Caudata macrogenetics: Species range size predicts intraspecific genetic variation in global salamanders

Luis Amador[1], Daniele L. F. Wiley[1], Irvin Arroyo-Torres[1], Chris X. McDaniels[1], Esteban O. Rosario-Sánchez[1], Hannah Farmer[2], Hannah Bradley[1], James Erdmann[1], Tara A. Pelletier[2], Lisa N. Barrow[1]

[1]Museum of Southwestern Biology and Department of Biology, University of New Mexico, Albuquerque, NM 87131, USA.

[2]Department of Biology, Radford University, Radford, VA 24142, USA.

Correspondence

Lisa N. Barrow

Acknowledgments

We thank Emily Anderson of Radford University for assistance and the members of the Amphibian & Reptile Biodiversity Lab (the University of New Mexico) for their feedback on earlier versions of this project.

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).

Tara A. Pelletier is an Associate Professor at Radford University and uses techniques in phylogeography and macrogenetics to study biodiversity patterns in a variety of organisms.

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.

Data availability statement

The data that support the results of this research are openly available in https://github.com/luchoamador/Caudata

Funding statement

Funding was provided by the National Science Foundation (awards DEB-2112946 granted to L.N. Barrow and DBI-1911293 granted to T.A. Pelletier).

Conflict of interest disclosure

The authors have no conflicts of interest to disclose.

Caudata macrogenetics: Species range size predicts intraspecific genetic variation in global salamanders

Short running title: Global salamander macrogenetics

Abstract

Aim: Genetic diversity contains information about adaptive potential and historical demography. Intraspecific genetic variation is shaped by species' natural history traits and by characteristics of geography and climate within their ranges. Salamanders (Order Caudata) are amphibians of ecological and conservation interest because of their global distribution, trait diversity, and roles within ecological communities. Here, we studied genetic variation within global salamanders to investigate the most important predictors of genetic structure and genetic diversity in this group.

Location: Global.

Time Period: Present.

Major Taxa Studied: Salamanders.

Methods: We repurposed mitochondrial DNA sequences, natural history traits, and geographic and environmental data from open-access databases for 214 salamander species. We calculated nucleotide diversity (π) and tested for isolation by distance (IBD) and isolation by environment (IBE) within each species. We analyzed these three response variables with random forest analyses and phylogenetic comparative methods using 27 predictors expected to be associated with genetic variation (e.g., number of offspring, total range size).

Results: We recovered 8,175 Cytb sequences with their associated geographic coordinates, of which 7,007 sequences were manually curated by us. Species range size was an important predictor of genetic variation (π, IBD, and IBE). We found higher diversity in regions including the Neotropics and central-eastern Europe. The absence of strong phylogenetic signal in π, IBD, and IBE suggests that genetic variation is shaped by local ecological and geographical factors rather than by shared ancestry.

Main Conclusions: Our finding that range size was an important predictor of genetic variation aligns with theoretical expectations that species with larger ranges tend to harbor more genetic diversity. Our results underscore the importance of integrating spatial data (i.e., geographic coordinates associated with DNA sequences) into macrogenetic studies, providing valuable

information for future studies and conservation strategies targeting regions with high or low genetic diversity.

Keywords

Amphibians, isolation-by-distance, isolation-by-environment, mitochondrial DNA, nucleotide diversity.

Introduction

Genetic variation plays a foundational role in maintaining biological diversity (Kardos et al., 2021). One significant evolutionary factor that predicts genetic variation within populations is the number of individuals that contribute DNA to the next generation (the effective population size; Wright, 1931). Under the neutral theory of molecular evolution, genetic variation is proportional to population size (Kimura, 1968). Migration between populations is another key factor for genetic variation because it can add and maintain novel alleles, enabling population and species persistence (Park et al., 2024). Understanding this variation and which factors influence it across different geographic and environmental contexts is essential for elucidating population dynamics and will help us identify populations with conservation needs (Pauls et al., 2013). For example, declining levels of genetic variation may be further exacerbated under global change scenarios due to the reduction of suitable habitat for many species (Schierenbeck, 2017).

Climatic and geographic factors influence dispersal patterns and population sizes, thereby shaping genetic variation within species (e.g., Hanson et al., 2017). Current spatial patterns of genetic variation can be detected as isolation-by-distance (IBD), where individuals become more genetically dissimilar with increasing geographic distance, (Wright, 1943), and isolation-by-environment (IBE), which factors in the effects of landscape heterogeneity on gene flow (Wang and Bradburd, 2014). Climatic history and topographical features have been linked to elevated levels of genetic variation; for example, in the alpine toad (*Scutiger ningshanensis*), climatic and geologic events in the Pleistocene contributed to the high genetic variation observed in the species (Meng et al., 2014). Life history traits, particularly those related to reproduction (e.g., number of offspring or egg size), have also been linked with intraspecific genetic variation (Ellegren and Galtier, 2016). In two species of marine fishes from the genus *Centropomus*, high levels of genetic diversity were associated with migratory behavior and reproductive strategies such as external fertilization (Malcher et al., 2023). A combination of life history and climatic variables predicted genetic diversity in plants (384 species) and animals (343 species) (De Kort et al., 2021), suggesting that some predictors are reliable indicators of genetic variation across different taxonomic groups.

Studies on focal taxa within a limited geographic range provide the resolution needed to understand local genetic patterns and idiosyncrasies (e.g., Lee-Yaw and Irwin, 2012; Fonseca et al., 2021). But it is equally important to investigate genetic variation on broader geographic and taxonomic scales, to have a global perspective of the role of ecological and geographic factors on evolutionary patterns. Macrogenetics is an approach to analyze open-access DNA sequences across large scales and investigate patterns of intraspecific genetic diversity across higher taxonomic groups (Blanchet et al. 2017). Macrogenetic research has gained popularity over the past decade, driven by the increasing availability of molecular resources (e.g., GenBank), computational tools (e.g., machine learning), and trait databases (Leigh et al., 2021). Efforts to understand patterns and predictors of intraspecific genetic diversity have been undertaken across various taxonomic groups. Lawrence et al. (2023) used microsatellite data to evaluate the impact of environmental variables on vertebrate genetic diversity patterns across the Americas. A study of global genetic diversity in insects attributed high levels of genetic diversity in the subtropics to temperature and climatic stability (French et al., 2023). Another study of nucleotide diversity in over 38,000 animal and plant species found that latitude significantly correlates with genetic diversity, with tropical species exhibiting higher intraspecific diversity (Fonseca et al., 2023). These studies highlight the potential of macrogenetics to uncover important predictors of genetic diversity across broad sets of taxa globally.

Amphibians contain remarkable diversity in life history strategies and the geographic and climatic variability of their habitats. Recent studies of amphibian genetic diversity indicate the biogeographic realm under investigation can influence which factors predict intraspecific genetic variation. For example, in Nearctic amphibians, taxonomic family, the number of sequences, and latitude were key predictors of intraspecific variation (Barrow et al., 2021), whereas range size, elevation, latitude, and precipitation predicted genetic variation in Neotropical amphibians (Amador et al., 2024). In the Americas, Lawrence et al. (2023) found a strong relationship between environmental variables and amphibian expected heterozygosity (H_E), and that amphibian H_E decreased with sample size (number of individuals).

Within amphibians, salamanders deserve further attention because of their largely global distribution (except Oceania, the Afrotropics, and Antarctica), relatively manageable number of species, and unique traits. They display remarkable ecological diversity, such as the variation in body size, with species as small as 20 mm (e.g., the arboreal minute salamander, *Thorius arboreus*) to the Japanese giant salamander (*Andrias japonicus*) that can grow up to 1.8 meters. Salamanders include generalist species (e.g., the tiger salamander, *Ambystoma tigrinum*) and others completely adapted to aquatic, terrestrial, or arboreal environments. Some salamanders exhibit paedomorphosis, a reproductive strategy in which adults retain larvae characteristics, allowing them to exploit different habitats and increase their chances to survive and reproduce (Denoël et al., 2005). Determinants of genetic diversity in Neotropical salamanders were recently explored by Segovia-Ramírez et al. (2023), who used genomic data from 62 species (one individual per species) to demonstrate that precipitation and snout-vent length were significant predictors of genomic diversity. In another recent study, Parsons et al. (2024) combined genetic, geographic, climatic, and life history data from salamander species and used a machine learning model to identify predictors of unrecognized genetic lineages. They found that Caudata hidden diversity is the result of variation in climatic variables. A comprehensive, georeferenced view of intraspecific genetic variation in salamanders, encompassing more species and potential predictors (e.g., natural history traits), could provide useful insights into the underlying drivers of this variation.

Here, we combine molecular, phenotypic, and environmental open-access data to investigate the determinants of intraspecific genetic variation in global salamanders. We used machine learning techniques and phylogenetic comparative methods to determine whether natural history traits or climatic or environmental variables can predict genetic variation within species on a global scale. We hypothesized that traits associated with reproductive strategies, such as total number of eggs (i.e., a proxy for fitness) or reproductive mode (e.g., larval or direct development), would predict intraspecific genetic variation in salamanders. Fitness and genetic diversity are expected to be positively correlated (Reed and Frankham, 2003), and we expected that reproductive mode could relate to dispersal and population connectivity. We also predicted that characteristics associated with population size (e.g., species range size) would be positively correlated with genetic variation (Frankham, 1996). Finally, we expected to find high spatial autocorrelation in

nucleotide diversity of salamanders overall based on their limited dispersal capabilities, which can lead to genetic clustering at small scales (Fusco et al., 2021).

