

ARTICLE



Testing assertions of widespread introgressive hybridization in a clade of neotropical toads with low mate selectivity (*Rhinella granulosa* species group)

Danielle Rivera^{1,2✉}, Ivan Prates³, Janalee P. Caldwell⁴, Miguel Trefaut Rodrigues⁵ and Matthew K. Fujita^{2,6}

© The Author(s), under exclusive licence to The Genetics Society 2022

Discordance between different genomic regions, often identified through multilocus sequencing of selected markers, presents particular difficulties in identifying historical processes which drive species diversity and boundaries. Mechanisms causing discordance, such as incomplete lineage sorting or introgression due to interspecific hybridization, are better identified based on population-level genomic datasets. In the toads of the *Rhinella granulosa* species group, patterns of mito-nuclear discordance and potential hybridization have been reported by several studies. However, these patterns were proposed based on few loci, such that alternative mechanisms behind gene-tree heterogeneity cannot be ruled out. Using genome-wide ddRADseq loci from a subset of species within this clade, we found only partial concordance between currently recognized species-level taxon boundaries and patterns of genetic structure. While most taxa within the *R. granulosa* group correspond to clades, genetic clustering analyses sometimes grouped distinct taxonomic units into a single cluster. Moreover, levels of admixture between inferred clusters were limited and restricted to a single taxon pair which is best explained by incomplete lineage sorting as opposed to introgressive hybridization, according to D-statistics results. These findings contradict previous assertions of widespread cryptic diversity and gene flow within the *R. granulosa* clade. Lastly, our analyses suggest that diversification events within the *Rhinella granulosa* group mostly dated back to the early Pliocene, being generally younger than species divergences in other closely related clades that present high levels of cross-species gene flow. This finding uniquely contradicts common assertions that this young clade of toads exhibits interspecific hybridization.

Heredity; <https://doi.org/10.1038/s41437-022-00571-9>

INTRODUCTION

Discordance between different genomic regions has been increasingly observed in numerous species over the past decades (Firneno et al. 2020; Ivanov et al. 2018; Toews and Brelsford 2012). This phenomenon has been explored more recently with high throughput sequencing, allowing the ability to analyze the nuclear genome at a larger scale compared to previous single or multi-locus studies (Firneno et al. 2020; Firneno and Townsend 2019). Different patterns of differentiation between mitochondrial and nuclear genomes are commonly observed when comparing relationships among species or populations in phylogeographic studies (Firneno et al. 2020; Rivera et al. 2020; Arntzen et al. 2017). Discordance can be caused by a variety of mechanisms, including differing selective pressures, sex-biased dispersal, neutral demographic processes, introgression or hybridization, incomplete lineage sorting, or a combination of mechanisms (Chiocchio et al. 2021; Dufresnes et al. 2020; Firneno et al. 2020; Ivanov et al. 2018; Thielsch et al. 2017; Toews and Brelsford 2012). Gene tree discordance has made phylogenetic reconstruction challenging and presents particular difficulties in the identification of historical

demographic processes that are driving species diversity and delimitation (Degnan and Rosenberg 2009; Firneno et al. 2020; Fujita et al. 2012). The use of larger genomic datasets has helped determine the potential causes of mito-nuclear discordance, even among species with relatively large genomes (Firneno et al. 2020; Hill et al. 2019). Moreover, the use of demographic model testing and tests for phylogenetic introgression between species or populations using this type of data provide the potential to uncover the root causes of genomic incongruence (Leaché et al. 2019; Portik et al. 2017). These methods can help to identify potential biogeographic, selective, or neutral processes that might be affecting phylogenetic structure in nuclear versus mitochondrial genomes.

True toads (Bufonidae) are among the groups of organisms reported to have high levels of mito-nuclear discordance (Azevedo et al. 2003; Firneno et al. 2020; Pereyra et al. 2016; Sequeira et al. 2011). Widespread mito-nuclear discordance in toads was initially attributed to introgression of entire loci due to species hybridization, in agreement with the low selectivity of mating partners by these toads during their typically explosive breeding events

¹Department of Applied Ecology, North Carolina State University, Raleigh, NC, USA. ²Amphibian and Reptile Diversity Research Center, University of Texas at Arlington, Arlington, TX, USA. ³Department of Ecology and Evolutionary Biology and Museum of Zoology, University of Michigan, Ann Arbor, MI, USA. ⁴Sam Noble Museum and Department of Biology, University of Oklahoma, Norman, OK, USA. ⁵Departamento de Zoologia, Instituto de Biociências, Universidade de São Paulo, São Paulo, SP, Brazil. ⁶Department of Biology and Amphibian and Reptile Diversity Research Center, University of Texas at Arlington, Arlington, TX, USA. Associate editor: Ben Evans. ✉email: rivera2288@gmail.com

Received: 21 July 2021 Revised: 20 October 2022 Accepted: 21 October 2022

Published online: 05 November 2022

(Pereyra et al. 2016; Abreu et al. 2021). However, a recent analysis using genome-wide loci found evidence of incomplete lineage sorting and not hybridization to explain discordance in toads which exhibit mito-nuclear discordance (Firneno et al. 2020). This study demonstrates the benefit of large genomic datasets when attempting to discern which mechanisms or processes may be driving patterns of among-locus gene tree heterogeneity in natural populations (Degnan and Rosenberg 2009).

Within the true toad radiation, the *Rhinella granulosa* species group has been hypothesized to involve high levels of hybridization and poorly resolved intraspecific relationships due to mito-nuclear incongruencies (Narvaes and Rodrigues 2009; Pereyra et al. 2016). Within this group of South American toads, 13 nominal species are currently recognized: *R. granulosa*, *R. pygmaea*, *R. bergi*, *R. major*, *R. mirandaribeiroi*, *R. azarai*, *R. nattereri*, *R. fernandezae*, *R. dorbignyi*, *R. merianae*, *R. humboldti*, *R. bernardoi*, and *R. centralis* (Narvaes and Rodrigues 2009). These species are distributed across much of the Neotropics, with partially overlapping ranges among species with adjacent distributions (Simon et al. 2016). Although studies incorporating multi-locus datasets have found marked genetic structure often corresponding to proposed taxon boundaries, several relationships remain poorly resolved, with conflicting patterns among loci. While these conflicts have at times been attributed to limited variability in the loci targeted or limited taxonomic sampling, some studies have proposed that poorly resolved relationships may stem from conflicting signals across markers due to hybridization between species (Pereyra et al. 2016; Simon et al. 2016). Within the clade, potential hybridization has been hypothesized to occur between the taxon pairs *R. bergi* and *R. major*, *R. granulosa* and *R. mirandaribeiroi*, and *R. fernandezae* and *R. dorbignyi*. However, there has been no formal attempt to test the hybridization scenario versus alternative gene discordance-generating mechanisms based on an adequate number of independent genomic regions.

