Received: 20 June 2023 | Revised: 21 December 2023 | Accepted: 22 December 2023

DOI: 10.1111/jbi.14795

RESEARCH ARTICLE

Journal of Biogeography

WILEY

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

Luis Amador | Irvin Arroyo-Torres | Lisa N. Barrow

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

Correspondence
Lisa N. Barrow, Museum of Southwestern Biology and Department of Biology, University of New Mexico, Albuquerque, NM 87131, USA.
Email: lnbarrow@unm.edu

Funding information
U.S. National Science Foundation, Grant/Award Number: DEB-2112946

Abstract

Aim: Intraspecific genetic variation is key for adaptation and survival in changing environments and is known to be influenced by many factors, including population size, dispersal and life-history traits. We investigated genetic variation within Neotropical amphibian species to provide insights into how natural history traits, phylogenetic relatedness, climatic and geographic characteristics can explain intraspecific genetic diversity.

Location: Neotropics.

Taxon: Amphibians.

Methods: We assembled data sets using open-access databases for natural history traits, genetic sequences, phylogenetic trees, climatic and geographic data. For each species, we calculated overall nucleotide diversity (π) and tested for isolation by distance (IBD) and isolation by environment (IBE). We then identified predictors of π, IBD and IBE using random forest (RF) regression or RF classification. We also fitted phylogenetic generalized linear mixed models (PGLMMs) to predict π, IBD and IBE.

Results: We compiled 4052 mitochondrial DNA sequences from 256 amphibian species (230 frogs and 26 salamanders), georeferencing 2477 sequences from 176 species that were not linked to occurrence data. RF regressions and PGLMMs were congruent in identifying range size and precipitation (σ) as the most important predictors of π, influencing it positively. RF classification and PGLMMs identified minimum elevation as an important predictor of IBD; most species without IBD tended to occur at higher elevations. Maximum latitude and precipitation (σ) were the best predictors of IBE, and most species without IBE occur at lower latitudes and in areas with more variable precipitation.

Main Conclusions: This study identified predictors of genetic variation in Neotropical amphibians using both machine learning and phylogenetic methods. This approach was valuable to determine which predictors were congruent between methods. We found that species with small ranges or living in zones with less variable precipitation tended to have low genetic diversity. We also showed that Western Mesoamerica, Andes and Atlantic Forest biogeographic units harbour high diversity across many species that should be prioritized for protection. These results could play a key role in the development of conservation strategies for Neotropical amphibians.

KEYWORDS
Anura, Caudata, elevation, latitude, phylogeny, precipitation, random forest, range size

1 | INTRODUCTION

Characterizing genetic variation within species and understanding the processes that maintain that diversity are important goals in evolutionary biology (Ellegren & Galtier, 2016). Genetic variation is essential for natural populations to develop new traits and adapt to future environmental changes (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 and the distribution of spatial genetic variation between populations. 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). Therefore, investigating the forces that drive genetic diversity within species and how this diversity is distributed and influenced spatially can help explain the processes underlying the patterns of genetic variation that we observe in nature.

Levels of genetic variation vary greatly in natural populations and among species (see 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). 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 genetic diversity (e.g. Grundler et al., 2019; Hague & Routman, 2016). However, genetic diversity 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, genetic variation within species can be influenced by several additional factors, including environmental and intrinsic characteristics (e.g. Nevo, 1978).

Distinct geographic, climatic and life-history factors can determine intraspecific genetic variation and influence the diversification and extinction of populations and species. For example, species with larger ranges are expected to have higher genetic diversity, low inbreeding and reduced genetic drift because of the direct relationship between geographic range size and population size (see Leffler et al., 2012). García-Rodríguez et al. (2021) found that abiotic factors and geographical features affected the genetic diversity of nine co-distributed amphibian species in Isthmian Central America. In addition, habitat fragmentation can lead to reduced gene flow and increased genetic differentiation among populations, ultimately reducing genetic diversity within populations because of genetic drift (see Dixo et al., 2009). Life-history traits influence genetic variation, providing a connection between different demographic processes (Duminil et al., 2007). For example, species with shorter life spans and higher fecundity may have higher genetic diversity because there are more chances for mutation and recombination, and to cover the full gradient of environmental pressures experienced by the species (Romiguier et al., 2014). Evaluating body size as a predictor of genetic variation, Brüniche-Olsen et al. (2019) found that genetic diversity decreases with increasing size in Darwin's finches, with the possible explanation that species abundance is expected to decrease with increasing size (White et al., 2007). Larger bodied species may also have higher dispersal ability resulting in differences in spatial genetic variation, as has been demonstrated in bees (López-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. One approach is to collect new genetic data for a set of taxa of interest and analyse linear models of genetic variation with possible predictors (e.g. Dixo et al., 2009). An alternative approach, which can enable comparisons of many more species, is to use repurposed data that were previously collected for a different research purpose and that can be reanalysed in a common framework. These two approaches converge in macrogenetics (sensu Blanchet et al., 2017), a field that focuses on the integration of genetic data sets from multiple species at large scales with environmental data sets to identify drivers of intraspecific genetic variation (Leigh et al., 2021). Recently, Pelletier and Carstens (2018) applied a machine learning framework to repurpose georeferenced DNA sequences from more than 8000 species and found that geographic range size and latitude were the most important predictors of genetic structure. Another approach to address how landscapes contribute to the evolution of genetic variation quantifies the effects of geography and ecology using multiple matrix regression with randomization (MMRR; Wang, 2013). For example, Wieringa et al. (2020) used this approach and found that geographic and environmental distance were significant factors in several species in the southeastern United States.

Amphibians 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) and distributional patterns (e.g. species with limited dispersal capabilities that lead to high levels of genetic differentiation). Many studies have explored the patterns of diversity in amphibians (e.g. 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 genetic variation. In a study of Nearctic amphibian species, the most important predictors of genetic diversity were taxonomic family, number of sequences, and for salamander species ($N=98$), those at more northern latitudes had lower genetic diversity (Barrow et al., 2021). In the same region, Schmidt et al. (2022) analysed microsatellite data from 19 amphibian species; they found that genetic diversity was not predicted by the environmental variables they used and that areas with high species richness also had high genetic structure, but low genetic diversity.

The Neotropical Region includes almost 50% of the world's total amphibian species, which is greater than any other comparable area on the planet (Menéndez-Guerrero et al., 2020). This exceptionally high diversity is possibly due to a combination of factors, including the geological history of the region (e.g. the formation of the Isthmus of Panama, Andean uplift), climate change, ecological interactions and biotic diversification (see Antonelli, 2022; Elmer et al., 2013). For decades, the origins and maintenance of diversity in this region have been an important focus of research in amphibian ecology and systematics (e.g. Castroviejo-Fisher et al., 2014). 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 genetic variation in Neotropical amphibians.