Methods

Data collection

We repurposed Caudata DNA sequences, natural history traits, and geographic and environmental data from open-access databases. Following the approach of recent research on the determinants of genetic diversity of multiple amphibian species (e.g., Barrow et al., 2021; Amador et al., 2024), we assembled sequences of the cytochrome-b (Cytb) mitochondrial gene. This gene is suitable for genetic diversity studies because it contains both conserved and variable regions, is commonly used in amphibian single-locus phylogenetic and population genetic studies and is therefore the most abundant gene in open-access databases for amphibians. Sequences were obtained from GenBank (National Center for Biotechnology Information, NCBI), phylogatR (Pelletier et al. 2022), the *phruta* R package (Román-Palacios, 2023), and Amador et al. (2024). We included species with at least 10 sequences, seeking a better representation of intraspecific genetic variation. Alignments were generated, visualized, and edited with the MUSCLE aligner v.3.8.31 (Edgar, 2004) in AliView v.1.28 (Larsson, 2014) and saved in FASTA format.

The geographic coordinates (latitude and longitude in decimal degrees format) previously associated with each Cytb sequence were recovered from GenBank, phylogatR, and Amador et al., (2024); corresponding to 31 species (14.5% of the total species) and 1,169 sequences (14.3% of the total sequences). We then manually retrieved the remaining geographic coordinates associated with the sequences (corresponding to 183 species and 7,007 sequences) following a tutorial prepared for this purpose (see Supplementary Material 1). We georeferenced localities using either the function geocode() of the *tidygeocoder* R package (Cambon et al., 2021) or the GEOLocate web application (Rios and Bart, 2010). We only retained sequences that were georeferenced with an error of less than 20 km, resulting in 12 species with fewer than 10 georeferenced sequences.

We calculated both the total range size and the sampling range size of DNA sequences for each species as follows. For total range size, we obtained the geographic range map in shapefile

format for each species from the International Union for Conservation of Nature (IUCN) portal using the getIUCN() function of the *rasterSp* R package (RS-eco, 2023). Species ranges not available in the IUCN portal (24 species) were generated with minimum convex polygons using the sequence coordinates and the function mcp() in the R package *adehabitatHR* (Calenge, 2024; see Supplementary Material). This same approach was also used to obtain the geographic range of the sampled sequences for all species (i.e., using only the sequence occurrences). Range size was calculated from shapefiles using the areaPolygon() function of the *geosphere* R package (Hijmans, 2022). Elevation and latitude (min, max, and mean) were obtained from the retrieved geographic coordinates using the function get_elev_point() of the *elevatr* R package (Hollister, 2023) for elevation, and custom R scripts (R Core Team, 2023) for latitude (see Supplementary Material 2). We also obtained precipitation and temperature data (mean and standard deviation) from WorldClim v2.1 (Fick and Hijmans, 2017) using the function worldclim_global() of the *geodata* R package v0.6-2 (Hijmans et al., 2024).

We used the same natural history and geographic traits as in Amador et al., (2024) with slight additions and omissions, choosing predictors we expected to be related to genetic variation (e.g., body size, reproductive mode, elevation, latitude, species range size, precipitation and temperature; Table S1). We omitted the variable activity because nearly all salamanders (except for three of the 214 species) included in our dataset were considered to be nocturnal. Six new predictors were added to the matrix (27 predictors in total), which included sampling effort/coordinates (i.e., number of sequences with associated coordinates), sequence length/coordinates (i.e., number of base pairs of sequences with associated coordinates), snout-vent length in millimeters (SVL), paedomorphism (whether or not juvenile features are retained as an adult), litter size min (minimum number of offspring or eggs per clutch), and litter size max (maximum number of offspring or eggs per clutch) (Table S1). We obtained trait information from the open-access database AmphiBIO (Oliveira et al. 2017), species accounts in AmphibiaWeb (AmphibiaWeb, 2024), and relevant literature.

Measures of genetic variation

We evaluated overall genetic diversity (range-wide nucleotide diversity) and spatial patterns of genetic variation (isolation by distance and environment) within salamander species. We calculated nucleotide diversity (π) with the function nuc.div() of the *vegan* R package (Oksanen et al., 2024) for all sequences (π_A) and sequences with associated coordinates only (π_S). Genetic distances (*gendist*) for each species were calculated using the dist.dna() function of the *ape* R package (Paradis and Schliep, 2019). We tested isolation by distance (IBD) and isolation by environment (IBE) for each salamander species as follows. Geographic distance (*geodist*) was calculated between each pair of sequence coordinates using Euclidean distance with the dist() function of the *stats* package (R Core Team, 2023) in R. Before calculating environmental distances, we used the function scale() in R to standardize variables measured in different units (e.g., temperature - °C or precipitation - mm). Environmental distance (*envdist*) was calculated between each pair of sequence coordinates based on the 19 variables of the WorldClim v2.1 database using the vegdist() function of the *vegan* package. This function calculates a distance matrix, selecting all environmental variables using Euclidean distances (Oksanen et al., 2024). We performed Mantel tests for IBD and IBE separately. We compared the observed correlation between *geodist* vs. *gendist* (IBD), and *envdist* vs. *gendist* (IBE) to the correlation from 1,000 permutations using the mantel() function in the *vegan* package. Species were coded as 'yes' if they did present IBD and IBE or 'no' if they did not based on the significance ($p < 0.05$) of the model. We used the output (yes or no IBD; yes or no IBE) as binary response variables for further analyses.

We mapped π_S to visualize the distribution of genetic diversity across the globe and assess differences among biogeographic realms (Nearctic, Neotropic, Oriental, and Palearctic). Briefly, following Amador et al. (2024), we combined sequences within a species based on a maximum distance of 100 km between coordinates, and calculated nucleotide diversity for each locality (see Supplementary Material 3). We then mapped π_S for four different grid sizes (1–4), which correspond with one to four decimal degree grids, or approximately 110 km^2 to 440 km^2. We also visualized the number of sequences per grid cell for the four resolutions. For these visualizations, we used the R packages *ggplot2* (Wickham, 2016), *ggspatial* (Dunnington, 2023), and *sf* (Pebesma, 2018; Pebesma and Bivand, 2023). Finally, we tested for spatial autocorrelation for π

and the number of sequences separately with the Moran's I test using the moran.test() function in the *spdep* R package (Bivand et al., 2013).

Predicting genetic variation using random forests (RF)

We used random forest (RF) regression to predict π and RF classification to predict IBD and IBE based on a set of geographic range characteristics and ecological variables. We built RF models with the function randomForest() in the *randomForest* R package (Liaw & Wiener, 2002), with 10,000 trees and 100 permutations per model. We used the tuneRF() function to obtain the optimal mtry parameter (number of variables that are randomly sampled as candidates at each split) and then ran a new RF regression or classification analysis using the best mtry value. The data was split into a training set (70% of the data) for building models and a testing set (30%) for making predictions. For RF regression, we assessed models using the R-squared (R^2) statistic, which represents the proportion of the variance explained by the model, with a higher R^2 value indicating a better fit and more predictive power. Relative importance of the predictor variables for π_A and π_S was evaluated based on the percentage increase in mean squared error (%IncMSE).

In RF classification, in addition to the original models, we also tried upsampling and downsampling to address the large class imbalance within our datasets (yes IBD = 143 species, no IBD = 68 species; yes IBE = 154 species, no IBE = 58 species). For downsampling, the majority cases (yes IBD/IBE) were randomly subsampled to equal the number of minority cases (no IBD/IBE). For upsampling, minority cases were randomly duplicated to equal the number of majority cases. We chose the best model based on the lowest out-of-bag (OOB) error rate. To assess overfitting and improve cross-validation in the model, we split our data into a training set and a test set (70% train, 30% test). To evaluate predictions of the RF classification models, we created a confusion matrix with the confusionMatrix() function of the *caret* R package (Kuhn, 2008) and tested whether model accuracy was significantly better than the no information rate (NIR). Finally, we assessed the relative importance of the variables based on the Mean Decrease Accuracy (MDA) for the RF classification models of IBD and IBE.

Evaluating predictors of genetic variation with phylogenetic comparative methods

For phylogenetic analyses, we downloaded a phylogeny subset from VertLife.org for the species in our dataset, which represents phylogenetic relationships inferred by Jetz and Pyron (2018). We used a subset of 180 species for these analyses because 34 species from our dataset are not included in the available phylogeny. To assess the potential effect of phylogenetic history in our models, we tested for phylogenetic signal in π using Pagel's λ (lambda; Pagel, 1999) and Blomberg's K (Blomberg et al., 2003) implemented in the *phytools* R package with the function phylosig(). For IBD and IBE, we used the *D* statistic implemented in the package *caper* with the function phylo.d() to assess phylogenetic signal.

We used phylogenetic independent contrasts (PIC; Felsenstein, 1985) to investigate the evolutionary correlation between π and the most important predictors identified in the RF analyses. We used the pic() function of the *ape* R package to compute PIC for both response and explanatory variables. We then fit linear models in R for both the uncorrected and phylogenetically corrected (PIC) response and predictor variables. We also used phylogenetic generalized linear mixed models (PGLMMs) to analyze binary traits as described in Ives and Helmus (2011). This method was used to analyze IBD and IBE as binary dependent variables (0 for no IBD/IBE, 1 for yes IBD/IBE) versus the most important predictors based on RF results. We compared IBD vs. sampling range size, conservation status, and mean precipitation, and IBE vs. sampling range size and mean temperature. We implemented this analysis using the function pglmm() in the *phyr* R package (Li et al., 2020).