In this contribution, we implement a phylogenomic approach to formally test the hypothesis of introgressive hybridization as a source of mito-nuclear discordance across taxa within the *Rhinella granulosa* species group. Based on comprehensive geographic sampling of genome-wide patterns of variation in multiple taxa across South America's major biomes, we infer the frequency and extent of introgression based on a comparison of mitochondrial versus nuclear genetic signal, gene flow estimates and alternative models of historical demography. Based on this approach, we seek to address the following questions: Is there introgression between species within this clade, and how widespread is it? Are zones of introgression restricted to certain geographic regions, and where are they located? Lastly, how do patterns of introgression within the *R. granulosa* clade compare to other similarly distributed toad clades where introgression has recently been demonstrated to be rampant?

METHODS

Sampling of molecular data

Our genetic sampling included 45 individuals belonging to a subset of taxa in the *Rhinella granulosa* species group, namely seven individuals of *R. granulosa*, eight *R. major*, 14 *R. mirandaribeiroi*, 13 *R. merianae*, one *R. centralis*, and two *R. humboldti*. As outgroups, we included four samples from the *Rhinella margaritifera* species group. Tissue samples were obtained from the Museum of Zoology of the University of São Paulo (MZUSP), the Amphibian and Reptile Diversity Research Center (ARDRC) at the University of Texas in Arlington, and the Louisiana State University Museum of Natural Science (LSUMNS).

We extracted genomic DNA using a phenol-chloroform extraction protocol with SeraPure SpeedBead cleanup (Sambrook and Russell 2006). Fragments of the mitochondrial 16S rRNA gene were PCR-amplified using the 16Sar and 16Sbr primers and sequenced on an ABI 3730xL sequencer following Rivera et al. (2021). We also supplemented the 16S dataset with

additional sequences from *R. bernardoi*, *R. merianae*, *R. mirandaribeiroi*, *R. pygmaea*, *R. major*, *R. bergi*, *R. centralis*, *R. dorbignyi*, *R. fernandezae*, *R. granulosa*, and *R. humboldti* obtained from GenBank. The only species not sampled from this group for the 16S phylogeny was *R. nattereri*. Sequences were edited and aligned in Geneious Prime 2020.0.4 (Identification and Accession numbers in Supplementary Table S1).

We generated a double-digest restriction-site associated DNA (ddRAD) dataset following Peterson et al. (2012). Briefly, 200–500 ng of genomic DNA were digested using the *SbfI* (restriction site 5'-CCTGCAGG-3') and *MspI* (restriction site 5'-CCGG-3') enzymes in a single reaction using the manufacturer's recommended buffer (New England Biolabs) for 5 h at 37 °C. Digested DNA was bead-purified before ligating barcodes and index adapters, then samples with the same index were pooled and size-selected (415–515 bp) on a Blue Pippin Prep size selector (Sage Science). The quality and concentration of final libraries were analyzed and quantified on a BioAnalyzer (Agilent) and Qubit Fluorometer 4 (Thermo Fisher Scientific). The resulting 100 bp single-end libraries were sequenced at MedGenome on an Illumina HiSeq2500.

We used the command line version of ipyrad v. 0.9.45 (Eaton and Overcast 2020) (available at <https://ipyrad.readthedocs.io>) to de-multiplex and assign reads to individuals based on sequence barcodes allowing no mismatches from individual barcodes, perform reference read assembly using a minimum clustering similarity threshold of 0.90, align the reads into loci, and call single nucleotide polymorphisms (SNPs). As a reference, we used the *Rhinella marina* genome (Edwards et al. 2018). A minimum Phred quality score (=33), sequence coverage (=6x), read length (=35 bp), and maximum proportion of heterozygous sites per locus (=0.5) were enforced, while ensuring that variable sites had no more than two alleles (i.e., a diploid genome). Following the initial assembly, we used Matrix Condenser (de Medeiros and Farrell 2018) to assess levels of missing data across samples and then re-assembled our dataset to ensure a maximum of 25% missing data within each locus (i.e., each retained locus was sequenced in at least 25% of the sampled individuals). This strategy resulted in 45 ingroup and four outgroup samples in a matrix with less than 15% total missing data. The final dataset was composed of 16,455 SNPs over 1542 RAD loci.

Phylogenetic relationships

We inferred maximum likelihood phylogenies for both the mitochondrial dataset and a concatenated ddRAD loci dataset using IQTREE v2.1.2, utilizing the built-in model selection tool ModelFinder Plus, implementing 1000 ultrafast bootstraps only testing models of evolution and partition schemes available in MrBayes (Hoang et al. 2018; Kalyaanamoorthy et al. 2017; Minh et al. 2020). In addition to the maximum likelihood analyses, we performed phylogenetic inference for both the mitochondrial dataset and a concatenated ddRAD loci dataset under a Bayesian framework using MrBayes 3.2.6 (Ronquist et al. 2012), implementing three independent runs of four Markov chains of 10 million generations each and sampling every 1000 generations with the first 25% generations discarded as burn-in. We used the same best-fit model found by ModelFinder Plus in the Bayesian analyses. We used Tracer 1.7 (Rambaut et al. 2018) to assess whether Markov chain mixing was adequate (effective sample sizes >200) and to visually assess model parameter stationarity and convergence between runs. We then summarized a 50% majority-rule consensus tree and used iTol to edit and visualize trees (Letunic and Bork 2019).

We also compared the nuclear IQTREE results to a species-tree inferred under a multispecies coalescent framework utilizing the nuclear data in SVD Quartets (Chifman and Kubatko 2014) as implemented in the command line version of PAUP v. 4 (Swofford 2002). In this analysis, we sampled all possible quartets and estimated node support based on 1000 bootstrap replicates.

Population genetic structure

To determine populations and admixture in the *R. granulosa* species group, we filtered SNPs as described above but excluding outgroup samples using ipyrad and VCFtools (Danecek et al. 2011), resulting in a dataset composed of 20,527 SNPs. SNPs were filtered further to only include one random SNP per RAD locus, resulting in a final dataset of 1527 unlinked SNPs. We then determined the best-fit number of genetic clusters (K) using the maximum likelihood method ADMIXTURE with 20 replicates per K and a 10-fold crossvalidation to determine the best-fit K (Alexander et al. 2009). Specifically, the best-fit K was determined as the replicate with the lowest cross-validation error. In addition to our ADMIXTURE analysis, we used TESS3r in R, running 50 replicates per K, to both assess genomic admixture

as well as estimate and visualize spatially informed population structure in geographic space (Caye et al. 2016, 2018).

To further characterize population structure, we used the non-parametric method of discriminant analysis of principal components (DAPC), implemented in the R package *ade4* (Jombart and Ahmed 2011; Jombart et al. 2010). The *find.clusters* function was used to test the fit of 1–20 clusters (K). The K with the lowest Bayesian information criterion (BIC) score was considered the best-fit number of clusters.