Here, we investigated what factors predict genetic variation within Neotropical frog and salamander species. We gathered genetic sequences, natural history traits, phylogenetic relationships and climatic and geographic information for each species from open-access databases and literature. We estimated nucleotide diversity (π) and tested for isolation-by-distance (IBD) and isolation-by-environment (IBE) from repurposed DNA sequences. To identify potential hotspots of intraspecific diversity, we built maps of nucleotide diversity across the Neotropics. We then applied machine learning and phylogenetic linear models to investigate the predictors of π, IBD and IBE within Neotropical amphibians.

2 | MATERIALS AND METHODS

2.1 | DNA sequences and associated geographic coordinates

We obtained sequences of the mitochondrial gene cytochrome-b (Cytb) from open access databases. We chose Cytb since this was the most abundant gene in Neotropical amphibian studies and it has informative variation within species (van den Burg et al., 2020; Zeisset & Beebee, 2008). 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 v.1.28 (Larsson, 2014) with the MUSCLE aligner v.3.8.31 (Edgar, 2004) using default settings. Geographic coordinates from each sequence were obtained from phylogatR and GenBank when available, which corresponded to 1575 sequences (38.87% of the total sequences) from 80 species (31.25% of the total species). We linked 2477 additional sequences from 176 species 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).

2.2 | Metrics of genetic variation

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 spatial genetic variation within amphibian species, we calculated IBD and IBE for each species. The 'raw' genetic distance (*gendist*; the proportion of sites that differ between each pair of sequences) was calculated using the function dist.dna() in the *ape* R package (Paradis & Schliep, 2019). Topographic distance (*geodist*) was calculated between coordinates associated with each sequence using the function topoDist() implemented in the *topoDistance* R package (Wang, 2020). Environmental distance (*envdist*) was calculated based on 19 bioclimatic variables from the WorldClim database (Hijmans et al., 2005) and tree cover data derived from the ESA WorldCover database (Zanaga et al., 2021). These layers were retrieved at 30-s spatial resolution using the function landcover() of the *geodata* R package (Hijmans et al., 2023). We extracted values of each variable in each locality for all species in the analyses using the extract() function in the *raster* R package (Hijmans, 2022). Then using the scale() function in R, we standardized the data to have comparable values to analyse. Finally, we performed a principal components analysis with the prcomp() function in R, and calculated *envdist* between localities using the first principal component. We conducted Multiple Matrix Regression with Randomization analyses with the MMRR function in R (Wang, 2013) using the three distances, *gendist*, *geodist* and *envdist*, to determine whether each species showed significant IBD and IBE. Regression coefficients of geography (IBD, β_D) and ecology (IBE, β_E) and their significance were calculated after 10,000 permutations. MMRR analyses require data sets with $n > 4$ and assume that >3 samples show differences in their distances. Based on this premise, we used 194 species for this analysis. After classifying each species as having significant IBD or not and significant IBE or not, we used these binary classifications as response variables in subsequent analyses.

2.3 | Trait and geographic data compilation

We obtained information on traits from the AmphiBIO data set (Oliveira et al., 2017), AmphibiaWeb (2022) and specific data from manuscripts (Appendix S1; Table 1). Body size (snout–vent length), 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, the International Union for Conservation of Nature (IUCN) and WorldClim (Fick & Hijmans, 2017). From the latter, we also obtained temperature (BIO1 = Annual Mean Temperature) and precipitation (BIO12 = Annual Precipitation). 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

TABLE 1 Variables used in this study to predict nucleotide diversity, isolation-by-distance and isolation-by-environment in Neotropical amphibians. Variables with an asterisk (*) indicate non-biologically motivated variables and were not used in all analyses. The rationale for considering each variable is provided as an example prediction and is not meant to be an exhaustive explanation.

Variable	Value	Rationale
Habitat	Aquatic, arboreal, burrowing, semi-aquatic, semi-arboreal, terrestrial	Some habitats may have more natural barriers (e.g. fully aquatic species would have lower dispersal), leading to genetic differentiation of populations
Activity	Diurnal, nocturnal	Diurnal or nocturnal activity could restrict or promote gene flow, leading to genetic variation between populations
Body size	Millimetre (mm)	Larger bodied species are expected to have higher dispersal ability, facilitating gene flow between populations. Larger bodied species also may have limits on population abundance
Development mode	Direct, larval	Species with larval development may have lower dispersal ability since they are tied to water bodies
Sampling effort*	Number of sequences >5	Sampling effort can be considered as a proxy of abundance (population size). Species with more sequences sampled are expected to have higher genetic variation
Sequence length*	Number of base pairs	Species with longer cytochrome-b fragments may have more variable sites sampled
Elevation (mean, minimum, maximum)	Metres above sea level (m a.s.l.)	Species at higher elevations may experience physical barriers or occur in more isolated populations, increasing genetic differentiation
Latitude (mean, minimum, maximum)	Decimal degrees	Latitudinal gradients in climatic variables can influence patterns of distribution and abundance within species
Precipitation (mean, σ)	Precipitation millimetre (mm)	Amphibian diversity is directly affected by precipitation, e.g. species living in areas with high precipitation may have higher abundances
Temperature (mean, σ)	Average temperature (°C)	Temperature may not have strong influence on Neotropical species due to the relatively stable and mild temperatures of the region
Range size	Area in square kilometres (km^2)	Species with larger ranges are expected to have higher genetic variation due to larger population sizes and more chances of dispersal
Order	Anura, Caudata	Order-level classification may be associated with characteristics (e.g. reproductive mode, body size) that influence genetic diversity
Family	Anura (17 families), Caudata (Plethodontidae)	Family-level classification may be associated with characteristics (e.g. reproductive mode, distribution) that influence genetic diversity
Biogeographic unit	17 units shown in Figure 1a	Biogeographic units are associated with different climates, geologic features or total area that can affect population sizes and gene flow between populations leading to genetic differentiation

Conservation status	NE = not evaluated
	DD = data deficient
	LC = least concern
	NT = near threatened
	VU = vulnerable
	EN = endangered
	CR = critically endangered
	EX = extinct
		Threatened species with small population sizes are expected to have low genetic diversity; least concern species are expected to have large range sizes and associated high genetic diversity

Land cover (mean)	0/1, 1 indicates forest presence	Species living in forested areas (greater number of trees) are expected to have higher abundances and more genetic variation than species living in deforested areas

