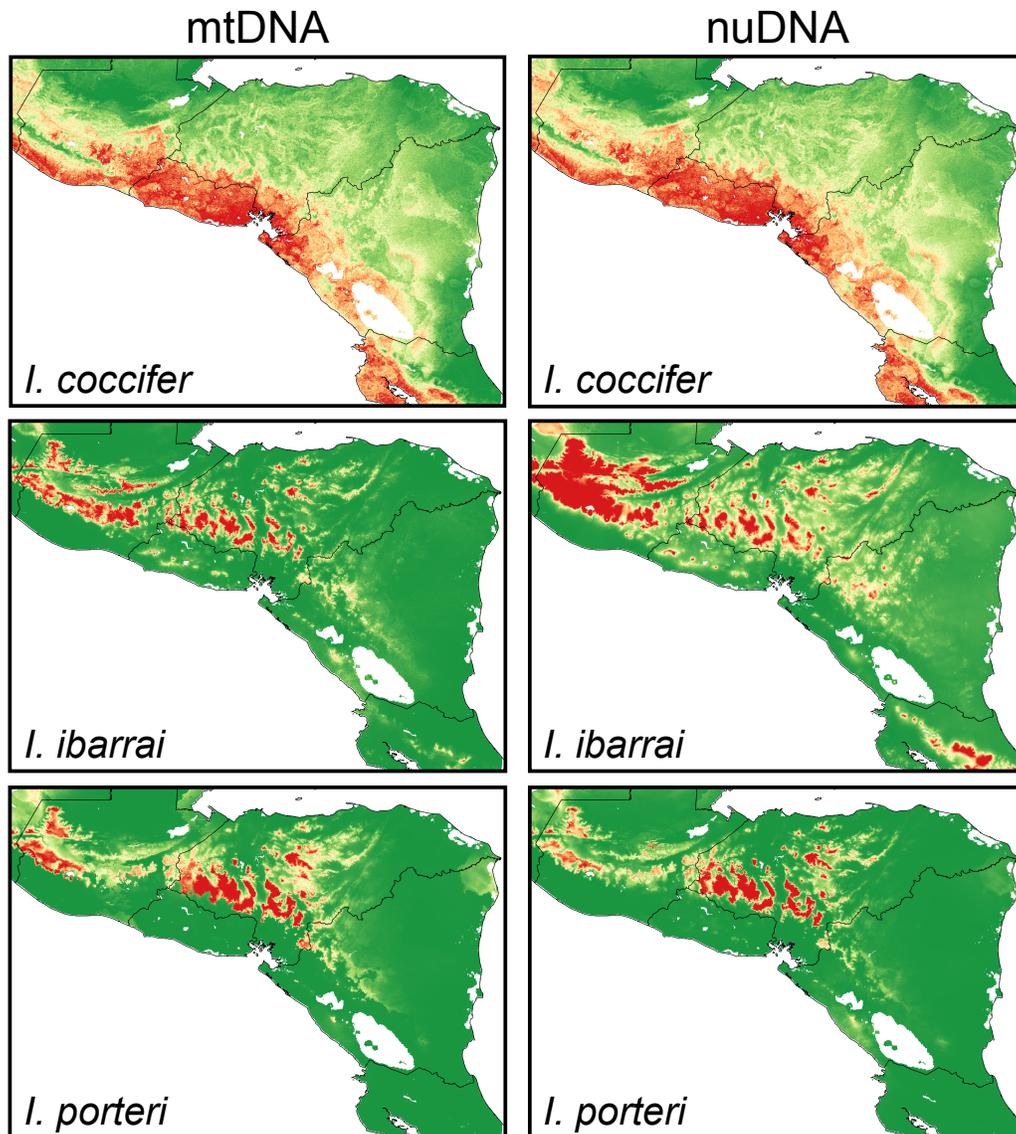
Figure S5. Map of the proposed distributions for the focal taxa (*I. coccifer* = red; *I. ibarrai* = yellow; *I. porteri* = blue) with sampling localities that correspond to our genetic, morphological, and macroecological modeling analyses. Dashed lines represent the borders of the Chortís Block, which serve as biogeographic barriers to highland taxa *I. ibarrai* and *I. porteri*.
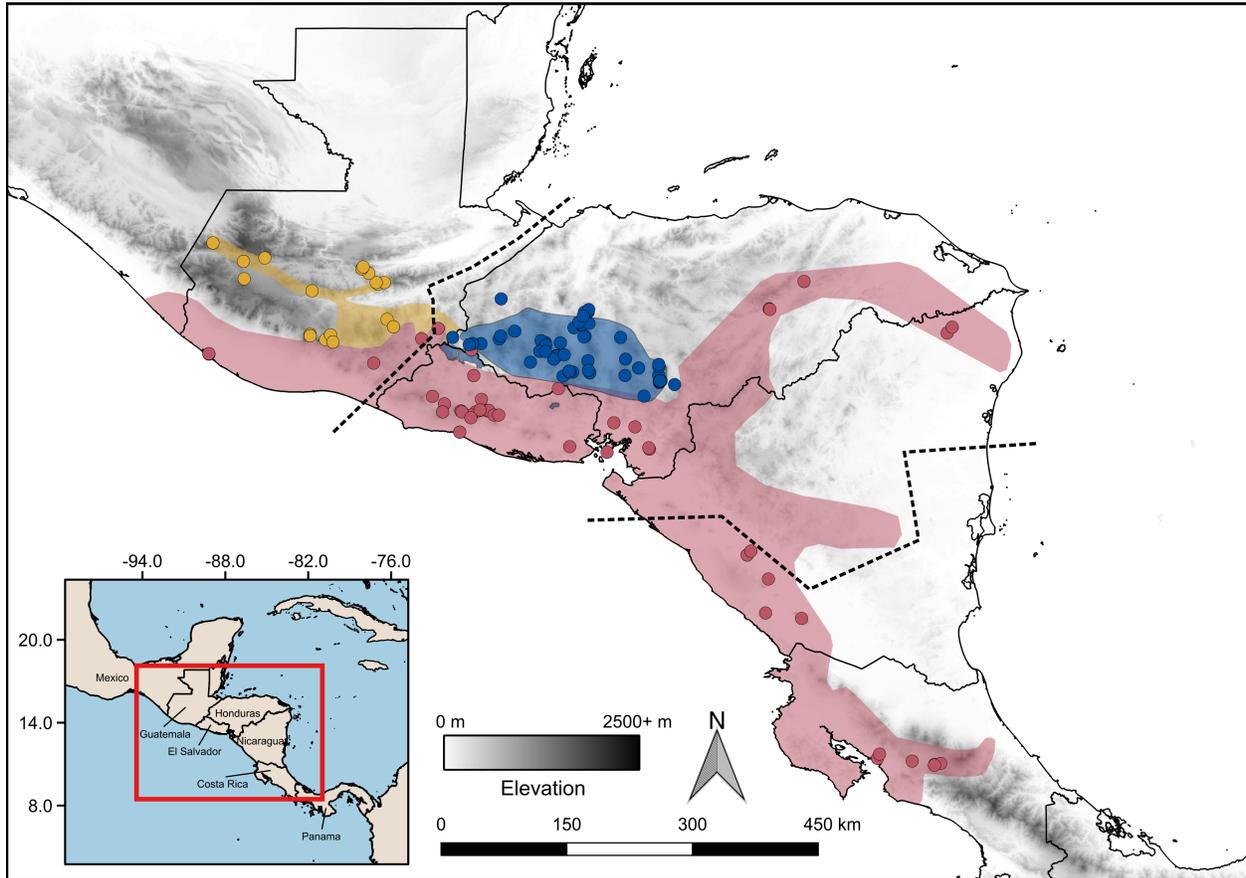
CHAPTER 4:

TRANSCRIPTOMIC ANALYSIS REVEALS POTENTIAL CANDIDATE PATHWAYS AND GENES INVOLVED IN TOXIN BIOSYNTHESIS IN TRUE TOADS

Thomas J. Firneno, Jr.[1,2,*], Balan Ramesh[1], Jose A. Maldonado[1,2], Alejandro I. Hernandez-Briones[1], Alyson H. Emery[1], Corey E. Roelke[1,2], and Matthew K. Fujita[1,2]

[1]*Department of Biology, University of Texas, Arlington, Texas, 76019-0498, USA*

[2]*Amphibian and Reptile Diversity Research Center, Department of Biology, University of Texas, Arlington, Texas, 76019-0498, USA*

ABSTRACT

 Synthesized chemical defenses have broadly evolved across countless taxa and are important in shaping evolutionary and ecological interactions within ecosystems. However, the underlying genomic mechanisms by which these organisms synthesize and utilize their toxins are relatively unknown. Herein, we use comparative transcriptomics to uncover potential toxin synthesizing genes and pathways, as well as interspecific patterns of toxin synthesizing genes across ten species of North American true toads (Bufonidae). Upon assembly and annotation of our ten transcriptomes, we explored patterns of relative gene expression, expression divergence, and possible protein-protein interactions across our species to determine what genes and/or pathways may be responsible for toxin synthesis. We also tested our transcriptome dataset for signatures of positive selection to aid in knowing how selection may be acting upon potential toxin producing genes. We were able to produce high quality transcriptome assemblies and were able to find several genes and pathways that may be involved in toxin synthesis pathways. However, we did not find any specific patterns of expression or expression divergence across our species, nor did we find any potential toxin producing genes exhibiting signatures of positive selection. We hope that our study lays a foundation for future studies to explore the genomic underpinnings and specific pathways of toxin synthesis in toads, as well as at the macroevolutionary scale across numerous taxa that produce their own defensive toxins.

INTRODUCTION

Chemical defenses are an important evolutionary trait that has evolved throughout the tree of life (Dumbacher et al. 2002; Mebs 2001; Wortman-Wunder and Vivanco 2011; Zhang 2015). Organisms that utilize these defenses can synthesize their own chemicals, sequester them from their environment (e.g. diet), or exploit mutualisms with toxic symbionts (Noguchi et al. 1986; Daly et al. 1986; Laurent et al. 2005). These organisms often have complicated evolutionary histories that may require unique morphological or physiological adaptations, ecological conversions, and multiple origins of toxin acquisition and self-resistance (Mebs 2001; Tarvin et al. 2016). Furthermore, these chemical defenses can play a large role in shaping the ecosystems in which the organisms that use them inhabit (Phillips and Shine, 2004; Bidart-Bouzat and Imeh-Nathaniel 2008; Crossland et al. 2008; Ujvari et al. 2015). Many organisms that synthesize their own defensive toxins (as opposed to venoms) do so in the form of small, secondary metabolites or small peptides (Mebs 2001; Agrawal et al. 2012; Speed et al. 2012). The knowledge of the chemical ecology of these defensive toxins is becoming more prevalent and known in a variety of organisms (Agrawal et al. 2012; Saporito et al. 2012; Rodríguez et al. 2017; Bókony et al. 2019). However, the underlying genomic mechanisms that are responsible for their synthesis are still relatively unknown or just beginning to be understood. Though, strides have been recently been made to begin to understand these genomic underpinning in plants that use chemical defenses. One of the first steps to understanding the genomic mechanisms of toxin synthesis is to identify those genes or pathways involved in toxin metabolism. Our primary aim toward understanding the underlying mechanism of synthesized chemical defenses is to identify potential underlying genes or pathways that may be involved in

the synthesis of toxin metabolites in a relatively understudied group of organisms and one of the few vertebrate taxa that synthesize their own defensive toxins: true toads (Anura: Bufonidae).

Many amphibians employ some form of chemical defense including toxic alkaloids (e.g. *Taricha*, Dendrobatidae, *Mantella*, *Atelopus*; Fuhrman et al. 1969; Garraffo et al. 1993; Hanifin et al. 1999; Daly et al. 2005), cardiogenic steroids (e.g. bufadienolides in Bufonidae; Siperstein et al. 1957; Rodríguez et al. 2017), antimicrobial peptides (e.g. opioid peptides in *Phyllomedusa*; Erspamer et al. 1985; Raaymakers et al. 2017), and/or proteolytic enzymes (e.g. *Aparasphenodon*; Jared et al. 2015). These toxins are often found on or stored within the skin of the amphibians that use them. A majority of amphibians that employ these defensive toxins sequester them from their diet (e.g. dendrobatid frogs; Neuwirth et al. 1979; Saporito et al. 2012). However, some amphibians synthesize their own toxins through endogenous metabolic processes (e.g. bufonid toads; Siperstein et al. 1957; Cei et al. 1968). While the study of sequestered toxins in amphibians has boomed, especially within the realm of understanding the genomic basis of sequestration and self-resistance (Saporito et al. 2012; Savitsky et al. 2012; Tarvin et al. 2016, 2017), the underlying genomic mechanisms of synthesized defensive toxins in amphibians still remains a mystery, most likely in large part due to the complexity of trying to uncover the metabolic pathways responsible for these synthesis reactions.

True toads (Anura: Bufonidae) synthesize their own suite of chemical defenses that can include cardiac glycosides (e.g. bufadienolides/bufotoxins), alkaloids and biogenic amines (e.g. bufotenine), and/or antimicrobial peptides (reviewed in Rodríguez et al. 2017). These chemicals are secreted through the skin and are primarily concentrated within the parotoid gland that is located posterior to the eye (Hostetler and Cannon, 1974; Cannon and Hostetler, 1976). These toxins have a variety of deterrent effects on predators including causing a distastefulness of the

defended organism, hallucinogenic or neurological issues, or inducing cardiac arrest and death (Erspamer et al. 1965; Lichtstein et al. 1986). While these toxins are used for defense by the animal, these toxins have been used for medicinal and pharmacological purposes around the world, showing signs of being effective treatments for chronic pain, heart disorders, and cancer (reviewed in Rodriguez et al. 2017). Differences in the composition of these suites of toxins have been found at the levels of genus and species (e.g. through electrophoretic profiling of toxin secretions; Sciani et al. 2013). However, it is still not understood by what metabolic pathways and what underlying genomic mechanisms these toxins are synthesized by bufonids.

Here we utilize RNA sequencing of the parotoid gland from ten North American bufonid species to identify potential genes and/or pathways that may be involved in toxin synthesis by (1) assembling the parotoid gland transcriptomes of our ten focal species, (2) annotating and characterizing the parotoid gland transcriptomes, (3) identifying gene and pathway expression, expression divergence, and protein-protein interaction patterns within and across each species, and (4) if genes are under positive selection which may indicate their evolution for toxin synthesis. From prior literature on the toxins metabolites that bufonids synthesize, we know that they produce and possess steroid-derived cardiac glycosides (bufadienolides), tryptophan derived tryptamines (bufotenins), and other potential metabolites that may make the toad distasteful. Therefore, we hypothesize that: (1) we will observe genes and/or pathways involved in steroid biosynthesis/degradation/metabolism, tryptophan metabolism, or other biosynthetic/metabolic pathways that generate metabolites that cause distastefulness of the organisms being more highly upregulated or expressed within our species; and (2) that we will see certain species that are known to generate higher levels of certain toxin metabolites (e.g. *Incilius alvarius* synthesizes

higher levels of bufotenins) more highly express genes involved in these pathways than other species.