The resulting ancestry coefficient matrices (Q-matrices) from the ADMIXTURE analysis were then imported into QGIS (QGIS Geographic Information System. QGIS Association. <http://www.qgis.org>) to allow visualizing patterns of genetic clustering and admixture in geographic space.

Gene flow

To infer potential hybridization or gene flow among taxa, we calculated Patterson's D statistic and the related admixture fraction estimates, or f_4 -ratio statistics, using the program Dsuite (Patterson et al. 2012; Malinsky et al. 2020). This approach, also known as an ABBA-BABA test, uses a four-taxon fixed phylogeny—in the form (((P1,P2)P3)O)—to quantify the proportion of shared alleles that can be attributed to horizontal transfer among the populations considered. This approach assumes that a typical ancestral ("A") and derived ("B") allele pattern among four terminal taxa should generate a BBAA structure. Under incomplete lineage sorting, conflicting ABBA and BABA patterns should occur in equal frequencies, resulting in a D statistic = 0. If, however, introgression between P3 and P1 or P2 has occurred, there should be an excess of one of these two patterns, and thus a D statistic significantly different from 0 (Durand et al. 2011; Patterson et al. 2012). We used the f_4 -branch or $f_b(C)$ metric to tease apart potentially correlated f_4 -ratio statistics and estimate gene flow events between internal branches on the phylogeny (Martin et al. 2013; Malinsky et al. 2018). Dsuite uses a VCF file and a jackknifing approach to assess correlations in allele frequencies between closely related species (Malinsky et al. 2020). Within Dsuite, we used the *Dtrios* and *Fbranch* programs to identify introgression between all combinations of species, using *R. margaritifera* as the outgroup taxa, and applied the Benjamini-Hochberg (BH) correction to control for false discovery rate using a p value of 0.05.

Demographic modeling

We use the full-likelihood, multispecies coalescent method Generalized Phylogenetic Coalescent Sampler (G-PhoCS) v.1.3.2 on the CIPRES Science Gateway to estimate demographic parameters such as effective population sizes and divergence times along the evolutionary history of the *Rhinella granulosa* species group (Miller et al. 2010; Gronau et al. 2011). For this analysis, we used the phylogenetic topology inferred by our maximum likelihood and Bayesian analyses on the ddRAD data (see Results). To reduce computing time, we used a maximum of 10 randomly chosen individuals per delimited genetic cluster and re-ran this dataset through the filtering pipeline resulting in 2744 unlinked SNPs. G-PhoCS ran an automatic fine-tuning for 500 steps, running the entire analysis twice independently for 500,000 generations, then again for one million generations, sampling every 1000 generations. We then merged runs using LogCombiner and discarded the first 10% of samples as burn-in (Pramuk et al. 2007; Drummond and Rambaut 2007; Alonso et al. 2012). The Dsuite analysis determined no significant gene flow between populations (see Results), so we did not apply migration bands in GPhoCS analyses. We followed Prates et al. (2018) to select distributions for the priors of the θ and τ parameters (scripts available at https://github.com/ivanprates/2018_Anolis_EcolEvol), applying a gamma distribution with parameters $\alpha = 2.0$ and $\beta = 30$. We used TRACER 1.7 to assess proper chain mixing and convergence based on the log output files (Rambaut et al. 2018). We then converted θ estimates to N_e , in number of individuals, using the relationship $\theta = 4N_e\mu$, and τ estimates to T , in years, using the relationship $\tau = T\mu/g$. To apply these conversions, we used a nuclear mutation rate (μ) of 2.4×10^{-9} (Prates et al. 2016, 2018) and a generation time (g) of 2 years for bufonid toads (Lever 2001).

RESULTS

Phylogenetic relationships

Phylogenetic analyses of the 16S mitochondrial gene based on both maximum likelihood and Bayesian approaches yielded nearly identical relationships (Fig. 1). Overall, most species-level taxa currently recognized within the *R. granulosa* species group were

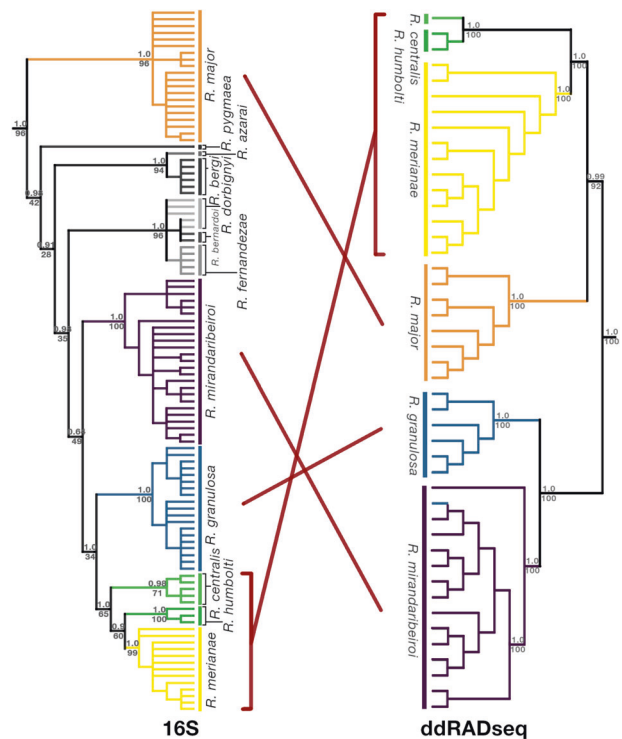


Fig. 1 A comparison of the *Rhinella granulosa* clade 16S and ddRAD phylogenies (IQTree/MrBayes). Support values are indicated above (PP) and below (BS) each node in gray. Red lines highlight mito-nuclear discordance.

inferred as monophyletic. Mitochondrial analyses inferred 10 major clades, most of which corresponded to currently recognized taxa. Moreover, most of these clades were geographically coherent, with samples assigned to the same taxon generally clustering in geographic space, as follows: *R. dorbignyi* and *R. fernandezae* in the southern Pampas; *R. major* across southern and eastern Amazonia; *R. granulosa* in the Caatinga and northern Atlantic Forest; *R. centralis* in the northern Andes, *R. humboldti* in northern Amazonia; *R. merianae* in central Amazonia; and *R. mirandaribeiroi* across the Cerrado (Locality map characterized by 16S data is available in Supplementary Fig. S1).

The Bayesian mitochondrial analysis resulted in a phylogeny with higher overall relative support for most of the internal nodes (PP = 0.90–1.0) that correspond to relationships between species-level taxa. The only relationship that received lower support (PP = 0.64) was the position of *R. mirandaribeiroi* and the clade composed of *R. granulosa*, *R. centralis*, *R. humboldti*, and *R. merianae*. (Fig. 1). This phylogeny placed *R. major* as the sister to all other remaining taxa within this clade; within this clade, *R. pygmaea*, *R. azarai* and *R. bergi* formed their own respective clades, and *R. dorbignyi* and *R. fernandezae* formed a subclade. The other major clade within the species group included *R. mirandaribeiroi*, *R. granulosa*, *R. centralis*, *R. humboldti*, and *R. merianae*, nested in this order relative to the root.