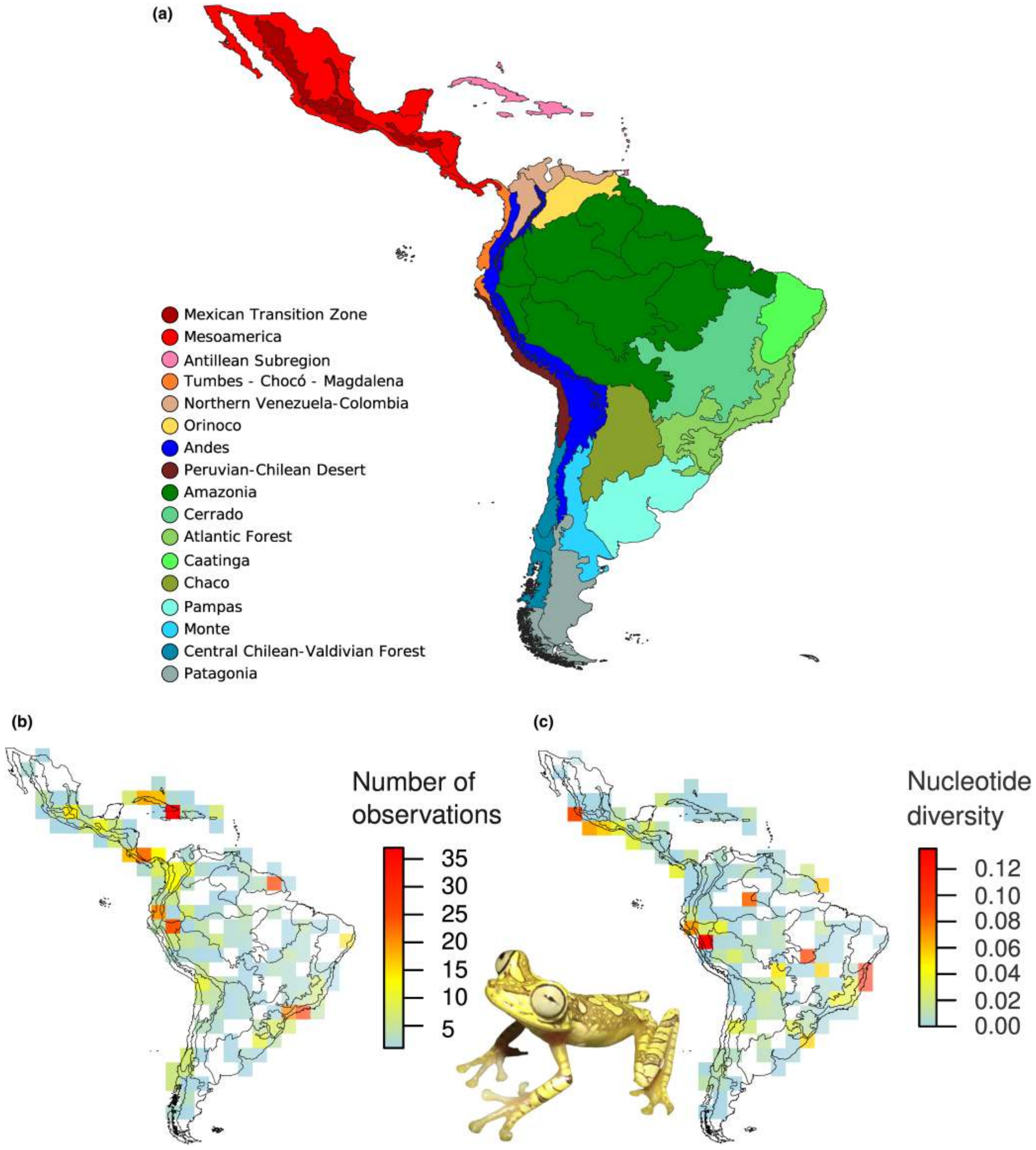

FIGURE 1 (a) Neotropical biogeographic units used in this study. Limits of the biogeographic regionalization of the neotropical region follows Morrone (2014), biogeographic unit classification was adapted from Josse et al. (2003), Morrone (2014) and Antonelli et al. (2018). (b) Number of observations/sequences per grid cell. (c) Amphibian genetic diversity patterns in the Neotropics. The (b) and (c) maps use equal-area grid cells of 350 km.

downloaded from the IUCN portal. We obtained the shapefile for 24 species (e.g. *Allobates hodli*, *Bolitoglossa awajun*) not included in IUCN by calculating minimum convex polygons based on the sequence coordinates and using the function mcp() with the *adehabitatHR* R package (Calenge & Fortmann-Roe, 2023). When necessary, we corrected the polygon for species recently split and not included in IUCN, for example, *Rhinella marina–Rhinella horribilis* (Acevedo et al., 2016). The range area was calculated in square kilometres (km^2) using the functions shapefile() of the *raster* package and areaPolygon() of the *geosphere* package (Hijmans, 2021) in

R 4.2.2 (R Core Team, 2022). The IUCN Red List Category was also recorded for each species (IUCN, 2022). We tested for multicollinearity of the continuous variables using variance inflation factors (VIF) implemented in the R package *car* (Fox & Weisberg, 2019) using a threshold of VIF < 5 to detect evidence of collinearity between variables.

2.4 | 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 adapting the classifications proposed by Morrone (2014), Antonelli et al. (2018) and Josse et al. (2003) (Figure 1a). We calculated and mapped intraspecific nucleotide diversity first with a single value of π per species (mean value of π) and second with a value of π per locality within each species. Each locality consisted of at least two individuals, and we assigned sequences to the same locality when they shared the same coordinates. In cases when the sequences were from different geographic coordinates, we combined sequences based on the distance between points (150 km or less) and shared habitat characteristics (e.g. occurring in the same mountain range or river). Although this approach involved arbitrarily choosing a somewhat coarse resolution, it allowed us to maximize the species and geographic area included to provide an initial picture of spatial genetic diversity given the data available. We visualized π for three different resolutions (grid cell size of 150, 250 and 350 km^2; Figure 1c; Figure S1b,d) which corresponds to 3640, 1326 and 673 grid cells, 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. All mapping analyses were performed in R. To test for spatial autocorrelation in sampling effort (number of observations) and π, we computed the Moran's *I* statistic by using the moran.test() function of the *spdep* package (Bivand, 2022).

2.5 | Identifying predictors of genetic variation 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 π, IBD and IBE 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 RF 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 and IBE) models were created using the randomForest() function in the *randomForest* R package (Liaw & Wiener, 2002), with 5000 trees and 100 permutations. We split the data into training (90%) and test (10%) data sets and created RF models with the training data. We used the tuneRF() function to find the optimal mtry value (number of variables to randomly sample as candidates at each split), and a new model was built using the best mtry value. We made predictions on the training and test (unseen data values in the models) data sets that were evaluated with the mean squared error and R^2 metrics in the RF regression analysis and with the confusion matrix in RF classification analyses. Relative importance for each independent variable was measured and printed using the importance() function in the *randomForest* package, and visualized with the vip() function of the *vip* R package (Greenwell & Boehmke, 2020) for each model. To evaluate the effect 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 two different reduced data sets: (1) a data set with at least 10 sequences per species (114 species) and (2) a data set with at least 400 bp in each species alignment (197 species).