MATERIALS AND METHODS

Taxon sampling, tissue collection, and molecular data acquisition

Ten samples were collected from ten different bufonid species across North America in Summer 2018. Toads were caught by hand and then had their parotoid glands expressed to empty the parotoid gland. They were then kept in isolated containers for 48 hours to allow them to recover their toxins and express genes involved in that recovery. Toads were then humanely euthanized under our IACUC approved protocol. We harvested the parotoid glands for all specimens and stored the tissues in RNAlater (Sigma-Aldrich, St. Louis, MO). We preserved the specimens with 10% formalin and deposited them in the Amphibian and Reptile Diversity and Research Center (ARDRC) at The University of Texas at Arlington (UTA) (Supplementary Table S1).

We extracted total RNA from parotoid gland tissue of 2-3 replicates for each species (27 samples total) using a Promega SV Total RNA Isolation kit (Promega, Madison, WI) following the manufacturer's protocol. We quantified RNA extractions on a Qubit 2.0 Flurometer (Life Technologies, Carlsbad, CA) and assessed RNA quality and size distribution on an Agilent 2100 Bioanlayzer (Agilent Technologies, Santa Clara, CA). We then prepared complimentary DNA (cDNA) libraries using a KAPA mRNA Hyper Prep kit (Roche, Indianapolis, IN), indexed each sample, and multiplexed twelve samples to a pool. Pools were then sent to NovoGene for sequencing on a HighSeq4000 (Ilumina, San Diego, CA) for 150bp paired-end reads or to the North Texas Genome Center (https://northtexasgenomecenter.com) for sequencing on a NovaSeq6000 (Ilumina, San Diego, CA) for 100bp paired-end reads (Table S1).

Transcriptome assembly and quality assessment

The raw reads were processed to trim barcode sequences and remove low quality reads using Trimmomatic v0.39 (Bolger et al. 2014) before being assembled. We used a 4-base sliding window, trimming nucleotides with a Q score <5 and discarding reads <25 base pairs (bp) long to trim the raw reads and remove failed reads (Macmanes 2014). To evaluate read quality, length, and number of reads retained, we processed the filtered and trimmed reads on a per sample basis using FASTQC v0.10.1 (Babraham Bioinformatics). We carried out *de novo* assemblies of each sample using Trinity v2.9.1 (Grabherr et al. 2011; Haas et al. 2013) by combining the data from species replicates to create more complete transcriptomes per species.

We used Benchmarking Universal Single Copy Orthologs (BUSCO) (Simäo et al. 2015) and tetrapoda lineage (odb10) to assess the quality and completeness of our *de novo* assemblies. To improve the quality of the assemblies and to reduce sequence duplication, we clustered sequences that are 95% similar using CD-Hit v4.8.1 (cd-hit-est; Li and Godzik 2006) and then reran BUCSO to reassess determine if the quality and completeness of our assemblies improved after removing redundant transcripts. We used these clustered sequences to predict open reading frames and conserved single copy orthologs among our ten species as detailed in the steps below.

Identification of orthologs

We identified candidate coding genes from our assembled transcript sets for all ten bufonid species using the TransDecoder v3.0.1 pipeline (Haas et al. 2015). Though the default length for opening reading frame (ORF) prediction is set to ≥100 amino acids in TransDecoder to reduce the false discovery rates (Haas et al. 2015), we set our ORF length to ≥40 amino acids due to the potential presence of antimicrobial peptides in our samples, which can range between 12-50 amino acids (Simmaco et al. 1998; Conlon 2011; Ladram and Nicolas, 2016). To

maximize the number of ORFs captured and ensure that we did not lose any potential coding genes that did not meet the TransDecoder 40 amino acids long criteria, we used the TransDecoder optional homology search against the PFAM database to increase the sensitivity of detecting ORFs that have functional importance (Finn et al. 2016).

We used OrthoFinder v2.2.3 (Emms and Kelly, 2015) to identify orthologous genes between (1) all ten bufonid species and (2) all ten species, plus an outgroup (*Nanorana parkeri* (PRJNA344660)) from the protein coding gene sets generated by TransDecoder. These orthologous gene sets were used for (1) gene expression analyses and (2) tests of gene family size change and selection, respectively (detailed below).

Transcript annotation, gene expression, and expression divergence

We annotated and assigned functions to our complete protein coding gene set for each species (from TransDecoder) using the Blast2Go v5 pipeline (Götz et al. 2008). We ran the Blastp function to blast our transcript set against the NCBI NR-protein database using an e-value cutoff of 1e-5 and all other parameters set to default. Annotation was performed using an e-value cutoff of 1e-3, annotation score of 45, and a GO weight of 5. To generate functional annotations (biological processes, molecular functions, and cellular components), we first grouped our GO annotation into GO-Slim terms to simplify the input and filtered out nodes containing >10 sequences. To determine metabolic pathways, we used transcript sequences with corresponding enzyme codes (ECs) obtained from Blast2GO and mapped them to the Kyoto Encyclopedia of Genes and Genomes (KEGG) metabolic pathway database (Kanehisa and Goto, 2000; Kanehisa, 2019).

We used the candidate coding sequences from TransDecoder for each species as a reference transcriptome to quantify expression using salmon v1.1.0 (Patro et al. 2017)

accounting for GC content and sequence specific biases in the reads. To obtain normalized expression across species we used DESeq2 v1.26.0 (Love et al. 2014) and edgeR v3.28.1 (Robinson et al. 2010; McCarthy et al. 2012). Briefly, the estimated raw counts of the genes from DESeq2 were passed to edgeR to normalize for library size and the length of the transcripts. Using the normalized expression and conserved single copy orthologs, we identified transcripts involved in various metabolic pathways and annotated using Blast2GO and KEGG analyses outlined above. With a minimum expression threshold of >1 TMM, we visualized the expression pattern of different genes and pathways removing the effect of lowly expressed genes.

To determine expression divergence across our ten bufonid species, we determined which genes/transcripts demonstrated species-specific expression shifts. To do this we used the expression quantification counts generated by salmon of the 1-to-1 orthologs across our ten bufonid species. These expression profiles were then combined across species, filtered for low abundance (<10 average counts) and rlog normalized in DESeq2. We quantified species-level expression shifts using the R package EVE using a β shared test (Rohlfs & Nielsen, 2015). Briefly, this program identifies orthogroups with both high levels of between species expression variance represented by low β values, which indicate divergent shifts in expression, versus orthogroups with high levels of within species expression variance represented by high β values, which indicate orthogroups with evolutionary conserved expression profiles. We -log10 transformed our β values to generate our expression divergence metric. We used our Blast2GO and KEGG annotations to determine what patterns of expression divergence any potential toxin producing genes may exhibit.

Protein-protein interaction network construction and functional enrichment analysis

145

We used the Search Tool for the Retrieval of Interacting Genes/Proteins (STRING v11.0; https://string-db.org/) to construct a protein-protein interaction (PPI) network among all expressed genes that met our TMM filtering criteria. Interaction sources selected for generation of the PPI network included text mining, experiments, databases, co-expression, neighborhood, gene fusion, and cooccurrence. PPIs that possessed at least a high confidence score of ≥0.700 were considered for network generation. Functional enrichment analysis was also carried out by STRING to identify significantly enriched KEGG pathways within our network of expressed genes set.

Tests for positive selection

To identify genes that may be under positive selection in our toad species and therefore potentially being involved in toxin synthesis, we estimated dN/dS in all 881 1-to-1 orthologous genes found by OrthoFinder across our ten bufonid species and *Nanorana parkerii* following the recommendations of Yang (2006). We used the branch-sites model (codeml: M2 and NSites2) from the PAML v4.9 package (Yang 2007). Because this analysis requires an *a priori* phylogenetic tree to test for selection in foreground branches, we used a tree generated by OrthoFinder, which merges orthologous gene trees into a species tree. We designated the Bufonidae lineage (our ten focal taxa) as foreground branches and the *Nanorana* branch as a background branch. We compared the likelihood of two different models for each orthologous gene: (1) a simplistic null model that has a fixed $\omega_2 = 1$ on the foreground branches, and all other branches having a $\omega_0 = 1$ and $\omega_1 < 1$; and (2) an alternative model that allows for a proportion of the sites to be under positive selection ($\omega_2 \geq 1$) along the bufonid branches, and the background branches having a proportion of sites being under purifying selection ($\omega_1 < 1$) or neutrally evolving ($\omega_0 = 1$) (Zhang et al. 2005; Yang 2007). After obtaining the likelihood values for each

orthologous gene run under both models, we used a likelihood ratio test (2ΔlnL) between the

models to evaluate whether the alternative model outperformed the null model. Finally, we

performed a Bonferroni correction to account for multiple comparisons. A significant result from

the branch-sites model indicates that a subset of the sites in that coding gene has gone through

episodic positive selection, with the selected sites providing an advantage to the bufonid lineage.

To identify what genes and pathways, may be under, we annotated the assigned orthogroups this

analysis using our GO annotations and KEGG assignments from Blast2GO. Figure 1 is an

overview of our entire methodological pipeline as described above.

RESULTS

Sequencing and *de novo* assembly quality assessment

Our sequenced cDNA libraries yielded an average total of 101,589,456 paired-end reads

between all 27 samples, and we retained 66.9% of our reads after filtering out low-quality reads

from our raw dataset. FastQC verified that only high-quality reads with a Q score >30 were kept

and assembled by Trinity. The total number of assembled transcripts was between 163,758–

342,030 for all ten species, with the average mean transcript length being ~611.159 bp.

Assembly statistics for all ten species and replicates are listed in Table 1.

To quantify the completeness of our ten transcript sets, we ran our protein coding gene

set for each assembly though CD-HIT to remove redundant transcripts from the assemblies. Our

protein coding transcript sets for our ten focal species (between 163,758-342,030) decreased by

15.3-16.2% to generate a non-redundant transcript set of 138,693-286,545 sequences. We ran our

both our redundant and non-redundant transcript sets for all ten species against a conserved set of

5,310 universal tetrapod single-copy orthologs using BUSCO. Our redundant transcript set

recovered 29-45% complete single-copy, 29-43% complete duplicated, 2-7% partial orthologs,

and 13-21% missing orthologs, whereas our non-redundant transcript set recovered 55-64% complete single-copy, 13-19% complete duplicated, 2-7% partial orthologs, and 13-23% missing orthologs (Figure S1). The number of complete conserved orthologs present in each transcript set is about average with other high-quality, single tissue transcriptome assemblies for anurans (summarized in Richardson et al., 2018).

Identification of putative orthologs

To identify the orthologous gene groups and 1-to-1 orthologs among our coding genes identified by TransDecoder between our ten bufonid species we ran OrthoFinder. TransDecoder found between 31,627–51,581 protein coding genes with the longest open reading frame from our assembled transcript sets across the ten species. The identified protein coding genes ranged from 40–7,695 amino acids in length. We submitted a total of 407,473 coding transcripts to OrthoFinder to identify orthogroups. Orthofinder identified 46,572 orthogroups (including both orthologs and paralogs) containing 284,441 transcripts. We found that 11.95% (5566) of the orthogroups had all ten individuals present, and 99.76% (46,459) had at least two individuals present (Figure S2). We found 1,454 transcripts (14,540 orthologs) that were true 1-to-1 orthologs between all ten species, and 881 transcripts that were 1-to-1 orthologs between our ten bufonid species and *Nanorana parkeri*.

Gene ontology and annotation

Out of a total of 403,590 transcripts across our ten species submitted to Blast2GO, around 43.5% were successfully blasted against the NCBI NR-protein database, with the top BLAST hit being *Nanorana parkeri* and the remaining orthologs blasting over other amphibian species (e.g. *Xenopus tropicalis*, *X. laevis*, *Lithobates catesbianus*). We successfully annotated a total of 175,533 transcripts into 31 functional categories using GO-slim assignments within the three

categories of the GO classification system: biological process, cellular component, and molecular function. The primary categories that our coding genes were clustered into were ion binding (average of 4185.7), organic cyclic compound binding (2506.9), heterocyclic compound binding (2506.9), cellular metabolic process (6784.8), nitrogen compound metabolic process (6471.8), and primary metabolic process (5507.7). An average breakdown of GO terms for each category for the ten species is shown in Figure 2 (individual assignments are shown in Figure S3). We were able to connect our annotated transcripts to an average of 141.3 KEGG pathways, representing an average of 659.4 enzymes.

Gene and pathway expression and expression evolution

We used relative expression analysis to determine which genes and pathways might be expressed across our ten bufonid species and, therefore, involved in toxin synthesis in the parotoid gland. After filtering our genes and pathways for a minimum expression threshold of ≥1 TMM, we identified 562 annotated genes (of our 1,454 1-to-1 orthologs; Figure S4) and 76 pathways that had genes that were expressed across our species (Figure 3). The genes that showed relatively high expression across all of our species included adenosylhomocysteinase (*ahcy*), hydroxyacyl−coenzyme A dehydrogenase (*hadh*), basigin (*bsg*), macrophage migration inhibitory factor (mif), L−lactate dehydrogenase B chain (*ldhb*), and diphosphomevalonate decarboxylase (*mvd.L*) (Figure S4). The pathways that showed relatively high expression across all of our species included melanogenesis, bisphenol degradation, fatty acid biosynthesis, N-glycan biosynthesis, and glucosinolate biosynthesis (Figure 3). We did not see any species-specific patterns of relative gene/pathway expression when we compared across our ten focal species.

We quantified expression patterns across our ten bufonid to demonstrate species-specific expression shifts using EVE. Overall, 28 of our 1454 1-to-1 orthologs demonstrated higher expression divergence between species than expression divergence within species ($p<0.01$) and are considered to have adaptive expression patterns (Figure 4). None of these 28 orthologs annotated to pathways or genes that might be involved in toxin synthesis.

PPI network and functional enrichment

We extracted protein-protein association data from our expressed genes within our parotoid gland transcriptome datasets. The mined PPI data was comprised of a total of 954 PPIs (Figure 5). Manual curation of the data revealed that 43 proteins out of 954 PPIs possessed ≥4 interacting partners and 17 proteins were associated with only one interacting partner. Network statistics obtained by STRING can be found in Table S2. We also identified significantly enriched KEGG pathways and the genes linked to these pathways within our network using STRING (Table S3). We found several enriched pathways that also had a number of genes with interacting partners in our network that may be involved in toxin synthesis in the parotoid gland including propanoate metabolism (7 genes in network), geraniol metabolism (2 genes), terpenoid backbone biosynthesis (4 genes), tyrosine metabolism (4 genes), tryptophan metabolism (4 genes), glucosinolate biosynthesis (1 gene), steroid hormone biosynthesis (1 gene), and general metabolic pathways (110 genes). Several of the genes associated with these pathways were also found to be interacting partners with each other in our network (e.g. *EHHADH* and *hadh* of the geraniol metabolism pathway interact with *ldh*, *hadha*, *suclg2*, and *acads* of the propanoate metabolism pathway and *bact2* of the glucosinolate biosynthesis pathway) and also seem to have proposed interaction associations with a number of other metabolic pathway genes (Figure 5).

