Machine learning and phylogenetic models identify predictors of genetic variation in Neotropical amphibians

RUNNING TITLE: Neotropical amphibian genetic variation

Luis Amador (laamador@unm.edu; https://orcid.org/0000-0003-2638-4068),

Irvin Arroyo-Torres (irtorre5@unm.edu; https://orcid.org/0000-0001-6811-7422),

Lisa N. Barrow (lnbarrow@unm.edu; https://orcid.org/0000-0001-7081-2432)

Museum of Southwestern Biology and Department of Biology, University of New Mexico, Albuquerque, NM, USA, 87131

Abstract

Aim. Many factors influence genetic variation within species, which is key for adaptation and survival in changing environments. Using amphibians as a study system, we investigated genetic diversity and structure within Neotropical species to provide insights into how natural history traits, phylogeny, and geographic characteristics influence intraspecific genetic variation.

Location. Neotropical Region.

Taxon. Amphibians.

Methods. We assembled datasets for Neotropical amphibians using open-access databases for natural history traits, genetic sequences, phylogenetic trees, and geographic data. For each species, we calculated overall nucleotide diversity (π) and tested for isolation by distance (IBD). We then identified predictors of genetic diversity (π) and genetic structure (IBD) using Random Forest (RF) regression or RF classification, respectively. To more explicitly incorporate phylogenetic relationships, we also used phylogenetic generalized linear mixed models (PGLMMs), and phylogenetic linear regression or phylogenetic logistic regression models to predict π and IBD, respectively.

Results. We compiled > 4,000 mitochondrial DNA sequences from 256 amphibian species (230 frogs and 26 salamanders), georeferencing 2,468 sequences from 176 species that were not already linked to occurrence data. RF regressions, phylogenetic linear regression, and PGLMMs were congruent in identifying range size and elevation as the most important predictors of π. Species with larger ranges and at lower elevations tended to have higher overall genetic diversity. RF classification and phylogenetic comparative methods identified minimum elevation as the most important predictor of IBD, with species at lower elevations being more likely to exhibit genetic structure.

Main conclusions. These results could play a key role in the development of conservation strategies for amphibians. For example, species living at high elevations or with small ranges need substantial

protection due to their low genetic diversity. We also show that the Chocó, Andes, and Atlantic Forests biogeographic units harbor high diversity across many species that should be prioritized for protection.

Keywords

Anura, Caudata, elevation, phylogeny, Random Forest, range size.

Introduction

Characterizing genetic variation within species and understanding the processes that maintain it are important goals in evolutionary biology (Ellegren & Galtier, 2016). The maintenance of genetic variation is essential for natural populations to develop new traits and adapt to future environmental changes (Sgrò et al., 2011; Stange et al., 2021), making it a fundamental aspect of biodiversity conservation efforts (DeWoody et al., 2021). Intraspecific genetic variation includes both the variation within populations (hereafter, genetic diversity, or GD) and the distribution of genetic variation between populations (hereafter, genetic structure, or GS). These two components provide insights into demographic history and evolutionary processes such as changes in effective population size (N_e) or the pattern of genetic exchange between populations (Paz-Vinas et al., 2018; Templeton, 2006). Therefore, investigating both GD and GS within species can help explain the processes underlying genetic variation that we observe in nature and how that variation is maintained.

Several evolutionary mechanisms influence genetic variation in natural populations. This variation can be affected by (1) genetic drift, which can lead to the loss of genetic variation over time; (2) migration, which can increase genetic variation within populations (high GD) and reduce genetic differentiation

between populations (low GS); (3) mutation, with higher mutation rates increasing genetic variation; or (4) natural selection, which can increase or decrease genetic variation depending on the type of selection (e.g., Lacy, 1987; Sunyaev et al., 2003). Levels of genetic variation vary greatly in natural populations and among species (Romiguier et al., 2014), however, determinants of intraspecific genetic variation remain poorly understood. For decades, many studies have considered N_e to be the most important evolutionary parameter that has a significant impact on genetic variation within and between populations (e.g., Frankham, 1996; Mackintosh et al., 2019; Peart et al., 2020). For example, small, isolated populations tend to have low genetic variation due to increased genetic drift and reduced gene flow. Under neutral theory, genetic variation is expected to increase with population size (Kimura, 1983), due mainly to reduced genetic drift in large populations (Buffalo, 2021). Several empirical studies have shown results consistent with the neutral theory, in which species with higher population abundances have higher GD (e.g., Grundler et al. 2019; Hague and Routman, 2016). However, GD does not necessarily correlate with population size in all cases (e.g., Bazin et al. 2006), and genetic variation levels across species have been observed to be much narrower than their variation in population size (the so-called Lewontin's paradox; Lewontin, 1974). In nature, GD and GS within species can be influenced by several additional factors, including geographic and ecological characteristics (e.g., Gibbs, 2001; Nevo, 1978).

Distinct geographic, life history, and ecological factors can determine intraspecific genetic variation and influence the diversification and extinction of populations and species. The total range size and geographic distribution of a species can influence genetic variation; for example, species with larger ranges are expected to have higher GD, low inbreeding, and reduced genetic drift because of the direct relationship between geographic range size and population size (see Leffler et al., 2012 and Mace et al., 2010). In addition, habitat fragmentation can lead to reduced gene flow and increased genetic differentiation (GS) among populations, ultimately reducing GD within populations because of genetic

drift (see Dixo et al., 2009 for an example in Brazilian toads). It is important to consider the complete geographic distribution of species when sampling is performed, especially in those species that have experienced recent range expansion. For example, a study of genetic variability in 25 Australian scincid lizards found that most taxa follow the central-marginal hypothesis, which predicts that core populations have greater GD than edge populations within their geographic distributions (Singhal et al., 2022). Life history traits also have an impact on GD and GS, providing a connection between different demographic processes (Ellegren and Galtier, 2016; Duminil et al., 2007). For example, species with shorter lifespans and higher fecundity may have higher GD because there are more chances for mutation and recombination (Romiguier et al. 2014). Body size can be another important predictor of both GD and GS within some taxonomic groups. For example, Brüniche-Olsen et al. (2019) found that GD decreases with increasing body size in Darwin's finches, with the possible explanation that species abundance is expected to decrease with increasing body size (White et al. 2007). Larger-bodied species may also have higher dispersal ability resulting in lower GS, a pattern demonstrated in bees (Lopez-Uribe et al. 2019) and frogs (Paz et al. 2015).

An improved understanding of these factors can provide insights into how genetic variation is shaped in natural populations. Given the variety of factors and their complex relationships, the approach used to assess potential predictors of GD and GS is no less important. One approach is to collect new genetic data for a set of taxa of interest and analyze linear models of GD/GS and possible predictors (e.g., Barry et al., 2022; Dixo et al., 2009; Paz et al. 2015). An alternative approach, which can enable comparisons of many more species, is to use repurposed data that was previously collected for a different research purpose and that can be reanalyzed in a common framework. Machine learning (ML) algorithms provide a powerful framework for detecting patterns in very large datasets with many possible predictors that may be correlated. For example, Pelletier and Carstens (2018) used a ML framework and georeferenced DNA sequences recovered from more than 8000 species and found that geographic range size and latitude were

the most important predictors of GS. Phylogenetic Comparative Methods (PCMs) provide another important framework when species are the unit of study. These models consider phylogenetic non-independence across taxa to account for the shared evolutionary history of species (Felsenstein, 1985). For example, Romiguier et al. (2014) used phylogenetic independent contrasts to investigate the relationships between genome-wide GD and life-history traits in 76 metazoan species. The application of both ML and PCMs to repurposed data can enable identification of important predictors of genetic variation within species, while incorporating important biological information (i.e., species relationships) into our models.

Amphibians (frogs, salamanders, and caecilians) have been a preferred study system for several ecological and evolutionary studies because they exhibit a wide variety of natural history traits (e.g., complex life cycles that involves both aquatic and terrestrial stages) and distributional patterns (e.g., many species have limited dispersal capabilities that lead to high levels of GS). Many studies have explored patterns of species richness in amphibians. For example, Wiens (2007) found that diversification rates in frogs and salamanders increased with decreasing latitude, following the latitudinal gradient of species richness. On the other hand, Marin and Hedges (2016) found that species richness of global amphibians was best explained by time (lineage age). On a regional scale, a positive latitudinal gradient of taxonomic, functional, and phylogenetic diversity in the Americas has been demonstrated in amphibians (Ochoa-Ochoa et al., 2020). Global genetic diversity patterns also appear to follow a latitudinal gradient in amphibians (Miraldo et al. 2016), but less is known about the predictors of intraspecific variation. In a study of Nearctic amphibian species, the most important predictors of GD were taxonomic family, number of sequences, and for salamander species (N = 98), those at more northern latitudes had lower GD (Barrow et al., 2021). In the same region, Schmidt et al. (2022) analyzed microsatellite data from 19 amphibian species; they found that GD was not predicted by the environmental variables they used and that areas with high species richness also had high GS, but low

GD. Amphibians are the most threatened group of vertebrates on the planet (Harfoot et al., 2021), and are also particularly vulnerable to habitat loss, pollution, climate change, and disease. This vulnerability makes them an important study system for understanding their genetic variation and adaptive potential.

The Neotropical Region includes almost 50% of the world's total amphibian species, which is greater than any other comparable area on the planet (Bolaños et al., 2008; Urbina-Cardona, 2008). This exceptionally high species diversity is possibly due to a combination of ecological and evolutionary factors, including the geological history of the region (e.g., Andean uplift), climate change, ecological interactions, and biotic diversification (see Antonelli, 2022). For decades, the origins and maintenance of diversity in this region has been an important focus of research in amphibian ecology and systematics. For example, Gonzalez-Voyer et al. (2011) used PCMs and found that high-altitude ranges, vascularized ventral skin, and vegetation type correlate significantly with species richness of Brachycephaloidea, the New World direct-developing frogs. Using glassfrogs as a study model, Castroviejo-Fisher et al. (2014) showed how the climatic and geological history of the Neotropics shaped the diversity of this group, demonstrating that initial uplifts of the Andes and marine incursions are coincidental with divergence times of most glassfrog clades. Recently, Tobar-Suárez et al. (2022) found that amphibian species richness in Neotropical cloud forests increased towards the equator, with frog and caecilian species richness increasing towards lower latitudes, while salamanders showed the opposite pattern. Despite this active research focus on species-level diversity, very little is known about the determinants of intraspecific GD and GS in amphibians within the Neotropical Region.