Phylogenetic relationships based on the ddRADseq dataset—which included only a subset of the taxa represented in the mitochondrial trees—were overall highly supported (Figs. 1 and 2A). Similar to the mitochondrial analyses, all of the currently recognized species were inferred to be monophyletic and geographically coherent, except for one *R. granulosa* individual, which was inferred as nested within the *R. mirandaribeiroi* clade (Figs. 1 and 2A). The ddRAD phylogeny included two clades: one composed of *R. granulosa* and *R. mirandaribeiroi*, and another composed of *R. major*, *R. merianae*, *R. humboldti*, and *R. centralis* (Figs. 1 and 2A). Nuclear phylogenetic analyses under both the

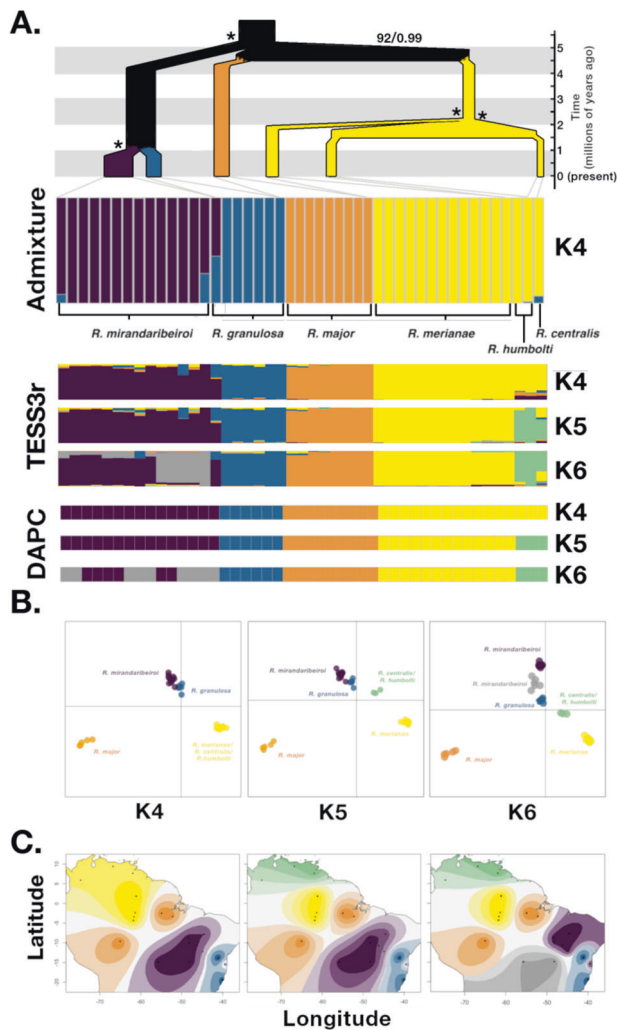


Fig. 2 Population structure of the *Rhinella granulosa* clade. **A** Maximum likelihood/Bayesian phylogeny incorporating data from GPhoCS for *R. granulosa* clade focal species using ddRADseq data, and corresponding ADMIXTURE, DAPC, and TESS3r plots. Branch widths are scaled to estimated population sizes (G-PhoCS) and asterisks represent BS/PP > 95/0.99. **B** DAPC plots for $K=4-6$, **C** TESS3r plots for $K=4-6$.

concatenated (IQTREE, MrBayes; Fig. 1) and multispecies coalescent (SVD Quartets; Supplementary Fig. S2) approaches yielded the same species topology.

While the sample composition of clades was similar between mitochondrial and nuclear analyses, the relationships between taxon clades differed between marker types. Most notably, the affinity between *R. major* and the clade composed of *R. centralis*, *R. humboldti*, and *R. merianae* varied by marker type. *Rhinella major* was sister to all other species in the 16S phylogeny, while in the nuclear phylogeny it was inferred to form a clade with *R. centralis*, *R. humboldti*, and *R. merianae* (Fig. 1). Additionally, the placement of *R. mirandaribeiroi* in relation to *R. granulosa* differed between mitochondrial and nuclear markers in the 16S phylogeny. *R. mirandaribeiroi* was sister to a clade formed by *R. granulosa*, *R. centralis*, *R. humboldti*, and *R. merianae*, while in the nuclear phylogeny, *R. mirandaribeiroi* shared a clade with *R. granulosa* only (Fig. 1).

Population structure

ADMIXTURE results indicated the K value with the lowest crossvalidation error was four, which was a cluster formed by

samples assigned to *R. major* (orange), one *R. granulosa* cluster (blue), one *R. mirandaribeiroi* cluster (purple), and a cluster combining samples assigned to *R. merianae*, *R. humboldti*, and *R. centralis* (yellow; Fig. 2A). The only individuals that were found to be admixed had alleles that traced back to the *R. granulosa* and the *R. mirandaribeiroi* clusters (Fig. 2A). The DAPC analysis yielded similar BIC scores for $K=4-6$ (Supplementary Fig. S3). $K=4$ resulted in the same clusters identified by ADMIXTURE, while $K=5$ split the ADMIXTURE cluster composed of *R. merianae*, *R. humboldti*, and *R. centralis* into two clusters, one composed of *R. merianae* samples and the other of *R. humboldti* and *R. centralis* samples (Fig. 2). $K=6$ further subdivided *R. mirandaribeiroi* into two clusters, one of which included the *R. granulosa* individual inferred to be nested within the *R. mirandaribeiroi* cluster (Fig. 2A, B). ADMIXTURE results for $K=5-6$ yielded the same clusters as the results from the DAPC analysis (Supplementary Fig. S3; cross-validation plots: Supplementary Fig. S4). TESS3r results were very similar to both ADMIXTURE and DAPC results (Fig. 2A, B). Overall, there was slight amounts of admixture among *R. mirandaribeiroi*, *R. granulosa*, *R. humboldti*, and *R. centralis* for $K=4-6$ (Fig. 2A).

Similar to the phylogenetic analyses, major genetic clusters inferred by ADMIXTURE were geographically coherent, each restricted to a certain portion of South America. Specifically, the cluster composed of samples assigned to *R. major* was restricted to the Amazon region; the cluster composed of *R. merianae*, *R. humboldti*, and *R. centralis* samples was distributed across the Amazon and the Northern Andes; the cluster composed of samples assigned to *R. granulosa* was distributed across the Northern Atlantic Forest in eastern Brazil; and the cluster composed of samples assigned to *R. mirandaribeiroi* was distributed across the Cerrado in central Brazil (Fig. 3). The spatially explicit TESS3r resulting maps showing the diffusion of ancestry coefficients across geographic space shows genetic groups are geographically clustered for $K=4-6$ (Fig. 2C).