Results

Data summary

We obtained 13,265 Cytb sequences from 214 species distributed globally, of which 8,963 sequences were retrieved directly from GenBank, 2,349 sequences from phylogatR, 1,657 from the *phruta* package, and 296 from Amador et al. (2024). Of these, we obtained geographic coordinates for 8,175 sequences from 213 species. A single species, *Plethodon sherando*, did not have associated coordinates, therefore was not included in IBD and IBE analyses; and 11 more species had between four and nine sequences with coordinates. The final dataset included 214 species in nine families (representing >25% of species and 90% of families globally), three response variables (π, IBD, and IBE), and 26 predictors (Table S2). All sequences associated with geographic coordinates and R scripts used to retrieve sequences and coordinates are available as Supplementary Material 3.

Patterns of genetic variation in global salamanders

Nucleotide diversity within species ranged from $\pi_A = 0$ to $\pi_A = 0.087$ (*Pseudoeurycea lineola*) when we analyzed all sequences, and from $\pi_S = 0$ to $\pi_S = 0.089$ (*Desmognathus amphileucus*) using only the sequences with associated coordinates. We found that nucleotide diversity for all sequences (π_A) and for the sequences with associated coordinates only (π_S) were highly correlated (Figure S1; Pearson's correlation $r = 0.94$, $t(212) = 40.12$, p-value < 0.001). Based on this validation, we conducted further analyses using only sequences with associated coordinates (π_S). *Bolitoglossa yariguiensis* was the only species with a value of $\pi = 0$ (Table S2; Figure S2). We found that nucleotide diversity and sampling effort (i.e., number of sequences) were not randomly distributed across the world map (Figure 1, Figures S3 – S5). For example, regions on the map with high genetic diversity included the Ecuadorian Amazon and central-southern Mexico in the Neotropics; and central-eastern Europe in the Palearctic. Areas with low genetic diversity were identified in northern North America (Nearctic), western Europe (Palearctic), and central-western China in the Oriental realm (Figure 1). Furthermore, we found statistically

significant, but weak, positive spatial autocorrelation in both π_s (Moran's I = 0.1173, p-value = 0.0003) and the number of sequences (Moran's I = 0.2053, p-value < 0.0001).

Most species presented significant IBD and IBE patterns, with 143 species (68%) showing IBD and 154 species (73%) showing IBE (Table S2; Figure S6). We found that 134 species had significant patterns for both IBD and IBE (Table S2). The Nearctic and Neotropical realms had the lowest percentages of species that showed IBD (62.1% and 62.5%, respectively; Figure S7). The same biogeographic realms also had fewer species with IBE patterns than the other realms (Neotropic = 62.5%, Nearctic = 70.1%). The Palearctic region had the highest percentage of species that showed IBD (74.3%) and IBE (77.1%) patterns (Figure S7).

Genetic variation was randomly distributed across the phylogeny (Figure 2). For example, we found species in the same genus with dissimilar π values, such as *Plethodon* (range = 0.0008 – 0.0698) and *Eurycea* (range = 0.0015 – 0.0768) in Plethodontidae, *Paramesotriton* (range = 0.0041 – 0.0767) in Salamandridae, and *Hynobius* (range = 0.0025 – 0.0853) in Hynobiidae. We found no or low phylogenetic signal in π based on the method used, for example, with Pagel's λ = 0.029, p-value (based on LR test) = 0.425; and Blomberg's K = 0.106, p-value (based on 1000 randomizations) = 0.046 (Figure S8). Similarly, we found no significant phylogenetic signal in IBD and IBE (Estimated D_{IBD} = 0.982, p-value = 0.376; Estimated D_{IBE} = 1.042, p-value = 0.651) (Figure S9).

Predictors of genetic variation in global salamanders

In our initial random forest analyses, we found an influence of sampling effort on genetic diversity estimates, with the number of sequences as a top predictor of π, IBD, and IBE (Figure S10). We then performed subsequent RF analyses excluding sampling effort to better understand the relative roles of the biological variables of interest. Additionally, we compared both species range size variables in our dataset, total range size versus sampling range size, and we found a strong, positive linear relationship (Figure S11; Pearson's correlation r = 0.78, $t(212)$ = 18.41, p-value < 0.001). These results suggested the sampled range for most species adequately represented the total range size. Nevertheless, we ran RF analyses using either total range size or

sampling range size and found similar results between the two variables (Figures 3, S10, S12). Because the range size variables were highly correlated with one another and captured similar aspects of the species' biology, we subsequently describe the results using the sampling range size.

In our final RF analyses, we found that sampling range size was the best predictor of π, IBD, and IBE (Figure 3). Random forest regressions showed that taxonomic family was the second-best predictor of nucleotide diversity in global salamanders (Figure 3a, S13). Validation of the RF regression results for π as the response variable based on the R^2 score indicated good model performance ($R^2 = 0.92$). Conservation status and mean precipitation were other important predictors of IBD (Figure 3b, S14). For IBE, RF analyses showed that mean temperature, conservation status, body size, and snout-vent length were also important predictors (Figure 3c, S15). In RF classification analyses, the best IBD and IBE RF models were obtained when we downsampled our dataset so the response classes were matched. Our IBD RF model had high accuracy (87.5%; 95% Confidence Interval CI: 75.9% – 94.8%) and sensitivity (100%), but lower specificity (65%). Similarly, the IBE RF model had high accuracy (85.7%; CI: 73.8% – 93.6%) and sensitivity (100%), but slightly lower specificity (61.9%).

Sampling range size had a significant, positive relationship with π, showing that genetic diversity within species increases with the geographic area sampled (Figure 4A; $R^2 = 0.09$; p-value < 0.001). We found a similar pattern using phylogenetic independent contrasts (PIC), where sampling range size and π have a strong, positive relationship (Figure 4B; $R^2 = 0.25$; p-value < 0.001). We found a significant, positive relationship between mean temperature and π, suggesting that species living in areas with higher temperatures have more genetic diversity (Figure S16A). However, PIC analyses did not show a significant correlation between temperature or other predictors (body size and elevation) and π after correcting for phylogenetic relatedness (Figures S16 – S19).

Although sampling range size was a top predictor for IBD in RF models, there was no clear difference (p-value = 0.19) in sampling range size for species that did or did not exhibit an IBD pattern (Figure S20A). Species with significant IBE (p-value = 0.01) had slightly larger sampling

range sizes on average (Figure S20B). When temperate and tropical species were considered separately, we found a trend in which tropical species with larger ranges were more likely to exhibit IBD and IBE, but there were no differences in range size for temperate species with or without either pattern (Figure 5). Phylogenetic generalized linear mixed models (PGLMM) found that sampling range size had a statistically significant, negative relationship with IBD ($\beta = -9.58 \times 10^{-7}$, $SE = 4.15 \times 10^{-7}$, $p = 0.021$), but the relationship between sampling range size and IBE was not significant ($\beta = -1.98 \times 10^{-7}$, $SE = 1.84 \times 10^{-7}$, $p = 0.283$). These results suggest that after incorporating phylogenetic relatedness, as sampling range size increases, the likelihood of having an IBD pattern decreases. Coefficients, standard errors, Z-scores, and p-values of the other predictors (i.e., conservation status and mean precipitation for IBD, and mean temperature for IBE) are shown in Table S3. Taken together, these weak relationships and inconsistent results suggest limited signal in the data that should be interpreted with caution.

Discussion

In this study, we analyzed 13,265 Cytb sequences from 214 global salamander species and identified key predictors of intraspecific genetic variation (π, IBD, and IBE). Species range size was an important predictor of all three aspects of genetic variation. Specifically, species occupying larger geographic areas tended to have greater overall nucleotide diversity, and tropical species with larger ranges tended to be geographically and environmentally structured. The results for π are consistent with population genetics theory (Wright, 1931), in that large geographic ranges can support larger populations that harbor more genetic diversity because of reduced genetic drift (Lacy, 1987). The relationship between range size and genetic structure (i.e., IBD and IBE) could be explained by the greater potential for geographic isolation between populations at distant locations in species with large geographic ranges (Eckert et al., 2008). However, these results were relatively weak and only apparent for tropical species, highlighting the challenge of disentangling the many processes that could influence connectivity within species.

Global patterns of nucleotide diversity (π)

Our results provide evidence of the uneven distribution of nucleotide diversity (π_s) across the globe. We observed that certain hotspots of genetic diversity, like the Ecuadorian Amazon and central Europe, southwestern and central North America, and southeast Asia, align with regions of high biodiversity (e.g., Mittermeier et al., 2011). Contrastingly, low genetic diversity in regions such as northern North America or western Europe could reflect recent recolonizations or historical bottlenecks. Our results contrasted with Miraldo et al. (2016), who found that amphibian genetic diversity decreased at higher latitudes. In our study, latitude variables (mean, maximum, or minimum) were not important predictors of genetic diversity in salamanders. The high diversity we found in the Neotropics (although with only 16 species) could be explained by a combination of factors including tropical climate stability, high topographic heterogeneity, or long evolutionary history in the Neotropics (Antonelli, 2022). Another plausible explanation for the observed high levels of π in some regions is that they coincide with glacial refugia, areas where species persisted and maintained high levels of genetic diversity through time (e.g.,

Ramírez-Barahona and Eguiarte, 2013). The Andean foothills were a major glacial refugium, particularly the central Andes in the Neotropical region (e.g., Escobar et al., 2021). In the Nearctic, the Southern Appalachians and coastal areas of the Pacific Northwest are considered potential glacial refugia (e.g., Shafer et al., 2010). In the Palearctic, the Caucasus Mountain range has been identified as a potential refugium during glacial periods (e.g., *Triturus* newts; Wielstra et al., 2013). The Korean Peninsula and the mountainous regions of China are two examples of glacial refugia in the Oriental region (Fu and Wen, 2023). Low levels of π in northern North America are consistent with the impact of historical and demographic processes in the region, such as Pleistocene glaciations and associated range contractions in salamanders (Rovito and Schoville, 2017; López-Delgado and Meirmans, 2022).