Tests for positive selection

To determine if any of our 1-to-1 orthologous genes had undergone positive selection we carried out selection tests in PAML and performed likelihood-likelihood tests to assess significance. We also adjusted out p-values by performing a Bonferonni correction test to account for multiple comparisons and committing a type I error. We found 66 genes from our 1-to-1 ortholog set that had undergone positive selection. However, after our p-value adjustment, we found 4 genes with a significance value < 0.05. Table 2 has a complete list of these genes with the null and alternative likelihood values, unadjusted and adjusted p-values, and the Bayes Empirical Bayes (BEB) scores for sites under selection that were ≥0.9.

DISCUSSION

The primary aim of our study was to lay a foundation to begin to examine defensive toxin evolution both within bufonid toads and broader taxonomic organisms in a genomic context by identifying potential genes and/or pathways that may be involved in toxin biosynthesis in these animals. Our results have revealed several genes and their respective pathways as being expressed across a majority of our species, indicating the possibility that they are involved in the biosynthesis of these toxin metabolites or peptides. However, we do not see any relative patterns of gene expression or adaptive expression divergence across our ten focal species, nor do we see any potential toxin producing genes showing any signatures of positive selection. Below, we summarize our findings and discuss the influence of this study in a broader evolutionary context of studying the genomics of synthesized defensive toxins.

Transcriptomics of the bufonid parotoid gland

Currently, there is a relative lack genomic or transcriptomic resources for the family Bufonidae compared to other amphibian species, with the exception of the *Rhinella marina* draft genome and some transcriptomic studies where bufonids are mostly used secondarily to the focal

taxa (Gerchen et al. 2016; Pastenes et al. 2017; Ceschin et al. 2020). There are even fewer resources in relation to comparative studies across bufonid species and with respect to synthesized toxin evolution in these organisms. Our study offers the first comprehensive and comparative look at the bufonid parotoid gland transcriptome across ten North American bufonid species with the goal of identifying potential underlying genes and/pathways involved in toxin metabolite biosynthesis.

We were able to generate high quality transcriptomes for our ten focal species as indicated our assembly statistics and BUSCO results. These metrics are able to serve as a guide to understand and evaluate sources of error in our assemblies and provide evidence for the quality of our assembled transcriptomes. We also assessed the assemblies by implementing the hierarchical clustering tool CD-HIT to address the possible generation of chimaeras, redundant transcripts, and fragmented assemblies that are common when carrying out *de novo* assemblies (Fu et al. 2012). The BUSCO assessments of our assemblies showed the number of our complete single-copy orthologs increase from between 29-45% to 55-64% and our complete duplicate orthologs decrease from between 29-43% to 13-19%, representing between 70.6-83.0% completeness. This indicates that CD-HIT improved our transcriptome assemblies as indicated by the improvement in quality and completeness scores.

In terms of quality and completeness, our assemblies are comparable to other studies that use a single or select number tissues (Theissinger et al. 2016) and *de novo* assembly methods (Grabherr et al. 2011; De Oliveria et al. 2016; Theissenger et al. 2016). As far as we are aware, our study is the first to generate transcriptome assemblies from these bufonid species, as well as from parotoid gland tissue. Future studies that may want to explore toxin evolution in bufonids using RNA-seq data might want to include other tissues that may be responsible for generating

these toxins (e.g. liver) and across different time points for more in depth expression profiles over time (discussed more below).

Gene expression and protein-protein interactions identify potential genes and pathways that may be involved in toxin synthesis within species

As previously mentioned, bufonid toads synthesize a large suite of toxins consisting primarily of metabolite alkaloids, biogenic amines, steroids, as well as some peptides and proteins (e.g. antimicrobial peptides) (reviewed in Rodriguez et al. 2017). While there is a lack of resources for the underlying genomic mechanisms of the synthesis of these toxins in toads, there have been relatively great strides made in plants that synthesize similar toxins (Agrawal et al. 2012; Pandey et al. 2016). Within our study we attempted to identify potential candidate genes/pathways that may be involved in toxin synthesis in toads by (1) identifying genes/pathways that are expressed and may be interacting with each other in the parotoid gland, (2) identifying any patterns of gene expression between our ten focal bufonid species from our parotoid gland transcriptomes, and (3) identifying genes that may be under positive selection.

Alkaloids and biogenic amines comprise a relatively large portion of the suite of toxins synthesized by bufonids (Zhang et al. 2005; Saporito et al. 2012; Rodriguez et al. 2017). These alkaloids can include guanidine alkaloids (found only in *Atelopus*; Fuhrman et al. 1969; Daly et al. 1994; Yotsu and Tateki, 2010), lipophilic alkaloids (found in *Melanophryniscus*; Daly et al. 1986; Saporito et al. 2012), and indole alkaloids (found broadly across Bufonidae, but includes our two focal genera *Anaxyrus* and *Incilius*; Erspamer et al. 1965; Cei et al. 1968). We found four pathways that were expressed at low to high in our transcriptomes that may be implicated in alkaloid and biogenic amine biosynthesis: (1) isoquinoline alkaloid biosynthesis (*bsg*, *Ddc*), (2) tryptophan metabolism (*ap3s1*), (3) tyrosine metabolism (*hgd.L*), and (4)

chloroalkane/chloroalkene degradation (*Lta4h*, *dhrs9*, *adh1*) (Figure 3, S5). Isoquinoline

alkaloids are primarily found in several families of plants (Berberidaceae, Fumariaceae,

Menispermaceae, Papaveraceae, and Ranunculaceae) and generally have sedative, psychotropic

and/or analgesic properties (Phillipson et al. 2012; Matsuura and Fett-Neto 2015; Srivastava et

al. 2015). Both tryptophan and tyrosine metabolism have been linked to the synthesis of

tryptamines and are most likely involved as precursors or primary metabolites in the synthesis of

those tryptamines found in bufonid toad toxins. One of the most commonly known tryptamines

found in bufonid toxin secretions is 5-methyoxy-*N*,*N*-dimethyletryptamine (5-MeO-DMT) from

Incilius alvarius (Erspamer et al. 1965). While we expected to see tryptophan metabolism to be

more highly expressed in *Incilius alvarius* because of this, we do not see this pattern compared to

the other nine species (Figure 3). Finally, chloroalkane/chloroalkene are common

organochlorides that may be involved in the biosynthesis of alkaloid or steroid based toxins in

bufonids. A common toxic chlorinated alkaloid is epibatidine, which is found in the skin of

Epipedobates poison frogs (Daly et al. 1986). The chloroalkane/chloroalkene pathway being

upregulated in our focal species could indicate that these species produce some organochloride

alkaloids or analogs to them.

In addition to alkaloids and biogenic amines, steroids comprise the other large portion of

the suite of toxins in bufonids. A great diversity of these steroids (the most common being

bufadienolides and cardenolides) have been found across many taxa within Bufonidae (with the

exception of some species of *Melanophryniscus*) (Mebs et al. 2007; Meng et al. 2016; Rodriguez

et al. 2017; Botelho et al. 2019) and are very similar in both structure and function to those found

in several plants (e.g. digoxin, digitoxin, ouabain, oleandrin; Newman et al. 2008; Botelho et al.

2019). These bufadienolides and cardenolides often act as ouabain binding inhibitors in the

Na$^+$/K$^+$-dependent ATPase pump (Flier et al. 1980). While primarily being found within the skin and parotoid glands, these steroids can also be present within the ovaries, eggs, bile, and larval life stages (Lichtstein et al. 1992; Lee et al. 1994; Matsukawa et al. 1998; Hayes et al. 2009). Cholesterol has been identified as the precursor for the synthesis of these steroids in toads and plants (Siperstein et al. 1957; Wickramasinghe et al. 1969); however, the route by which they are biosynthesized is believed to be different (Porto and Gros, 1970; Porto et al. 1972; Stuhlemmer and Kreis 1996; Zhong et al. 2020; Pandey et al. 2016). We found four pathways that were expressed at low to high levels in our transcriptomes that may be implicated in steroid toxin (bufadienolide) biosynthesis: (1) geraniol metabolism (*EHHADH*, *Hadh*), (2) terpenoid backbone biosynthesis (*icmt*, *PCYOX1L*, *hoxb2*, *ggps1*, *mvd.L*, *pdss2*), (3) steroid biosynthesis (*bpmer*, *unc93A*), and (4) steroid hormone biosynthesis (*TLCD3A*, *arsa.1*) (Figure 3, S5). Enzymes related to terpenoid backbone biosynthesis and geraniol metabolism (known to be important in the biosynthesis of terpenes) have been identified in plant cardenolide biosynthesis and have been suggested to potentially be involved in the formation of precursor or intermediate metabolites (specifically the formation of squalene) in that pathway (Pandey et al. 2016). Though we do not know if terpene backbone biosynthesis is involved in bufadienolide biosynthesis in toads, we do see these pathways upregulated in the parotoid gland transcriptome and could be important to the synthesis of cholesterol or other steroid precursors or intermediates. As we expected/hypothesized, we also see genes/pathways involved in steroid and steroid hormone biosynthesis upregulated. However, we do not see them as highly expressed as we would have expected seeing as steroids compose a large portion of bufonids toxins. This was most likely due to waiting too long post-secretion extraction to harvest the parotoid gland.

155

While alkaloids and steroids seem to make up a majority of bufonid toxins and are often the focus of most biochemical studies concerning these toxins (Rodriguez et al. 2017), there may also be other metabolic compounds that have not been identified in these studies but may be distasteful or odiferous and function in deterring predation. We identified three pathways that were expressed at moderate to high levels in our transcriptomes that may be implicated in these deterant properties: (1) phenylpropanoid biosynthesis (*sf3b1*, *zfp91*, *duox2*), (2) glucosinolate biosynthesis (*bcat2*), and (3) propanoate metabolism (*ldhb*, *hadha*, *suclg2*, *acads*, *epn3*) (Figure 3, S5). Phenylpropanoid synthesis is relatively widespread throughout the plant kingdom for several functions, one of which is defense against herbivores and pathogens (Dixon et al. 2002; Hossain et al. 2018). Glucosinolates are natural components of many plants (e.g. Brassicaeae, Capparaceae, Caricaceae) that are responsible for their bitter or pungent flavor and are used by the plants as deterrents against pests (Shroff et al. 2008; Sønderby et al. 2010). Finally, propanoates are esters derived from propionic acid, which is often produced by some bacteria and plant volatiles, and causes bad smells (e.g. *Propionibacterium* causing body odor). It is possible that these metabolic compounds, or analogs to them, are responsible for the bitter taste and odd odors that are attributed to bufonid parotoid secretions (Chen and Chen 1933; Mailho-Fontana et al. 2014).

Of genes that were not mapped to a metabolic pathway, one gene that was of interest was a cysteine-rich secretory protein (CRISP). CRISPs are glycoproteins that largely associated with the function of the mammalian reproductive system but are also found in a variety of snake venoms. They typically target and block calcium channels and reduce the function of potassium-induced smooth muscles contractions (Hill and Mackessy, 2000; Yamazaki and Morita, 2004).

Overall, we were able to identify several pathways and their respective genes that may be involved in the biosynthesis of toxin metabolites in the bufonid parotoid gland, as well as their potential interactions with genes in similar or potentially toxin producing pathways. We were not able to identify or annotate any possible antimicrobial peptides from our transcriptomes. We also expected to see some genes/pathways expressed more highly than they were in our data (e.g. steroid biosynthesis, tryptophan metabolism); however, we did not see this in either our relative expression results or our expression divergence analysis. This may be due to a number of factors including: (1) only using a single tissue (the parotoid gland), when some of these toxins may be produced elsewhere in the body and then transported to the parotoid gland/skin (Junqueira-de-Azevedo et al. 2015; Macrander et al. 2016; Surm et al. 2019); (2) only using a single timepoint post-secretion extraction, and this timepoint (48 hours) being too long after extraction to capture the upregulation or expression of genes that are involved in toxin synthesis (Lau and Sattely 2015; Hodgson et al. 2019); and/or (3) toxins/toxin production differs at the level of the population or individual, and we are unable to capture the true expression of genes/pathways due to differences between the individuals we sampled (Sciani et al. 2013; Amazonas et al. 2018).

Tests for signatures of positive selection indicate that potential toxin producing genes may not be undergoing positive selection

There are several pervading and tested hypotheses of how selection acts upon and how offensive toxins and their composition have evolved through gene duplications and losses toward dietary specialization (Mebs 2001; Casewell et al. 2013; Holding et al., 2016). Defensive toxins, however, lack the same tested hypotheses for how they and their variation have evolved throughout multiple kingdoms in the tree of life. This is primarily due to the complexity of underlying biochemistry and physiology required for the acquisition (both acquired and

synthesized) and maintenance of defensive toxins (Mebs 2001; Zhang et al. 2015), as well as their taxonomic diversity and intraspecific variation throughout the tree of life (e.g. animals, plants, fungi, bacteria; Mebs 2001; Zagrobleny et al. 2008; Jousset et al. 2009; Speed et al. 2012; Zhang et al. 2015). Using PAML, we assessed if any potentially toxin producing genes may be under positive selection in the bufonid parotoid gland in order to broadly infer what genes/pathways may be implicated in toxin biosynthesis in toads.

We tested our 1-to-1 ortholog dataset (with an outgroup) using PAML to determine if we could see any signatures of positive selection among any potential toxin producing genes. We only found four genes that were under positive selection within this dataset, all of which do not seem to belong to pathways or are genes that seem to be involved in toxin synthesis. One gene that we identified as being under positive selection, apolipoprotein O (*APOO*), may be of interest due to its involvement in the storage and metabolism of lipids. Though not directly involved in toxin synthesis, the parotoid gland secretion of toads has been described as being highly concentrated with lipids, which may be involved in the storage and maintenance of these toxin metabolites (Mailho-Fontana et al., 2018). Therefore, apolipoproteins, such as *APOO* could be involved in the storage and maintenance of these toxins within the bufonid parotoid gland. However, we are not sure why this gene would be under selection when other potential toxin producing genes are not under selection.

