The Influence of Environmental Variation on the Genetic Structure of a Poison Frog Distributed Across Continuous Amazonian Rainforest

Anthony S. Ferreira, Albertina P. Lima, Robert Jehle, Miquéias Ferrão, and Adam Stow

From the Programa de Capacitação Institucional, Instituto Nacional de Pesquisas da Amazônia, Manaus, Amazonas, Brazil (Ferreira); Coordenacão de Biodiversidade, Instituto Nacional de Pesquisas da Amazônia, Manaus, Amazonas, Brazil (Lima); School of Science, Engineering and Environment, University of Salford, M5 4WT Salford, UK (Jehle); Museum of Comparative Zoology, Harvard University, Cambridge, MA, USA (Ferrão); and Department of Biological Sciences, Macquarie University, Sydney, NSW, Australia (Stow).

Address correspondence to A.S. Ferreira at the address above, or e-mail: anthonyferreira@gmail.com.br

Received April 22, 2020; First decision June 16, 2020; Accepted August 20, 2020.

Corresponding Editor: Ryan Garrick

Abstract

Biogeographic barriers such as rivers have been shown to shape spatial patterns of biodiversity in the Amazon basin, yet relatively little is known about the distribution of genetic variation across continuous rainforest. Here, we characterize the genetic structure of the brilliant-thighed poison frog (Allobates femoralis) across an 880-km-long transect along the Purus-Madeira interfluve south of the Amazon river, based on 64 individuals genotyped at 7609 single-nucleotide polymorphism (SNP) loci. A population tree and clustering analyses revealed 4 distinct genetic groups, one of which was strongly divergent. These genetic groups were concomitant with femoral spot coloration differences, which was intermediate within a zone of admixture between two of the groups. The location of these genetic groups did not consistently correspond to current ecological transitions between major forest types. A multimodel approach to quantify the relative influence of isolation-by-geographic distance (IBD) and isolation-by-environmental resistance (IBR) nevertheless revealed that, in addition to a strong signal of IBD, spatial genetic differentiation was explained by IBR primarily linked to dry season intensity ($r^2 = 8.4\%$) and canopy cover ($r^2 = 6.4\%$). We show significant phylogenetic divergence in the absence of obvious biogeographical barriers and that finer-scaled measures of genetic structure are associated with environmental variables also known to predict the density of A. femoralis.

Subject area: molecular adaptation and selection

Keywords: RADseq, genetic clusters, landscape genetics, Amazonia, amphibians

Introduction

A key goal in ecology and evolutionary studies is to understand the processes that explain contemporary patterns of genetic diversity. Based on the classic allopatric speciation model, genetic divergence is a consequence of geographic isolation (Wallace 1852; Mayr 1963; Coyne and Orr 2004). However, divergence can also arise when isolation is incomplete, under scenarios that may include ecologically mediated selection triggered by environmental heterogeneity (Nosil
Neutral genetic population structure arises through the interplay of drift, mutation, and migration. Disentangling the legacy of historical events on patterns of genetic structure from more contemporary effects needs to account for the sensitivity of the molecular assays, the analytical approaches employed, as well as recognizing the time required for causal processes to shape genetic structure (Stow et al. 2001; Anderson et al. 2010; Epps and Keyghobadi 2015). Although isolation-by-geographic distance (IBD; Wright 1943; Slatkin 1987) is revealed by most empirical studies (for summaries, see, e.g., Jenkins et al. 2010; Sexton et al. 2014), gene flow can be further influenced by the landscape matrix where habitat heterogeneity results in different levels of resistance to migration (Manel et al. 2003; Storfer et al. 2010). Because patterns of isolation-by-environmental resistance (IBR) are influenced by species-specific life-history attributes and ecological preferences, such as propensity and ability for migration through given environments, they reveal essential information about habitat relationships of the studied taxa (Balkenhol et al. 2017; Armansin et al. 2020). The spatial scale of sampling is an especially important consideration when testing for IBD and IBR. If the scale of sampling is too small relative to the scale of gene flow of the target species, gene flow from beyond the study area may overwhelm patterns of genetic structure mediated by local environmental variables (Anderson et al. 2010). On the other hand, observed genetic discontinuities may also have arisen from past events rather than contemporary landscapes, due to a time lag between demographic processes and their consequences for population genetic structure (Epps and Keyghobadi 2015).