2.6 | Testing predictors of genetic variation with phylogenetic comparative methods

RF can handle large numbers of variables to build predictive models, but do not explicitly incorporate phylogenetic relationships (other than taxonomic level as a possible predictor). Therefore, we also fitted PGLMMs to determine relationships between natural history traits, geographic and climatic variables and genetic variation of Neotropical amphibians. We pruned the maximum clade credibility tree from Jetz and Pyron (2018) to create a phylogeny that includes only the species in our data set using the phylo4() function in the *phylobase* R package (Hackathon et al., 2020). This resource is the most complete phylogeny available (7238 species), covering ≈83% of the known amphibian diversity. Of the 256 species in our data set, 15 species were not included in the phylogeny of Jetz and Pyron (Appendix S2); therefore for phylogenetic analyses, we used a subset consisting of 241 species. We mapped π as a continuous trait using the contMap() function in the *phytools* R package (Revell, 2012; Figure S14), and the distribution of IBD and IBE was mapped at the tips on the tree. We tested for phylogenetic signal in π using two metrics, Pagel's λ (lambda; Pagel, 1999) and Blomberg's *K* (Blomberg et al., 2003), calculated with the phylosig() function implemented in *phytools*. We tested for phylogenetic signal in IBD and IBE using the phylo.d() function that calculates the *D* statistic, a measure of phylogenetic signal in a binary trait (Fritz & Purvis, 2010), implemented in the R package *caper* (Orme et al., 2018).

Finally, we used PGLMMs (Ives & Helmus, 2011) implemented in the *MCMCglmm* R package (Hadfield, 2010) to investigate relationships between a subset of the predictors and three responses: (1) π, (2) IBD and (3) IBE, while accounting for phylogeny. 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). We defined priors for models with no random effects, with either species or phylogenetic random effects, and with both species and phylogenetic random effects. For each response variable (π, IBD and IBE), we compared 10 models that were generated with different combinations of predictors based on the results of RF analyses (Table S2). 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 six predictors and no random effect, (4) a model with six predictors and species as a random effect, (5) a model with six predictors and a phylogenetic random effect, (6) a model with six predictors and species and phylogenetic random effects, (7) a model with 12 predictors and no random effects, (8) a model with 12 predictors and a species random effect, (9) a model with 12 predictors and a phylogenetic random effect and (10) a model with 12 predictors and with both species and phylogenetic random effects. We ran each model four times with 2×10^6 MCMC iterations, thinning interval = 100 and burn-in = 200,000. For Bayesian model selection, we used the deviance information criterion. *MCMCglmm* results were verified by checking model diagnostics using trace and density plots of the MCMC samples.

3 | RESULTS

3.1 | Data compilation

We compiled 4052 Cytb mtDNA sequences from 256 Neotropical amphibians (Figure S2), 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 by retrieving coordinates from published literature or by georeferencing localities. To visualize sampling effort across the region, we mapped the number of sequences per grid cell (Figure 1b; Figure S1a,c). There was weak positive spatial autocorrelation of number of sequences (Moran's $I = 0.0569$, p-value = 0.019), suggesting that the number of sequences sampled is somewhat clustered together. The total occurrences represented all 17 Neotropical biogeographic units (Figure S3). The final data set included 230 frogs in 17 families and 26 salamanders in family Plethodontidae, three response variables (π, IBD and IBE) and 22 predictor variables (Table 1; Table S1). We did not detect correlation among predictor variables in the model (VIF values <5) except for the three elevation variables (mean, minimum and maximum) (Figure S4).

3.2 | Distribution of genetic variation 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). When we visualized nucleotide diversity per locality, we found higher π values in western Mesoamerica, central Andes, Atlantic Forest and a few areas of the Amazon region, compared to their adjacent biogeographic units. The southern Andes and the southern portion of South America showed low genetic diversity (Figure 1c). Mean values of π per species were higher in southern Mesoamerica, the Chocó region, northern Andes and Atlantic Forest region (Figure S5). Spatial autocorrelation analysis suggests that π values are not randomly distributed in geographic space (Moran's $I = 0.0662$, p-value = 0.009). With respect to spatial genetic variation, we found that 89 Neotropical amphibian species showed significant IBD, while 105 species did not (Table S1). Species following an IBD pattern were mostly found in Mesoamerica ($n = 18$), the Antillean Subregion ($n = 13$), the Atlantic Forest ($n = 13$) and the Andes region ($n = 13$). The Amazon included the most species that did not present IBD ($n = 28$), followed by the Atlantic Forest and Andes region (Figure S6). We found that most species did not show IBE ($n = 124$) and only 70 species showed significant IBE. Species with IBE were found mostly in the Antillean region ($n = 14$), Andes ($n = 13$) and Amazonia ($n = 12$). The Amazon region also included the most species with no IBE patterns ($n = 27$), followed by the Atlantic Forest and Mesoamerica (Figure S7).

3.3 | Important predictors of genetic variation based on RFs

RF regressions showed that range size was the most important predictor of π for Neotropical amphibians (Figure 2a). Precipitation standard deviation (σ) was another important predictor of π, followed by body size and mean temperature (Figure 2a). The overall variance explained by the RF regression model shown was 20.31%. RF classification found that body size was the most important predictor of IBD, followed by mean precipitation, IUCN rank and minimum elevation (Figure 2b). The most important predictor for IBE was maximum latitude, followed by precipitation (σ) and latitude (mean and minimum) (Figure 2c). The model error for the RF classification model for IBD was 37.21% and for IBE was 40.21%. When we evaluated the effect of the sampling effort (number of sequences) per species (at least 10 sequences), and the sequence length (number of 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 analysed (Figure S8) as in the model using the complete data set. For IBD, body size remained one of the most important predictors for the 'at least 10 sequences' data set, but not the 'at least 400 bp' data set. Instead, mean temperature was the most important predictor of IBD in all the models using reduced data sets (Figure S9). Mean temperature was also the best predictor of IBE in the 'at least 400 bp' data set, while microhabitat was the best predictor for the 'at least 10 sequences' data set (Figure S10). For IBE, latitude was still among the most important predictors for both reduced data sets.