There are a number of potential caveats to why we did not find many genes or any toxin producing genes that exhibited signatures of selection. The first having to do with the quality of our dataset including: (1) we did not capture any potential toxin producing genes within our 1-to-1 ortholog dataset and, therefore, they were not included in the actual tests for selection; (2) the time interval at which we extracted the parotoid gland did not allow for the identification of toxin

producing genes; and/or (3) the quality of our assemblies somehow skewed our data and did not allow for the identification of toxin genes or those that may be undergoing positive selection (though this caveat is more detrimental for giving false positives of genes undergoing positive selection). Aside from data quality, there may be some evolutionary explanations to our lack of genes under positive selection, including: (1) that toxin producing genes are under purifying selection in that these genes have such a strong role in deterring predation in toads that changing them in some way would be deleterious or detrimental; (2) that selection is somehow acting on the composition of defensive toxins to maximize the deterrence of all possible enemies (Sime, 2002; Gross et al., 2008; Speed et al., 2012); and/or (3) that the quantitative variation in toxin levels amongst prey is cost free and their defensive effects saturate after some threshold, and that above this threshold differing phenotypes would be selectively neutral (equally good at deterring predation) (Speed et al., 2012).

Using a transcriptomics, functional annotation, relative expression and expression divergence, and quantitative tests for selection, we were able to identify potential genes and pathways that may be involved in the biosynthesis of metabolic defensive toxins in bufonid toads. We hope that this study can be used as a foundation or guide for future work focused on specific metabolic pathways and the underlying genomic mechanisms that are responsible for toxin metabolite production in bufonid toads. We also hope that this study fuels the necessity of studying these defensive toxins in more animal taxa so that broader inferences can be made, and studies can be undertaken that compare defensive toxins at macroevolutionary scales.

ACKNOWLEDGEMENTS

DATA ACCESSIBILITY

All raw, unprocessed sequences were deposited in the Sequence Read Archive. We have included a large data package on DRYAD (https://datadryad.org/stash/share/lL2VOf5ouSLmWWcWzpZg_gNkjwFUIPKGLJdqtZ-WqZM) that includes the assemblies for all ten transcriptomes, inputs and codes for EVE, STRING, and PAML selection tests. The scripts for all of our analyses can also be found at https://github.com/rameshbalan.

LITERATURE CITED

Agrawal AA, Petschenka G, Bingham RA, Weber MG, Rasmann S. 2012. Toxic cardenolides: chemical ecology and coevolution of specialized plant–herbivore interactions. New Phytologist. 194(1):28-45.

Amazonas DR, Portes-Junior JA, Nishiyama-Jr MY, Nicolau CA, Chlkidis HM, et al. 2018. Molecular mechanisms underlying intraspecific variation in snake venom. Journal of Proteomics. 181:60-72.

Bidart-Bouzat MG, Imeh-Nathaniel A. 2008. Global Change Effects on Plant Chemical Defenses against Insect Herbivores. Journal of Integrative Plant Biology. 50(11):1339-1354.

Bókony V. Üveges B, Verebélyi V, Ujhegyi N, Móricz AM. 2019. Toads phenotypically adjust their chemical defences to anthropogenic habitat change. Scientific Reports. 9:3163(2019)

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

Cannon MS, Hostetler JP. 1976. The anatomy of the parotoid gland in Bufonidae with some histochemical findings† II. *Bufo alvarius*. Journal of Morphology. 148(2):137-159.

Casewell NR, Wüster W, Vonk FJ, Harrison RA, Fry BG. 2013. Complex cocktails: the evolutionary novelty of venoms. Trends in Ecology and Evolution. 28(4):219-229.

Cei JM, Erspamer V, Roseghini M. 1968. Taxonomic and Evolutionary Significance of Biogenic Amines and Polypeptides in Amphibian Skin. II. Toads of the Genera *Bufo* and *Melanophryniscus*. Systematic Zoology. 17(3):232-245.

Ceschin DG, Pires NS, Mardirosian MN, Lascano CI, Venturino A. 2020. The *Rhinella arenarum* transcriptome: de novo assembly, annotation and gene prediction. Scientific Reports. 10:1053(2020).

Chen KK, Chen AL. 1933. A study of the poisonous secretions of five North American species of toads. Journal of Pharmacology and Experimental Therapeutics. 49(4):526-542.

Conlon JM. 2011. The contribution of skin antimicrobial peptides to the system of innate immunity in anurans. Cell and Tissue Research. 343:201-212.

Crossland MR, Brown GP, Anstis M, Shilton C, Shine R. 2008. Mass mortality of native anuran tadpoles in tropical Australia due to the invasive cane toad (*Bufo marinus*). Biological Conservation. 141(9):2387-2394.

Daly JW, Spande TF. 1986. Amphibian alkaloids: chemistry, pharmacology, and biology. In: Pelletier SW, editor. Alkaloids: chemical and biological perspectives. New York: John Wiley & Sons. p. 1–274.

Daly JW, Gusovsky F, Myers CW, Yotsu-Yamashita M, Yasumoto T. 1994. First occurrence of tetrodotoxin in a dendrobatid frog (*Colostethus inguinalis*), with further reports for the bufonid genus *Atelopus*. Toxicon. 32(3):279-285.

Daly JW, Spande TF, Garraffo HM. 2005. Alkaloids from amphibian skin: a tabulation of over eight-hundred compounds. J Nat Prod. 68:1556–1575.

De Oliveira AL, Wollesen T, Kristof A, Scherholz M, Redl E, Todt C, et al. 2016. Comparative transcriptomics enlarges the toolkit of known developmental genes in mollusks. BMC Genomics.17:905.

Dixon RA, Achnine L, Kota P, Liu C, Srinivasa MS, et al. 2002. The phenylpropanoid pathway and plant defence—a genomics perspective. Molecular Plant Pathology. 3(5):371-390.

Dumbacher JP, Spande TF, Daly JW. 2000. Batrachotoxin alkaloids from passerine birds: a second toxic bird genus (Ifrita kowaldi) from New Guinea. Proc Natl Acad Sci U S A. 97:12970–12975.

Emms DM, Kelly S. 2015. OrthoFinder: solving fundamental biases in whole genome

   comparisons dramatically improves orthogroup inference accuracy. Genome Biology.

   16:157(2015).

Erspamer V, Melchiorri P, Falconieri Erspamer G, Montecucchi PC, de Castiglione R. 1985.

   Phyllomedusa skin: a huge factory and store-house of a variety of active peptides. Peptides

   6(Suppl 3):7–12.

Finn RD, Coggill P, Eberhardt RY, Eddy SR, Mistry J, et al. 2016. The Pfam protein families

   database: towards a more sustainable future. Nucleic Acids Research. 44(D1):D279-D285.

Botelho AFM, Pierezan F, Soto-Blanco B, Melo MM. 2019. A review of cardiac glycosides:

   Structure, toxicokinetics, clinical signs, diagnosis and antineoplastic potential. Toxicon.

   158:63-68.

Fu L, Niu B, Zhu Z, Wu S, Li W. 2012. CD-HIT: accelerated for clustering the next-generation

   sequencing data. Bioinformatics. 28(23):3150-3152.

Fuhrman FA, Fuhrman GJ, Mosher HS. 1969. Toxin from skin of frogs of the genus *Atelopus*:

   differentiation from Dendrobatid Toxins Science. 165:1376-1377.

Garraffo HM, Caceres J, Daly JW, Spande TF, Andriamaharavo NR, Andriantsiferana M. 1993.

   Alkaloids in Madagascan frogs (*Mantella*): pumiliotoxins, indolizidines, quinolizidines, and

   pyrrolizidines. J Nat Prod. 56:1016–1038.

Gerchen JF, Reichert SJ, Röhr JT, Dieterrich C, Kloas W, et al. 2016. A Single Transcriptome of

   a Green Toad (*Bufo viridis*) Yields Candidate Genes for Sex Determination and -

   Differentiation and Non-Anonymous Population Genetic Markers. PLoS ONE. 11(5):

   e0156419.

Götz S, García-Gómez JM, Terol J, Williams TD, Nagaraj SH, et al. 2008. High-throughput functional annotation and data mining with the Blast2GO suite. Nucleic Acids Research. 36(10):3420-3435.

Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, et al. 2011. Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. Nature Biotechnology. 29(7):644-652.

Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, et al. 2013. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nature Protocols. 8:1494-1512.

Hanifin CT, Yotsu-Yamashita M, Yasumoto T, Brodie ED III, Brodie ED Jr. 1999. Toxicity of dangerous prey: variation of tetrodotoxin levels within and among populations of the newt *Taricha granulosa*. Journal of Chemical Ecology. 25:6–11.

Hayes RA, Crossland MR, Hagman M, Capon RJ, Shine R. 2009. Ontogenetic Variation in the Chemical Defenses of Cane Toads (*Bufo marinus*): Toxin Profiles and Effects on Predators. Journal of Chemical Ecology. 35:391-399.

Hill RE, Mackessy SP. 2000. Characterization of venom (Duvernoy's secretion) from twelve species of colubrid snakes and partial sequence of four venom proteins. Toxicon. 38(12):1663-1687.

Hodgson H, de la Peña R, Stephenson MJ, Thimmappa R, Vincent JL, et al. 2019. Identification of key enzymes responsible for protolimonoid biosynthesis in plants: Opening the door to azadirachtin production. Proceedings of the National Academy of Sciences. 116(34): 17096-17104.

Holding ML, Biardi JE, Gibbs HL. 2016. Coevolution of venom function and venom resistance in a rattlesnake predator and its squirrel prey. Proceedings of the Royal Society B. 283:20152841.

Hossain MZ, Ishiga Y, Yamanaka N, Ogiso-Tanaka E, Yamaoka Y. 2018. Soybean leaves transcriptomic data dissects the phenylpropanoid pathway genes as a defence response against *Phakopsora pachyrhizi*. Plant Physiology and Biochemistry. 132:424-433.

Hostetler JR, Cannon MS. 1974. The anatomy of the parotoid gland in Bufonidae with some histochemical findings. I. *Bufo marinus*. Journal of Morphology. 142(2):225-239.

Jared C, Mailho-Fontana PL, Antoniazzi MM, Mendes VA, Barbaro KC, et al. 2015. Venomous Frogs Use Heads as Weapons. Current Biology. 25(16):2166-2170.

Jousset A, Rochat L, Péchy-Tarr M, Keel C, Scheu S, et al. 2009. Predators promote defence of rhizosphere bacterial populations by selective feeding on non-toxic cheaters. ISME Journal. 3:666-674.

Junqueira-de-Azevedo ILM, Bastos CMV, Ho PL, Luna MS, Yamanouye N, et al. 2015. Venom-Related Transcripts from *Bothrops jararaca* Tissues Provide Novel Molecular Insights into the Production and Evolution of Snake Venom. Molecular Biology and Evolution. 32(3): 754-766.

Kanehisa M, Sato Y, Furumichi M, Morishima K, Tanabe M. 2019. New approach for understanding genome variations in KEGG. Nucleic Acids Research. 47(D1):D590-D595.

Kanehisa. M, Goto S. 2000. KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids. Research. 28(1):27-30.

Ladram A, Nicolas P. 2016. Antimicrobial peptides from frog skin: biodiversity and therapeutic promises. Frontiers in Bioscience. 21:1341-1371.

Laurent P, Braekman JC, Daloze S. 2005. Insect chemical defense. Top Current Chemistry. 240:167–229.

Lee SS, Derguini F, Breuning RC, Nakanishi K, Wallick ET, et al. 1994. Digitalis-like compounds of toad bile: sulfation and reduction of bufadienolides decrease potency of Na+, K+-ATPase Inhibition. Heterocycles. 2:669-686.

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

Lichtstein D, Kachalsky S, Deutsch J. 1986. Identification of an ouabain-like compound in toad skin and plasma as a bufodienolide derivative. Life Science. 38:1261-1270.

Lichtstein D, Gati I, Haver E, Katz U. 1992. Digitalis-like compounds in the toad *Bufo viridis*: tissue and plasma levels and significance in osmotic stress. Life Science. 51:119-128.

Lau W, Sattely ES. 2015. Six enzymes from mayapple that complete the biosynthetic pathway to the etoposide aglycone. Science. 349(6253):1224-1228.

Macmanes MD. 2014. On the optimal trimming of high-throughput mRNA sequence data. Frontiers in Genetics. 5:13.

Macrander J, Broe M, Daly M. 2016. Tissue-Specific Venom Composition and Differential Gene Expression in Sea Anemones. Genome Biology and Evolution. 8(8)2358-2375.

Mailho-Fontana PL, Antoniazzi MM, Toledo LF, Verdade VK, Sciani JM, et al. 2014. Passive and active defense in toads: The parotoid macroglands in *Rhinella marina* and *Rhaebo guttatus*. Journal of Experimental Zoology – A: Ecological and Integrative Physiology. 321(2):65-77.

Mailho-Fontana PL, Antoniazzi MM, Sciani JM, Pimenta DC, Barbaro KC, et al. 2018.
Morphological and biochemical characterization of the cutaneous poison glands in toads
(*Rhinella marina* group) from different environments. Frontiers in Zoology. 15:46.

Matsukawa M, Mukai T, Akizawa T, Miyatake S, Yoshioka M, Morris JF, et al. 1998. Isolation
and Characterization of Novel Endogenous Digitalis-like Factors in the Ovary of the Giant
Toad, *Bufo marinus*. Journal of Natural Products. 61(12):1476-1481.

Matsuura HN, Fett-Neto AG. 2015. Plant Alkaloids: Main Features, Toxicity, and Mechanisms
of Action. In Gopalakrishnakone, P., Carlini, C.R., Ligabue-Braun, R., editors. *Plant Toxins*.
Dordrecht: Springer.

Mebs D. 2001. Toxicity in animals. Trends in evolution? Toxicon. 39(1):87-96.

Mebs D, Wagner MG, Pogoda W, Manyero R, Kwet A, et al. 2007. Lack of bufadienolides in the
skin secretion of red bellied toads, *Melanophryniscus* spp. (Anura, Bufonidae), from
Uruguay. Comparative Biochemistry and Physiology Part C: Toxicology & Pharmacology.
144(4):398-402.

Meng Q, Yau L, Lu J, Wu Z, Zhang B, et al. 2016. Chemical profiling and cytotoxicity assay of
bufadienolides in toad venom and toad skin. Journal of Ethnopharmacology. 187(1):74-82.

Neuwirth M, Daly JW, Myers CW, Tice LW. 1979. Morphology of the granular secretory glands
in skin of poison-dart frogs (*Dendrobatidae*). Tissue Cell 11:755–771.

Newman RA, Yang P, Pawlus AD, Block KI. 2008. Cardiac Glycosides as Novel Cancer
Therapeutic Agents. Molecular Interventions. 8(1): 36-49.

Noguchi T, Jeon J, Arakawa O, Sugita H, Deguchi Y, et al. 1986. Occurrence of Tetrodotoxin
and Anhydrotetrodotoxin in Vibrio sp. Isolated from the Intestines of a Xanthid Crab,
Atergatis floridus.Journal of Biochemistry. 99(1):311-314.