Global patterns of isolation-by-distance (IBD) and environment (IBE)

The prevalence of IBD and IBE across species (134 species, or 64%, showed both patterns) suggests that both geographic and environmental factors are important in shaping genetic variation in salamanders, compared to vertebrates where 29% and 21% of vertebrates displayed IBD and IBE, respectively (Pelletier and Carstens, 2018). These results may relate to low dispersal abilities in salamanders and fragmented habitats in the regions these species inhabit (see Auteri et al., 2022 for an example with *Dicamptodon* species). A previous meta-analysis found that ectotherms were more likely to exhibit IBD than endotherms, consistent with our detection of frequent IBD patterns in salamanders (Jenkins et al., 2010). The proportions of species showing IBD and IBE differed slightly across biogeographic realms, with the Palearctic including the most species with these patterns. One example is the fire salamander (*Salamandra salamandra*), one of the most common and widespread salamander species in Europe, which was found to be strongly structured in a landscape genetics analysis (Bani et al., 2015). The Neotropical and Nearctic regions had lower proportions of species with IBD or IBE patterns compared to the other regions, but these differences were relatively minor. Species without these spatial patterns may experience uniformly high gene flow that can homogenize genetic variation between geographically distant or environmentally disparate populations, or they may have been strongly influenced by historical processes that disrupt patterns of genetic isolation associated with IBD and IBE (Sexton et al., 2014). For example, montane populations isolated in separate

refugia in response to climatic changes might differentiate and exhibit highly restricted gene flow even at short distances or among similar environments, as reported in plethodontid salamanders (Kozak and Wiens, 2006; Rovito, 2017).

Predictors of genetic variation in global salamanders

Two of the best predictors of nucleotide diversity in this study, the number of sequences sampled per species and range size, could both be considered proxies of demographic processes. Sampling more sequences may simply provide a better opportunity to detect more genetic variation, or species with more available sequences could truly be more abundant in nature, reflecting sampling bias towards species that are more frequently encountered. Either explanation is consistent with the positive correlations found between number of sequences, range size, and π in our dataset (Figure S21). Species with larger geographic ranges can have larger and more balanced populations, reflecting the relationship between effective population size and genetic variation (Waples, 2022). Geographic range size was also the best predictor of IBD and IBE, similar to previous findings in a wide range of taxonomic groups (Pelletier and Carstens, 2018). A large geographic range can harbor more complex population structure, where population subdivision could be influenced by different ecological and evolutionary pressures and undergo contrasting demographic histories (Lowe et al., 2017). The relationship between range size and genetic diversity we identified in global salamanders is partially consistent with previous regional studies, such as Amador et al. (2024), which found a similar pattern in Neotropical amphibians for π, but not for IBD or IBE. In the present study, most salamander species had IBD and IBE patterns, making it difficult to detect a relationship between presence or absence of genetic structure and range size (Figure S20; Table S3). Nevertheless, when we analyzed species separately by biogeographic zones, we found differences in the tropical zone where species with larger geographic ranges tended to have IBD and IBE. Similar to previous meta-analyses of IBD and IBE (Crispo and Hendry 2005; Jenkins et al. 2010), we found that identifying global predictors of spatial genetic variation is challenging because of the many potential factors that influence the presence and detection of these patterns.

Climatic variables, including different aspects of temperature and precipitation, were among the top ten predictors for the three response variables in RF analyses, suggesting an important role of climate on genetic variation within salamander. Mean temperature was the second best predictor for IBE, the fourth best predictor for IBD, and the tenth best predictor of π. The association between higher genetic variation and warmer temperatures is consistent with the evolutionary speed hypothesis (Wright et al., 2010; Gillman and Wright, 2014), where genetic diversity is expected to increase with temperature. Variation in temperature and precipitation were also important predictors for each response. A possible explanation for this association is that current stable climates, such as those in tropical environments, promote genetic variation, thus supporting the climate stability hypothesis (Janzen, 1967; Stevens, 1989). Although the relationships between these climatic variables and the response variables we examined were relatively weak, previous work has also highlighted the role of climate in explaining variation within salamanders. Specifically, phylogeographic structure measured using species delimitation methods was best explained by variation in climate (Parsons et al., 2024). Like other studies on birds (Smith et al., 2017) and amphibians (Barrow et al., 2021), our analyses suggest that life history and ecological traits are less important than geographic or climatic variation for predicting intraspecific diversity in salamanders.

To further understand the potential link between historical demography and the patterns of genetic variation we found, we performed a post hoc summary of a statistic commonly used to test for population size change, Tajima's D (Tajima, 1989). We did not use Tajima's D for formal analyses within our study because this test is compatible with different interpretations related to natural selection versus demographic change, making it difficult to distinguish among hypotheses and present results (Yang, 2014). However, we showed that species experiencing expansion, based on Tajima's D, presented low levels of nucleotide diversity (Figure S22). This result is expected when expansion from a small population occurred after a bottleneck, leading to a limited gene pool carried by a few founding individuals, such as range expansions after the last glacial maximum (Mayr, 2001). For IBD and IBE, we found that species presenting either pattern were more likely to have experienced population contraction than species that did not exhibit IBD/IBE. This trend makes sense because when a population contracts, individuals are more likely to mate within a smaller geographic area, enhancing IBD patterns; in the case of

IBE, localized environmental pressures can become more prominent, leading to genetic differentiation between subpopulations. Species interpreted as experiencing population expansion had a similar proportion of IBD/no IBD and IBE/no IBE patterns, suggesting there is not a consistent link between recent expansion and genetic structure (Figure S23).

Phylogenetic, taxonomic, and conservation implications

We did not find significant phylogenetic signal in π, although taxonomic family was the second-best predictor of π in the final RF model, suggesting that some variation can be explained by evolutionary history, but not all. Sample sizes were limited within some families, but diversity estimates in those families were consistently high (e.g., Sirenidae) or low (e.g., Cryptobranchidae), indicating family can be a useful predictor of intraspecific variation. On the other hand, discrepancies in π within genera of family Plethodontidae, such as observed in *Plethodon* or *Eurycea*, emphasize the influence of local ecological and recent demographic processes over shared ancestry for influencing variation within species. Processes leading to homoplasy (e.g., convergence, parallel evolution) dilute phylogenetic signal (Klingenberg and Gidaszewski, 2010). Homoplasy is a common phenomenon in salamander evolution; for example, paedomorphosis has evolved independently several times (Wake, 2009). Likewise, levels of genetic variation may be similar in salamander species that do not share a recent common ancestor (Kozak and Wiens, 2006). Phylogenetic signal in IBD and IBE were not statistically significant either, suggesting other factors independent of phylogenetic history influence these patterns.

Species with higher π values could contain cryptic diversity and would be useful target species for exploring species boundaries in future studies (e.g., Parsons et al. 2022). Using a smaller dataset than the present study (83 salamander species total, most from the COI mitochondrial gene), Parsons et al. (2024) found that ~2/3 of species in the dataset showed strong phylogeographic structure, determined by single locus molecular species delimitation methods. Several species with high values of π in our dataset match species from their study, indicating they have hidden genetic lineages that should be further explored (*Mertensiella caucasica, Bolitoglossa rufescens, Plethodon cinereus, Eurycea cirrigera, Batrachoseps attenuatus,*

Ommatotriton vittatus, *Aneides flavipunctatus*), ideally incorporating nuclear genomic data and thorough geographic sampling.

Salamanders are within one of the most threatened vertebrate groups (Luedtke et al., 2023) and require innovative syntheses of available information to highlight priorities for conservation strategies. Nucleotide diversity could be a useful proxy for conservation status (Petit-Marty, 2021). The relationship we identified between species range size and genetic variation highlights the vulnerability of species with small ranges. Conservation status was an important predictor for each response variable we examined, but sample sizes were limited within several categories. Seventy-five (35%) of the 214 species in our dataset are threatened based on the IUCN Red List; 11 as Critical Endangered (CR), 29 Endangered (EN), and 35 Vulnerable (VU). These salamanders should receive special attention, especially if they occupy small ranges and have low genetic diversity.

Limitations and challenges

One major limitation for macrogenetic studies is the uneven spatial sampling of available sequences, which can lead to an incomplete picture of the distribution of genetic diversity. There are also sampling biases among taxa, such that some species are well characterized while others are omitted from global comparisons. Another limitation is the use of a single mitochondrial gene (e.g., Cytb) to calculate genetic diversity and genetic distances. Although mitochondrial DNA is widely used in phylogenetic and population genetic studies because it evolves quickly (Galtier et al., 2009), it provides an incomplete picture of intraspecific diversity because it is a single locus, haploid, and solely maternally inherited (Moore, 1995). However, single-locus macrogenetic studies are still valuable because the amount and availability of data elucidate patterns associated with historical processes and variability within species. Furthermore, the large genome size of salamanders hampers the use of nuclear genomic data in studies on this scale. These limitations reinforce the need for additional geographic sampling of high-quality tissues that include broader taxonomic coverage and facilitate new genomic sequencing efforts.