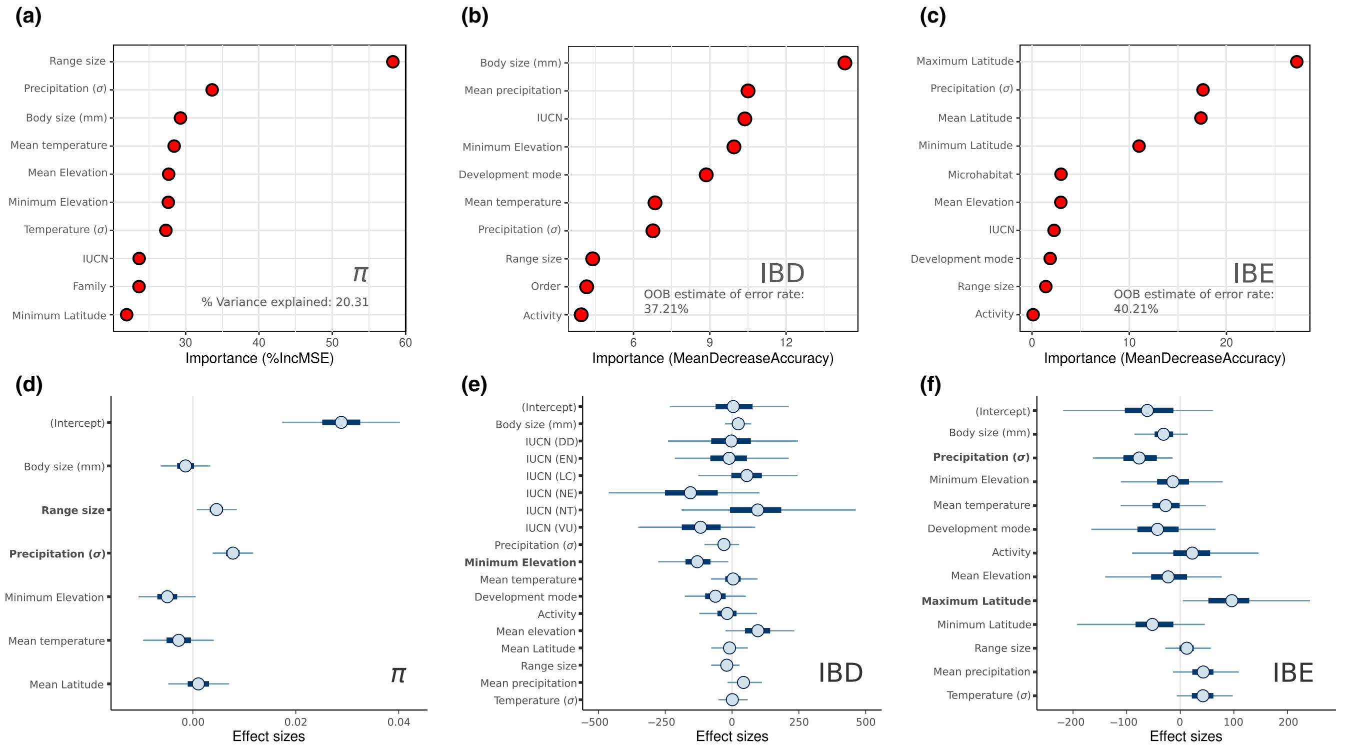

FIGURE 2 Top ranked predictors according to their variable importance for random forest models in explaining (a) π, (b) isolation-by-distance (IBD) and (c) isolation-by-environment (IBE). Effect sizes of π (d), IBD (e) and IBE (f) across predictors in *MCMCglmm* analyses. The associated 95% credible intervals do not cross zero for range size and precipitation in π, for minimum elevation in IBD and precipitation and maximum latitude in IBE; indicating that these predictors are statistically significant.

TABLE 2 Model comparison with deviance information criterion (DIC) scores from each MCMCglmm model for nucleotide diversity (π), isolation-by-distance (IBD) and isolation-by-environment (IBE). Comparisons were made with a major and 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). Bold DIC scores indicate the best model for each response.

Model detail					Response variables		
Predictors			Random effects				
None	Reduced	Major	Species	Phylo	π—DIC (SD)	IBD—DIC (SD)	IBE—DIC (SD)
X					−1002.391 (0.013)	250.5857 (0.012)	244.9121 (0.009)
X			X	X	−1101.188 (0.399)	35.5618 (16.544)	83.9517 (31.095)
	X				−1041.873 (0.028)	247.7395 (0.054)	245.0202 (0.033)
	X		X		−1043.356 (0.124)	5.3151 (0.497)	8.0073 (0.599)
	X			X	−1051.133 (0.035)	248.1518 (0.020)	245.8697 (0.008)
	X		X	X	**−1125.022 (0.302)**	5.4907 (0.756)	8.6062 (3.430)
		X			−1027.762 (0.039)	253.1265 (0.073)	257.0998 (4.340)
		X	X		−1031.574 (0.219)	4.8898 (0.578)	**4.6792 (0.359)**
		X		X	−1032.643 (0.036)	252.0281 (0.032)	259.9906 (0.037)
		X	X	X	−1116.827 (0.698)	**4.6857 (0.269)**	5.1812 (0.501)

3.4 | Relationships between predictors and genetic variation based on PGLMMs

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). This model indicated that range size and precipitation (σ) predicted π, consistent with our RF results. Both range size and precipitation (σ) had positive relationships with π (Figure 2d). Neotropical amphibian species with larger ranges and living in areas with more variable precipitation tended to have higher π (Figure 3a–c). For IBD, the best-fit *MCMCglmm* model was the full model with 12 predictors and species and phylogenetic random effects (Table 2). *MCMCglmm* also identified minimum elevation as a significant predictor of IBD, and their relationship was negative (Figure 2e). Neotropical amphibians living at higher elevations tended to have no IBD, and those with significant IBD tended to live at lower elevations (Figure 3d). Similar to RF analyses, precipitation (σ) and maximum latitude predicted IBE in PGLMM analyses (Figure 2f; Table 2). Species living at southern latitudes in the Neotropics tended to have no IBE, contrary to amphibians in northern latitudes that tended to exhibit IBE (Figure 3e). Species following an IBE pattern mostly occur in areas with lower precipitation (σ) (Figure 3f).

3.5 | Testing phylogenetic signal

We found significant phylogenetic signal in π (Pagel's $\lambda = 0.7684$, p-value (based on LR test) < 0.0001; Blomberg's $K = 0.1509$, p-value (based on 1000 randomizations) $= 0.002$) (Figure 4; Figure S11a,b). We found clusters of closely related genera that had dissimilar π values; for example, within Bufonidae, *Atelopus* species had 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 showed high π (average $\pi = 0.024$; range: 0.003–0.059) in contrast to species in the genus *Andinobates* which had very low π (average $\pi = 0.003$; range: 0.000–0.009). We did not find significant phylogenetic signal in IBD or IBE, with values of D that were greater than 1 and were overdispersed compared to a Brownian threshold model (Estimated $D_{IBD} = 1.008$, p-value $= 0.53$; Estimated $D_{IBE} = 1.047$, p-value $= 0.685$) (Figure S11c,d).