Pandey A, Swarnkar V, Pandey T, Srivastava P, Kanojiya S, et al. 2016. Transcriptome and
Metabolite analysis reveal candidate genes of the cardiac glycoside biosynthetic pathway
from *Calotropis procera*. Scientific Reports. 6:34464.

Pastenes L, Valdivieso C, Genova, AD, Travisany D, Hart A, et al. 2017. Global gene expression
analysis provides insight into local adaptation to geothermal streams in tadpoles of the
Andean toad *Rhinella spinulosa*. Scientific Reports. 7:1966(2017).

Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. 2017. Salmon provides fast and bias-
aware quantification of transcript expression. Nature Methods. 14:417-419.

Phillips BL, Shine R. 2004. Adapting to an invasive species: toxic cane toads induce
morphological change in Australian snakes. Proceedings of the National Academy of
Sciences. 101(49):17150-17155.

Phillipson JD, Roberts, MF, Zenk MH. 2012. *The Chemistry and Biology of Isoquinoline
Alkaloids*. Springer, Berlin.

Porto AM, Gros EG. 1970. Biosynthesis of animal and plant bufadienolides. Parallel experiments
with pregn-5-en-3-β-ol-20-one-20-14C in *Scilla maritima* and *Bufo paracnemis*. Experientia.
26:10.1007/BF01900357.

Porto AM, Baralle FE, Gros EG. 1972. Biosynthesis of bufadienolides in toads. Journal of
Steroid Biochemistry. 3:11-17.

Raaymakers C, Verbrugghe E, Hernot S, Hellebuyck T, Betti C, et al. 2017. Antimicrobial
peptides in frog poisons constitute a molecular toxin delivery system against predators.
Nature Communications. 8:1495(2017).

Richardson MF, Sequeira F, Selechnik D, Carneiro M, Vallinoto M, et al. 2018. Improving amphibian genomic resources: a multitissue reference transcriptome of an iconic invader. Giga Science. 7(1): gix114.

Rodríguez C, Rollins-Smith L, Ibáñez R, Durant-Archibold AA, Gutiérrez M. 2017. Toxins and pharmacologically active compounds from species of the family Bufonidae (Amphibia, Anura). Journal of Ethnopharmacology. 198(23):235-254.

Rohlfs RV, Nielsen R. 2015. Phylogenetic ANOVA: The Expression Variance and Evolution Model for Quantitative Trait Evolution. Systematic Biology. 64(5):695-708.

Saporito RA, Donnelly MA, Spande TF, Garraffo HM. 2012. A review of chemical ecology in poison frogs. Chemoecology 22:159–168.

Savitzky AH, Mori A, Hutchinson DA, Saporito RA, Burghardt GM, et al. 2012. Sequestered defensive toxins in tetrapod vertebrates: principles, patterns, and prospects for future studies. Chemoecology 22:141–158.

Sciani JM, Angeli CB, Antoniazzi MM, Jared C, Pimenta DC. 2013. Differences and similarities among parotoid macrogland secretions in South American toads: a preliminary biochemical delineation. Scientific World Journal. 2013:937407.

Shroff R, Vergara F, Muck A, Svatoš A, Gershenzon J. 2008. Nonuniform distribution of glucosinolates in *Arabidopsis thaliana* leaves has important consequences for plant defense. Proceedings of the National Academy of Sciences. 105(16):6196-6201.

Simäo F, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics. 31(19):3210-3212.

Sime K. 2002. Chemical defence of *Battus philenor* larvae against attack by the parasitoid

    Trogus pennator. Ecological Entomology 27:337-345

Simmaco M, Mignogna G, Barra D. 1998. Antimicrobial peptides from amphibian skin: What do

    they tell us?. Peptide Science. 47(6):435-450.

Siperstein MD, Murray AW, Titus E. 1957. Biosynthesis of cardiotonic sterols from cholesterol

    in the toad, *Bufo marinus*. Archives of Biochemistry and Biophysics. 67(1):154-160.

Speed MP, Ruxton GD, Mappes J, Sherrat TN. 2012. Why are defensive toxins so variable? An

    evolutionary perspective. Biological Reviews. 87(4):874-884.

Sønderby IE, Geu-Flores F, Halkier BA. 2010. Biosynthesis of glucosinolates – gene discovery

    and beyond. Trends in Plant Science. 15(5):283-290.

Srivastava S, Srivastava M, Misra A, Pandey G, Rawat AKS. 2015. A review on biological and

    chemical diversity in Berberis (Berberidaceae). Journal of Experimental and Clinical

    Sciences. 14:247-267.

Stuhlemmer U, Kreis W. 1996. Cardenolide formation and activity of pregnane-modifying

    enzymes in cell suspension cultures, shoot cultures and leaves of *Digitalis lanata*. Plant

    Physiology and Biochemistry. 43(1):85-91.

Surm JM, Smith HL, Madio B, Undheim EAB, King GF, et al. 2019. A process of convergent

    amplification and tissue-specific expression dominates the evolution of toxin and toxin-like

    genes in sea anemones. Molecular Ecology. 28(9):2272-2289.

Tarvin RD, Santos JC, O'Connell LA, Zakon HH, Cannatella, DC. 2016. Convergent

    Substitutions in a Sodium Channel Suggest Multiple Origins of Toxin Resistance in Poison

    Frogs. Molecular Biology and Evolution. 33(4):1068=1080.

Tarvin RD, Borghese CM, Sachs W, Santos JC, Lu Y, et al. 2017. Interacting amino acid replacements allow poison frogs to evolve epibatidine resistance. Science. 357(6357):1261-1266.

Theissinger K, Falckenhayn C, Blande D, Tolijamo A, Gutenkunst J, et al. 2016. De Novo assembly and annotation of the freshwater crayfish *Astacus astacus* transcriptome. Marine Genomics. 28:7-10.

Ujvari B, Casewell NR, Sunagar K, Arbuckle K, Wüster W, et al. 2015. Widespread convergence in toxin resistance by predictable molecular evolution. Proc Natl Acad Sci U S A. 112:11911–11916.

Wickramasinghe JAF, Burrows EP, Sharma RK, Greig JB, Caspi, GE. 1969. The obligatory involvement of a C21 intermediate in the biosynthesis of cardenolides from cholesterol. Phytochemistry. 8(8):1433-1440.

Wortman-Wunder E, Vivanco JM. 2011. Chemical Ecology: Definition and Famous Examples. In Vivianco J, Weir T, editors. *Chemical Biology of the Tropics. Signaling and Communication in Plants,Vol 8.* Springer, Berlin, Heidelberg.

Yamazaki and Morita, 2004

Yang Z. 2006. *Computational Molecular Evolution*. Oxford, New York.

Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. Molecular Biology and Evolution. 24:1586-1591.

Yotsu M, Tateki E. 2010. First report on toxins in the Panamanian toads Atelopus limosus, *A. glyphus* and *A. certus*. Toxicon. 55:153-156.

Zagrobleny M, Bak S, Lindberg B, Møller L. 2008. Cyanogenesis in plants and arthropods. Phytochemistry. 69(7):1457-1468.

Zhang J, Nielsen R, Yang Z. 2005. Evaluation of an Improved Branch-Site Likelihood Method for Detecting Positive Selection at the Molecular Level. Molecular Biology and Evolution. 22(12):2472-2479.

Zhang Y. 2015. Why do we study animal toxins?. Zoological Research. 36(4):183-222.

Zhong Y, Zhao C, Wu W, Fan T, Lí N, et al. 2020. Total synthesis, chemical modification and structure-activity relationship of bufadienolides. European Journal of Medicinal Chemistry. 189:112038.

TABLES

Table 1. Assembly statistics for our ten bufonid species.

	Average raw paired-end reads	Average cleaned paired-end reads	N50	Number of transcripts	Max. transcript length	Min. transcript length	Mean transcript length	GC Content
Incilius alvarius	33161928	20706357.67	855	231583	16957	179	589.32	43.73
Anaxyrus americanus	36790027	28440591	1136	201813	15445	180	666.59	45.20
Anaxyrus cognatus	40488144.5	21642307.67	932	268343	24315	183	615.86	44.93
Anaxyrus fowleri	41699161.33	24739897.33	920	265224	21029	178	613.89	44.94
Anaxyrus microscaphus	30060419.33	16024484.33	965	213157	17415	174	619.6	44.82
Incilius nebulifer	37080453	23644778.33	700	290104	17283	168	551.76	44.79
Anaxyrus punctatus	40020982.33	26690504.33	640	342030	20151	177	534.1	44.98
Anaxyrus speciosus	38635852.5	19778744.33	1210	163758	16865	178	682.39	44.65
Anaxyrus terrestris	25230673.33	16762736.67	1006	217966	17132	169	615.57	44.95
Anaxyrus woodhousii	41838543	28425276	958	207028	16200	180	622.51	45.08

Table 2. Alternative and null model likelihoods for the genes that show signatures of positive selection after p-value correction. BEB scores are only shown for sites ≥0.9.

Gene	Alternative Log Likelihood	Null Log Likelihood	Unadjusted P-Values	Adjusted P-values	BEB scores for positive sites
APOO	-1740.984920	-1747.708240	2.45E-04	1.61E-02	none
USP10	-4612.624708	-4621.957690	1.56E-05	1.02E-03	K 0.957 V 0.952 S 0.956 S 0.952 S 0.969 V 0.955
MON1B	-5074.060504	-5082.413883	4.36E-05	2.87E-03	none
PNN	-4780.236623	-4797.738475	3.29E-09	2.17E-07	none