The most important challenge for our study was the time required to retrieve the geographic coordinates associated with each Cytb sequence. Most sequences in open access databases such as GenBank do not have their latitude and longitude associated with them, which limits studies of spatial genetic variation. Efforts such as phylogatR (Pelletier et al. 2022) aggregate genetic sequences associated with georeferenced occurrences from global databases using automated pipelines, but researchers providing the sequences must include the proper metadata when uploading their data for this approach to work. We encourage scientists to share spatial data and museum catalogue numbers when applicable along with sequences when they are published, ensuring data reproducibility and extendibility.

Conclusions

Our study underscores the value of integrating spatial information (i.e., geographic coordinates associated with DNA sequences) into macrogenetic studies. Our results support the hypotheses that variables associated with population size (e.g., species range size) are positively correlated with genetic variation and that salamanders tend to exhibit spatial patterns of genetic variation that also differ across biogeographic realms. Based on our results, we reject the hypothesis that traits associated with reproductive strategies predict intraspecific genetic variation in salamanders. Our results suggest the importance of range-wide climatic variation, local environmental heterogeneity, and historical demographic events for influencing genetic variation. Future amphibian macrogenetic and genomic studies can incorporate more explicit tests of range expansion and contraction to better understand these processes. Finally, our findings have broad implications for establishing conservation priorities and highlighting regions where sampling has been historically scarce, particularly in biodiversity hotspots (e.g., Neotropical forests) that warrant further attention.

References

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*, 51(5), 909–923. https://doi.org/10.1111/jbi.14795

AmphibiaWeb. (2024). <https://amphibiaweb.org> University of California, Berkeley, CA, USA. Accessed 3 October 2024.

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

Auteri, G. G., Marchán-Rivadeneira, M. R., Olson, D. H., & Knowles, L. L. (2022). Landscape connectivity among coastal giant salamander (*Dicamptodon tenebrosus*) populations shows no association with land use, fire frequency, or river drainage but exhibits genetic signatures of potential conservation concern. *PLOS ONE*, *17*(6), e0268882. https://doi.org/10.1371/journal.pone.0268882

Bani, L., Pisa, G., Luppi, M., Spilotros, G., Fabbri, E., Randi, E., & Orioli, V. (2015). Ecological connectivity assessment in a strongly structured fire salamander (*Salamandra salamandra*) population. *Ecology and Evolution*, *5*(16), 3472–3485. https://doi.org/10.1002/ece3.1617

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

Bivand R, Pebesma E, Gómez-Rubio V. (2013). *Applied spatial data analysis with R, Second edition*. Springer, NY. https://asdar-book.org/.

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

Calenge C (2024). adehabitatHR: Home Range Estimation. R package version 0.4.22, <https://CRAN.R-project.org/package=adehabitatHR>.

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.

Cambon, J., Hernangómez, D., Belanger, C., & Possenriede, D. (2021). tidygeocoder: An R package for geocoding. *Journal of Open Source Software*, *6*(65), 3544. https://doi.org/10.21105/joss.03544

Crispo, E., & Hendry, A. P. (2005). Does time since colonization influence isolation by distance? A meta-analysis. *Conservation Genetics*, *6*(5), 665–682. https://doi.org/10.1007/s10592-005-9026-4

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

Denoël, M., Joly, P., & Whiteman, H. H. (2005). Evolutionary ecology of facultative paedomorphosis in newts and salamanders. *Biological Reviews*, *80*(4), 663–671. https://doi.org/10.1017/S1464793105006858

Dunnington, D. (2023). ggspatial: Spatial Data Framework for ggplot2. R package version 1.1.9, <https://CRAN.R-project.org/package=ggspatial>.

Eckert, C. G., Samis, K. E., & Lougheed, S. C. (2008). Genetic variation across species' geographical ranges: The central–marginal hypothesis and beyond. *Molecular Ecology*, *17*(5), 1170–1188. https://doi.org/10.1111/j.1365-294X.2007.03659.x

Edgar, R. C. (2004). MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics*, *5*(1), 113. https://doi.org/10.1186/1471-2105-5-113

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

Escobar, S., Helmstetter, A. J., Jarvie, S., Montúfar, R., Balslev, H., & Couvreur, T. L. P. (2021). Pleistocene climatic fluctuations promoted alternative evolutionary histories in *Phytelephas aequatorialis* , an endemic palm from western Ecuador. *Journal of Biogeography*, *48*(5), 1023–1037. https://doi.org/10.1111/jbi.14055

Felsenstein, J. (1985). Phylogenies and the Comparative Method. *The American Naturalist*, *125*(1,), 1–15.

Fick, S.E. and R.J. Hijmans, (2017). WorldClim 2: new 1km spatial resolution climate surfaces for global land areas. *International Journal of Climatology* 37 (12): 4302-4315.

Fonseca, E. M., Garda, A. A., Oliveira, E. F., Camurugi, F., Magalhães, F. D. M., Lanna, F. M., Zurano, J. P., Marques, R., Vences, M., & Gehara, M. (2021). The riverine thruway hypothesis: Rivers as a key mediator of gene flow for the aquatic paradoxical frog *Pseudis tocantins* (Anura, Hylidae). *Landscape Ecology*, *36*(10), 3049–3060. https://doi.org/10.1007/s10980-021-01257-z

Fonseca, E. M., Pelletier, T. A., Decker, S. K., Parsons, D. J., & Carstens, B. C. (2023). Pleistocene glaciations caused the latitudinal gradient of within-species genetic diversity. *Evolution Letters*, *7*(5), 331–338. https://doi.org/10.1093/evlett/qrad030

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

French, C. M., Bertola, L. D., Carnaval, A. C., Economo, E. P., Kass, J. M., Lohman, D. J., Marske, K. A., Meier, R., Overcast, I., Rominger, A. J., Staniczenko, P. P. A., & Hickerson, M. J. (2023). Global determinants of insect mitochondrial genetic diversity. *Nature Communications*, *14*(1), 5276. https://doi.org/10.1038/s41467-023-40936-0

Fu, J., & Wen, L. (2023). Impacts of Quaternary glaciation, geological history and geography on animal species history in continental East Asia: A phylogeographic review. *Molecular Ecology*, *32*(16), 4497–4514. https://doi.org/10.1111/mec.17053

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

Galtier, N., Nabholz, B., Glémin, S., & Hurst, G. D. D. (2009). Mitochondrial DNA as a marker of molecular diversity: A reappraisal. *Molecular Ecology*, 18(22), 4541–4550. https://doi.org/10.1111/j.1365-294X.2009.04380.x

Gillman, L. N., & Wright, S. D. (2014). Species richness and evolutionary speed: The influence of temperature, water and area. *Journal of Biogeography*, *41*(1), 39–51. https://doi.org/10.1111/jbi.12173

Hanson, J. O., Rhodes, J. R., Riginos, C., & Fuller, R. A. (2017). Environmental and geographic variables are effective surrogates for genetic variation in conservation planning. *Proceedings of the National Academy of Sciences*, *114*(48), 12755–12760. https://doi.org/10.1073/pnas.1711009114

Hijmans R (2022). geosphere: Spherical Trigonometry. R package version 1.5-18, <https://CRAN.R-project.org/package=geosphere>.

Hijmans RJ, Barbosa M, Ghosh A, Mandel A (2024). geodata: Download Geographic Data. R package version 0.6-2, <https://CRAN.R-project.org/package=geodata>.

Ives, A. R., & Helmus, M. R. (2011). Generalized linear mixed models for phylogenetic analyses of community structure. *Ecological Monographs*, *81*(3), 511–525. https://doi.org/10.1890/10-1264.1

Janzen, D. H. (1967). Why Mountain Passes are Higher in the Tropics. *The American Naturalist*, *101*(919), 233–249. https://doi.org/10.1086/282487

Jenkins, D. G., Carey, M., Czerniewska, J., Fletcher, J., Hether, T., Jones, A., Knight, S., Knox, J., Long, T., Mannino, M., McGuire, M., Riffle, A., Segelsky, S., Shappell, L., Sterner, A., Strickler, T., & Tursi, R. (2010). A meta-analysis of isolation by distance: Relic or reference standard for landscape genetics? *Ecography*, *33*(2), 315–320. https://doi.org/10.1111/j.1600-0587.2010.06285.x

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*(5), 850–858. https://doi.org/10.1038/s41559-018-0515-5

Kardos, M., Armstrong, E. E., Fitzpatrick, S. W., Hauser, S., Hedrick, P. W., Miller, J. M., Tallmon, D. A., & Funk, W. C. (2021). The crucial role of genome-wide genetic variation in conservation. *Proceedings of the National Academy of Sciences*, *118*(48), e2104642118. https://doi.org/10.1073/pnas.2104642118

Kimura, M. (1968). Evolutionary rate at the molecular level. *Nature*, 217(5129), 624-626. https://doi.org/10.1038/217624a0

Klingenberg, C. P., & Gidaszewski, N. A. (2010). Testing and Quantifying Phylogenetic Signals and Homoplasy in Morphometric Data. *Systematic Biology*, 59(3), 245–261. https://doi.org/10.1093/sysbio/syp106

Kozak, K. H., & Wiens, J. J. (2006). Does niche conservatism promote speciation? A case study in North American salamanders. *Evolution*, *60*(12), 2604–2621. https://doi.org/10.1111/j.0014-3820.2006.tb01893.x

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.