4 | DISCUSSION

In this study, we investigated the predictors of genetic variation (π, IBD and IBE) 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, precipitation, elevation and latitude were significant predictors of different aspects of genetic variation within species. Specifically, we found that amphibian species inhabiting smaller ranges and places with lower variation in precipitation had lower intraspecific π; species living at higher elevations

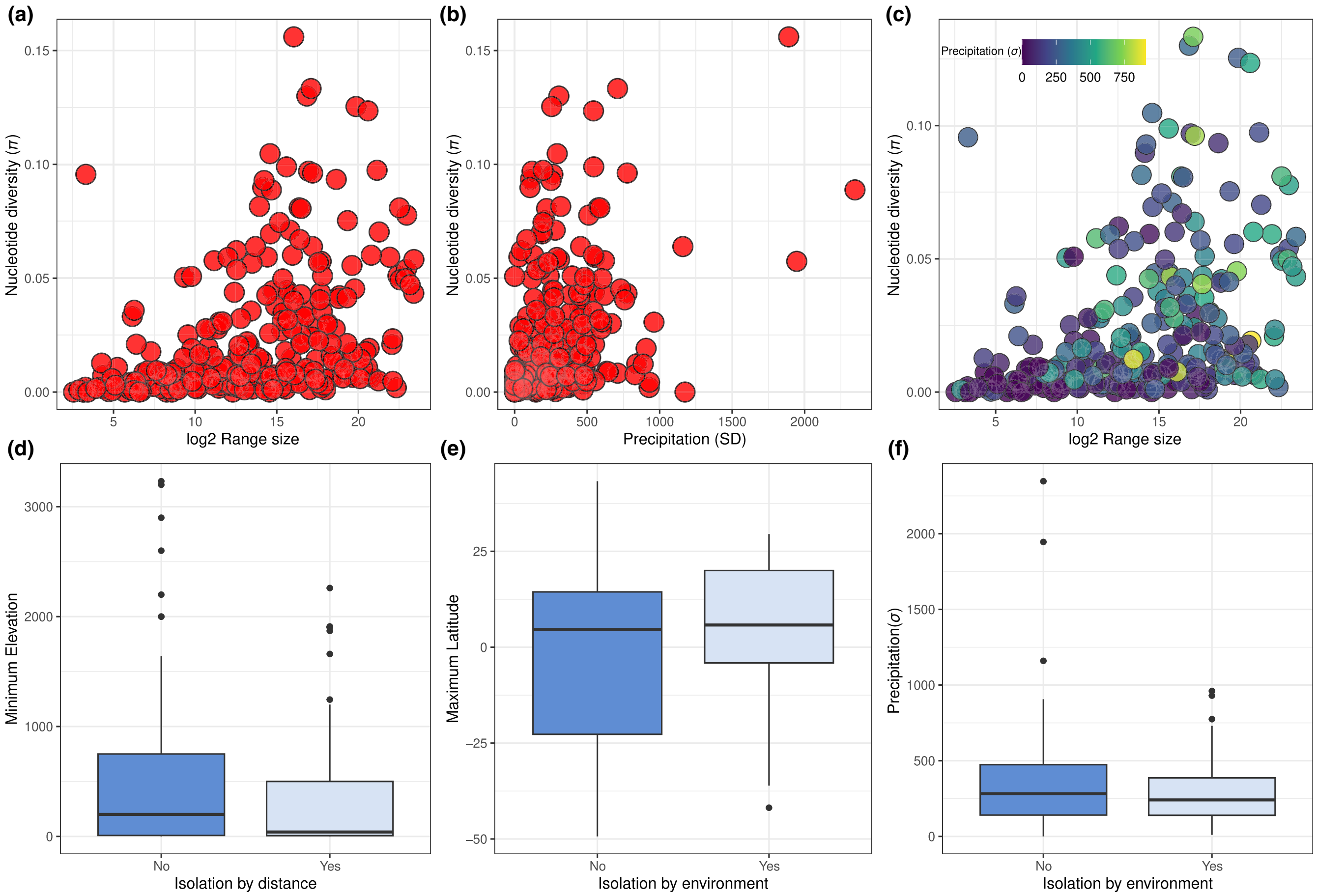

FIGURE 3 (a) Range size (log2) and mtDNA π for all amphibian species in our data set; (b) precipitation (σ) and π relationship; (c) range size (log2) and π relationship without outliers, the colours indicate precipitation (σ) of occurrences in each species. (d) Minimum elevation for species with and without of isolation-by-distance. (e, f) Maximum latitude and precipitation (σ) for species with and without isolation-by-environment. Each box–whisker plot indicates the median (bold lines), the interquartile range (boxes) and dots represent outliers.