FIGURES

Figure 1. Overview of our workflow to assemble and analyze the parotoid transcriptomes of our ten focal bufonid species.
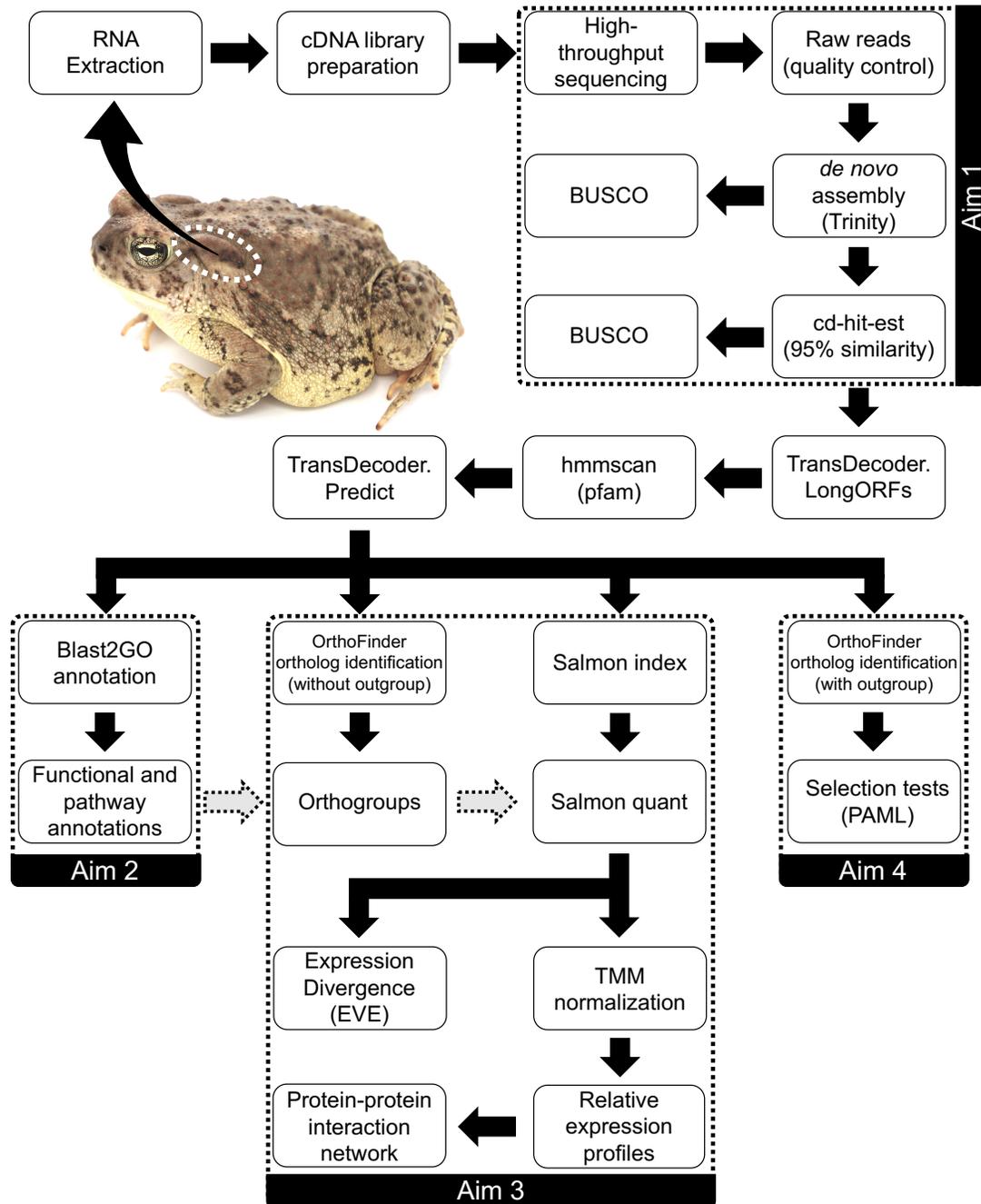
Figure 2. Average Gene Ontology (GOslim) assignments for transcripts across ten bufonid species. Level 4 annotations are shown for the cellular component graph, and level 3 annotations for the molecular function and biological process graphs.
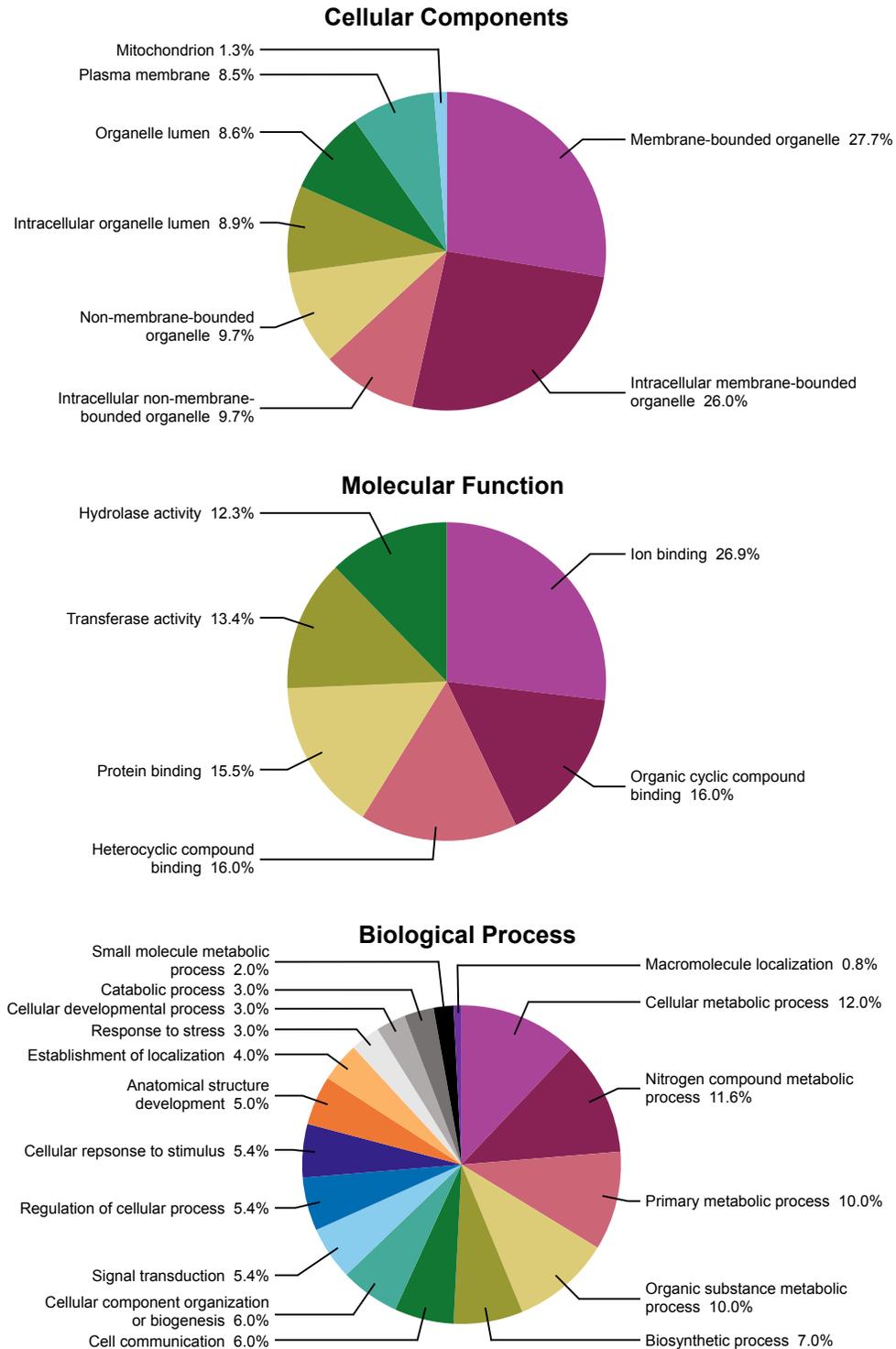
Figure 3. Relative expression profiles for metabolic pathways across our ten focal bufonid species. A species tree generated from OrthoFinder showing the genetic relationships between our species is shown above then graph. Color gradient indicates low (white) to high (dark blue) expression of pathways.
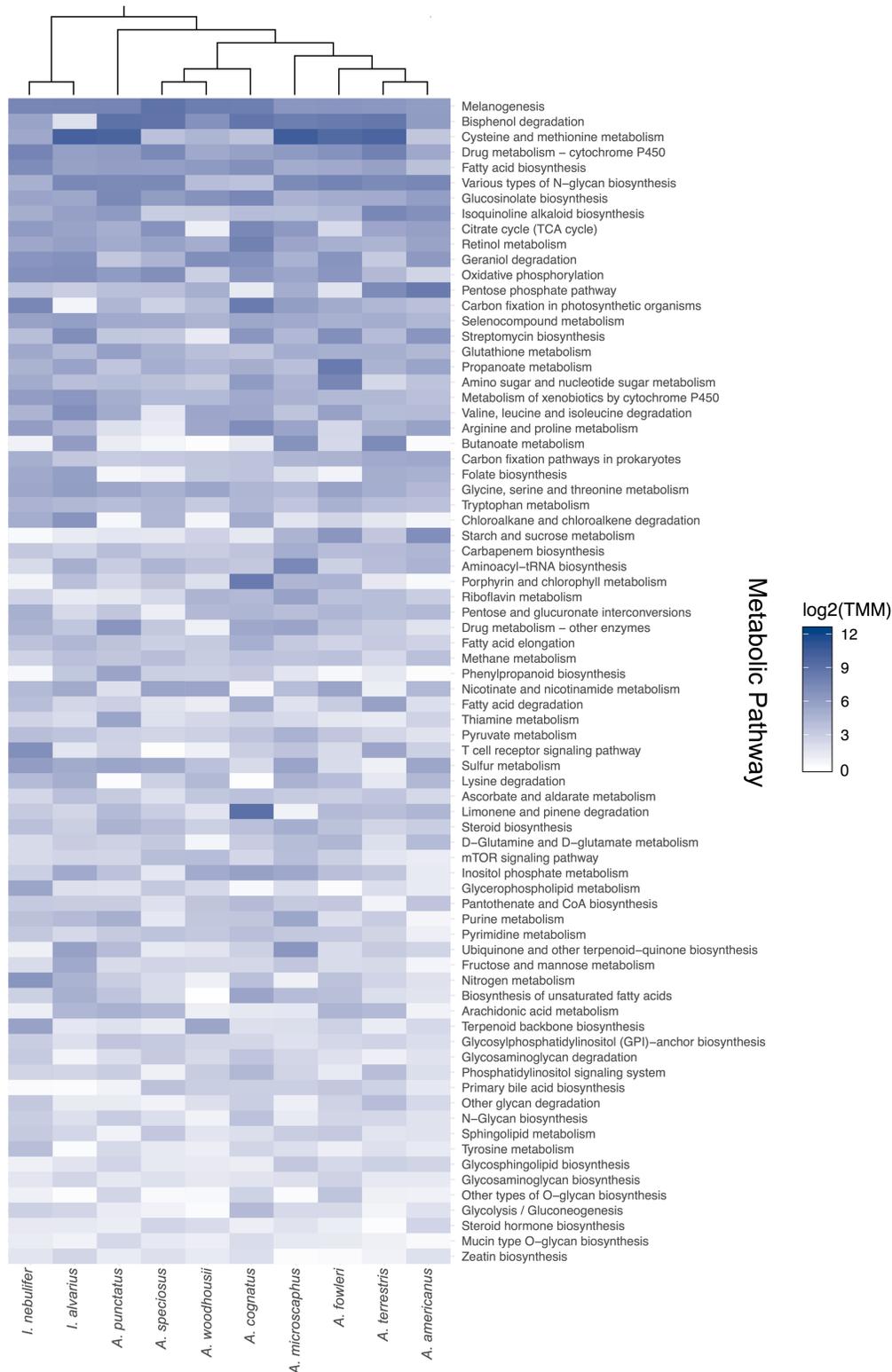
Figure 4. Expression divergence results from EVE. Genes annotated to potential toxin producing pathways (blue dots) do not show signatures of adaptive expression divergence (above dashed red line (p<0.01) and positive expression divergence) of our 1454 1-to-1 orthologs (gray dots).
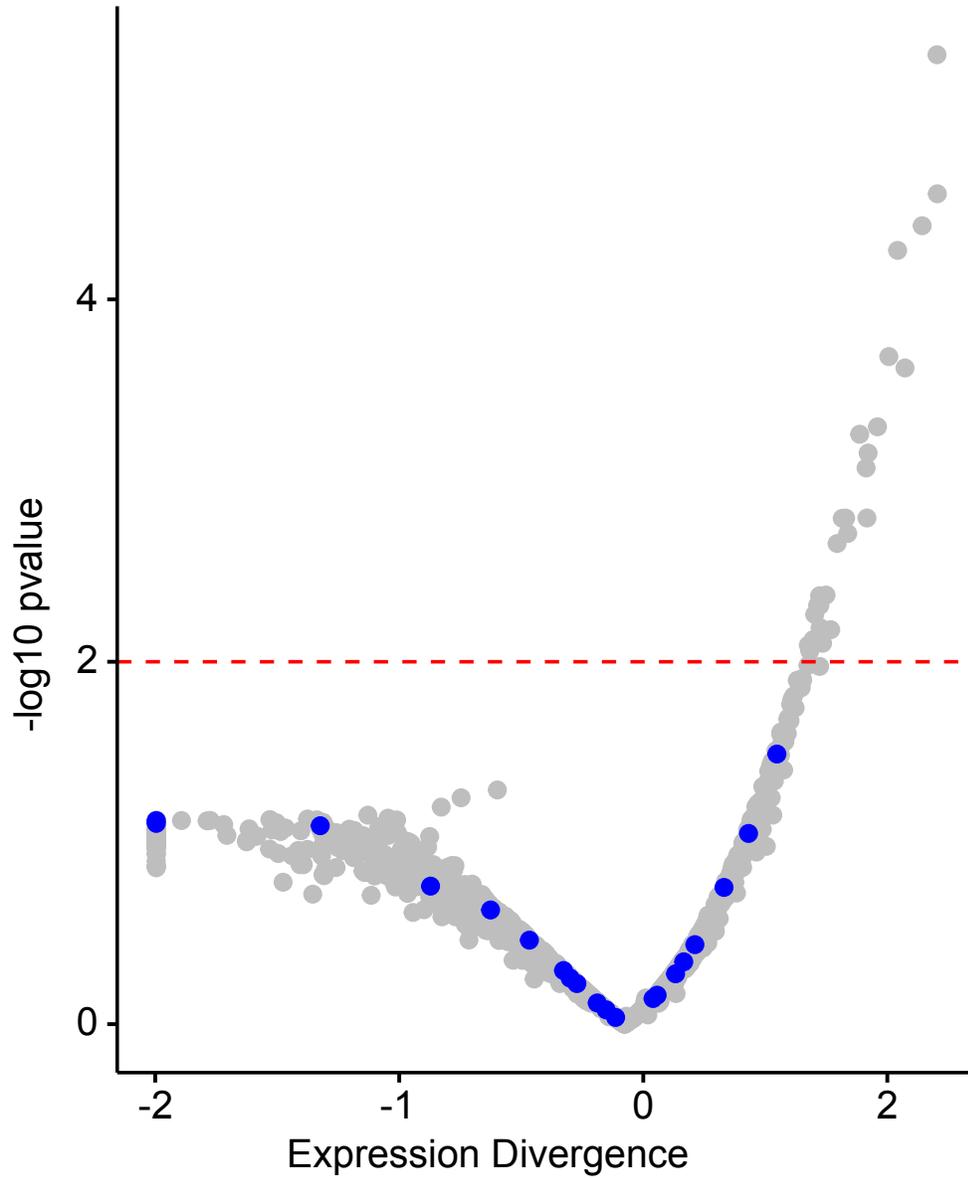
Figure 5. Protein-protein interaction (PPI) network obtained and visualized by STRING v11.0 for expressed genes within the parotoid gland transcriptome. Nodes depict proteins and PPIs are represented by edges in the network.

SUPPLEMENTARY INFORMATION

Table S1. Samples used for molecular data analyses, with locality data (State Abbreviation: County: Specific Locality), sex of the specimen, and museum voucher numbers.