D-statistics

Analyses of introgression based on the ddRAD data using Dsuite resulted in only one trio ((*R. mirandaribeiroi*, *R. granulosa*) *R. major*) out of 20 trios tested that had a significant D-statistic (0.41; Supplementary Table S2), pointing to a larger proportion of shared alleles between *R. granulosa* and *R. major* than expected based solely on incomplete lineage sorting. However, after applying the Benjamini-Hochberg (BH) correction to control for the false discovery rate, none of the trios tested were significant (Supplementary Table S2). These results were corroborated using the $f_b(C)$ metric analysis, which showed an f_b statistic of 0.15 between *R. granulosa* and *R. major*, but it was not significant (Fig. 4 and Supplementary Table S3). Taken together, these results support that introgression is not an important factor behind shared allele patterns among any of the groups considered.

Demographic inference

The G-PhoCS analysis estimated the *Rhinella granulosa* species group to date back to about 5 million years before present (mybp) (mean value; 95% highest posterior density (HPD) = 4.817–5.092; Fig. 2 and Supplementary Table S4). Sister species divergences within the species group dated back to the mid-Pleistocene (Fig. 2 and Supplementary Table S4). The divergence between *R. granulosa* and *R. mirandaribeiroi* was estimated to date back to 1.026 mybp (HPD = 0.983–1.075), while the divergence between *R. humboldti* and *R. centralis* dated back to 1.496 mybp (HPD = 1.258–1.733). In turn, the most recent common ancestor of *R. merianae*, *R. humboldti*, and *R. centralis* dated back to 2.263 mybp (2.158–2.358 95% HPD), while the ancestor of these three species and *R. major* dated back to 4.826 mybp (HPD = 4.625–4.983 95% HPD). The effective population size estimates resulted in a root population size of ~1.3 million,

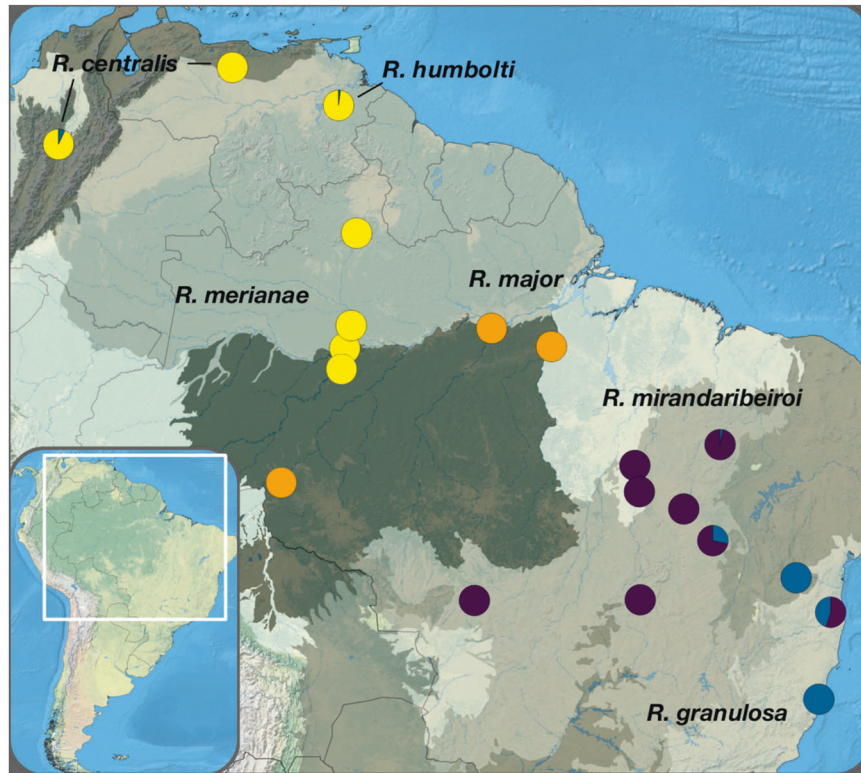


Fig. 3 Locality map of focal species depicting average ADMIXTURE cluster assignments per locality ($K = 4$). Map partitioned into biomes (Central Andes, Northern Amazonia, Western Amazonia, Eastern Amazonia, Southern Amazonia, Pantanal, Chaco, Cerrado, Caatinga, Northern Atlantic Forest, Southern Atlantic Forest).

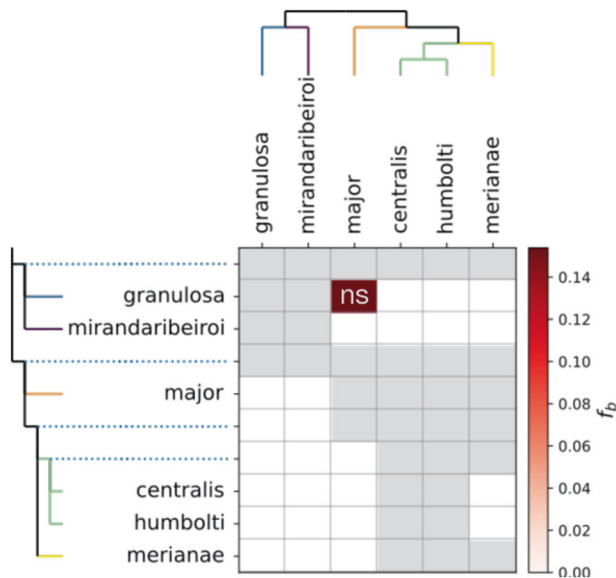


Fig. 4 The f_b statistic (summary of f_4 admixture ratios). Gray color corresponds to tests that are not possible because of constraints on the phylogeny. ns = not significant.

remaining stable along the *R. granulosa*-*mirandaribeiroi* branch. This lineage massively expanded to Amazonia (along the *R. major*-*merianae*-*humbolti*-*centralis* branch), then experienced a bottleneck after *R. major* diverged, and then expanded again after *R. merianae* diverged (Fig. 2 and Supplementary Table S4). Estimates of current population sizes for each species ranged from ~145,833–1,093,750 (Supplementary Table S4).

DISCUSSION

Based on range-wide geographic sampling of mitochondrial and nuclear genetic variation in a widespread species group of neotropical toads, we found only partial concordance between currently recognized species-level taxon boundaries and patterns of genetic structure. While most taxa within the *R. granulosa* species group corresponded to clades, genetic clustering analyses sometimes grouped distinct taxonomic units into a single cluster. Moreover, levels of admixture between inferred clusters were limited and restricted to a single sister taxon pair (*R. granulosa* and *R. mirandaribeiroi*), while D-statistics estimation provided no support for the hypothesis that allele sharing between non-sister taxa is a result of horizontal transfer. These findings contradict previous assertions of widespread cryptic diversity and of gene flow and introgression across taxa in the *R. granulosa* clade, a proposed scenario based on small numbers of loci (Guerra et al. 2011; Pereyra et al. 2016). Lastly, our analyses suggest that diversification events within the *Rhinella granulosa* clade mostly dated back to the early Pliocene, which is later than species divergences in other closely related clades, as observed in the closely-related *R. marina* species group (Rivera et al. 2022).