For the world’s largest area of continuous rainforest in the Amazon basin, the main processes responsible for spatial patterns of biodiversity remain debated (Moritz et al. 2000; Hoorn et al. 2010; Ribas et al. 2012; Leite and Rogers 2013). The majority of empirical studies demonstrate that the retraction of past environmental barriers in the Holocene resulted in range expansions of lineages that diverged in isolation up to about 0.8 million years ago (Ma), with major rivers often acting as local biogeographic boundaries (e.g., Naka et al. 2012; Nazarenko et al. 2017; Ribas et al. 2018; Thom et al. 2020). The vast, forested areas between major rivers of the Amazon basin are however also characterized by open lowland rainforest with a mean tree basal area of approximately 19.31 m²/ha, podzolic soils with high clay content, and small temporary rivers filled during the rainy season (Cintra et al. 2013; Schietti et al. 2016). Southwestern and central parts are characterized by open lowland rainforest with a mean tree basal area of 19.31 m²/ha, podzolic soils with high clay content, and small temporary rivers filled during the rainy season (Cintra et al. 2013; Ferreira et al. 2018). Considerable areas of savanna are also present between these 2 forested regions (IBGE 1997; Figure 1).

Materials and Methods

Study Area and Sampling

The PMI is situated south of the Amazon River and covers approximately 15.4 million hectares, with vegetation, soil, and climatic conditions gradually changing along a latitudinal gradient (Cintra et al. 2013; Schietti et al. 2016). The mean annual precipitation varies from 2200 to 2800 mm and is highest in central areas (Alves et al. 2013; Fick and Hijmans 2017). The northeast of the PMI is characterized by dense lowland rainforest with a mean tree basal area of 56.45 m²/ha, plinthosols with a predominance of silt, and a complex hydrography with large seasonally flooded areas (Fan and Miguez-Macho 2010; Schietti et al. 2016). Southwestern and central parts are characterized by open lowland rainforest with a mean tree basal area of 19.31 m²/ha, podzolic soils with high clay content, and small temporary rivers filled during the rainy season (Cintra et al. 2013; Ferreira et al. 2018). Considerable areas of savanna are also present between these 2 forested regions (IBGE 1997; Figure 1).

Between November and March 2010–2015, we collected a total of 66 A. femoralis individuals from 13 localities along an established 880 km transect which runs in parallel to a federal highway (BR-319), and spans the entire length of the PMI (Figure 1; Supplementary Table S1). Sampling was carried on regularly spaced biodiversity monitoring plots (modules) constructed by the Rapid Assessment for Long Duration Ecological Projects (RAPELD) system (for details, see Magnusson et al. 2013). The same sampling design has previously been used to quantify environmental correlates for the occurrence and abundance of A. femoralis (Ferreira et al. 2018).
and revealed that the species is present in all but 3 modules (M3–M5, see Figure 1). Allobates femoralis was sampled by acoustic and visual surveys during the daily periods of peak vocalization (7:00–10:00 a.m. and 14:00–18:00 p.m.). We captured frogs by hand and maintained them in sealed plastic bags until arrival in the laboratory, where they were sacrificed and fixed after tissue (leg muscle) was removed for genetic analyses and stored in 96% ethanol. For each captured individual, the femoral spot coloration was noted as yellow, red, or orange.

DNA Extraction, Genotyping, and Initial Filtering

Extraction of DNA and single-nucleotide polymorphism (SNP) discovery was carried out at Diversity Arrays Technology sequencing Pty. Ltd. (DArTseq) facility (Canberra, Australia; more detail in Supplementary Information Text S1). A modified double-digest restriction-site associated DNA sequencing protocol was performed on libraries prepared using a combination of PstI-HpaII restriction enzymes (Kilian et al. 2012). The PstI enzyme adaptor also contained an Illumina adaptor sequence, a primer sequence and a variable-length barcode as described by Elshire et al. (2011). The HpaII adaptor contained an Illumina flow cell attachment and overhang sequence. Following enzymatic digestion, fragments were amplified and sequenced on an Illumina HiSeq2500. DNA sequences were aligned via BLAST using the Nanorana parkeri reference genome (Sun et al. 2015). To check for contamination, sequences were also blasted to bacterial and fungal genomes (NCBI).