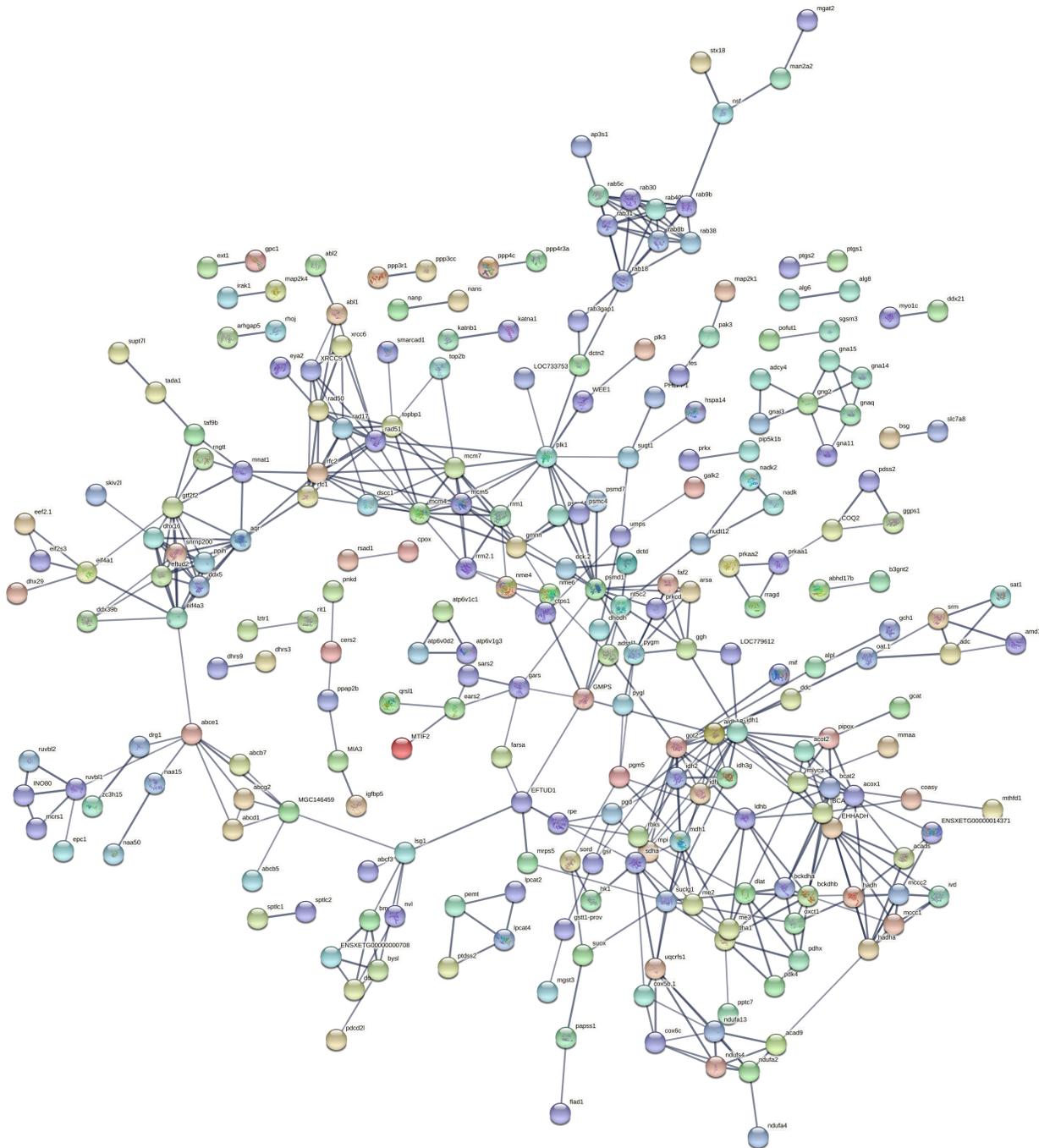| Species	Museum Number	Field Number	Sex	Locality	Latitude	Longitude	Sequencing Platform
Anaxyrus americanus	UTA A-65801	TJF036	F	PA: Allegheny: Frick Park Environmental Center	40.4369	-79.9073	NovaSeq6000
Anaxyrus americanus	UTA A-65802	TJF038	F	PA: Allegheny: Frick Park Environmental Center	40.4369	-79.9073	HiSeq2500
Anaxyrus americanus	UTA A-65803	TJF039	M	PA: Allegheny: Frick Park Environmental Center	40.4369	-79.9073	HiSeq2500
Anaxyrus cognatus	UTA A-65804	TJF042	M	TX: Clay: Byers – Hwy 171	34.08333	-98.2344	NovaSeq6000
Anaxyrus cognatus	UTA A-65806	TJF046	F	TX: Clay: Byers – Hwy 171	34.08061	-98.27361	HiSeq2500
Anaxyrus fowleri	UTA A-65810	TJF094	F	AL: Lee: Northeast of Auburn University	32.639719	-85.440642	HiSeq2500
Anaxyrus fowleri	UTA A-65811	TJF096	F	AL: Macon: Tuskegee National Forest	32.510877	-85.58406	HiSeq2500
Anaxyrus fowleri	UTA A-65812	TJF097	F	AL: Macon: Tuskegee National Forest	32.510877	-85.58406	HiSeq2500
Anaxyrus microscaphus	UTA A-65813	TJF087	F	AZ: Gila: Road to Lower Canyon Creek	34.2489	-110.8037	HiSeq2500
Anaxyrus microscaphus	UTA A-65814	TJF090	F	AZ: Gila: Road to Lower Canyon Creek	34.252	-110.8078	HiSeq2500
Anaxyrus microscaphus	UTA A-65815	TJF091	F	AZ: Gila: Road to Lower Canyon Creek	34.266	-110.8178	HiSeq2500
Anaxyrus punctatus	UTA A-65816	TJF012	F	TX: Brewster: Lajitas – Hwy 170	29.262	-103.7756	NovaSeq6000
Anaxyrus punctatus	UTA A-65817	TJF023	F	TX: Brewster	29.59978	-103.5605	HiSeq2500
Anaxyrus punctatus	UTA A-65818	TJF024	F	TX: Brewster	29.59978	-103.5605	HiSeq2500
Anaxyrus speciosus	UTA A-65819	TJF021	F	TX: Brewster: Fort Davis/Alpine – Hwy 118	29.5286	-103.54191	NovaSeq6000
Anaxyrus speciosus	UTA A-65820	TJF016	M	TX: Jeff Davis: Sergeant Mulhoorn Loop	30.567559	-103.879575	HiSeq2500
Anaxyrus terrestris	UTA A-65822	TJF099	F	FL: Nassau: Egans Creek Park	30.669021	-81.440135	HiSeq2500
Anaxyrus terrestris	UTA A-65824	TJF102	F	FL: Nassau: Egans Creek Park	30.669021	-81.440135	HiSeq2500
Anaxyrus woodhousii	UTA A-65825	TJF033	F	TX: Cottle	33.930916	-100.13472	NovaSeq6000
Anaxyrus woodhousii	UTA A-65826	TJF027	F	TX: Cottle	33.930916	-100.13472	HiSeq2500
Anaxyrus woodhousii	UTA A-65827	TJF028	M	TX: Cottle	33.930916	-100.13472	HiSeq2500
Incilius alvarius	UTA A-65828	TJF072	F	AZ: Cochise: Double Adobe Road	31.4518	-109.6305	HiSeq2500
Incilius alvarius	UTA A-65829	TJF073	F	AZ: Cochise: North Central Highway	31.5562	-109.6979	HiSeq2500
Incilius alvarius	UTA A-65830	TJF074	F	AZ: Pima: Sabino Canyon Visitor Center	32.3096	-110.8224	HiSeq2500
Incilius nebulifer	UTA A-65831	TJF052	F	TX: Tarrant: Elmer Oliver Nature Park	32.5879	-97.098	NovaSeq6000
Incilius nebulifer	UTA A-65832	TJF053	F	TX: Tarrant: Elmer Oliver Nature Park	32.5886	-97.1007	HiSeq2500
Incilius nebulifer	UTA A-65833	TJF054	F	TX: Tarrant: Elmer Oliver Nature Park	32.5884	-97.1007	HiSeq2500

Table S2. STRING network statistics.

Number of nodes	386
Number of edges	477
Average node degree	2.47
Average local clustering coefficient	0.436
Expected number of edges	272
PPI enrichment p-value	0.00

Table S3. Functional enrichment of pathways from STRING.

#term ID	Term description	Observed gene count	Background gene count	Strength	False discovery rate
xtr01100	Metabolic pathways	108	1110	0.67	2.25E-38
xtr01200	Carbon metabolism	16	105	0.86	1.72E-07
xtr00280	Valine, leucine and isoleucine degradation	12	48	1.08	1.72E-07
xtr00230	Purine metabolism	12	150	0.58	0.0012
xtr00240	Pyrimidine metabolism	10	81	0.77	0.00018
xtr00190	Oxidative phosphorylation	10	108	0.65	0.0013
xtr00270	Cysteine and methionine metabolism	9	41	1.02	1.48E-05
xtr00480	Glutathione metabolism	9	45	0.98	2.51E-05
xtr01230	Biosynthesis of amino acids	9	63	0.83	0.00018
xtr00020	Citrate cycle (TCA cycle)	8	26	1.17	1.06E-05
xtr00330	Arginine and proline metabolism	8	43	0.95	0.00012
xtr00983	Drug metabolism - other enzymes	8	46	0.92	0.00017
xtr04146	Peroxisome	8	81	0.67	0.0034
xtr01210	2-Oxocarboxylic acid metabolism	7	17	1.29	1.06E-05
xtr00640	Propanoate metabolism	7	29	1.06	0.00012
xtr01212	Fatty acid metabolism	7	47	0.85	0.0011
xtr00564	Glycerophospholipid metabolism	7	80	0.62	0.0115
xtr04371	Apelin signaling pathway	7	105	0.5	0.0287

xtr04068	FoxO signaling pathway	7	117	0.46	0.0408
xtr00071	Fatty acid degradation	6	42	0.83	0.0032
xtr04912	GnRH signaling pathway	6	69	0.62	0.02
xtr04270	Vascular smooth muscle contraction	6	79	0.56	0.0285
xtr00650	Butanoate metabolism	5	20	1.08	0.0012
xtr00760	Nicotinate and nicotinamide metabolism	5	31	0.89	0.0053
xtr00620	Pyruvate metabolism	5	36	0.82	0.0088
xtr00970	Aminoacyl-tRNA biosynthesis	5	38	0.8	0.0104
xtr00600	Sphingolipid metabolism	5	49	0.69	0.0216
xtr00310	Lysine degradation	5	52	0.66	0.0239
xtr04012	ErbB signaling pathway	5	66	0.56	0.0433
xtr00900	Terpenoid backbone biosynthesis	4	18	1.03	0.0061
xtr00062	Fatty acid elongation	4	24	0.9	0.0129
xtr00790	Folate biosynthesis	4	24	0.9	0.0129
xtr01040	Biosynthesis of unsaturated fatty acids	4	26	0.87	0.0148
xtr00410	beta-Alanine metabolism	4	30	0.8	0.0216
xtr03030	DNA replication	4	30	0.8	0.0216
xtr00980	Metabolism of xenobiotics by cytochrome P450	4	37	0.71	0.031
xtr03050	Proteasome	4	37	0.71	0.031
xtr00350	Tyrosine metabolism	4	41	0.67	0.0408
xtr00380	Tryptophan metabolism	4	41	0.67	0.0408
xtr00520	Amino sugar and nucleotide sugar metabolism	4	43	0.65	0.0433
xtr03450	Non-homologous end-joining	3	11	1.11	0.013
xtr00770	Pantothenate and CoA biosynthesis	3	16	0.95	0.0244
xtr00360	Phenylalanine metabolism	3	17	0.93	0.0275
xtr00040	Pentose and glucuronate interconversions	3	18	0.9	0.0291
xtr00532	Glycosaminoglycan biosynthesis - chondroitin sulfate / dermatan sulfate	3	18	0.9	0.0291
xtr00290	Valine, leucine and isoleucine biosynthesis	2	4	1.38	0.0238

Figure S1. Benchmarking Universal Single-copy Orthologs (BUSCO) summary of complete, duplicated, fragmented, and missing orthologs search against the 5310 single-copy orthologs before and after removing redundant transcripts.
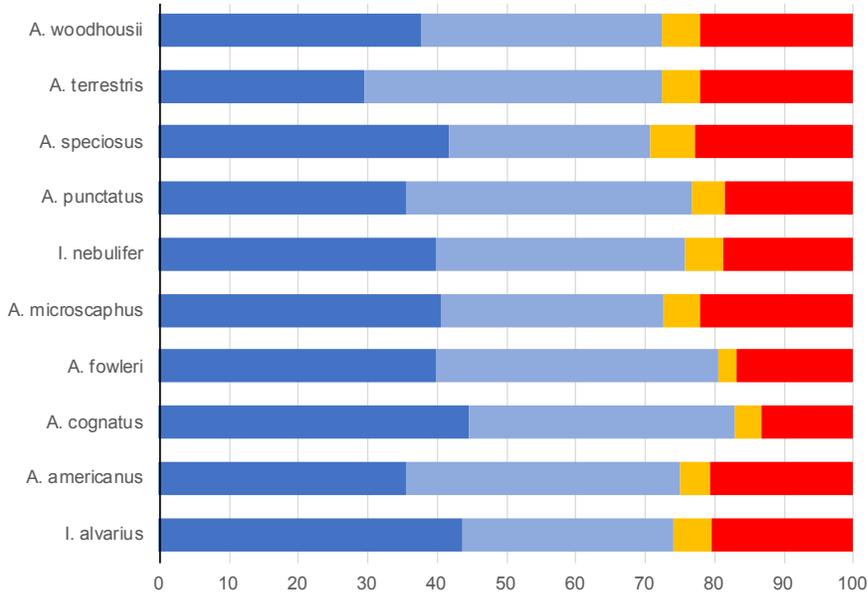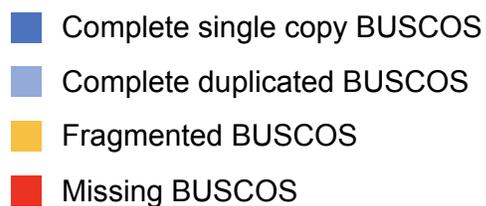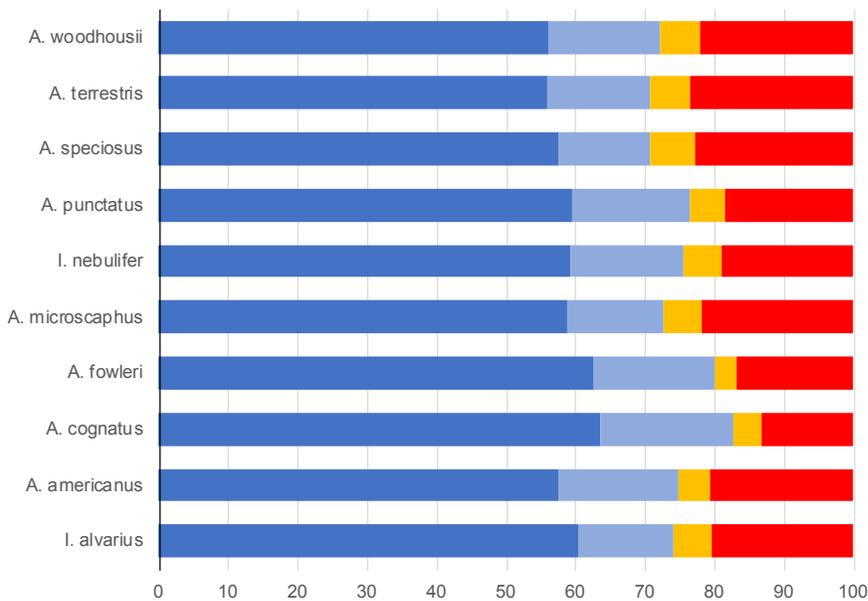
Figure S2. Plot of the number of orthogroups across our ten bufonid species.
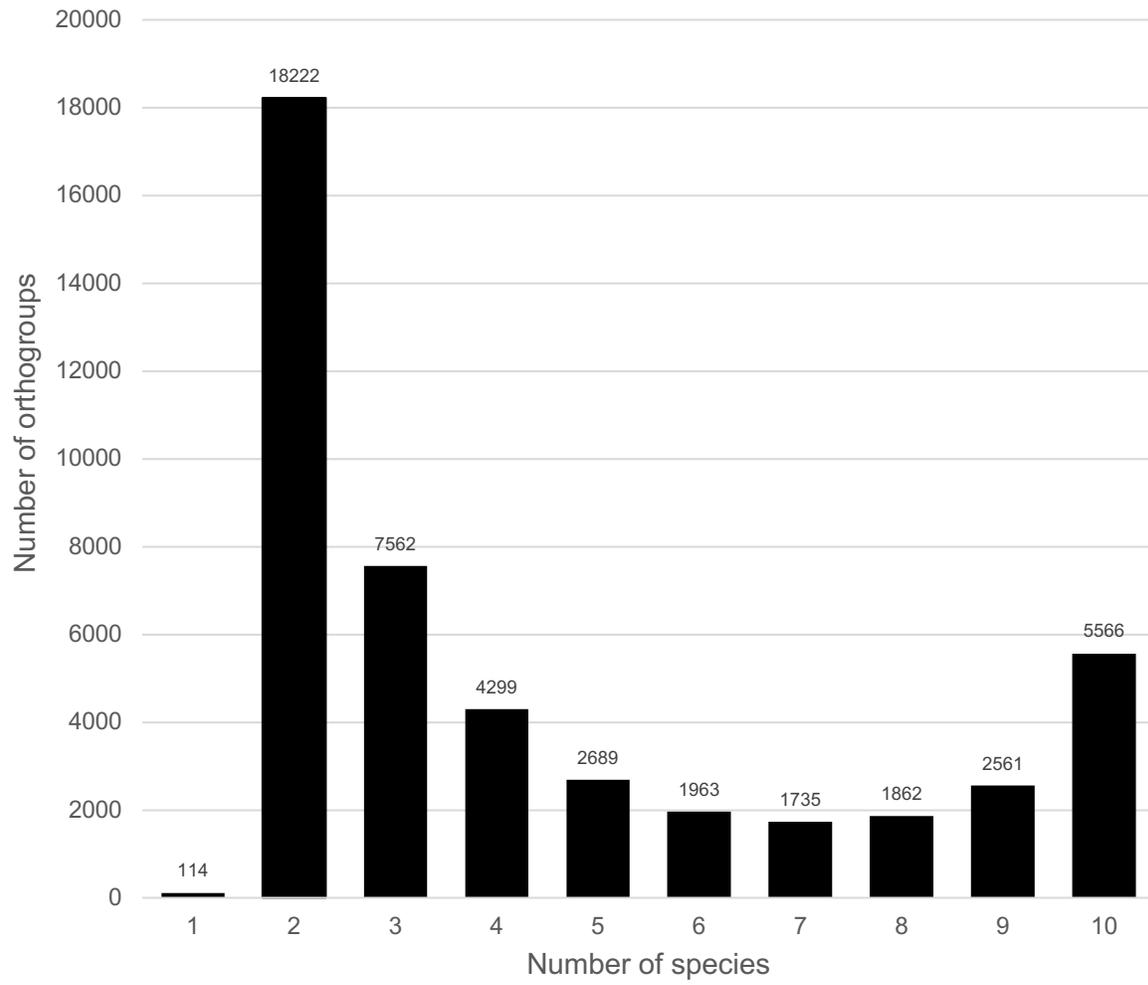
Figure S3. Gene ontology annotations for the parotoid gland transcriptomes of the individual ten bufonid species.