Here, we investigated what factors predict genetic diversity and structure in Neotropical frogs and salamanders. First, we estimated genetic diversity and genetic structure from repurposed DNA sequences and their associated geographic coordinates. We also gathered natural history trait and geographic

information for each species from open-access databases and literature. Next, we built maps of GD per species and per locality across different Neotropical biogeographic units to identify potential hotspots of intraspecific diversity. We then applied a machine learning algorithm (Random Forests, RF) to investigate the relative importance of predictors of GD and GS. RF can handle large numbers of variables to build predictive models, but does not explicitly incorporate phylogenetic history (other than including taxonomic Order and Family as predictors). Therefore, we also used phylogenetic comparative methods (PCMs) to test for phylogenetic signal and to determine relationships between natural history traits, geographic variables, and GD or GS of Neotropical amphibians. In addition to identifying important predictors of genetic variation in Neotropical amphibians, our study examines whether RF and PCMs provide similar insights.

Materials and Methods

DNA sequences and associated geographic coordinates

We obtained sequences of the mitochondrial (mtDNA) gene cytochrome-b (Cytb) from open access databases since this was the most abundant gene in Neotropical amphibian studies. We downloaded sequences from three sources: phylogatR (Pelletier et al., 2022), ACDC (van den Burg et al., 2020) and GenBank (National Center for Biotechnology Information). Although these sources include Cytb sequences for hundreds of species, we chose those species with at least five sequences for further analyses to more adequately represent genetic variation within species (Barrow et al. 2021). Alignments were saved in FASTA format and were edited and aligned using AliView version 1.28 (Larsson 2014) with the MUSCLE aligner version 3.8.31 (Edgar, 2004). Geographic coordinates from each sequence were

obtained from phylogatR and GenBank when available, which corresponded to 1,575 sequences (38.96% of the total sequences) from 80 species (31.25% of the total species). We linked 2,468 additional sequences to geographic coordinates using either (1) the GeoNames (geonames.org) and GEOLocate (geo-locate.org) databases by entering the associated location name of the sequence to obtain an approximate occurrence for that sequence; or (2) information from manuscripts (e.g., the original species description, distributional records, or systematic and phylogeographic works; see Appendix S1).

Metrics of genetic diversity and structure

To evaluate genetic diversity within species, we calculated nucleotide diversity (π) from each mtDNA alignment of species and localities using the function nuc.div() in the *pegas* R package (Paradis, 2010). To evaluate genetic structure within amphibian species, we calculated isolation-by-distance (IBD) for each species. The K80 genetic distance (Kimura, 1980) was calculated using the function dist.dna() in the *ape* R package (Paradis and Schliep, 2019); this function computes a matrix of pairwise distances from DNA sequences. Euclidean geographic distance was calculated between coordinates associated with each sequence. We conducted Mantel tests using the function mantel.randtest() in the *ade4* R package (Dray and Dufour, 2007), the genetic and geographic distance matrices described above, and 999 permutations. To determine whether a species exhibits genetic structure or not, we visualized the histogram plot from the Mantel test analysis and the significance of the model (p-value < 0.05).

Trait and geographic data compilation

We obtained information on traits that may influence genetic variation in amphibian species from the AmphiBIO dataset (Oliveira et al., 2017), AmphibiaWeb (2022), and specific data from manuscripts (Appendix S1; Table 1). Body size (snout-vent length – SVL), the type of habitat predominantly used by adults, activity (diurnal or nocturnal), and development mode (larval or direct development) were included since these traits are involved in the dispersal capacity of amphibian species and can impact intraspecific genetic differentiation (Hillman et al., 2014). We collected elevational and latitudinal data (mean, maximum, and minimum) from Rolland et al. (2018), who generated this information from the Global Biodiversity Information Facility (GBIF), the International Union for Conservation of Nature (IUCN), and WorldClim (Fick and Hijmans, 2017). We also obtained latitudinal and elevational data from specific manuscripts. Elevational data were also recovered from species accounts in the Amphibian Species of the World 6.1 Online Reference (ASW database; Frost, 2021). The geographical range as a shapefile (.shp) for each species was obtained from the IUCN Red List of Threatened Species (IUCN, 2022) and the range area was calculated in square kilometers (km^2) using the packages *raster* and *geosphere* (Hijmans, 2021; 2022) in R 4.2.2 (R Core Team, 2022). The IUCN Red List Category was also recorded for each species (IUCN, 2022).

Geographic framework

To evaluate and visualize how amphibian genetic diversity is distributed in the Neotropical Region, we divided our study area into 17 biogeographic units based on the regions proposed and modified by Antonelli et al. (2018), Morrone (2014), Josse et al. (2003), and Olson et al. (2001) (Figure 1a). We calculated and mapped intraspecific genetic diversity first with a single value of π per species, and second with a value of π per locality/alignment within each species. Each locality consisted of at least two individuals, and we assigned sequences to the same locality when they were within 150 kilometers of

each other. To map GD, we rounded latitude/longitude coordinates up to two decimals. In cases where all sequences in an alignment shared the same geographic coordinate, the alignment was defined as a single locality. We visualized genetic diversity for three different resolutions (grid cell size of 150 km^2, 250 km^2, and 350 km^2; Figures 1b, S1.1b,d) which corresponds to 3640, 1326, and 673 grid cells within the Neotropical Region, respectively. We also visualized the number of sequences per grid cell for the three resolutions to present the sampling effort and distribution of sequences available. In total, we analyzed 4,403 sequences from 669 localities (some species have sequences from the same locality; Table S1.1). All mapping analyses were performed in R 4.2.2 (R Core Team, 2022).

Identifying predictors of genetic diversity and structure with Random Forests

We used the Random Forests (RF) machine learning algorithm (Breiman, 2001) to build predictive models and identify variables (e.g., body size, habitat, elevation, range size; see Table 1) that are important predictors of GD (π) and GS (IBD) in Neotropical amphibians. RF uses independent variables to create many individual decision trees (a forest) that act as an ensemble to predict a response. Each decision tree in the random forest uses a subset of the independent variables and returns a response, and variable importance is determined based on the increase in model error when that variable is not included (Kabacoff, 2015). RF regression (for π) and classification (for IBD) models were created using the randomForest() function in the *randomForest* R package, with 1000 trees and 10 permutations. Relative importance for each independent variable was measured, printed, and visualized in the *randomForest* package using the importance() and varImplot() functions. RF models were created for the complete dataset, and for anurans and caudates separately. To evaluate the effect of the number of sequences per species and the number of base pairs (bp) in each alignment might have on the findings, we also created RF models with three different reduced datasets: (1) a dataset with at least 10 sequences per species (113

species), (2) a dataset with at least 400 bp in each species alignment (197 species), and (3) a dataset with both at least 10 sequences per species and 400 bp in each species alignment (99 species).

Testing predictors of genetic diversity and structure with phylogenetic comparative methods

Since the RF models do not explicitly take phylogenetic relationships into account (other than a taxonomic level as a possible predictor), but we know that species are non-independent, we implemented phylogenetic comparative methods (PCMs). We subset the phylogenetic tree from Jetz and Pyron (2018) to create a phylogeny that includes only the species in our dataset using the phylo4() function in the *phylobase* R package (Hackathon et al., 2020). Of the 256 species in our dataset, 15 species are not found in the phylogeny of Jetz and Pyron (Appendix S1), therefore for phylogenetic analyses we used a subset consisting of 241 species. In addition to combined analyses of both Anura and Caudata, we extracted portions of the tree to include only Anura or Caudata species with the mrca() and extract.clade() functions in the *ape* R package (Paradis and Schliep, 2019). We mapped π as a continuous trait and IBD as a binary trait on the tree using the contMap() and make.simmap() functions in the *phytools* R package (Revell, 2012). Before starting our statistical analysis using phylogenetic information, we considered the assumptions of the models. We assume that the topology and the branch lengths of our phylogeny are correct, and we visualized the branches of our tree using the function fastBM() in *phytools* to determine whether they evolve in the manner of the Brownian motion model (Figure S1.2). We tested for phylogenetic signal in π using the phylosig() function with the method lambda implemented in *phytools*. We tested for phylogenetic signal in IBD using the phylo.d() function that calculates the D statistic, a measure of phylogenetic signal in a binary trait (Fritz and Purvis, 2010), implemented in the R package *caper* (Orme et al., 2018).

We used phylogenetic generalized least squares (PGLS) models (Grafen, 1989; Ives and Garland, 2010), which are suitable for testing evolutionary regression between two characters (Harmon, 2019), to assess relationships between possible predictors and genetic diversity or structure while accounting for phylogeny. We fitted phylogenetic linear regression models for π, and phylogenetic logistic regression models, which allow for phylogenetic analysis including binary traits as described in Ives and Garland (2010), for IBD as a binary dependent variable (0 for no IBD, 1 for IBD). For these models, we used the phylolm() and phyloglm() functions respectively in the *phylolm* R package (Ho and Ane, 2014). For each response variable (π and IBD), we ran models using 13 predictors separately (body size, mean and minimum elevation, mean, minimum and maximum latitude, range size, development mode, microhabitat, activity, number of sequences, IUCN conservation status, and biogeographic unit).

Finally, we used phylogenetic generalized linear mixed models (PGLMM; Ives and Helmus, 2011) implemented in the *MCMCglmm* R package (Hadfield, 2010) to investigate relationships between a subset of the predictors and two responses: (1) π, and (2) IBD. In these models, due to computational constraints, we analyzed nine cofactor variables (body size, minimum elevation, mean latitude, range size, development mode, IUCN conservation status, microhabitat and biogeographic unit). For these models, we included species 'random effects', which account for the variability caused by species-specific effects, and phylogenetic random effects, which consider the phylogenetic relationship between species, by transforming the phylogeny into a variance-covariance matrix of relatedness between species (Garamszegi, 2014). For each response variable (π and IBD), we compared six models that were generated with a combination of the nine predictors above and a reduced set of predictors based on the results of other PCM and RF analyses. These models included: (1) an intercept-only null model, (2) a model with both species and phylogenetic random effects and no predictors, (3) a model with all predictors and both species and phylogenetic random effects, (4) a model with a reduced number of predictors and no random effects, (5) a model with a reduced number of predictors and only species as a

random effect, and (6) a model with a reduced number of predictors and with both species and phylogenetic random effects. For Bayesian model selection, we used the deviance information criterion (DIC), which is a generalization of the Akaike information criterion (AIC) and makes use of the posterior distributions obtained from MCMC output (Spiegelhalter et al., 2002).

Results

Data compilation

We compiled 4,043 Cytb mtDNA sequences from 256 Neotropical amphibian species (Figure S2.1), of which only 80 species had associated occurrences (we retrieved occurrences for 50 species from phylogatR and 30 species from GenBank). Occurrences for the sequences of the other 176 species were recovered by us in the present work. The total occurrences represented all 17 Neotropical biogeographic units (Figure S2.2). The final dataset is composed of 230 frogs in 17 families and 26 salamanders of the family Plethodontidae, two response variables (π and IBD), and 16 predictor variables (Table 1, S1.1).