Previous analyses of phylogenetic relationships in the *R. granulosa* clade based on multi-locus datasets have resulted in unresolved interspecific relationships and evidence of mitochondrial nuclear discordance (Azevedo et al. 2003; Sequeira et al. 2011; Pereyra et al. 2016). By incorporating thousands of ddRADseq nuclear loci, however, we were able to improve phylogenetic resolution within this clade (Fig. 2). Previous studies have, in several cases, inferred relationships similar to ours. Specifically, both our study and that of Pereyra et al. (2016) found a sister relationship between *R. granulosa* and *R. mirandaribeiroi*, as well as phylogenetic clustering of *R. centralis*, *R. humboldti*, and *R. merianae*. However, while Pereyra et al. (2016) found *R. major* to be the sister of all other species, our analysis inferred *R. major* as sister to the

clade composed of *R. centralis*, *R. humboldti*, and *R. merianae* with high support (Fig. 2). These discrepancies between studies could have originated from differences in the genetic markers used. For instance, although we did not observe a clear signal of mitochondrial introgression across taxa based on 16S, an analysis by Pereyra et al. (2016) incorporating other mitochondrial genes found several taxa to be nonmonophyletic. This is the case, for instance, of *R. bergi*, *R. major*, and *R. fernandezae*, which were thus hypothesized to experience past introgressive events. By assessing the mitochondrial DNA in this study separate from the thousands of nuclear markers generated, our analysis of genetic structure and introgression based on extensive nuclear data found no support for the idea that hybridization played a major role in the evolutionary history of these toads.

Besides hybridization, another mechanism that may generate a pattern of allele sharing among non-sister taxa is incomplete lineage sorting, as reported in studies of several other clades, including birds, lizards, mammals, plants, and flies (Jakob and Blattner 2006; Pollard et al. 2006; McGuire et al. 2007; McKay and Zink 2010; Wang et al. 2018). Our analyses of the *R. granulosa* group are consistent with the idea that instances of mitochondrial discordance can be attributed to incomplete lineage sorting. For instance, *R. granulosa* and the genetic group formed by the samples originally assigned to *R. centralis*, *R. humboldti*, and *R. meriane* were inferred as sister clades in the mitochondrial phylogeny but not in the nuclear one. The D-statistics estimates supported that the observed proportion of shared nuclear alleles among these two groups is no different from that expected under incomplete lineage sorting. Incomplete lineage sorting has been similarly invoked to explain similar patterns of genealogical heterogeneity in phylogenetic studies across the tree of life (Liu et al. 2009; Blanco-Pastor et al. 2012; Doronina et al. 2015). Given our sampling of the species group, a full sampling of all ~13 species should be used to more robustly assess gene flow and demographic change among species in a hypothesis-testing framework.

Alternatively, a pattern of taxon paraphyly in gene genealogies can result from purifying selection, which reduces genetic diversity at linked sites, such as the different genes that comprise the mitochondrial genome (Cvijović et al. 2018). Accordingly, studies have found signatures of selection on the mitochondrial genes of species that show patterns of mitochondrial discordance (Morales et al. 2015). It is currently unclear whether selection may have contributed to a pattern of paraphyletic taxa as inferred by mitochondrial genes in the *R. granulosa* species group (Pereyra et al. 2016). Nevertheless, our analyses of thousands of nuclear loci suggest that shared alleles across non-sister taxa conform to expectations under incomplete lineage sorting and not introgressive hybridization. As such, our analyses challenge previous assertions of rampant hybridization in the evolutionary history of the *R. granulosa* species group, similar to recent proposals for other vertebrate clades (Wang et al. 2018). We note that our taxon sampling was more limited than that of previous studies of this clade due to challenges in accessing samples from throughout its vast distribution. Additionally, the relatively short 16S fragment sequenced in this study (564 base pairs) precludes accurate tests of selection among this group. It is possible that inclusion of additional taxa, as well as better sampling of each taxon across their respective ranges, especially near any potential contact zones, could reveal more complex patterns of admixture.

Our finding of negligible introgression across taxa in the *R. granulosa* species group contrasts with patterns observed in other closely related toad clades. For instance, the *Rhinella marina* species group experienced many instances of genetic admixture among species over its evolutionary history. Recent investigation based on thousands of genome-wide loci inferred multiple instances of admixture across and within taxa in the *R. marina*

group, as well as past and ongoing pulses of gene flow (Rivera et al. 2022). Why we observe such distinct patterns of introgression across taxa in two closely related toad clades is unclear, especially considering the *R. marina* group is similarly diverse, though this particular study had larger sample sizes in comparison. Interestingly, higher levels of introgression do not seem to be associated with recent population divergences, a pattern suggested for other species groups (Singhal and Moritz 2013; Singhal and Bi 2017). Instead, estimates for the *R. marina* clade suggest older divergences overall than those in the *R. granulosa* clade (Rivera et al. 2022).

Instead, it is possible that these contrasting patterns across clades reflect differences in organismal attributes. Toads are generally reported to be prone to low selectivity of mates during their typically explosive breeding events, which also seems to apply to the *R. granulosa* species group; for instance, interspecific amplexus between co-distributed *Rhinella granulosa* and other species from the *Rhinella* and *Ceratophrys* genera have been reported (Abreu et al. 2021; Gama et al. 2020). Nevertheless, our findings could suggest that these apparently permeable prezygotic reproductive barriers do not necessarily lead to gene transfer across interbreeding species. If that is the case, it is possible that species in the *R. granulosa* species group have developed post-zygotic barriers to hybridization that species in other closely related clades have not, as appears to be the case of the *R. marina* species group, though more evidence is required to test this assertion (Pereyra et al. 2016; Bonnet et al. 2017; Abreu et al. 2021). Future studies of genetic structure across clades will benefit from incorporating information on species' reproductive biology and natural history. Unfortunately, this much-needed information is largely lacking for species-rich clades of tropical organisms, as is the case of the intriguing true toads.

DATA AVAILABILITY

Sequence data are available on GenBank (ON692445–ON692509) and demultiplexed sequence files are available on the Sequence Read Archive (PRJNA858546). Associated scripts and input files have been made available on dryad: <https://doi.org/10.5061/dryad.p2ngf1vtc>.