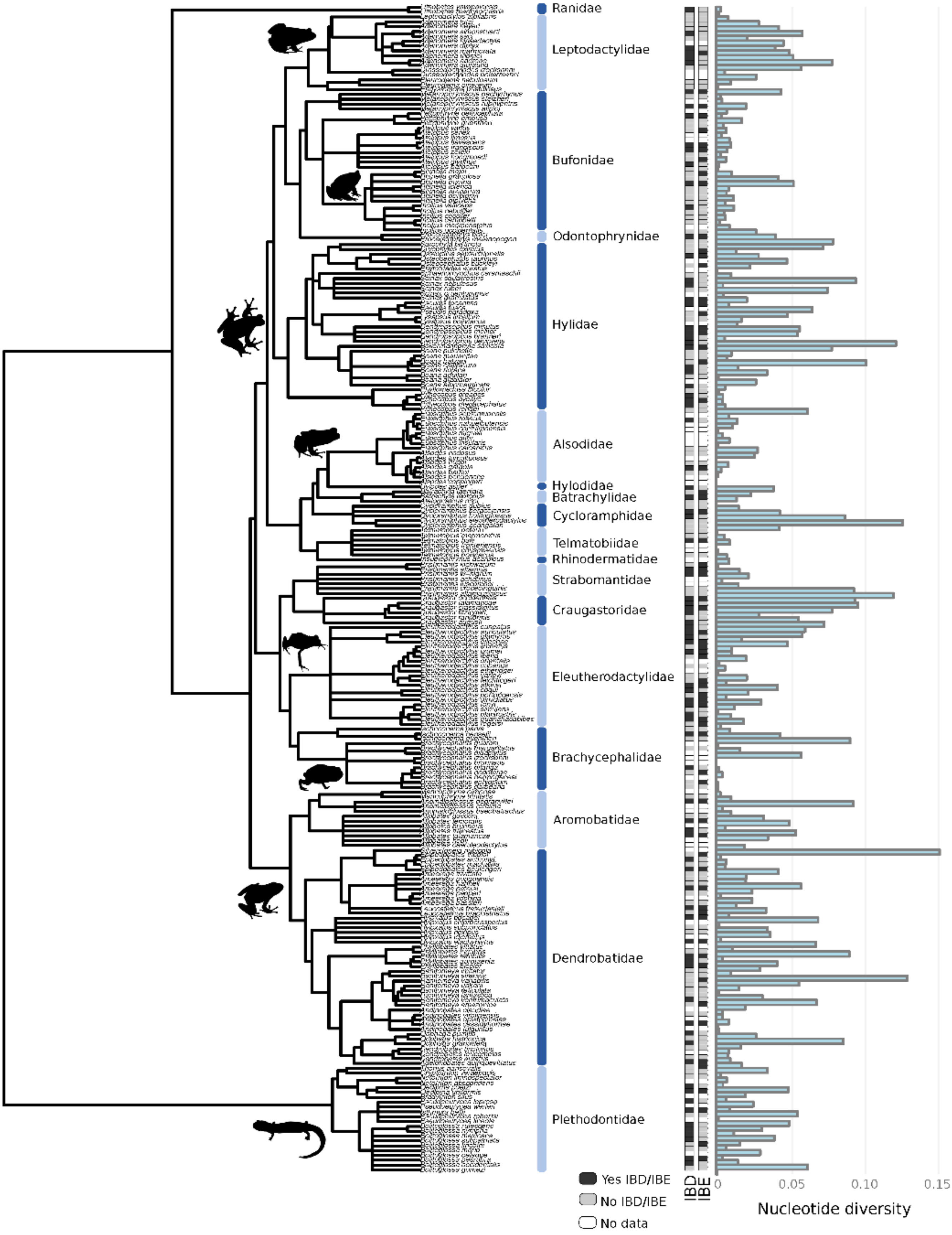

FIGURE 4 Phylogeny of neotropical amphibians (downloaded and subset from Jetz & Pyron, 2018) with values of cytochrome-b π presented as light blue bars, and presence (black dots) or absence (grey dots) of isolation-by-distance (IBD) and isolation-by-environment (IBE) in neotropical amphibians. Silhouettes of frogs and salamanders were obtained from phylopic.org.

(e.g. mountain ranges) tended not to exhibit IBD, and species living in southern latitudes tended not to exhibit IBE.

Geographic range size was the top predictor of π, but did not predict IBD and IBE, that is, Neotropical amphibian species with and without IBD or IBE can inhabit small or large geographic ranges. We suspect that environmental heterogeneity within the geographic range rather than range size may have a greater influence on IBD and IBE patterns of species in the Neotropics, as has been observed in other vertebrate studies (e.g. Quiroga-Carmona & D'Elía, 2022). Geographic range size has also emerged as an important predictor of genetic variation in other taxonomic groups such as squamates (Larkin et al., 2023) and Darwin's finches (Brüniche-Olsen et al., 2019). Interestingly, the relationship between genetic diversity 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 genetic variation, which may also indicate an increased risk of extinction (Levy et al., 2016). 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 amphibian species occurring in larger ranges and Data Deficient, Endangered and Critically Endangered species occurring in smaller ranges. Based on our estimates, however, IUCN conservation status did not have as clear of an association with π, suggesting that IUCN status may not currently capture this important parameter for population persistence (Appendix S4). Regardless, populations of threatened species with low π such as harlequin frogs of the genus *Atelopus* should be closely monitored.

Precipitation and temperature, along with associated gradients of elevation, have significant impacts on genetic variation across animal species (De Kort et al., 2021). Our study identified minimum elevation as a significant predictor of IBD (and it was also one important predictor of π), and precipitation as one of the best predictors of π, IBD and IBE. We observed that species inhabiting higher elevations did not usually exhibit IBD, while those living in lower elevations tended to have both significant IBD patterns and larger geographic range sizes. 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 IBD in several taxonomic groups. In our case, latitude (maximum) was the most important predictor of IBE. Species with a significant IBE pattern occurred at higher latitudes than species without IBE. Although latitude was not a significant predictor of π, analysing mean latitude with π 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 (Appendix S4). 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).

The nucleotide diversity map of Neotropical amphibians 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. Biogeographic units identified with higher π values are areas that also have high precipitation rates. Therefore, precipitation could explain genetic variation in the Neotropics; however, this hypothesis needs to be tested in future studies with more species homogeneously distributed throughout the region. 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, mainly because we do not have an equilibrated sampling and our repurposed data set has a high concentration of occurrence points in certain regions (e.g. Amazonia, Atlantic forests or Mesoamerica). In addition, we were not able to recover more than five CytB sequences (our minimum number of sequences per species) for any Gymnophiona species, nor were we able to recover genetic information for centrolenid frogs, 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 are lacking in 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 $\pi = 0.028$, $n = 137$; 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 data set. At least nine of the species with high π values 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*—Márquez et al., 2020; *Dendropsophus decipiens*—de Oliveira et al., 2021; see Appendix S3). 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 interpret the results of genetic variation of these species with caution because they may represent multiple taxa, for example, if we were able to split cryptic species in separate species, range sizes would be smaller impacting the levels of intraspecific genetic diversity.

We found that levels of π 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 (Appendix S4). In contrast, the presence of spatial genetic variation (IBD and IBE) within species appears to be more randomly distributed throughout the phylogeny of Neotropical amphibians. These findings highlight the value of employing different methodological frameworks as we did in this study. With the growing size and complexity of biodiversity data sets, machine learning methods such as RF prove valuable in identifying potential predictors, even using both complete and reduced data sets (see Barrow et al., 2021; Pelletier & Carstens, 2018). Combining these methods with phylogenetically informed models allowed us to gain a deeper understanding of the relationships between predictors and genetic variation within species. For example, our RF (regression and classification) and *MCMCglmm* models were consistent in identifying range size and precipitation as the best predictors of genetic variation.

Our study provides valuable insights into the distribution of genetic variation in Neotropical amphibians and identifies important predictors of intraspecific genetic variation. These findings underscore the importance of considering both nucleotide diversity and spatial genetic variation 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 π 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 groups, 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 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 Lisa Barrow from the U.S. National Science Foundation (DEB-2112946). The authors thank the associate editor and two reviewers for their suggestions that helped to improve this manuscript. We thank the members of the Amphibian & Reptile Biodiversity Lab (the 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. No permits were required in this research.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

Table S1 and Supporting Information (e.g. appendices, fasta sequences used for the analyses, R scripts) are available on Dryad Data Repository (https://doi.org/10.5061/dryad.cz8w9gj7s).