Distribution of genetic diversity and structure in the Neotropics

The calculated π ranged from 0 (10 species) to 0.156 (Boquete rocket frog, *Silverstoneia nubicola*, family Dendrobatidae) with a mean of $\pi = 0.025$ (Table S1.1). In our geographic context, π is higher in southern Mesoamerica, the Chocó region, northern Andes, and Atlantic Forest region, compared to their adjacent biogeographic units. On the other hand, the central and southern Andes and the southern portion of South

America show low genetic diversity (Figure 1c). With respect to genetic structure, we found that 116 Neotropical amphibian species show significant IBD, while 140 species do not (Table S1.1). Species following an IBD pattern were mostly found in the Antillean Subregion, Mesoamerica, and the Amazonian region. The Amazon also included the most species that do not present IBD, followed by the Atlantic Forest and Andes region (Figure S2.3).

Important predictors of genetic diversity and structure based on Random Forests

Random forest regressions showed that range size was the most important predictor of π for all amphibians and for Anura alone (Figure 2, S2.4). Mean elevation and taxonomic family were other important predictors for the Anura data set, while the number of sequences was the least important predictor. For the Caudata dataset, minimum elevation and body size were the most important predictors; and minimum elevation was the third-best predictor for this data set (Figure S2.4). Random forest classification found that minimum elevation was the most important predictor of IBD for all amphibians and for Caudata alone (Figure 2, S2.5). On the contrary, development mode was the least important predictor. Activity pattern (diurnal vs. nocturnal) and maximum latitude were other important predictors for the complete dataset and for Anura only (Figure S2.5). For the Caudata data set, body size and biogeographic unit were other important predictors of IBD (Figure S2.5). When we evaluated the effect of the number of sequences per species (at least 10), and the number of base pairs (bp) in each alignment (at least 400 bp) on our results, we found that range size was the best predictor of π for all models analyzed (Figure S2.6) as in the model using the complete dataset. Biogeographic unit and minimum elevation were the most important predictors of IBD in the three different reduced datasets (Figure S2.7).

Phylogenetic signal in nucleotide diversity, but not IBD

According to the phylogenetic analysis of genetic diversity in Neotropical amphibians, π is not distributed uniformly across phylogeny. We found clusters of closely related genera that have dissimilar π values; for example, within Bufonidae, *Atelopus* species have low π (average = 0.006, range: 0.001-0.010) while toads of genus *Rhinella* presented a higher average π (average = 0.020, range: 0.006-0.054). The same pattern was also observed in poison frogs of the family Dendrobatidae, where *Ameerega* species show high GD (average π = 0.024; range: 0.003-0.059) in contrast to species in the genus *Andinobates* which had very low GD (average π = 0.003; range: 0.000-0.009) (Table S1.1; Figure 3, S2.8). On the other hand, genetic structure (IBD) is more homogeneously distributed along the phylogeny (Figure 3). We found strong phylogenetic signal in π (lambda = 0.7456, P-value (based on LR test) = 0.0002) (Figure S2.9a), but we did not find significant phylogenetic signal in IBD (Estimated D = 0.7632, P-value = 0.081) (Figure S2.9b). Results of measurements of phylogenetic signal on different traits (predictors) can be found in Appendix 2.

Relationships between predictors, nucleotide diversity, and IBD based on PCMs

Phylogenetic linear regression models indicated that range size (P-value = 0.0016) and minimum elevation (P-value = 0.0004) predicted π (Table S2.1), confirming our Random Forest results. Neotropical amphibian species with larger ranges tend to have higher π (Figure S2.10); on the contrary, π decreases when the elevation (minimum) of a species' range increases (Figure S2.11). Four of the important predictors identified with RF (maximum latitude, activity, body size and minimum elevation) were also significant predictors according to the phylogenetic logistic regression models; these models also

identified mean latitude, development mode, number of sequences, and mean elevation as significant predictors of IBD (Table S2.1). Species living at southern latitudes in the Neotropics tend to have no IBD, contrary to amphibians in northern latitudes that tended to exhibit IBD (Figure S2.12). Species following an IBD pattern are mostly nocturnal, on the other hand, species with no IBD do not show a preference for diurnal or nocturnal activities (Figure S2.13). We found significant differences in species with IBD and no IBD when analyzing body size (P-value = 0.0018). These differences were most apparent in salamanders, where species with IBD were larger on average (Figure S2.14). Neotropical amphibians with significant IBD tend to live at lower elevations compared to species with no IBD (mean elevation $P < 0.0001$; minimum elevation $P < 0.0001$) (Figure S2.15).

Phylogenetic Generalized Linear Mixed Models using *MCMCglmm* showed that the best fit model for π was the one with a reduced number of predictors and with both species and phylogenetic random effects (Table 2). Once again, range size and minimum elevation were the best predictors of π, with a positive relationship between range size and π and a negative relationship between minimum elevation and π (Figure 4a). *MCMCglmm* models showed that for IBD the best-fit model was the full model with all predictors and species and phylogenetic random effects (Table 2, Figure S2.16). The second best *MCMCglmm* model for IBD was, as for π, a model with a reduced number of predictors and with both species and phylogenetic random effects (Table 2). Consistent with RF and other PCM results, minimum elevation was the best predictor of IBD, and their relationship was negative (Figure 4b). *MCMCglmm* results were verified by checking model diagnostics using trace and density plots of the MCMC samples (Figures S2.17 - S2.18).

Discussion

We investigated the predictors of genetic variation in Neotropical amphibians using repurposed data including mtDNA sequences, natural history traits, and geographic information gathered from open-access databases. Our analyses revealed that geographic range size and elevation were the most significant predictors of overall genetic diversity within species. Specifically, we found that amphibian species inhabiting smaller ranges and higher elevations have lower intraspecific π (Figure 4c), in agreement with recent published results from De Kort et al. (2021) who found that GD, measured as expected heterozygosity, was higher at low elevations. Geographic range size has also emerged as an important predictor of genetic variation in other taxonomic groups such as squamates (Larkin et al., 2023), pinnipeds (Peart et al., 2020), and Darwin's finches (Brüniche-Olsen et al., 2019). Interestingly, the relationship between GD and geographic range is not always straightforward, as evidenced by the case of butterflies (Mackintosh et al., 2019) and other non-model animal species (Romiguier et al., 2014). These discrepancies highlight the need for continued study of a variety of taxonomic groups and regions to better understand the ecological and geographic context of genetic variation.

Geographically restricted species are expected to have less GD, which may also indicate an increased risk of extinction (Levy et al., 2016). For example, Doyle et al. (2015) found that threatened and endangered mammalian species with small range sizes have less GD than those with larger range sizes. In amphibians, Caviedes-Solis et al. (2020) found that Neotropical treefrogs living in high elevations were more likely to be classified with threatened status. Our analysis of Neotropical amphibians from several taxonomic families confirmed expected differences between geographic range sizes based on IUCN conservation status, with Least Concern (LC) amphibian species occurring in larger ranges and Data Deficient (DD), Endangered (EN) and Critically Endangered (CR) species occurring in smaller ranges (Figure S3.1.). Based on our estimates, however, IUCN conservation status did not have as clear of an association with π (Figure S3.2), suggesting that IUCN status may not currently capture this important

parameter for population persistence. Regardless, populations of threatened species with low GD such as harlequin frogs of the genus *Atelopus* should be closely monitored.

Elevation, along with associated gradients of precipitation and temperature, has significant impacts on genetic variation across animal species (De Kort et al., 2021). Our study identified elevation as the best predictor of IBD and revealed significant differences between species with and without IBD patterns. We observed that species inhabiting higher elevations did not usually exhibit genetic structure (see Figures 4d, S2.14), while those living in lower elevations tended to have both significant IBD patterns and larger geographic range sizes (as indicated in Figure 4c,d). The absence of IBD patterns, and in turn low π, in species living at higher elevations could be attributed to their smaller ranges, leading to smaller population sizes. Our findings partially agree with those reported by Pelletier and Carstens (2018), who found that geographic range size, elevation, and latitude predicted genetic structure (IBD) in several taxonomic groups. However, in our case, geographic range size alone did not predict IBD, i.e., Neotropical amphibian species with and without IBD can inhabit small or large geographic ranges; in the global analysis of eukaryotes by Pelletier and Carstens (2018), species with IBD tended to have larger range sizes. We suspect that environmental heterogeneity within the geographic range rather than range size may have a greater influence on IBD patterns of species in the Neotropics, as has been observed in other vertebrate studies (e.g., Quiroga-Carmona and D'Elía, 2022).

A few other geographic and life history traits were identified as important predictors of IBD based on Random Forests and phylogenetic linear regression models (Table S2.1; Figure 2b). For example, the activity pattern (diurnal or nocturnal) of species and latitude were top predictors of IBD in Neotropical amphibians and in the Anura-only dataset (Figures S2.5, S2.11, S2.12). Interestingly, of the 116 species with significant IBD patterns in our dataset, approximately two-thirds (n = 77; 66%) were nocturnal

species. This result could suggest that a life adapted to having more activity at night also decreases the capacity for long-distance movements, i.e., nocturnal habits could promote low vagility. Regarding latitude, species with IBD occur at slightly higher latitudes than species without IBD. Although latitude was not a significant predictor of π, analyzing mean latitude with GD and geographical range size revealed that amphibians near the equator have larger ranges, and diversity increases near the equator and slowly decreases towards higher latitudes (Figure S3.3). This pattern of genetic diversity forms a plateau around the equator, a pattern noted by Pereira (2016) in his perspective about a global study mapping amphibian genetic diversity by Miraldo et al. (2016). Biogeographic unit was the third most important predictor of IBD in our RF analyses, following maximum latitude and activity pattern in anurans, and minimum elevation and body size in caudates (Figure S2.5). We found differences in terms of presence or absence of IBD patterns throughout the biogeographic units. For instance, many of the species from the Antillean Subregion and Mesoamerica, which are two of the northernmost Neotropical biogeographic units, showed IBD patterns (Figure S2.3). In this case, latitude may be confounded with regions of high topographic complexity that would be expected to promote high IBD, such as archipelagos or mountain ranges.