Larsson, A. (2014). AliView: A fast and lightweight alignment viewer and editor for large datasets. *Bioinformatics*, *30*(22), 3276–3278. https://doi.org/10.1093/bioinformatics/btu531

Lawrence, E. R., Pedersen, E. J., & Fraser, D. J. (2023). Macrogenetics reveals multifaceted influences of environmental variation on vertebrate population genetic diversity across the Americas. *Molecular Ecology*, *32*(16), 4557–4569. https://doi.org/10.1111/mec.17059

Lee-Yaw, J. A., & Irwin, D. E. (2012). Large geographic range size reflects a patchwork of divergent lineages in the long-toed salamander (*Ambystoma macrodactylum*). *Journal of Evolutionary Biology*, *25*(11), 2276–2287. https://doi.org/10.1111/j.1420-9101.2012.02604.x

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

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

Liaw, A., Wiener, M. (2002). Classification and Regression by randomForest. *R News*, 2(3), 18-22. <https://CRAN.R-project.org/doc/Rnews/>.

López-Delgado, J., & Meirmans, P. G. (2022). History or demography? Determining the drivers of genetic variation in North American plants. *Molecular Ecology*, 31(7), 1951–1962. https://doi.org/10.1111/mec.16230

Lowe, W. H., Kovach, R. P., & Allendorf, F. W. (2017). Population Genetics and Demography Unite Ecology and Evolution. *Trends in Ecology & Evolution*, *32*(2), 141–152. https://doi.org/10.1016/j.tree.2016.12.002

Luedtke, J. A., Chanson, J., Neam, K., Hobin, L., Maciel, A. O., Catenazzi, A., Borzée, A., Hamidy, A., Aowphol, A., Jean, A., Sosa-Bartuano, Á., Fong G., A., De Silva, A., Fouquet, A., Angulo, A., Kidov, A. A., Muñoz Saravia, A., Diesmos, A. C., Tominaga, A., … Stuart, S. N. (2023). Ongoing declines for the world's amphibians in the face of emerging threats. *Nature*, *622*(7982), 308–314. https://doi.org/10.1038/s41586-023-06578-4

Malcher, G., Amorim, A.L., Ferreira, P. *et al.* (2023). First evaluation of the population genetics and aspects of the evolutionary history of the Amazonian snook, *Centropomus irae*, and its

association with the Amazon plume. *Hydrobiologia* 850, 2115–2125. https://doi.org/10.1007/s10750-023-05223-5

Mayr, E. (2001). *What Evolution Is*. New York, NY: Basic Books.

Meng, H., Li, X., & Qiao, P. (2014). Population Structure, Historical Biogeography and Demographic History of the Alpine Toad *Scutiger ningshanensis* in the Tsinling Mountains of Central China. *PLoS ONE*, *9*(6), e100729. https://doi.org/10.1371/journal.pone.0100729

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

Mittermeier, R. A., Turner, W. R., Larsen, F. W., Brooks, T. M., & Gascon, C. (2011). Global Biodiversity Conservation: The Critical Role of Hotspots. In F. E. Zachos & J. C. Habel (Eds.), *Biodiversity Hotspots* (pp. 3–22). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-20992-5_1

Moore, W. S. (1995). Inferring phylogenies from mtDNA variation: mitochondrial-gene trees versus nuclear-gene trees. Evolution, 49(4), 718-726.

Oksanen J, Simpson G, Blanchet F, Kindt R, Legendre P, Minchin P, O'Hara R, Solymos P, Stevens M, Szoecs E, Wagner H, Barbour M, Bedward M, Bolker B, Borcard D, Carvalho G, Chirico M, De Caceres M, Durand S, Evangelista H, FitzJohn R, Friendly M, Furneaux B, Hannigan G, Hill M, Lahti L, McGlinn D, Ouellette M, Ribeiro Cunha E, Smith T, Stier A, Ter Braak C, Weedon J (2024). vegan: Community Ecology Package. R package version 2.6-8, <https://CRAN.R-project.org/package=vegan>

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

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

Paradis, E., & Schliep, K. (2019). ape 5.0: An environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics*, 35(3), 526–528. https://doi.org/10.1093/bioinformatics/bty633

Parsons, D. J., Pelletier, T. A., Wieringa, J. G., Duckett, D. J., & Carstens, B. C. (2022). Analysis of biodiversity data suggests that mammal species are hidden in predictable places. *Proceedings*

of the National Academy of Sciences, *119*(14), e2103400119. https://doi.org/10.1073/pnas.2103400119

Parsons, D. J., Green, A. E., Carstens, B. C., & Pelletier, T. A. (2024). Predicting genetic biodiversity in salamanders using geographic, climatic, and life history traits. *PLOS ONE*, *19*(10), e0310932. https://doi.org/10.1371/journal.pone.0310932

Park, K. Y., Lucas, M., Chaulk, A., Matter, S. F., Roland, J., & Keyghobadi, N. (2024). Immigration allows population persistence and maintains genetic diversity despite an attempted experimental extinction. *Royal Society Open Science*, *11*(7), 240557. https://doi.org/10.1098/rsos.240557

Pauls, S. U., Nowak, C., Bálint, M., & Pfenninger, M. (2013). The impact of global climate change on genetic diversity within populations and species. *Molecular Ecology*, *22*(4), 925–946. https://doi.org/10.1111/mec.12152

Pebesma, E., 2018. Simple Features for R: Standardized Support for Spatial Vector Data. The R Journal 10 (1), 439-446, https://doi.org/10.32614/RJ-2018-009

Pebesma, E., & Bivand, R. (2023). Spatial Data Science: With Applications in R. Chapman and Hall/CRC. https://doi.org/10.1201/9780429459016

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

Petit-Marty, N., Vázquez-Luis, M., & Hendriks, I. E. (2021). Use of the nucleotide diversity in COI mitochondrial gene as an early diagnostic of conservation status of animal species. *Conservation Letters*, *14*(1), e12756. https://doi.org/10.1111/conl.12756

R Core Team (2023). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.

Ramírez-Barahona, S., & Eguiarte, L. E. (2013). The role of glacial cycles in promoting genetic diversity in the Neotropics: The case of cloud forests during the Last Glacial Maximum. *Ecology and Evolution*, *3*(3), 725–738. https://doi.org/10.1002/ece3.483

Reed, D. H., & Frankham, R. (2003). Correlation between Fitness and Genetic Diversity. *Conservation Biology*, *17*(1), 230–237. https://doi.org/10.1046/j.1523-1739.2003.01236.x

Rios, N. E. & Bart, H. L. (2010). GEOLocate (Version 3.22). [Computer software]. Belle Chasse, LA: Tulane University Museum of Natural History.

Román-Palacios, C. (2023). The phruta r package: Increasing access, reproducibility and transparency in phylogenetic analyses. *Methods in Ecology and Evolution*, 14(9), 2284–2299. https://doi.org/10.1111/2041-210X.14147

Rovito, S. M. (2017). The Geography of Speciation in Neotropical Salamanders. *Herpetologica*, *73*(3), 229–241. https://doi.org/10.1655/HERPETOLOGICA-D-16-00077.1

Rovito, S. M., & Schoville, S. D. (2017). Testing models of refugial isolation, colonization and population connectivity in two species of montane salamanders. *Heredity*, *119*(4), 265–274. https://doi.org/10.1038/hdy.2017.31

RS-eco (2023). rasterSp: R Package to rasterize and summarise IUCN range maps. R package version 0.0.1, <https://github.com/RS-eco/rasterSp>.

Schierenbeck, K. A. (2017). Population-level genetic variation and climate change in a biodiversity hotspot. *Annals of Botany*, *119*(2), 215–228. https://doi.org/10.1093/aob/mcw214

Segovia-Ramírez, M. G., Ramírez-Sánchez, O., Decena Segarra, L. P., Rios-Carlos, H., & Rovito, S. M. (2023). Determinants of genetic diversity in Neotropical salamanders (Plethodontidae: Bolitoglossini). *Ecology and Evolution*, *13*(11), e10707. https://doi.org/10.1002/ece3.10707

Sexton, J. P., Hangartner, S. B., & Hoffmann, A. A. (2014). Genetic isolation by environment or distance: which pattern of gene flow is most common? Special Section. *Evolution*, *68*(1), 1–15. https://doi.org/10.1111/evo.12258

Shafer, A. B. A., Cullingham, C. I., Côté, S. D., & Coltman, D. W. (2010). Of glaciers and refugia: A decade of study sheds new light on the phylogeography of northwestern North America. *Molecular Ecology*, *19*(21), 4589–4621. https://doi.org/10.1111/j.1365-294X.2010.04828.x

Smith, B. T., Seeholzer, G. F., Harvey, M. G., Cuervo, A. M., & Brumfield, R. T. (2017). A latitudinal phylogeographic diversity gradient in birds. *PLOS Biology*, *15*(4), e2001073. https://doi.org/10.1371/journal.pbio.2001073

Stevens, G. C. (1989). The Latitudinal Gradient in Geographical Range: How so Many Species Coexist in the Tropics. *The American Naturalist*, *133*(2), 240–256. https://doi.org/10.1086/284913

Tajima, F. (1989). Statistical methods to test for nucleotide mutation hypothesis by DNA polymorphism. *Genetics* 123:585-595.