Cellular Component

Molecular Function

Incilius alvaius
Anaxyrus americanus
Anaxyrus cognatus
Anaxyrus fowleri
Anaxyrus microscaphus
Incilius nebulifer
Anaxyrus punctatus
Anaxyrus speciosus
Anaxyrus terrestris
Anaxyrus woodhousii

Biological Process
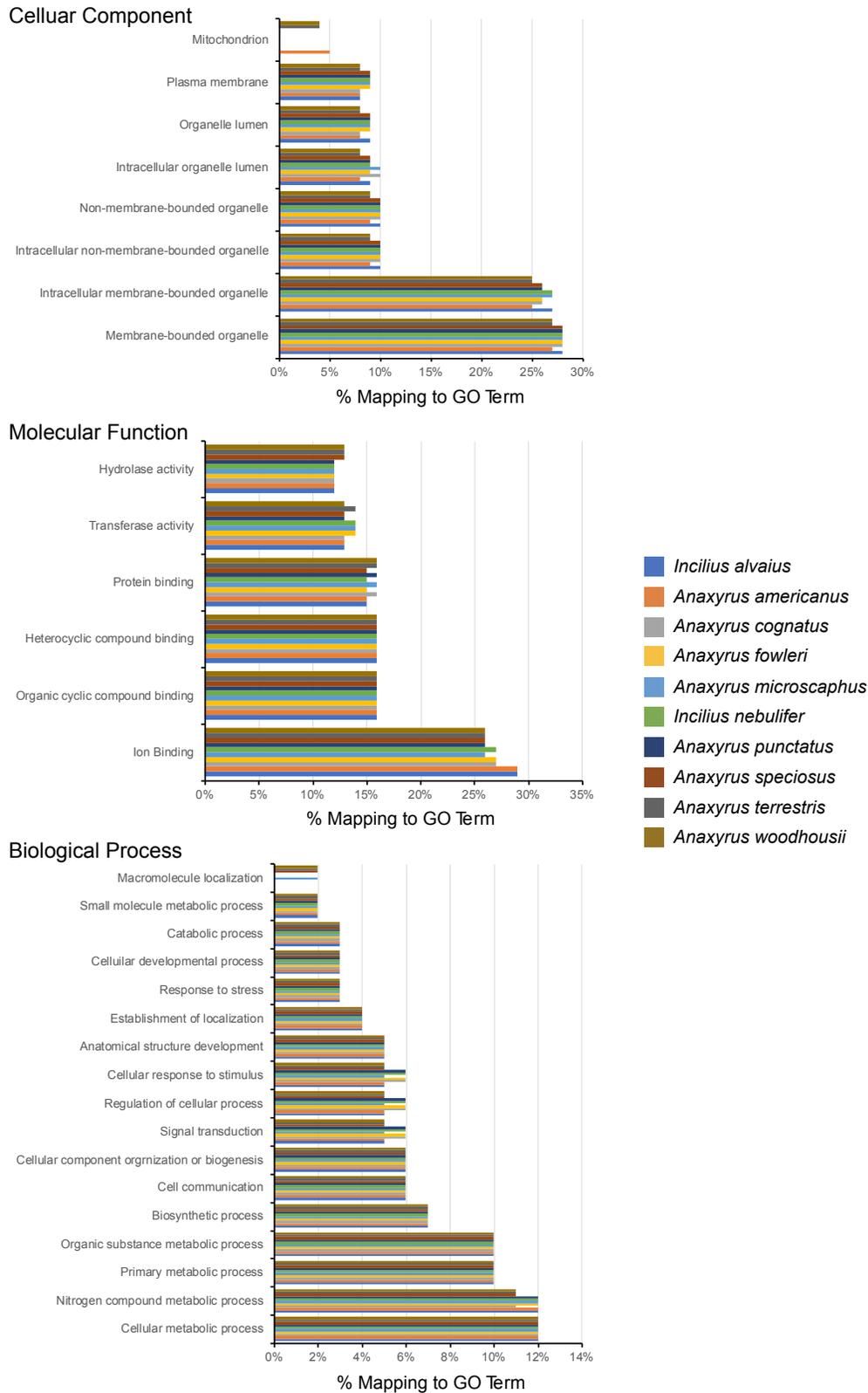
Figure S4. Relative expression of genes within their associated pathways across our ten focal species.

Figure S4. (cont.)

Butanoate metabolism

Carbon fixation in photosynthetic organisms

Carbon fixation pathways in prokaryotes

Chloroalkane and chloroalkene degradation

Citrate cycle (TCA cycle)

Cysteine and methionine metabolism

D–Glutamine and D–glutamate metabolism

Drug metabolism – cytochrome P450

Drug metabolism – other enzymes

Fatty acid biosynthesis

Fatty acid degradation

Fatty acid elongation

Drug metabolism – other enzymes

Figure S4. (cont.)

carnitine O–palmitoyltransferase 1, liver isoform

acyl–CoA dehydrogenase family member 9, mitochondrial

Fatty acid degradation

alpha/beta hydrolase domain–containing protein 17C

alpha/beta hydrolase domain–containing protein 17B

Alpha/beta hydrolase domain–containing protein 13

Fatty acid elongation

GTP cyclohydrolase 1–like

gamma–glutamyl hydrolase

Folate biosynthesis

multiple C2 and transmembrane domain–containing protein 2 isoform X1

mannose–6–phosphate isomerase

Fructose and mannose metabolism

peroxisomal bifunctional enzyme

hydroxyacyl–coenzyme A dehydrogenase, mitochondrial

Geraniol degradation

branched–chain–amino–acid aminotransferase, mitochondrial isoform X1

Glucosinolate biosynthesis

glutathione reductase, mitochondrial

Glutathione metabolism

tafazzin
protein ABHD16A
protein ABHD12B
phospholipid phosphatase 3
phosphatidylserine synthase 2
phosphatidylethanolamine N–methyltransferase
phosphatidate phosphatase LPIN1 isoform X2
phosphatidate cytidylyltransferase 2
manganese–dependent ADP–ribose/CDP–alcohol diphosphatase
lysophospholipid acyltransferase LPCAT4
lysophosphatidylcholine acyltransferase 2
endothelin–1 receptor
acylglycerol kinase, mitochondrial
acyl–protein thioesterase 2
acetyl–coenzyme A transporter 1

Glycerophospholipid metabolism

zinc finger CCCH domain–containing protein 15

cortactin–binding protein 2

2–amino–3–ketobutyrate coenzyme A ligase, mitochondrial

Glycine, serine and threonine metabolism

log2(TMM)

9

6

3

Glycolysis / Gluconeogenesis

Glycosaminoglycan biosynthesis – chondroitin sulfate / dermatan sulfate

I. nebulifer
I. alvarus
A. punctatus
A. speciosus
A. woodhousii
A. cognatus
A. microscaphus
A. fowleri
A. terrestris
A. americanus

Figure S4. (cont.)

Glycerophospholipid metabolism

Glycine, serine and threonine metabolism

Glycolysis / Gluconeogenesis

Glycosaminoglycan biosynthesis – chondroitin sulfate / dermatan sulfate

Glycosaminoglycan degradation

Glycosphingolipid biosynthesis – lacto and neolacto series

Glycosylphosphatidylinositol (GPI)–anchor biosynthesis

Inositol phosphate metabolism

Isoquinoline alkaloid biosynthesis

Limonene and pinene degradation

Lysine degradation

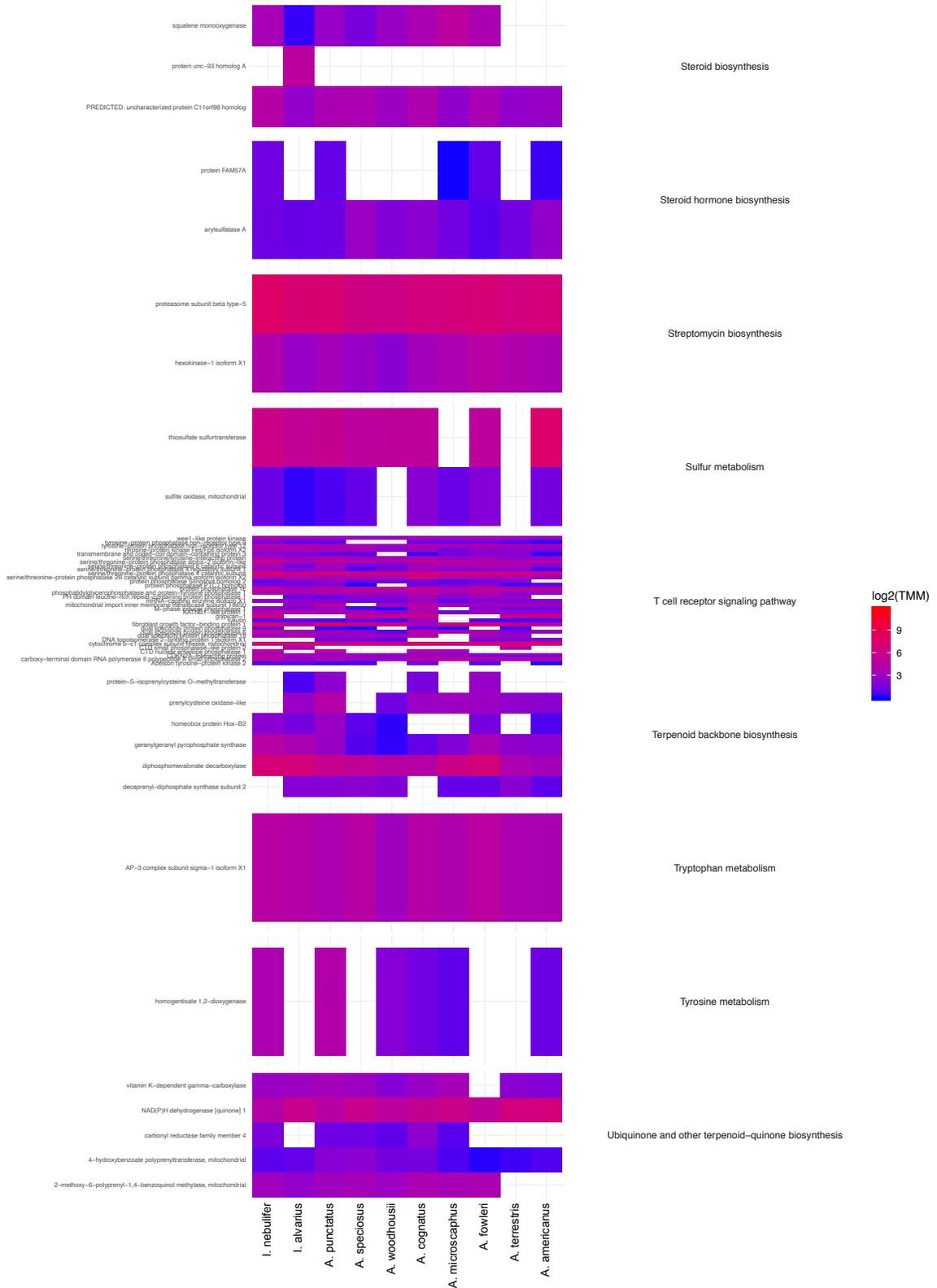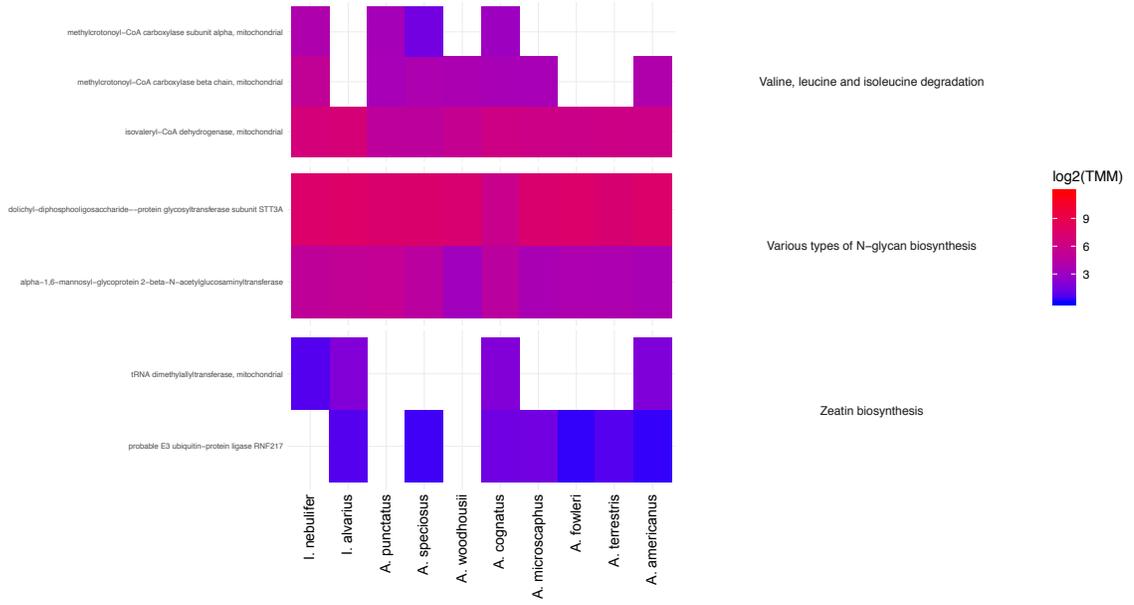Metabolism of xenobiotics by cytochrome P450

log2(TMM)

Figure S4. (cont.)

Figure S4. (cont.) Thiamine pathway only.