Neutral theory predicts that levels of genetic variation are related to population size, which is expected to be negatively correlated with body size in animals (Frankham, 1996). This prediction has been tested and confirmed in several previous studies with mammals (Doyle et al., 2015), birds (Brüniche-Olsen et al., 2019), butterflies (Mackintosh et al., 2019), and fishes (Barry et al., 2022). Our estimates indicated that body size was an important predictor of genetic variation (π and IBD) in salamanders (Table S2.1; Figures S2.4, S2.5). Interestingly, contrary to neutral theory, salamander species with larger body sizes had higher GD values (Figure S3.4). All Caudata species in our data set are lungless salamanders of the family Plethodontidae and all have direct development. The evolution of direct development and life history strategies in this group of salamanders has contributed not only to the extraordinary evolutionary

success of this species (Wake and Hanken, 1996), but might also result in patterns of genetic variation that contrast with expectations from other animal groups. Across animals, larger clutch sizes (i.e., more offspring) correspond with higher GD (Romiguier et al. 2014). Salamander species that undergo direct development have clutch sizes that are relatively smaller than those with larval development (Wake and Marks, 1993). Within direct-developing Plethodontidae, clutch size is constrained, but it is conceivable that larger-bodied species can produce more offspring, potentially leading to higher GD. Surprisingly, larger Neotropical salamanders not only have high GD, but also appear to have reduced gene flow (a significant IBD pattern, or high GS) between populations than smaller salamanders without IBD, suggesting that body size may not be a good indicator of dispersal ability in this group. Fusco et al. (2021) found similar patterns of high GD and high GS in the northern two-lined salamander (*Eurycea bislineata*). They reported reduced genetic connectivity between populations in urban areas, and contrary to their prediction due to a loss of connectivity, populations had high levels of GD. Despite these remarkable findings, we are cautious interpreting these results in a generalized way due to the small number (n = 26) of Caudata species we were able to include in this analysis.

The geographic patterns of π in Neotropical amphibians are illustrated in Figure 1. This genetic diversity map clearly shows areas of high π, such as Chocó, northern Andes, and Atlantic Forests in South America, and an area of high diversity in the Mesoamerica region located in Central America. Although we did not explicitly measure environmental variables, biogeographic units identified with higher π values are areas that also have high precipitation rates. Therefore, precipitation could explain and be positively correlated with genetic variation in the Neotropics; however, this hypothesis needs to be tested in future studies. Other studies have previously shown the important role of annual precipitation in explaining species richness, phylogenetic diversity, and functional diversity in Neotropical amphibians (Amador et al., 2019; Ochoa-Ochoa et al., 2019). Differences in research and collecting efforts throughout the Neotropical region could be complicating the interpretation of our genetic diversity map.

For example, we were not able to recover more than five CytB sequences (our minimum number of sequences per species) for any Gymnophiona species (caecilians), nor were we able to recover genetic information for centrolenid "glassfrogs", one of the most diverse taxa in the Neotropics. We were also unable to recover target sequences for marsupial frogs of the family Hemiphractidae or for Neotropical microhylid frogs; these two groups have high species richness, but a significant lack of available evolutionary studies. To alleviate some of these sampling issues, future studies should consider comparisons of other mitochondrial and nuclear genes and ideally standardize the markers sequenced across multiple species.

Comparing our results with those for Nearctic amphibians (Barrow et al. 2021), no differences in average π within species were evident (Nearctic amphibians: average $\pi = 0.028$, n = 299; Neotropical amphibians: average $\pi = 0.025$, n = 256). However, the highest values of intraspecific π were found in several Neotropical species; for example, only three Nearctic species had π values > 0.09 compared to 13 Neotropical species with similar or higher values. This disparity between Neotropical and Nearctic species could relate to differences in demographic history between regions or could be explained by bias in taxonomic practices (see Chek et al., 2003). It is possible that the higher genetic variation we observed in Neotropical amphibians is partially due to taxonomic under-splitting, which could lead to severe conservation implications for this group. We found very high values of π (e.g., > 0.09) for several species, suggesting the possibility of cryptic species in our dataset. At least nine of the species with high π values (Table S3.1) have been considered as species complexes or cryptic species in previous taxonomic studies (e.g., *Pristimantis altamazonicus* – Ortega-Andrade et al., 2017; *Anomaloglossus degranvillei* – Vacher et al., 2017; *Phyllobates lugubris* – Marquez et al., 2020; *Dendropsophus decipiens* – Oliveira et al., 2021). Higher environmental and climatic heterogeneity may lead to diversification dynamics with higher speciation and lower extinction rates supporting rapid evolution in the Neotropics (Brown, 2014). Under this scenario, species would have accumulated faster towards the present mainly due to recent geological

and climatic perturbations (e.g., the elevation of the Andes) (Meseguer et al., 2022). The potential under-split species in our data set are therefore expected to be young or recently diverged, which is one of the mechanisms causing cryptic diversity (Fišer et al., 2018).

We found that levels of genetic diversity varied among families, with species in certain families such as Dendrobatidae and Hylidae having the highest π estimates and Ranidae, Rhinodermatidae, and Telmatobiidae having the lowest (Figure S.2.8). Accordingly, we observed strong and significant phylogenetic signal in π. In contrast, the presence of genetic structure (IBD) within species appears to be more randomly distributed throughout the phylogeny of Neotropical amphibians. These findings highlight the significance of employing the approach utilized in this study. With the growing size and complexity of biodiversity datasets, machine learning methods such as Random Forests prove valuable in identifying potential predictors. Combining these methods with PCMs allowed us to use phylogenetically-informed models to gain a deeper understanding of the relationships between predictors and genetic variation within species. For example, our RF (regression and classification) and PCMs (phylolm and MCMCglmm) models were consistent in identifying range size and minimum elevation as the best predictors of nucleotide diversity and population structure, respectively (see Table S2.1, and Figures 2 and 4).

Our study provides valuable insights into the distribution of genetic diversity and structure in Neotropical amphibians and identifies important predictors of intraspecific genetic variation. These findings underscore the importance of considering both genetic diversity and population structure in the conservation and management of Neotropical amphibian populations. For example, the results demonstrate the importance of preserving forest areas, especially in biogeographic areas where the intraspecific genetic diversity is very low. This information is also valuable to assess the conservation

status of Neotropical amphibian species and consider the impact of threats these taxa could be facing in the future. The current distribution of genetic variation could play a key role in the development of targeted conservation strategies for amphibian species, particularly considering the diverse life histories observed among Neotropical amphibians. For example, certain species within the genera *Atelopus* or *Telmatobius*, which are highly endangered frog groups in the Neotropics, require specific conservation measures focused on preserving aquatic habitats (e.g., streams, ponds, or lakes). These habitats serve as crucial breeding grounds where frogs lay their eggs in shallow water, making their conservation of paramount importance. To address these and other subjects inherent to amphibian ecology and evolution, we suggest that future work, as performed here, should include more information, such as genome-scale data and where possible, add more species within the region and globally.

Acknowledgements

This work was supported by a grant to LNB from the U.S. National Science Foundation (DEB-2112946). We thank the members of the Amphibian & Reptile Biodiversity Lab (University of New Mexico) for their valuable comments on earlier versions of this manuscript, Bryan Carstens (The Ohio State University) for early access to the phylogatR database, and Jorge “Colo” Ortega for use of the photo of *Boana picturata* in Figure 1.

Conflict of Interest

The authors declare no conflict of interest.

Data Availability

Tables S1.1, S2.1 and additional files used for the analyses are available on Dryad.

References

Amador, L., Soto-Gamboa, M., & Guayasamin, J. M. (2019). Integrating alpha, beta, and phylogenetic diversity to understand anuran fauna along environmental gradients of tropical forests in western Ecuador. *Ecology and Evolution*, 9(19), 11040–11052. https://doi.org/10.1002/ece3.5593

Antonelli, A. (2022). The rise and fall of Neotropical biodiversity. *Botanical Journal of the Linnean Society*, 199(1), 8–24. https://doi.org/10.1093/botlinnean/boab061

Antonelli, A., Zizka, A., Carvalho, F. A., Scharn, R., Bacon, C. D., Silvestro, D., & Condamine, F. L. (2018). Amazonia is the primary source of Neotropical biodiversity. *Proceedings of the National Academy of Sciences*, *115*(23), 6034–6039. https://doi.org/10.1073/pnas.1713819115

Barry, P., Broquet, T., & Gagnaire, P. (2022). Age-specific survivorship and fecundity shape genetic diversity in marine fishes. *Evolution Letters*, 6(1), 46–62. https://doi.org/10.1002/evl3.265

Barrow, L. N., Masiero da Fonseca, E., Thompson, C. E. P., & Carstens, B. C. (2021). Predicting amphibian intraspecific diversity with machine learning: Challenges and prospects for integrating traits, geography, and genetic data. *Molecular Ecology Resources*, *21*(8), 2818–2831. https://doi.org/10.1111/1755-0998.13303

Bazin, E., Glémin, S., & Galtier, N. (2006). Population Size Does Not Influence Mitochondrial Genetic Diversity in Animals. *Science*, 312(5773), 570–572. https://doi.org/10.1126/science.1122033

Bolaños, F., Castro, F., Cortez, C., Hedges, B., Heyer, R., Ibañez, R., Silvano, D. L., Lotters, S., Olea, G. P., Reynolds, R., Rodriguez, L., Barrera, G. S., Ubeda, C., Veloso, A., Wilkinson, M., & Young, B. (2008). Chapter 9. Amphibians of the Neotropical Realm. In S. N. Stuart, M. Hoffmann, J. S. Chanson, N. A. Cox, R. J. Berridge, P. Ramani, & B. E. Young (Eds.), *Threatened Amphibians of the World* (pp. 92-105). Lynx Edicions, Barcelona, Spain; IUCN, Gland, Switzerland; and Conservation International.

Breiman, L. (2001). Random Forests. *Machine Learning 45*, 5–32. https://doi.org/10.1023/A:1010933404324

Brown, J. H. (2014). Why are there so many species in the tropics? *Journal of Biogeography*, 41(1), 8–22. https://doi.org/10.1111/jbi.12228

Brüniche-Olsen, A., Kellner, K. F., & DeWoody, J. A. (2019). Island area, body size and demographic history shape genomic diversity in Darwin's finches and related tanagers. *Molecular Ecology*, 28(22), 4914–4925. https://doi.org/10.1111/mec.15266

Buffalo, V. (2021). Quantifying the relationship between genetic diversity and population size suggests natural selection cannot explain Lewontin's Paradox. *ELife*, 10, e67509. https://doi.org/10.7554/eLife.67509

Castroviejo-Fisher, S., Guayasamin, J. M., Gonzalez-Voyer, A., & Vilà, C. (2014). Neotropical diversification seen through glassfrogs. *Journal of Biogeography*, 41(1), 66–80. https://doi.org/10.1111/jbi.12208

Caviedes-Solis, I. W., Kim, N., & Leaché, A. D. (2020). Species IUCN threat status level increases with elevation: A phylogenetic approach for Neotropical tree frog conservation. *Biodiversity and Conservation*, 29(8), 2515–2537. https://doi.org/10.1007/s10531-020-01986-8

Chazot, N., Condamine, F. L., Dudas, G., Peña, C., Kodandaramaiah, U., Matos-Maraví, P., Aduse-Poku, K., Elias, M., Warren, A. D., Lohman, D. J., Penz, C. M., DeVries, P., Fric, Z. F., Nylin, S., Müller, C., Kawahara, A. Y., Silva-Brandão, K. L., Lamas, G., Kleckova, I., … Wahlberg, N. (2021). Conserved ancestral tropical niche but different continental histories explain the latitudinal diversity gradient in brush-footed butterflies. *Nature Communications*, 12(1), 5717. https://doi.org/10.1038/s41467-021-25906-8

Chek, A. A., Austin, J. D., & Lougheed, S. C. (2003). Why is there a tropical–temperate disparity in the genetic diversity and taxonomy of species? *Evolutionary Ecology Research*, 5, 69 – 77.