ORCID

Luis Amador https://orcid.org/0000-0003-2638-4068
Irvin Arroyo-Torres https://orcid.org/0000-0001-6811-7422
Lisa N. Barrow https://orcid.org/0000-0001-7081-2432

REFERENCES

Acevedo, A. A., Lampo, M., & Cipriani, R. (2016). The cane or marine toad, *Rhinella marina* (Anura, Bufonidae): Two genetically and morphologically distinct species. *Zootaxa*, *4103*(6), 574. https://doi.org/10.11646/zootaxa.4103.6.7

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

AmphibiaWeb. (2022). AmphibiaWeb. University of California, Berkeley, CA, USA. Retrieved 2022 from https://amphibiaweb.org

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 of the United States of America*, *115*(23), 6034–6039. https://doi.org/10.1073/pnas.1713819115

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

Bivand, R. (2022). R packages for analyzing spatial data: A comparative case study with areal data. *Geographical Analysis*, *54*(3), 488–518. https://doi.org/10.1111/gean.12319

Blanchet, S., Prunier, J. G., & De Kort, H. (2017). Time to go bigger: Emerging patterns in macrogenetics. *Trends in Genetics*, *33*(9), 579–580. https://doi.org/10.1016/j.tig.2017.06.007

Blomberg, S. P., Garland, T., & Ives, A. R. (2003). Testing for phylogenetic signal in comparative data: Behavioral traits are more labile. *Evolution*, *57*(4), 717–745. https://doi.org/10.1111/j.0014-3820.2003.tb00285.x

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

Calenge, C., & Fortmann-Roe, S. (2023). *_adehabitatHR: Home Range Estimation_*. R package Version 0.4.21. https://CRAN.R-project.org/package=adehabitatHR

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

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

de Oliveira, R. F., Magalhães, F. d. M., Teixeira, B. F. d. V., de Moura, G. J. B., 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

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

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 Research*, *32*(5), 1792–1797. https://doi.org/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

Elmer, K. R., Bonett, R. M., Wake, D. B., & Lougheed, S. C. (2013). Early Miocene origin and cryptic diversification of South American salamanders. *BMC Evolutionary Biology*, *13*(1), 59. https://doi.org/10.1186/1471-2148-13-59

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

Fox, J., & Weisberg, S. (2019). *An R companion to applied regression* (3rd ed.). Sage. https://socialsciences.mcmaster.ca/jfox/Books/Companion/

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)*. American Museum of Natural History. https://doi.org/10.5531/db.vz.0001

Garamszegi, L. Z. (Ed.). (2014). *Modern phylogenetic comparative methods and their application in evolutionary biology: concepts and practice*. Springer. https://doi.org/10.1007/978-3-662-43550-2

García-Rodríguez, A., Guarnizo, C. E., Crawford, A. J., Garda, A. A., & Costa, G. C. (2021). Idiosyncratic responses to drivers of genetic differentiation in the complex landscapes of Isthmian Central America. *Heredity*, *126*(2), 251–265. https://doi.org/10.1038/s41437-020-00376-8

Greenwell, B. M., & Boehmke, B. C. (2020). Variable importance plots—An introduction to the vip package. *The R Journal*, *12*(1), 343–366. https://doi.org/10.32614/RJ-2020-013

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

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

Hijmans, R. J., Barbosa, M., Ghosh, A., & Mandel, A. (2023). *_geodata: Download geographic data_*. R package Version 0.5-9. https://CRAN.R-project.org/package=geodata

Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G., & Jarvis, A. (2005). Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, *25*(15), 1965–1978.

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

IUCN. (2022). *The IUCN red list of threatened species*. Version 2022-1. https://www.iucnredlist.org

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

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

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.

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

Kimura, M. (1983). *The neutral theory of molecular evolution*. Cambridge University Press.

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. https://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

Leigh, D. M., Van Rees, C. B., Millette, K. L., Breed, M. F., Schmidt, C., Bertola, L. D., Hand, B. K., Hunter, M. E., Jensen, E. L., Kershaw, F., Liggins, L., Luikart, G., Manel, S., Mergeay, J., Miller, J. M., Segelbacher, G., Hoban, S., & Paz-Vinas, I. (2021). Opportunities and challenges of macrogenetic studies. *Nature Reviews Genetics*, *22*(12), 791–807. https://doi.org/10.1038/s41576-021-00394-0

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.

Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. *R News*, *2*(3), 18–22.

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

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

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

Menéndez-Guerrero, P. A., Green, D. M., & Davies, T. J. (2020). Climate change and the future restructuring of Neotropical anuran biodiversity. *Ecography*, *43*(2), 222–235. https://doi.org/10.1111/ecog.04510

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

Nevo, E. (1978). Genetic variation in natural populations: Patterns and theory. *Theoretical Population Biology*, *13*(1), 121–177. https://doi.org/10.1016/0040-5809(78)90039-4

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

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

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

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=geodata

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.

Pagel, M. (1999). Inferring the historical patterns of biological evolution. *Nature*, *401*(6756), 877–884. https://doi.org/10.1038/44766

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. https://doi.org/10.1098/rspb.2017.2746

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. https://www.R-project.org/

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

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

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

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

Wang, I. J. (2013). Examining the full effects of landscape heterogeneity on spatial genetic variation: A multiple matrix regression approach for quantifying geographic and ecological isolation. *Evolution*, *67*(12), 3403–3411. https://doi.org/10.1111/evo.12134

Wang, I. J. (2020). Topographic path analysis for modeling dispersal and functional connectivity: Calculating topographic distances using the topoDistance R package. *Methods in Ecology and Evolution*, *11*, 265–272.

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

Wieringa, J. G., Boot, M. R., Dantas-Queiroz, M. V., Duckett, D., Fonseca, E. M., Glon, H., Hamilton, N., Kong, S., Lanna, F. M., Mattingly, K. Z., Parsons, D. J., Smith, M. L., Stone, B. W., Thompson, C., Zuo, L., & Carstens, B. C. (2020). Does habitat stability structure intraspecific genetic diversity? It's complicated.... *Frontiers of Biogeography*, *12*(2), e45377. https://doi.org/10.21425/F5FBG45377

Zanaga, D., Van De Kerchove, R., De Keersmaecker, W., Souverijns, N., Brockmann, C., Quast, R., Wevers, J., Grosu, A., Paccini, A., Vergnaud, S., Cartus, O., Santoro, M., Fritz, S., Georgieva, I., Lesiv, M., Carter, S., Herold, M., Li, L., Tsendbazar, N. E., … Arino, O. (2021). *ESA WorldCover 10 m 2020 v100*. https://doi.org/10.5281/zenodo.5571936

Zeisset, I., & Beebee, T. J. C. (2008). Amphibian phylogeography: A model for understanding historical aspects of species distributions. *Heredity*, *101*(2), 109–119. https://doi.org/10.1038/hdy.2008.30

BIOSKETCHES

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 N. 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. analysed the data; and L.A. and L.N.B. led the writing.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Amador, L., Arroyo-Torres, I., & Barrow, L. N. (2024). Machine learning and phylogenetic models identify predictors of genetic variation in Neotropical amphibians. *Journal of Biogeography*, *00*, 1–15. https://doi.org/10.1111/jbi.14795