REFERENCES

- Abreu LM, Santos JEM, Corrêa AT, Ribeiro MV (2021) Interspecific amplexus between two sympatric species, *Rhinella granulosa* (Spix, 1824) and *Rhinella crucifer* (Wied-Neuwied, 1821) (Anura: Bufonidae) in Ilhéus, Bahia, Brazil. HNO 14:869–871
- Alexander DH, Novembre J, Lange K (2009) Fast model-based estimation of ancestry in unrelated individuals. Genome Res 19:1655–1664
- Alonso R, Crawford AJ, Bermingham E (2012) Molecular phylogeny of an endemic radiation of Cuban toads (Bufonidae: *Peltophryne*) based on mitochondrial and nuclear genes: origin and diversification of Cuban toads. J Biogeogr 39:434–451
- Arntzen JW, de Vries W, Canestrelli D, Martínez-Solano I (2017) Hybrid zone formation and contrasting outcomes of secondary contact over transects in common toads. Mol Ecol 26:5663–5675
- Azevedo MFC, Foresti F, Ramos PRR, Jim J (2003) Comparative cytogenetic studies of *Bufo ictericus*, *B. paracnemis* (Amphibia, Anura) and an intermediate form in sympatry. Genet Mol Biol 26:289–294
- Blanco-Pastor JL, Vargas P, Pfeil BE (2012) Coalescent simulations reveal hybridization and incomplete lineage sorting in Mediterranean *Linaria*. PLoS ONE 7:e39089
- Bonnet T, Leblois R, Rousset F, Crochet P-A (2017) A reassessment of explanations for discordant introgressions of mitochondrial and nuclear genomes. Evolution 71:2140–2158
- Caye K, Jay F, Michel O, Francois O (2018) Fast inference of individual admixture coefficients using geographic data. Ann Appl Stat 12:586–608
- Caye K, Deist T, Martins H, Michel O, Francois O (2016) TESS3: fast inference of spatial population structure and genome scans for selection. Mol Ecol Resour 16:540–548
- Chifman J, Kubatko L (2014) Quartet inference from SNP data under the coalescent model. Bioinformatics 30:3317–3324

- Chiocchio A, Arntzen JW, Martínez-Solano I et al. (2021) Reconstructing hotspots of genetic diversity from glacial refugia and subsequent dispersal in Italian common toads (*Bufo bufo*). *Sci Rep* 11:260
- Cvijović I, Good BH, Desai MM (2018) The effect of strong purifying selection on genetic diversity. *Genetics* 209:1235–1278
- Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA et al. (2011) The variant call format and VCFtools. *Bioinformatics* 27:2156–2158
- Degnan JH, Rosenberg NA (2009) Gene tree discordance, phylogenetic inference and the multispecies coalescent. *Trends Ecol Evol* 24:332–340
- Doronina L, Churakov G, Shi J, Brosius J, Baertsch R, Clawson H et al. (2015) Exploring massive incomplete lineage sorting in Arctoids (Laurasiatheria, Carnivora). *Mol Biol Evol* 32:3194–3204
- Drummond AJ, Rambaut A (2007) BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol* 7:214
- Dufresnes C, Nicieza AG, Litvinchuk SN, Rodrigues N, Jeffries DL, Vences M et al. (2020) Are glacial refugia hotspots of speciation and cytonuclear discordances? Answers from the genomic phylogeography of Spanish common frogs. *Mol Ecol* 29:986–1000
- Durand EY, Patterson N, Reich D, Slatkin M (2011) Testing for ancient admixture between closely related populations. *Mol Biol Evol* 28:2239–2252
- Eaton DAR, Overcast I (2020) ipyrad: Interactive assembly and analysis of RADseq datasets. *Bioinformatics* 36:2592–2594
- Edwards RJ, Tuipulotu DE, Amos TG, O'Meally D, Richardson MF, Russell TL et al. (2018) Draft genome assembly of the invasive cane toad, *Rhinella marina*. *GigaScience* 7:1–13
- Firreno Jr TJ, Townsend JH (2019) Evaluation of species boundaries in sympatric and parapatric populations of Mesoamerican toads. *Zool Scr* 48:454–465
- Firreno Jr TJ, 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
- 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
- Gama V, dos Santos Protázio A, dos Santos, Protázio A (2020) Interspecific amplexus between male *Rhinella granulosa* (Spix, 1824) and a metamorph of *Ceratophrys joazeirensis* Mercadal, 1986 (Amphibia: Anura) in a temporary pond in the Caatinga, Bahia State, Northeastern Brazil. *Herpetol Notes* 13:749–751
- Gronau I, Hubisz MJ, Gulko B, Danko CG, Siepel A (2011) Bayesian inference of ancient human demography from individual genome sequences. *Nat Genet* 43:1031–1034
- Guerra C, Baldo D, Rosset S, Borteiro C, Kolenc F (2011) Advertisement and release calls in Neotropical toads of the *Rhinella granulosa* group and evidence of natural hybridization between *R. bergi* and *R. major* (Anura: Bufonidae). *Zootaxa* 3092:26–42
- Hill GE, Havird JC, Sloan DB, Burton RS, Greening C, Dowling DK (2019) Assessing the fitness consequences of mitonuclear interactions in natural populations. *Biol Rev Camb Philos Soc* 94:1089–1104
- Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS (2018) UFBoot2: improving the ultrafast bootstrap approximation. *Mol Biol Evol* 35:518–522
- Ivanov V, Lee KM, Mutanen M (2018) Mitonuclear discordance in wolf spiders: genomic evidence for species integrity and introgression. *Mol Ecol* 27:1681–1695
- Jakob SS, Blattner FR (2006) A chloroplast genealogy of *Hordeum* (Poaceae): long-term persisting haplotypes, incomplete lineage sorting, regional extinction, and the consequences for phylogenetic inference. *Mol Biol Evol* 23:1602–1612
- Jombart T, Ahmed I (2011) adegenet 1.3-1: new tools for the analysis of genomewide SNP data. *Bioinformatics* 27:3070–3071
- Jombart T, Devillard S, Balloux F (2010) Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genet* 11:94
- Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermini LS (2017) ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods* 14:587–589
- Leaché AD, Portik DM, Rivera D, Rödel MO, Gvoždík V, Penner J et al. (2019) Exploring rainforest diversification using demographic model testing in the African foam-nest treefrog (*Chiromantis rufescens*). *J Biogeogr* 46:2706–2721
- Letunic I, Bork P (2019) Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res* 47:W256–W259
- Lever C (2001) The cane toad: the history and ecology of a successful colonist. Westbury Academic and Scientific Publishing, Otley, Yorkshire, UK
- Liu L, Yu L, Kubatko L, Pearl DK, Edwards SV (2009) Coalescent methods for estimating phylogenetic trees. *Mol Phylogenet Evol* 53:320–328
- Malinsky M, Matschner M, Svardal H (2020) Dsuite—Fast D-statistics and related admixture evidence from VCF files. *Mol Ecol Resour* 21:584–595
- Malinsky M, Svardal H, Tyers AM, Miska EA, Genner MJ, Turner GF et al. (2018) Whole-genome sequences of Malawi cichlids reveal multiple radiations interconnected by gene flow. *Nat Ecol Evol* 2:1940–1955
- Martin SH, Dasmahapatra KK, Nadeau NJ, Salazar C, Walters JR, Simpson F et al. (2013) Genome-wide evidence for speciation with gene flow in *Heliconius* butterflies. *Genome Res* 23:1817–1828
- McGuire JA, Linkem CW, Koo MS, Hutchison DW, Lappin AK, Orange DI et al. (2007) Mitochondrial introgression and incomplete lineage sorting through space and time: phylogenetics of crotophytid lizards. *Evolution* 61:2879–2897
- McKay BD, Zink RM (2010) The causes of mitochondrial DNA gene tree paraphyly in birds. *Mol Phylogenet Evol* 54:647–650
- de Medeiros BAS, Farrell BD (2018) Whole-genome amplification in doubledigest RADseq results in adequate libraries but fewer sequenced loci. *PeerJ* 6:e5089
- Miller MA, Pfeiffer W, Schwartz T (2010) Creating the CIPRES Science Gateway for inference of large phylogenetic trees. 2010 Gateway Computing Environments Workshop (GCE), p 1–8
- Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A et al. (2020) IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol Biol Evol* 37:1530–1534
- Morales HE, Pavlova A, Joseph L, Sunnucks P (2015) Positive and purifying selection in mitochondrial genomes of a bird with mitonuclear discordance. *Mol Ecol* 24:2820–2837
- Narvaes P, Rodrigues MT (2009) Taxonomic revision of *Rhinella granulosa* species group (Amphibia, Anura, Bufonidae), with a description of a new species. *Arquivos De Zoologia* 40:1–73
- Patterson N, Moorjani P, Luo Y, Mallick S, Rohland N, Zhan Y et al. (2012) Ancient admixture in human history. *Genetics* 192:1065–1093
- Pereyra MO, Baldo D, Blotto BL, Iglesias PP, Thomé MTC, Haddad CFB et al. (2016) Phylogenetic relationships of toads of the *Rhinella granulosa* group (Anura: Bufonidae): a molecular perspective with comments on hybridization and introgression. *Cladistics* 32:36–53
- Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE (2012) Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLoS ONE* 7:e37135
- Pollard DA, Iyer VN, Moses AM, Eisen MB (2006) Widespread discordance of gene trees with species tree in *Drosophila*: evidence for incomplete lineage sorting. *PLoS Genet* 2:e173
- Portik DM, Leaché AD, Rivera D, Barej MF, Burger M, Hirschfeld M et al. (2017) Evaluating mechanisms of diversification in a Guineo-Congolian forest frog using demographic model selection. *Mol Ecol* 26:5245–5263
- Pramuk JB, Robertson T, Sites JW, Noonan BP (2007) Around the world in 10 million years: biogeography of the nearly cosmopolitan true toads (Anura: Bufonidae). *Glob Ecol Biogeogr* 17:72–83
- Prates I, Rivera D, Rodrigues MT, Carnaval AC (2016) A mid-Pleistocene rainforest corridor enabled synchronous invasions of the Atlantic Forest by Amazonian anole lizards. *Mol Ecol* 25:5174–5186
- Prates I, Penna A, Rodrigues MT, Carnaval AC (2018) Local adaptation in mainland anole lizards: Integrating population history and genome-environment associations. *Ecol Evol* 8:11932–11944
- Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA (2018) Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Syst Biol* 67:901–904
- Rivera D, Prates I, Rodrigues MT, Carnaval AC (2020) Effects of climate and geography on spatial patterns of genetic structure in tropical skinks. *Mol Phylogenet Evol* 143:106661
- Rivera D, Prates I, Firreno Jr. TJ, Rodrigues MT, Caldwell JP, Fujita MK (2022) Phylogenomics, introgression, and demographic history of South American true toads (*Rhinella*). *Mol Ecol* 31:978–992
- Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S et al. (2012) MRBAYES 3.2: Efficient Bayesian phylogenetic inference and model selection across a large model space. *Syst Biol* 61:539–542
- Sambrook J, Russell DW (2006) Purification of nucleic acids by extraction with phenol:chloroform. *Cold Spring Harb Protoc*. <https://doi.org/10.1101/pdb.prot4455>
- Sequeira F, Sodré D, Ferrand N, Bernardi JAR, Sampaio I, Schneider H et al. (2011) Hybridization and massive mtDNA unidirectional introgression between the closely related Neotropical toads *Rhinella marina* and *R. schneideri* inferred from mtDNA and nuclear markers. *BMC Evol Biol* 11:264
- Simon MN, Machado FA, Marroig G (2016) High evolutionary constraints limited adaptive responses to past climate changes in toad skulls. *Proc Biol Sci* 283:20161783
- Singhal S, Moritz C (2013) Reproductive isolation between phylogeographic lineages scales with divergence. *Proc Biol Sci* 280:20132246
- Singhal S, Bi K (2017) History cleans up messes: the impact of time in driving divergence and introgression in a tropical suture zone. *Evolution* 71:1888–1899
- Swofford DL (2002) PAUP: phylogenetic analysis using parsimony (and other methods), Version 4.0 Beta 10. Sinauer Associates, Sunderland