De Kort, H., Prunier, J. G., Ducatez, S., Honnay, O., Baguette, M., Stevens, V. M., & Blanchet, S. (2021). Life history, climate and biogeography interactively affect worldwide genetic diversity of plant and animal populations. *Nature Communications*, 12(1), 516. https://doi.org/10.1038/s41467-021-20958-2

DeWoody, J. A., Harder, A. M., Mathur, S., & Willoughby, J. R. (2021). The long-standing significance of genetic diversity in conservation. *Molecular Ecology*, 30(17), 4147–4154. https://doi.org/10.1111/mec.16051

Dixo, M., Metzger, J. P., Morgante, J. S., & Zamudio, K. R. (2009). Habitat fragmentation reduces genetic diversity and connectivity among toad populations in the Brazilian Atlantic Coastal Forest. *Biological Conservation*, 142(8), 1560–1569. https://doi.org/10.1016/j.biocon.2008.11.016

Doyle, J. M., Hacking, C. C., Willoughby, J. R., Sundaram, M., & DeWoody, J. A. (2015). Mammalian genetic diversity as a function of habitat, body size, trophic class, and conservation status. *Journal of Mammalogy*, 96(3), 564–572. https://doi.org/10.1093/jmammal/gyv061

Dray, S., Dufour, A. (2007). The ade4 Package: Implementing the Duality Diagram for Ecologists. *Journal of Statistical Software*, 22(4), 1–20. doi:10.18637/jss.v022.i04

Duminil, J., Fineschi, S., Hampe, A., Jordano, P., Salvini, D., Vendramin, G. G., & Petit, R. J. (2007). Can population genetic structure be predicted from life-history traits?. *The American naturalist*, *169*(5), 662–672. https://doi.org/10.1086/513490

Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res*. 32(5), 1792-1797. doi:10.1093/nar/gkh340

Ellegren, H., & Galtier, N. (2016). Determinants of genetic diversity. *Nature Reviews Genetics*, *17*(7), 422–433. https://doi.org/10.1038/nrg.2016.58

Felsenstein, J. (1985). Phylogenies and the Comparative Method. *The American Naturalist*, 125(1), 1–15. https://doi.org/10.1086/284325

Fick, S. E., & Hijmans, R. J. (2017). WorldClim 2: New 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology*, 37(12), 4302–4315. https://doi.org/10.1002/joc.5086

Fišer, C., Robinson, C. T., & Malard, F. (2018). Cryptic species as a window into the paradigm shift of the species concept. *Molecular Ecology*, 27(3), 613–635. https://doi.org/10.1111/mec.14486

Fouquet, A., Gilles, A., Vences, M., Marty, C., Blanc, M., & Gemmell, N. J. (2007). Underestimation of Species Richness in Neotropical Frogs Revealed by mtDNA Analyses. *PLoS ONE*, *2*(10), e1109. https://doi.org/10.1371/journal.pone.0001109

Frankham, R. (1996). Relationship of genetic variation to population size in wildlife. *Conservation Biology*, 10(6), 1500–1508. https://doi.org/10.1046/j.1523-1739.1996.10061500.x

Fritz, S. A., & Purvis, A. (2010). Selectivity in Mammalian Extinction Risk and Threat Types: A New Measure of Phylogenetic Signal Strength in Binary Traits: Selectivity in Extinction Risk. *Conservation Biology*, 24(4), 1042–1051. https://doi.org/10.1111/j.1523-1739.2010.01455.x

Frost, D. R. (2021). Amphibian Species of the World: an Online Reference. Version 6.1 (Date of access). Electronic Database accessible at https://amphibiansoftheworld.amnh.org/index.php. American Museum of Natural History, New York, USA. Doi.org/10.5531/db.vz.0001

Fusco, N. A., Pehek, E., & Munshi-South, J. (2021). Urbanization reduces gene flow but not genetic diversity of stream salamander populations in the New York City metropolitan area. *Evolutionary Applications*, 14(1), 99–116. https://doi.org/10.1111/eva.13025

Garamszegi, L. Z. (Ed.). (2014). Modern Phylogenetic Comparative Methods and Their Application in Evolutionary Biology: Concepts and Practice. Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-662-43550-2

Gonzalez-Voyer, A., Padial, J. M., Castroviejo-Fisher, S., De La Riva, I., & Vilà, C. (2011). Correlates of species richness in the largest Neotropical amphibian radiation: Correlates of species richness. *Journal of Evolutionary Biology*, 24(5), 931–942. https://doi.org/10.1111/j.1420-9101.2011.02243.x

Grafen, A. (1989). The phylogenetic regression. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 326,119–157

Grundler, M. R., Singhal, S., Cowan, M. A., & Rabosky, D. L. (2019). Is genomic diversity a useful proxy for census population size? Evidence from a species-rich community of desert lizards. *Molecular Ecology*, 28(7), 1664–1674. https://doi.org/10.1111/mec.15042

Hackathon, R., Bolker. B., Butler, M., Cowan, P., de Vienne, D., Eddelbuettel, D., Holder, M., Jombart, T., Kembel, S., Michonneau, F., Orme, D., O'Meara, B., Paradis, E., Regetz, J. & Zwickl D. (2020). phylobase: Base package for phylogenetic structures and comparative data. R package version 0.8.10. https://CRAN.R-project.org/package=phylobase

Hadfield, J. D. (2010). MCMC Methods for Multi-Response Generalized Linear Mixed Models: The MCMCglmm R Package. *Journal of Statistical Software*, 33(2), 1–22. https://doi.org/10.18637/jss.v033.i02

Hague, M. T. J., & Routman, E. J. (2016). Does population size affect genetic diversity? A test with sympatric lizard species. *Heredity*, 116(1), 92–98. https://doi.org/10.1038/hdy.2015.76

Harfoot, M. B. J., Johnston, A., Balmford, A., Burgess, N. D., Butchart, S. H. M., Dias, M. P., Hazin, C., Hilton-Taylor, C., Hoffmann, M., Isaac, N. J. B., Iversen, L. L., Outhwaite, C. L., Visconti, P., &

Geldmann, J. (2021). Using the IUCN Red List to map threats to terrestrial vertebrates at global scale. *Nature Ecology & Evolution*, *5*(11), 1510–1519. https://doi.org/10.1038/s41559-021-01542-9

Harmon, L. J. (2019). Phylogenetic Comparative Methods. Independent.

Hijmans, R. J. (2021). geosphere: Spherical Trigonometry. R package version 1.5-14. https://CRAN.R-project.org/package=geosphere

Hijmans, R. J. (2022). raster: Geographic Data Analysis and Modeling. R package version 3.6-3. https://CRAN.R-project.org/package=raster

Hillman, S. S., Drewes, R. C., Hedrick, M. S., & Hancock, T. V. (2014). Physiological Vagility: Correlations with Dispersal and Population Genetic Structure of Amphibians. *Physiological and Biochemical Zoology*, *87*(1), 105–112. https://doi.org/10.1086/671109

Ho, L. S. T. and Ane, C. (2014) A linear-time algorithm for Gaussian and non-Gaussian trait evolution models. *Systematic Biology*, 63(3), 397-408.

IUCN. 2022. The IUCN Red List of Threatened Species. Version 2022-1. https://www.iucnredlist.org. Accessed on 01 May 2022.

Ives, A. R., & Garland, T. (2010). Phylogenetic Logistic Regression for Binary Dependent Variables. *Systematic Biology*, 59(1), 9–26. https://doi.org/10.1093/sysbio/syp074

Ives, A.R., & Helmus, M.R. (2011). Generalized linear mixed models for phylogenetic analyses of community structure. *Ecological Monographs*, 81, 511–525.

Jenkins, C. N., Pimm, S. L., & Joppa, L. N. (2013). Global patterns of terrestrial vertebrate diversity and conservation. *Proceedings of the National Academy of Sciences*, 110(28). https://doi.org/10.1073/pnas.1302251110

Josse, C., Navarro, G., Comer, P., Evans, R., Faber-Langendoen, D., Fellows, M., Kittel, G., Menard, S., Pyne, M., Reid, M., Schulz, K., Snow, K., & Teague, J. (2003). Ecological Systems of Latin America and the Caribbean: A Working Classification of Terrestrial Systems. NatureServe, Arlington, VA.

Kabacoff, R. I. (2015). *R in Action* (2nd ed.). Manning Publications.

Kimura, M. (1983). *The Neutral Theory of Molecular Evolution.* Cambridge: Cambridge University Press.

Lacy, R. C. (1987). Loss of genetic diversity from managed populations: Interacting effects of drift, mutation, immigration, selection, and population subdivision. *Conservation Biology*, 1(2), 143–158. https://doi.org/10.1111/j.1523-1739.1987.tb00023.x

Larkin, I. E., Myers, E. A., Carstens, B. C., & Barrow, L. N. (2023). Predictors of genomic diversity within North American squamates. *Journal of Heredity*, 114(2), 131–142. https://doi.org/10.1093/jhered/esad001

Larsson, A. (2014). AliView: a fast and lightweight alignment viewer and editor for large data sets. *Bioinformatics*, *30*(22): 3276-3278. http://dx.doi.org/10.1093/bioinformatics/btu531

Leffler, E. M., Bullaughey, K., Matute, D. R., Meyer, W. K., Ségurel, L., Venkat, A., Andolfatto, P., & Przeworski, M. (2012). Revisiting an Old Riddle: What Determines Genetic Diversity Levels within Species? *PLoS Biology*, 10(9), e1001388. https://doi.org/10.1371/journal.pbio.1001388

Levy, E., Byrne, M., Coates, D. J., Macdonald, B. M., McArthur, S., & van Leeuwen, S. (2016). Contrasting Influences of Geographic Range and Distribution of Populations on Patterns of Genetic Diversity in Two Sympatric Pilbara Acacias. *PLOS ONE*, 11(10), e0163995. https://doi.org/10.1371/journal.pone.0163995

Lewontin, R. (1974). *The Genetic Basis of Evolutionary Change.* Columbia University Press.