Wake, D. B. (2009). What Salamanders Have Taught Us About Evolution. *Annual Review of Ecology, Evolution, and Systematics*, *40*(1), 333–352. https://doi.org/10.1146/annurev.ecolsys.39.110707.173552

Wang, I. J., & Bradburd, G. S. (2014). Isolation by environment. *Molecular Ecology*, *23*(23), 5649–5662. https://doi.org/10.1111/mec.12938

Waples, R. S. (2022). What is N_e, anyway? *Journal of Heredity*, 113(4), 371–379. https://doi.org/10.1093/jhered/esac023

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.

Wielstra, B., Crnobrnja-Isailović, J., Litvinchuk, S. N., Reijnen, B. T., Skidmore, A. K., Sotiropoulos, K., Toxopeus, A. G., Tzankov, N., Vukov, T., & Arntzen, J. W. (2013). Tracing glacial refugia of *Triturus* newts based on mitochondrial DNA phylogeography and species distribution modeling. *Frontiers in Zoology*, *10*(1), 13. https://doi.org/10.1186/1742-9994-10-13

Wright, S. (1931). Evolution in Mendelian populations. *Genetics*, *16*(2), 97–159. https://doi.org/10.1093/genetics/16.2.97

Wright, S. (1943). Isolation by distance. *Genetics*, *28*(2), 114–138. https://doi.org/10.1093/genetics/28.2.114

Wright, S. D., Gillman, L. N., Ross, H. A., & Keeling, D. J. (2010). Energy and the tempo of evolution in amphibians. *Global Ecology and Biogeography*, *19*(5), 733–740. https://doi.org/10.1111/j.1466-8238.2010.00549.x

Yang, Z. (2014). Molecular evolution: a statistical approach. Oxford University Press, Oxford.

Figures

Figure 1. Maps of A) mean nucleotide diversity (π) and B) mean number of sequences with coordinates (log scale) of global salamanders. Biogeographic realms are indicated with shading: Nearctic (brown), Neotropic (green), Oriental (blue), Palearctic (purple). Boxplots of nucleotide diversity for each biogeographic realm are shown in map A. The number of species and mean, minimum, and maximum values of sequences with coordinates for each biogeographic realm are shown in map B. For visualization purposes, we show grid cell = 4° (approximately 440 km^2) for mapping.

Figure 2. Phylogeny of Caudata species included in this study showing the values of nucleotide diversity (π) as barplots to the right and the presence (red) or absence (black) of isolation-by-distance (IBD)and isolation-by-environment (IBE) per species. Relationships are based on Jetz et al. (2018).

Figure 3. Random Forest (RF) results for A) nucleotide diversity (π, using RF regression), B) isolation-by-distance (IBD, using RF classification), and C) isolation-by-environment (IBE, using RF classification). All predictors except sampling effort variables (number of sequences, number of base pairs) and total range size were used in these RF models, but only the top 10 best predictors are shown here. Range size* = sampling range size based on localities with associated DNA sequences; SVL = snout-vent length; mm = millimeters; sd = standard deviation.

Figure 4. A) Comparison of sampling range size (in log scale) with nucleotide diversity (π) for each species. B) Comparison of phylogenetic independent contrast (PIC) results between the sampling range size (log scale) and nucleotide diversity (π).

Figure 5. Boxplots comparing the sampling range size (in log scale) between species with (yes) and without (no) A, B) isolation-by-distance (IBD) and C, D) isolation by environment (IBE). A, C) include temperate species combining the Nearctic and Palearctic biogeographic realms. B, D) include tropical species combining the Neotropic and Oriental biogeographic realms.

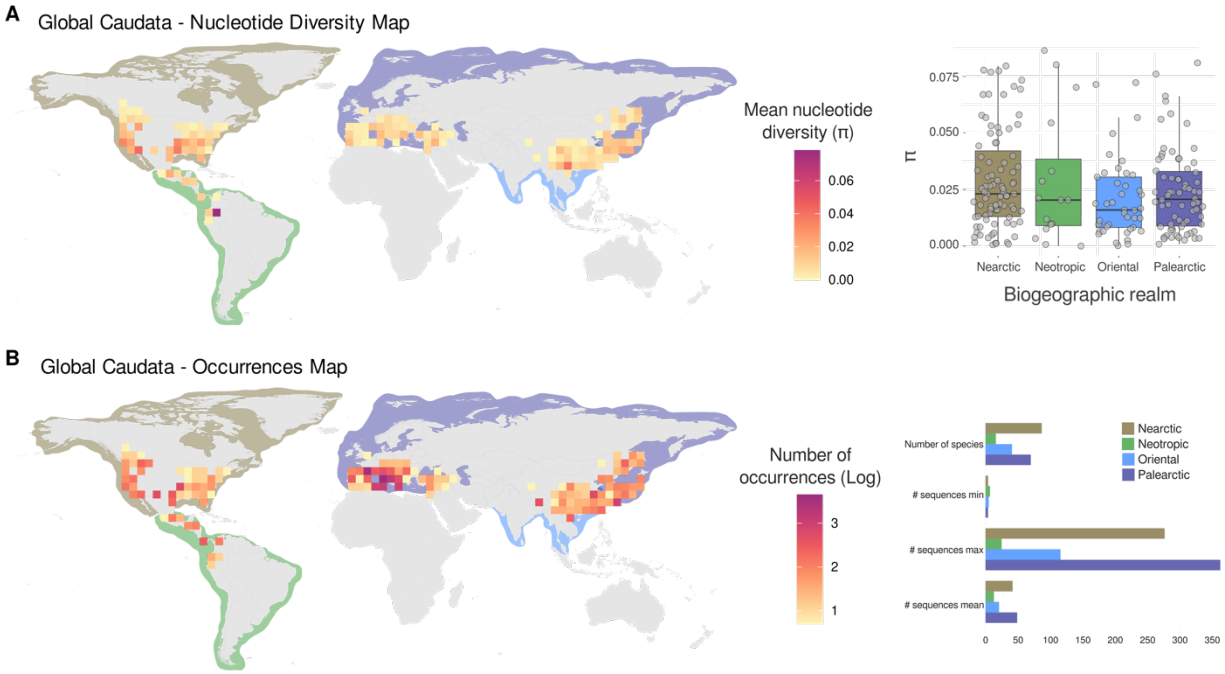

Figure 1. Maps of A) mean nucleotide diversity (π) and B) mean number of sequences with coordinates (log scale) of global salamanders. Biogeographic realms are indicated with shading: Nearctic (brown), Neotropic (green), Oriental (blue), and Palearctic (purple). Boxplots of nucleotide diversity for each biogeographic realm are shown in map A. The number of species and mean, minimum, and maximum values of sequences with coordinates for each biogeographic realm are shown in map B. For visualization purposes, we show grid cell = 4° (approximately 440 km^2) for mapping.