A raw data set of 147 595 SNPs was filtered for missing data using the filter_dart function of the R package RADIATOR v. 0.010 (Gosselin 2017). Only individuals and loci with ≥95% SNPs genotyped were retained. SNPs were also screened for allele coverage, with any SNPs displaying a local and global minor allele frequency threshold of less than 1% removed from the dataset. In cases where multiple SNPs were found within the same read, only one locus was retained (chosen randomly per RAD tag) to avoid statistical bias from physical linkage (Zheng et al. 2012; Lemay and Russello 2015). Two samples from M14 had <95% of loci genotyped and were removed, which resulted in 64 individuals from 13 populations genotyped at 10 275 SNPs (see Supplementary Table S2 for summary of filtering steps). File types required for downstream analyses were created using the RADIATOR package (Gosselin 2017), PGDSpider v. 2.1.1.3 (Lischer and Excoffier 2012), and PLINK v. 1.9 (Chang et al. 2015).

Phylogenomic Relationships

In order to evaluate the evolutionary relationships among A. femoralis possessing different femoral spot coloration, we constructed a population tree by coalescence using SNAPP v. 1.4.1 (Bryant et al. 2012) implemented in BEAST v. 2.5 (Bouckaert et al. 2014). This analysis assumes a lack of gene flow among lineages which is inferred by phenotypic distinctiveness and further tested using clustering analyses. To reduce computational requirements and run times, we selected 2–3 representative individuals per population.
without signatures of between-population admixture (assessed though femoral spot color). We used our data set of 10,275 SNPs, and mutation rates (υ and μ) as estimated by SNAPP, with the birth rate (κ) of the Yule prior based on the number of samples used. The trial run for each dataset used a chain length of 1,000,000 generations, sampling every 1,000 trees. We inspected final log files and created maximum clade credibility trees (median node heights) by combining 3 independent runs in TreeAnnotator v. 2.5 implemented in BEAST after discarding 25% as burn-in.

Detection of SNPs Associated With Selection
We removed SNPs with evidence of being associated with selection because our population and landscape genetic inferences assume neutral loci (see, e.g., Rellstab et al. 2015). Analyses to detect loci associated with selection were conducted on the full dataset using 2 different approaches. First, we detected SNPs under putatively positive or negative selection using \( F_{st} \) outlier analysis with BayeScan v.2.1 (Foll and Gaggiotti 2008), a Bayesian method based on a logistic regression model which is suited to detecting outliers in scenarios with low-admixed samples while taking into account sample size and genetic structure (De Villemereuil et al. 2014; Luu et al. 2017). We ran BayeScan using a prior model (prior odds parameterization) set to 100, thinning interval of 10–20 pilot runs of length 10,000, and burn-in of 50,000 steps. Second, we used environmental association analysis (EAA) with latent factor mixed models (LFMM), implemented in the R package LEA v. 2.1.0 (Frichot and François 2015). LFMM uses a hierarchical Bayesian mixed model based on residuals from principal component analysis (PCA) that take population genetic structure into account (e.g., Benestan et al. 2016). We ran LFMMs for each of the 4 environmental variables, which were previously identified as predictors of local abundance (Ferreira et al. 2018): land cover, silt content, temperature seasonality, and intensity of the dry season, separately using 10,000 iterations, a burn-in of 5000 steps, and 5 repetitions. We set both BayeScan and LFMM with a false discovery rate (FDR) of 0.05 (5%). We also investigated whether the SNPs identified as signaling selection could be attributed to a functional part of the genome in order to complement our tests of the influence of landscape variables on gene flow, as variables influencing connectivity may also impose selection (Armisén et al. 2020). Consequently, gene annotations were sought for RAD tags that contained SNPs identified with both BayeScan and LFMM using the NCBI BLAST platform (Johnson et al. 2008). Sequences were annotated to genes classified as “amphibians” (taxid:8292), “vertebrates” (taxid:7742), and aligned using the Nanorana parkeri (taxid:125878) reference genome (Sun et al. 2015), using BLAST with an E-value threshold of 0.0001. All SNPs that provided evidence for selection were removed from the data set for all downstream analyses of genetic structure. Summary statistics were calculated for each of the modules and any remaining loci that deviated from Hardy–Weinberg equilibrium at a Bonferroni-correction \( α = 0.004 \) (1000 simulations) were also excluded from the dataset. Estimates of observed \( (H_o) \) and expected \( (H_e) \) heterozygosity, inbreeding coefficients \( (F_{IS}) \), and private alleles were calculated using the R package diveRsity v. 1.9.90 (Keenan et al. 2013) with 95% confidence interval calculated with 1000 bootstraps.