Li, D., Dinnage, R., Nell, L. A., Helmus, M. R., & Ives, A. R. (2020). phyr: An r package for phylogenetic species-distribution modelling in ecological communities. *Methods in Ecology and Evolution*, 11(11), 1455–1463. https://doi.org/10.1111/2041-210X.13471

López-Uribe, M. M., Jha, S., & Soro, A. (2019). A trait-based approach to predict population genetic structure in bees. *Molecular Ecology*, 28(8), 1919–1929. https://doi.org/10.1111/mec.15028

Mace, G. M., Collen, B., Fuller, R. A., & Boakes, E. H. (2010). Population and geographic range dynamics: Implications for conservation planning. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 365(1558), 3743–3751. https://doi.org/10.1098/rstb.2010.0264

Mackintosh, A., Laetsch, D. R., Hayward, A., Charlesworth, B., Waterfall, M., Vila, R., & Lohse, K. (2019). The determinants of genetic diversity in butterflies. *Nature Communications*, *10*(1), 3466. https://doi.org/10.1038/s41467-019-11308-4

Marin, J., & Hedges, S. B. (2016). Time best explains global variation in species richness of amphibians, birds and mammals. *Journal of Biogeography*, *43*(6), 1069–1079. https://doi.org/10.1111/jbi.12709

Márquez, R., Linderoth, T. P., Mejía-Vargas, D., Nielsen, R., Amézquita, A., & Kronforst, M. R. (2020). Divergence, gene flow, and the origin of leapfrog geographic distributions: The history of colour pattern variation in *Phyllobates* poison-dart frogs. *Molecular Ecology*, 29(19), 3702–3719. https://doi.org/10.1111/mec.15598

Meseguer, A. S., Michel, A., Fabre, P.-H., Pérez Escobar, O. A., Chomicki, G., Riina, R., Antonelli, A., Antoine, P.-O., Delsuc, F., & Condamine, F. L. (2022). Diversification dynamics in the Neotropics through time, clades, and biogeographic regions. *ELife*, 11, e74503. https://doi.org/10.7554/eLife.74503

Miraldo, A., Li, S., Borregaard, M. K., Flórez-Rodríguez, A., Gopalakrishnan, S., Rizvanovic, M., Wang, Z., Rahbek, C., Marske, K. A., & Nogués-Bravo, D. (2016). An Anthropocene map of genetic diversity. *Science*, 353(6307), 1532–1535. https://doi.org/10.1126/science.aaf4381

Morrone, J. J. (2014). Biogeographical regionalisation of the Neotropical region. *Zootaxa*, *3782*(1), 1. https://doi.org/10.11646/zootaxa.3782.1.1

Ochoa-Ochoa, L. M., Mejía-Domínguez, N. R., Velasco, J. A., Marske, K. A., & Rahbek, C. (2019). Amphibian functional diversity is related to high annual precipitation and low precipitation seasonality in the New World. *Global Ecology and Biogeography*, 28(9), 1219–1229. https://doi.org/10.1111/geb.12926

Ochoa-Ochoa, L. M., Mejía-Domínguez, N. R., Velasco, J. A., Dimitrov, D., & Marske, K. A. (2020). Dimensions of amphibian alpha diversity in the New World. *Journal of Biogeography*, 47, 2293 – 2302. https://doi.org/10.1111/jbi.13948

Oliveira, B. F., São-Pedro, V. A., Santos-Barrera, G., Penone, C., & Costa, G. C. (2017). AmphiBIO, a global database for amphibian ecological traits. *Scientific Data*, *4*(1), 170123. https://doi.org/10.1038/sdata.2017.123

Oliveira, R. F. de, Magalhães, F. de M., Teixeira, B. F. da V., Moura, G. J. B. de, Porto, C. R., Guimarães, F. P. B. B., Giaretta, A. A., & Tinôco, M. S. (2021). A new species of the *Dendropsophus decipiens* Group (Anura: Hylidae) from Northeastern Brazil. *PLOS ONE*, 16(7), e0248112. https://doi.org/10.1371/journal.pone.0248112

Olson, D. M., Dinerstein, E., Wikramanayake, E. D., Burgess, N. D., Powell, G. V. N., Underwood, E. C., D'amico, J. A., Itoua, I., Strand, H. E., Morrison, J. C., Loucks, C. J., Allnutt, T. F., Ricketts, T. H., Kura, Y., Lamoreux, J. F., Wettengel, W. W., Hedao, P., & Kassem, K. R. (2001). Terrestrial Ecoregions of the World: A New Map of Life on Earth. *BioScience*, *51*(11), 933. https://doi.org/10.1641/0006-3568(2001)051[0933:TEOTWA]2.0.CO;2

Orme D, Freckleton R, Thomas G, Petzoldt T, Fritz S, Isaac N, Pearse W (2018). caper: Comparative Analyses of Phylogenetics and Evolution in R. R package version 1.0.1 https://CRAN.R-project.org/package=caper

Ortega-Andrade, H. M., Rojas-Soto, O. R., Espinosa de los Monteros, A., Valencia, J. H., Read, M., & Ron, S. R. (2017). Revalidation of *Pristimantis brevicrus* (Anura, Craugastoridae) with taxonomic comments on a widespread Amazonian direct-developing frog. Herpetological Journal 26, 81 – 97.

Paradis, E. (2010). pegas: an R package for population genetics with an integrated-modular approach. *Bioinformatics 26*, 419-420.

Paradis, E., & Schliep, K. (2019). ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* 35, 526 - 528.

Paz, A., Ibáñez, R., Lips, K. R., & Crawford, A. J. (2015). Testing the role of ecology and life history in structuring genetic variation across a landscape: A trait-based phylogeographic approach. *Molecular Ecology*, 24(14), 3723–3737. https://doi.org/10.1111/mec.13275

Paz-Vinas, I., Loot, G., Hermoso, V., Veyssiere, C., Poulet, N., Grenouillet, G., & Blanchet, S. (2018). Systematic conservation planning for intraspecific genetic diversity. *Proceedings of the Royal Society B 285*, 20172746. http://dx.doi.org/10.1098/rspb.2017.2746

Peart, C.R., Tusso, S., Pophaly, S.D. et al. (2020). Determinants of genetic variation across eco-evolutionary scales in pinnipeds. *Nature Ecology & Evolution 4*, 1095–1104. https://doi.org/10.1038/s41559-020-1215-5

Pelletier, T. A., & Carstens, B. C. (2018). Geographical range size and latitude predict population genetic structure in a global survey. *Biology Letters*, 14(1), 20170566. https://doi.org/10.1098/rsbl.2017.0566

Pelletier, T. A., Parsons, D. J., Decker, S. K., Crouch, S., Franz, E., Ohrstrom, J., & Carstens, B. C. (2022). PHYLOGATR: Phylogeographic data aggregation and repurposing. *Molecular Ecology Resources*, *22*(8), 2830–2842. https://doi.org/10.1111/1755-0998.13673

Pereira, H. M. (2016). A latitudinal gradient for genetic diversity. *Science*, 353(6307), 1494–1495. https://doi.org/10.1126/science.aah6730

Quiroga-Carmona, M., & D'Elía, G. (2022). Climate influences the genetic structure and niche differentiation among populations of the olive field mouse *Abrothrix olivacea* (Cricetidae: Abrotrichini). *Scientific Reports*, 12(1), 22395. https://doi.org/10.1038/s41598-022-26937-x

R Core Team (2022). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Raven, P. H., Gereau, R. E., Phillipson, P. B., Chatelain, C., Jenkins, C. N., & Ulloa Ulloa, C. (2020). The distribution of biodiversity richness in the tropics. *Science Advances*, 6(37), eabc6228. https://doi.org/10.1126/sciadv.abc6228

Revell, L. J. (2012). phytools: An R package for phylogenetic comparative biology (and other things): *phytools: R package*. *Methods in Ecology and Evolution*, *3*(2), 217–223. https://doi.org/10.1111/j.2041-210X.2011.00169.x

Rolland, J., Silvestro, D., Schluter, D., Guisan, A., Broennimann, O., & Salamin, N. (2018). The impact of endothermy on the climatic niche evolution and the distribution of vertebrate diversity. *Nature Ecology & Evolution*, *2*(3), 459–464. https://doi.org/10.1038/s41559-017-0451-9

Romiguier, J., Gayral, P., Ballenghien, M., Bernard, A., Cahais, V., Chenuil, A., Chiari, Y., Dernat, R., Duret, L., Faivre, N., Loire, E., Lourenco, J. M., Nabholz, B., Roux, C., Tsagkogeorga, G., Weber, A. A.-T., Weinert, L. A., Belkhir, K., Bierne, N., … Galtier, N. (2014). Comparative population genomics in animals uncovers the determinants of genetic diversity. *Nature*, 515(7526), 261–263. https://doi.org/10.1038/nature13685

Schmidt, C., Munshi-South, J., Dray, S., & Garroway, C. J. (2022). Determinants of genetic diversity and species richness of North American amphibians. *Journal of Biogeography*, *49*, 2005– 2015. https://doi.org/10.1111/jbi.14480

Sgrò, C. M., Lowe, A. J., & Hoffmann, A. A. (2011). Building evolutionary resilience for conserving biodiversity under climate change: Conserving biodiversity under climate change. *Evolutionary Applications*, *4*(2), 326–337. https://doi.org/10.1111/j.1752-4571.2010.00157.x

Singhal, S., Wrath, J., & Rabosky, D. L. (2022). Genetic variability and the ecology of geographic range: A test of the central-marginal hypothesis in Australian scincid lizards. *Molecular Ecology*, 31(16), 4242–4253. https://doi.org/10.1111/mec.16589

Smith, B. T., McCormack, J. E., Cuervo, A. M., Hickerson, Michael. J., Aleixo, A., Cadena, C. D., Pérez-Emán, J., Burney, C. W., Xie, X., Harvey, M. G., Faircloth, B. C., Glenn, T. C., Derryberry, E. P., Prejean, J., Fields, S., & Brumfield, R. T. (2014). The drivers of tropical speciation. *Nature*, 515(7527), 406–409. https://doi.org/10.1038/nature13687

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & Van Der Linde, A. (2002). Bayesian Measures of Model Complexity and Fit. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 64(4), 583–639. https://doi.org/10.1111/1467-9868.00353

Stange, M., Barrett, R. D. H., & Hendry, A. P. (2021). The importance of genomic variation for biodiversity, ecosystems and people. *Nature Reviews Genetics*, *22*(2), 89–105. https://doi.org/10.1038/s41576-020-00288-7

Sunyaev, S., Kondrashov, F. A., Bork, P., & Ramensky, V. (2003). Impact of selection, mutation rate and genetic drift on human genetic variation. *Human Molecular Genetics*, 12(24), 3325–3330. https://doi.org/10.1093/hmg/ddg359

Templeton, A. R. (2006). Modeling Evolution and the Hardy–Weinberg Law. In *Population Genetics and Microevolutionary Theory* (pp. 19–47). John Wiley & Sons, Ltd. https://doi.org/10.1002/0470047356.ch2

Tobar-Suárez, C., Urbina-Cardona, N., Villalobos, F., & Pineda, E. (2022). Amphibian species richness and endemism in tropical montane cloud forests across the Neotropics. *Biodiversity and Conservation*, 31(1), 295–313. https://doi.org/10.1007/s10531-021-02335-z

Urbina-Cardona, J. N. (2008). Conservation of Neotropical Herpetofauna: Research Trends and Challenges. *Tropical Conservation Science*, *1*(4), 359–375. https://doi.org/10.1177/194008290800100405

Vacher, J.-P., Kok, P. J. R., Rodrigues, M. T., Lima, J. D., Lorenzini, A., Martinez, Q., Fallet, M., Courtois, E. A., Blanc, M., Gaucher, P., Dewynter, M., Jairam, R., Ouboter, P., Thébaud, C., & Fouquet, A. (2017). Cryptic diversity in Amazonian frogs: Integrative taxonomy of the genus *Anomaloglossus* (Amphibia: Anura: Aromobatidae) reveals a unique case of diversification within the Guiana Shield. *Molecular Phylogenetics and Evolution*, 112, 158–173. https://doi.org/10.1016/j.ympev.2017.04.017

van den Burg, M. P., Herrando-Pérez, S., & Vieites, D. R. (2020). ACDC, a global database of amphibian cytochrome-b sequences using reproducible curation for GenBank records. *Scientific Data*, *7*(1), 268. https://doi.org/10.1038/s41597-020-00598-9

Wake, D. B., & Marks, S. B. (1993). Development and Evolution of Plethodontid Salamanders: A Review of Prior Studies and a Prospectus for Future Research. *Herpetologica*, 49(2), 194 – 203.