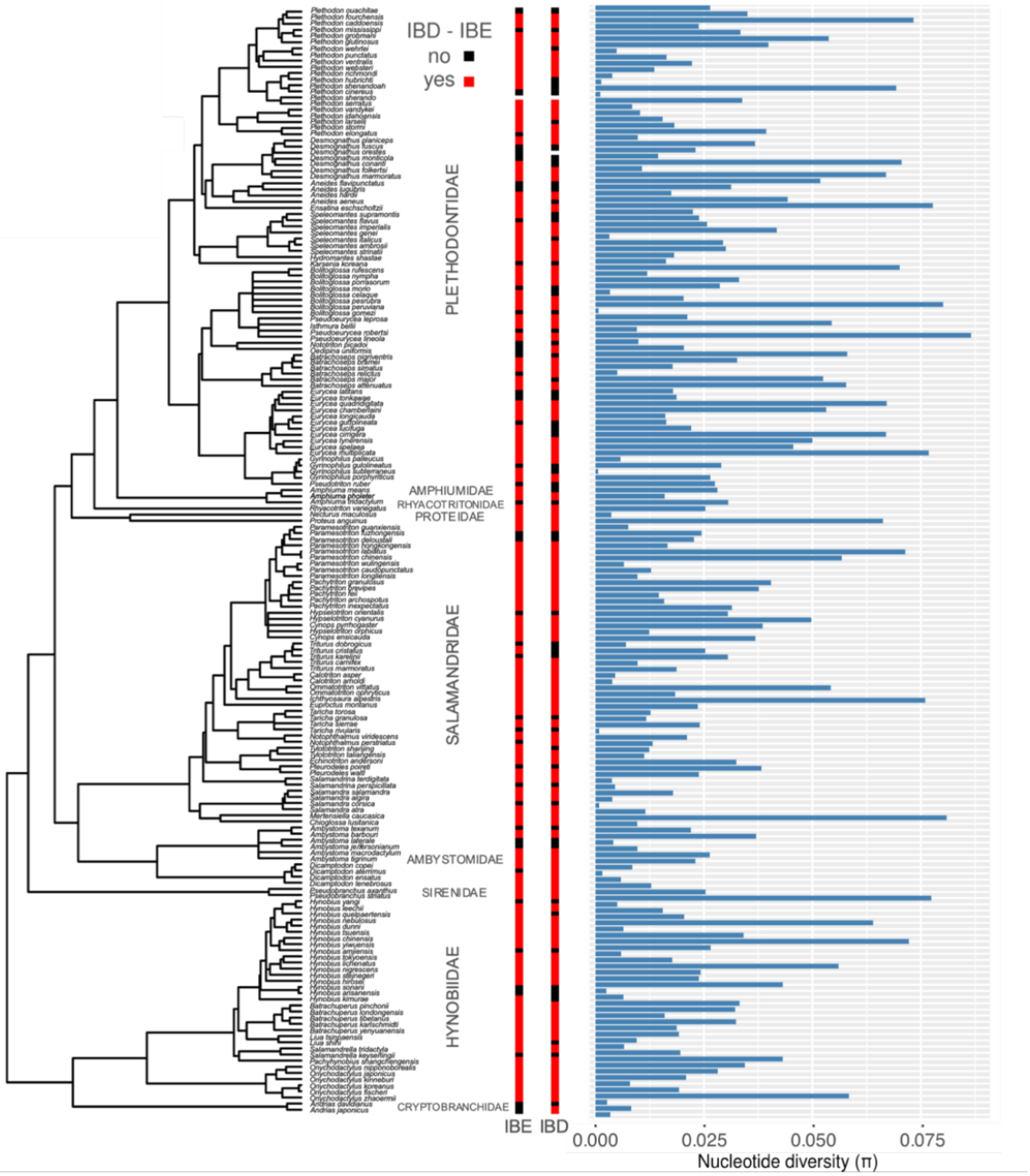

Figure 2. Phylogeny of Caudata species included in this study showing the values of nucleotide diversity (π) as barplots to the right and the presence (red) or absence (black) of isolation-by-distance (IBD) and isolation-by-environment (IBE) per species. Relationships are based on Jetz et al. (2018).

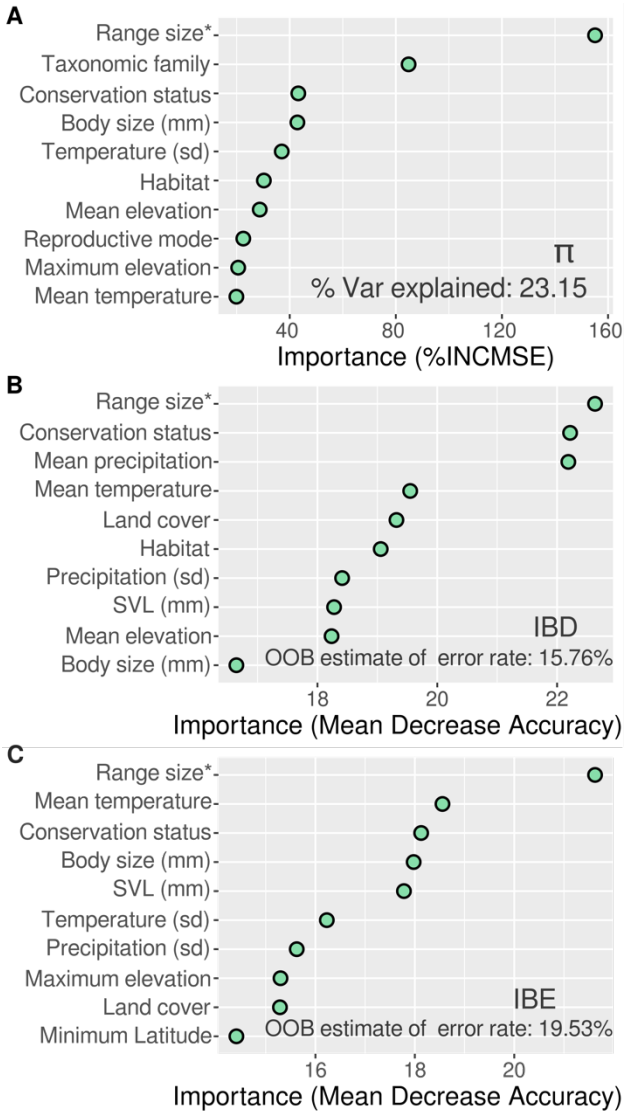

Figure 3. Random Forest (RF) results for A) nucleotide diversity (π, using RF regression), B) isolation-by-distance (IBD, using RF classification), and C) isolation-by-environment (IBE, using RF classification). All predictors except sampling effort variables (number of sequences, number of base pairs) and total range size were used in these RF models, but only the top 10 best predictors are shown here. Range size* = sampling range size based on localities with associated DNA sequences; SVL = snout-vent length; mm = millimeters; sd = standard deviation.

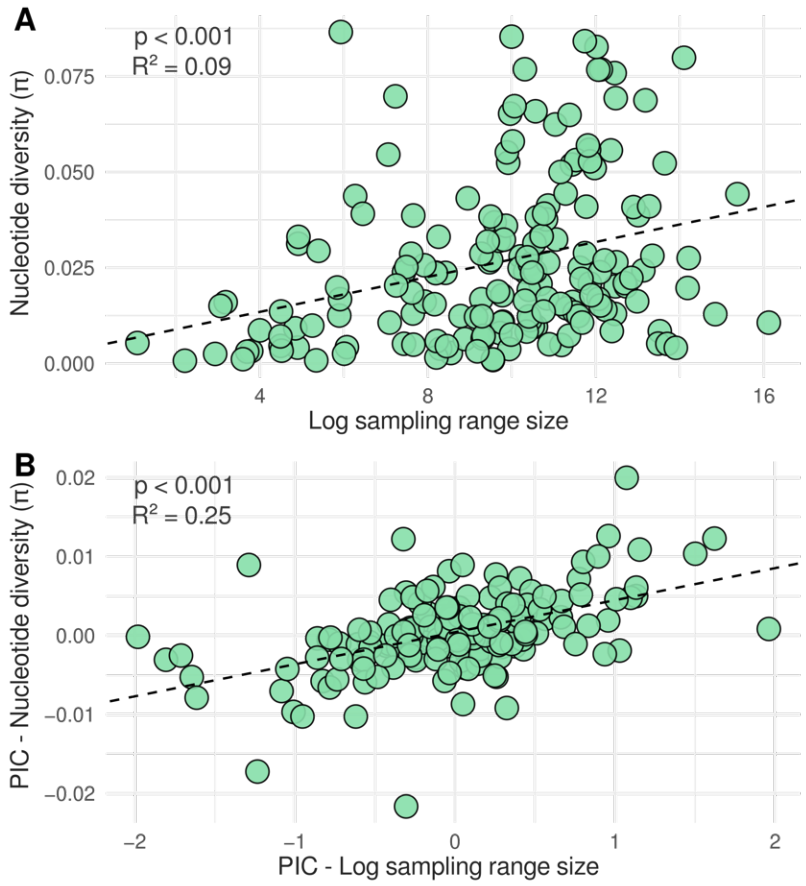

Figure 4. A) Comparison of sampling range size (in log scale) with nucleotide diversity (π) for each species. B) Comparison of phylogenetic independent contrast (PIC) results between the sampling range size (log scale) and nucleotide diversity (π).

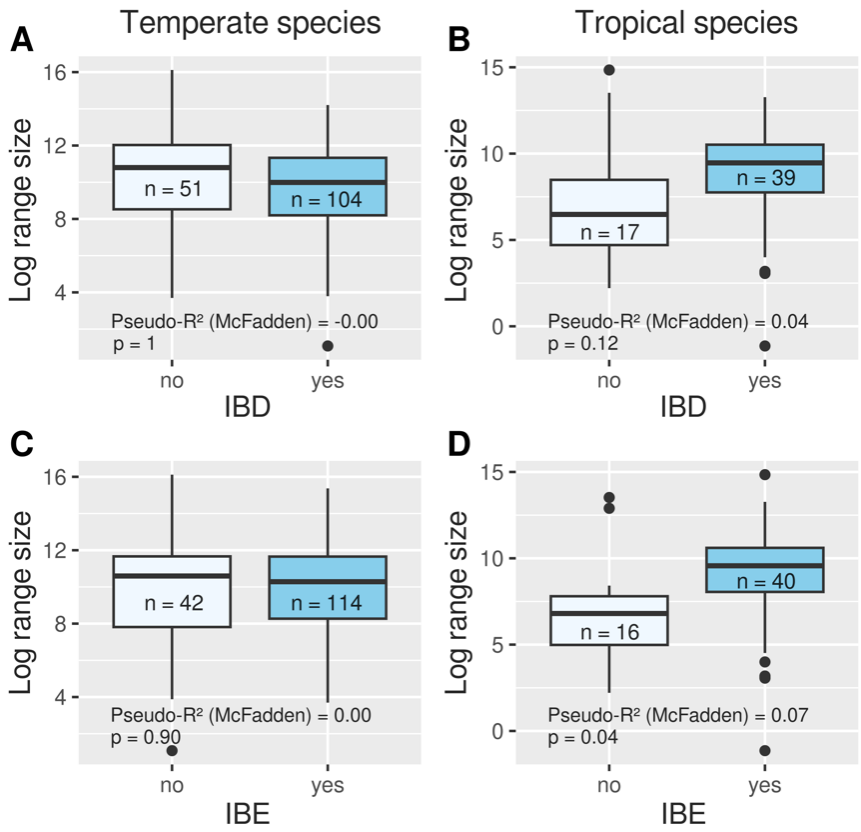

Figure 5. Boxplots comparing the sampling range size (in log scale) between species with (yes) and without (no) A,B) isolation-by-distance (IBD) and C,D) isolation by environment (IBE). A,C) include temperate species combining the Nearctic and Palearctic biogeographic realms. B,D) include tropical species combining the Neotropic and Oriental biogeographic realms.