- Thielsch A, Kneill A, Mohammadyari A, Petrusek A, Schwenk K (2017) Divergent clades or cryptic species? Mito-nuclear discordance in a *Daphnia* species complex. *BMC Evol Biol* 17:227
- Toews DPL, Brelsford A (2012) The biogeography of mitochondrial and nuclear discordance in animals. *Mol Ecol* 21:3907–3930
- Wang K, Lenstra JA, Liu L, Hu Q, Ma T, Qiu Q et al. (2018) Incomplete lineage sorting rather than hybridization explains the inconsistent phylogeny of the wisent. *Commun Biol* 1:169

ACKNOWLEDGEMENTS

We thank all the Brazilian collectors who made this study possible and particularly students in MTR's laboratory for field collection and support. Brazil's Instituto Chico Mendes de Conservação da Biodiversidade issued collecting permits (SISBIO 36753-1, 36753-4, and 27290-3). This work was co-funded by Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP; BIOTA 2013/50297-0), the National Science Foundation (DEB 1343578), and the National Aeronautics and Space Administration through the Dimensions of Biodiversity Program. JPC was funded by the National Science Foundation (DEB 9200779 and DEB 9505518). MTR acknowledges additional funding from FAPESP grants 2003/10335-8, 2011/50146-6, 2011/50206-9, 2012/15754-8, and 2017/08357-6. DR was funded by NSF GRFP. IP was funded by NSF grant DEB 1754398.

AUTHOR CONTRIBUTIONS

Contributions DR, IP, and MKF discussed and designed the study. IP provided expertise for many analyses. DR compiled the dataset, ran the analyses, and wrote

the first draft and revised versions of the manuscript. IP provided significant input to the first draft and revised versions of the manuscript. JPC, MTR, and MKF contributed to the manuscript.

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41437-022-00571-9>.

Correspondence and requests for materials should be addressed to Danielle Rivera.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.