Wake, D. B., & Hanken, J. (1996). Direct development in the lungless salamanders: What are the consequences for developmental biology, evolution and phylogenesis? *The International Journal of Developmental Biology*, 40, 859 – 869.

White, E. P., Ernest, S. K. M., Kerkhoff, A. J., & Enquist, B. J. (2007). Relationships between body size and abundance in ecology. *Trends in Ecology & Evolution*, 22(6), 323–330. https://doi.org/10.1016/j.tree.2007.03.007

Wiens, J. J. (2007). Global Patterns of Diversification and Species Richness in Amphibians. *The American Naturalist*, 170(S2), S86–S106. https://doi.org/10.1086/519396

Jetz, W., & Pyron, R.A. (2018). The interplay of past diversification and evolutionary isolation with present imperilment across the amphibian tree of life. *Nature Ecology & Evolution*, *2*, 850–858. https://doi.org/10.1038/s41559-018-0515-5

Biosketch

Luis Amador is an evolutionary biologist, currently studying the determinants of global amphibian genetic/genomic diversity as a postdoctoral fellow in the Amphibian and Reptile Biodiversity Lab (ARBL) at the University of New Mexico (UNM). Irvin Arroyo-Torres contributed to this work as an undergraduate researcher and is pursuing graduate studies in ecology and conservation at UNM. Lisa Barrow is Curator of Amphibians and Reptiles at the Museum of Southwestern Biology and principal investigator of ARBL at UNM, which focuses on research in evolution, ecology, and conservation of herpetofauna.

Author contributions: L.A. and L.N.B. conceived the ideas; L.A. and I.A.T. assembled the data; L.A. analyzed the data; and L.A. and L.N.B. led the writing.

Appendices

Appendix S1 Materials and methods – supplementary information.

Appendix S2 Supplementary results.

Appendix S3 Discussion – supplementary information.

Tables

Table 1. Variables used to predict nucleotide diversity in Neotropical amphibians

Variable	Value
Habitat	Aquatic, Arboreal, Burrowing, Semi-aquatic, Semi-arboreal, Terrestrial
Activity	Diurnal, Nocturnal
Body size	millimeter (mm)
Development mode	Direct, Larval
Number of sequences	>5
Mean Elevation	meters above sea level (masl)
Minimum Elevation	meters above sea level (masl)
Maximum Elevation	meters above sea level (masl)
Mean Latitude	decimal degrees
Minimum Latitude	decimal degrees
Maximum Latitude	decimal degrees
Range size	area in square kilometers (km2)
Order	Anura, Caudata
Family	Anura (17 families), Caudata (Plethodontidae)
Biogeographic Unit	Amazonia, Andes, Antillean, Atlantic Forest, Cerrado, Chaco, Chilean-Valdivian, Mexican Transition Zone, Mesoamerica, Monte, Northern Venezuela-Colombia, Orinoco, Pampas, Patagonia, Tumbes-Choco-Magdalena
Conservation Status	NE = Not Evaluated, DD = Data Deficient, LC = Least Concern, NT = Near Threatened, VU = Vulnerable, EN = Endangered, CR = Critically Endangered, EX = Extinct

Table 2. Model comparison with deviance information criterion (DIC) scores from each MCMCglmm model for both nucleotide diversity (π) and Isolation-by-Distance (IBD). Comparisons were made with all or a reduced number of predictors and without predictors; with no random effects and with species and phylogenetic (phylo) random effects. For each model, four different runs were performed; we present the average DIC score with the standard deviation (SD).

Model detail					Response variable	
Predictors			Random effects		Nucleotide diversity (π) – DIC (SD)	Isolation-by-Distance (IBD) – DIC (SD)
None	Reduced	All	Species	Phylo		
X					-1002.38 (0.01)	334.07 (0.01)
X			X	X	-1101.45 (1.52)	23.51 (14.96)
	X				-1013.37 (0.02)	330.56 (0.02)
	X		X		-1020.53 (0.20)	8.64 (2.95)
	X		X	X	**-1108.75 (1.50)**	7.67 (0.70)
		X	X	X	-1088.66 (1.21)	**4.02 (0.09)**

Figures

Figure 1. (a) 17 Neotropical biogeographic units were used in this study. Regionalization was modified from Olson et al. (2001), Josse et al. (2003), Morrone (2014), and Antonelli et al. (2018). (b) Number of observations/sequences per grid cell. (c) Amphibian genetic diversity patterns in the Neotropics. The map uses equal-area grid cells of 350 km.

Figure 2. Top predictors and variable importance for Random Forest models in analyses of all amphibians. (a) Random Forest regression for π, and (b) Random Forest classification for IBD.

Figure 3. Phylogeny of Neotropical amphibians (downloaded and subset from Jetz and Pyron, 2018) with (a) values of CytB π presented as blue bars, and (b) presence (black dots) or absence (yellow dots) of IBD in Neotropical amphibians. Silhouettes of frogs and salamanders were obtained from phylopic.org.

Figure 4. Posterior distribution of (a) nucleotide diversity and (b) Isolation-by-Distance across a reduced number of predictors, plotting predictor estimates and associated 95% credible intervals (CIs). The 95% CIs do not cross zero for range size and minimum elevation in A, and for minimum elevation in B, suggesting that these predictors are statistically significant. (c) Range size (log) and mtDNA π for all amphibian species in our dataset. The color scale indicates minimum elevation of occurrences in each species. (d) Minimum elevation as predictor of genetic structure or Isolation-by-Distance (IBD). No IBD (red), IBD (green). Each box-whisker plot indicates the median (bold lines), the interquartile range (boxes), white dots represent outliers.

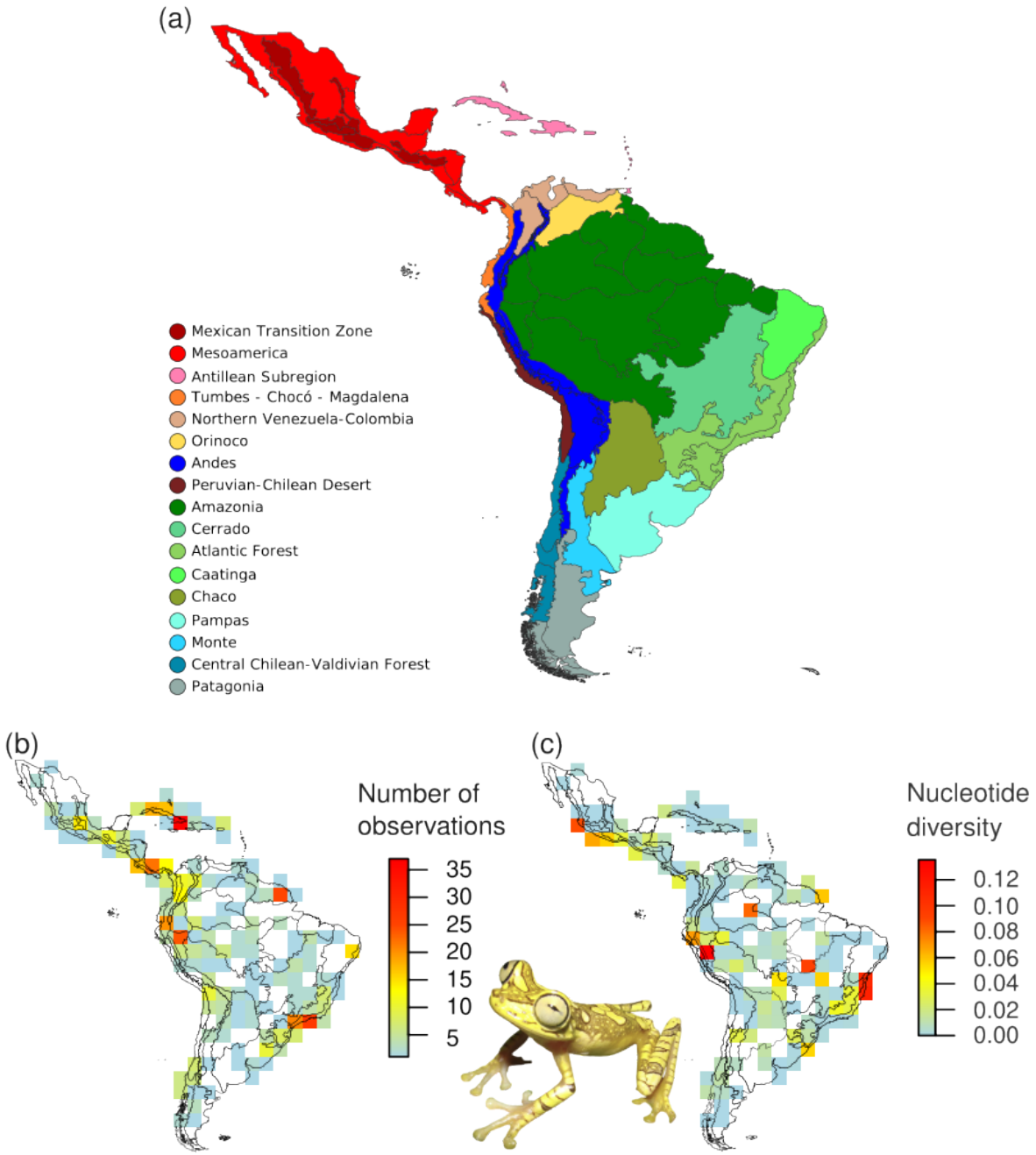

Figure 1. (a) 17 Neotropical biogeographic units were used in this study. Regionalization was modified from Olson et al. (2001), Josse et al. (2003), Morrone (2014), and Antonelli et al. (2018). (b) Number of

observations/sequences per grid cell. (c) Amphibian genetic diversity patterns in the Neotropics. The map uses equal-area grid cells of 350 km.

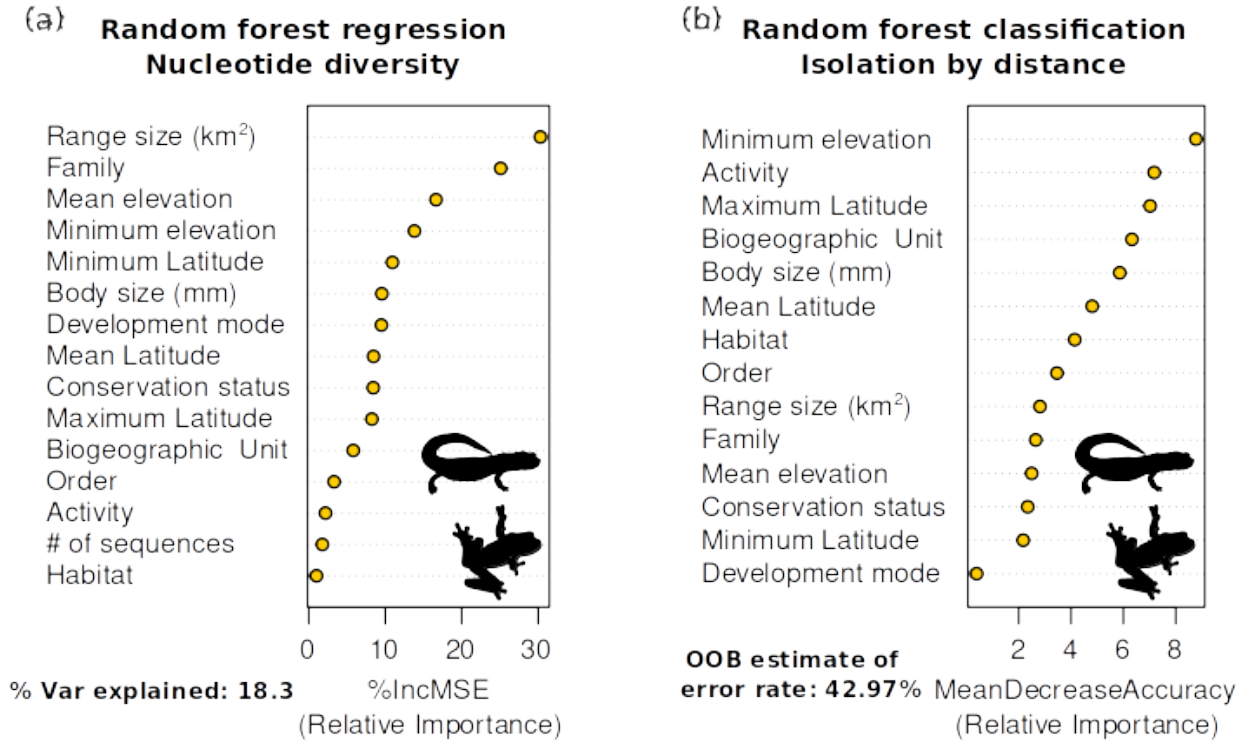

Figure 2. Top predictors and variable importance for Random Forest models in analyses of all amphibians. (a) Random Forest regression for π, and (b) Random Forest classification for IBD.

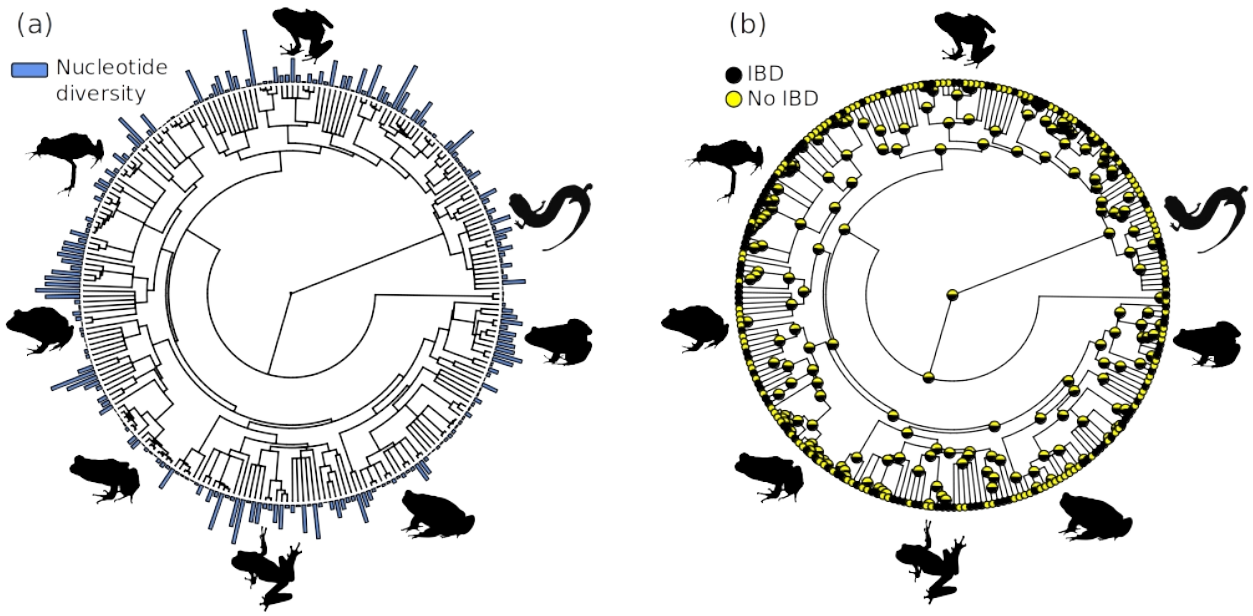

Figure 3. Phylogeny of Neotropical amphibians (downloaded and subset from Jetz and Pyron, 2018) with (a) values of CytB π presented as blue bars, and (b) presence (black dots) or absence (yellow dots) of IBD in Neotropical amphibians. Silhouettes of frogs and salamanders were obtained from phylopic.org.

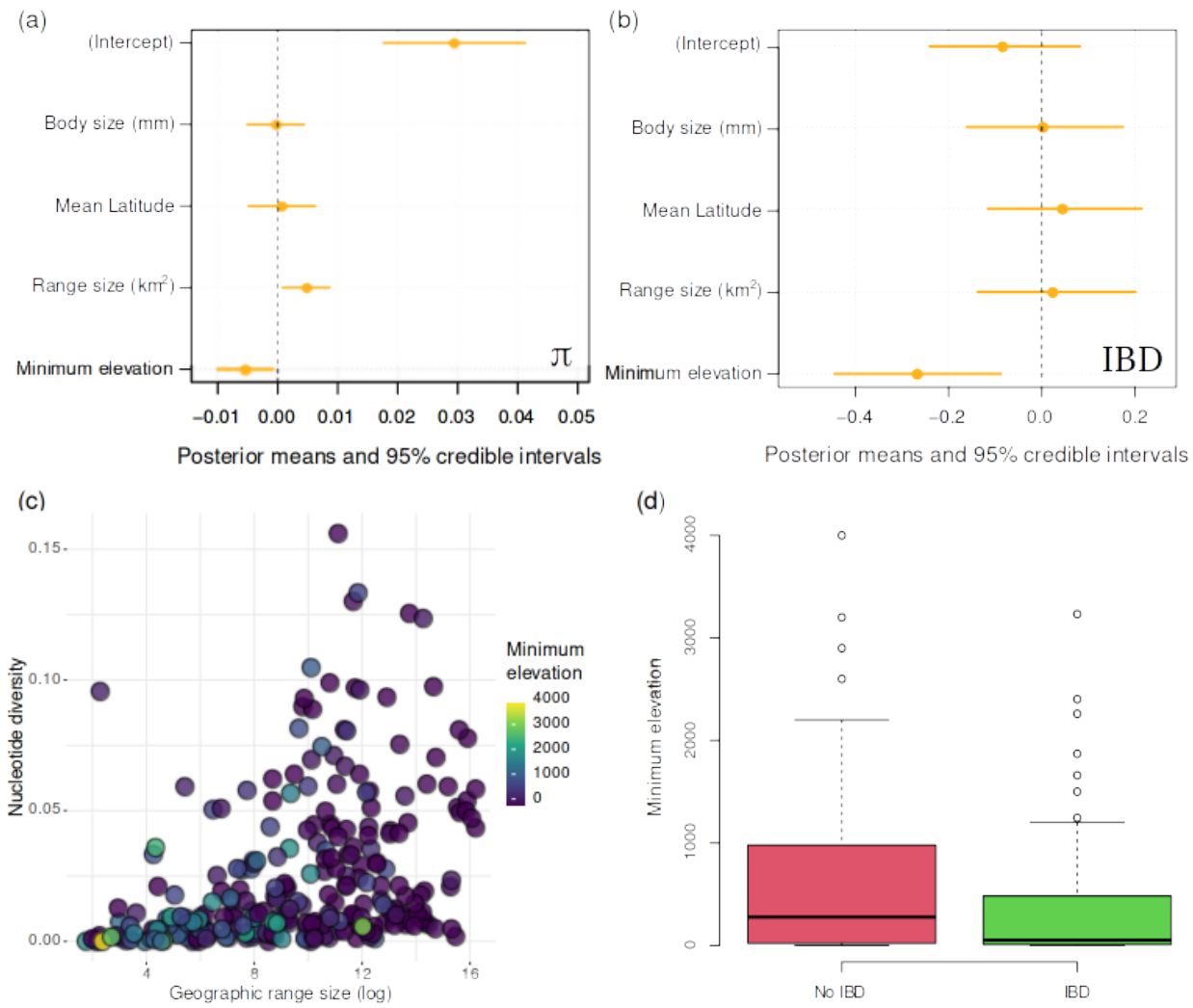

Figure 4. Posterior distribution of (a) nucleotide diversity and (b) Isolation-by-Distance across a reduced number of predictors, plotting predictor estimates and associated 95% credible intervals (CIs). The 95% CIs do not cross zero for range size and minimum elevation in A, and for minimum elevation in B, suggesting that these predictors are statistically significant. (c) Range size (log) and mtDNA π for all amphibian species in our dataset. The color scale indicates minimum elevation of occurrences in each species. (d) Minimum elevation as predictor of genetic structure or Isolation-by-Distance (IBD). No IBD (red), IBD (green). Each box-whisker plot indicates the median (bold lines), the interquartile range (boxes), white dots represent outliers.