Genetic Structure
Genetic structure was described with putatively neutral loci using the model-based clustering approaches implemented by ADMIXTURE (Alexander et al. 2009) and sNMF in the R package LEA v. 2.1.0 (Frichot et al. 2014). To ensure that the underlying genetic structure was not violating the assumptions of these models, we also carried out discriminant analysis of principal components (DAPC) calculated using the R package adegenet v. 2.1.1 (Jombart et al. 2010). Genetic partitioning was further described by calculating pairwise \( F_{st} \) between 11 sites in the R package adegenet v. 1.3.1 (Jombart and Ahmed 2011).

sNMF is a method based on sparse non-negative matrix factorization algorithms (NMF) and least-squares optimization (Frichot et al. 2014). We tested the number of genetic clusters \( (K) \) ranging from 1 to 11 (upper limit equal to the number of sampling localities) with 20 independent runs per test, alpha set at 100, a tolerance error of 0.00001, entropy set as true (where the cross-entropy criterion is calculated), a random seed of 50, and 10,000 interactions in the algorithm. The lowest best-supported \( K \) was determined by the lowest error value of ancestry through the cross-entropy criterion. ADMIXTURE simultaneously estimates the probability of the observed genotypes using ancestry proportions and population allele frequencies (Alexander et al. 2009). Significance was defined at \( P < 0.05 \), above which individuals were considered pure. We ran ADMIXTURE using a cross-validation with a random seed as 43, the block relaxation algorithm as the point estimation method, QuasiNewton as the convergence acceleration algorithm, and a delta of <0.0001 to terminate point estimations. The number of \( K \) was determined by the lowest cross-validation error value. DAPC is a multivariate method that performs discriminant functions to describe the relationships between clusters as well as membership probabilities of each individual for different groups, optimizing variance among groups while minimizing variance within groups (Jombart et al. 2010). We used cross-validation to define the number of principal components (PCs) retained in the analysis, identifying the optimal point in the trade-off between retaining too few and too many PCs in the model. We used the number of PCs associated with the lowest root mean squared error as the optimum number for the PCA in the DAPC analysis. Eight PCs and 2 DAs were retained for the analyses and explained 41% of the total variance. To test whether the number of sampled individuals in each module was sufficient for the inferences of genetic structure, we ran the above analyses with 2 alternative datasets: all individuals sampled and 3 randomly chosen individuals for each module only.

Construction of Environmental Resistance Surfaces
To test the effects of landscape variables on genetic connectivity, we used 4 environmental variables known to influence the occurrence and abundance of A. femoralis along our transect (see Ferreira et al. 2018): land cover, silt content, temperature seasonality (representing the annual range in temperatures), and the Walsh index, a measure of the intensity and duration of the dry season (Walsh 1996). Environmental data were obtained from the public repository Ambdata (www.dpi.inpe.br/Ambdata; Amaral et al. 2013) and converted to raster format using the R package raster v. 2.6.7 (Hijmans 2017) with a cell resolution of 30 arcseconds (1 km²). To avoid model overparameterization, we tested for collinearity between variables through pairwise Pearson’s correlations analyses based on values extracted of each sampling location. The 4 variables were not strongly correlated with each other (\( r < 0.65 \) in all cases) and were therefore retained. To facilitate comparisons among surfaces, we standardized all raster files to values between 1 and 100 (following Row et al. 2017, see Figure 2).
We generated multiple resistance surfaces from our environmental variables to test multiple hypotheses about their effects on genetic distance following Yadav et al. (2019), evaluating each resistance surface model separately. We assumed that resistance in each raster cell was a function of environmental variables as follows:

$$r_i = 1 + \alpha \left( \frac{v_i - 1}{\max - 1} \right)^\gamma,$$

where $$r_i$$ is the resistance of raster cell $$i$$, $$v_i$$ is the environmental variables value in cell $$i$$, and max is the maximum value of the raster surface (in our case 100, see above). Furthermore, $$\alpha$$ is a parameter that determines the maximum possible resistance value, and $$\gamma$$ is an exponent that determines the shape of the relationship (slope) between environmental variable values ($$v_i$$) and resistance ($$r_i$$), being linear when $$\gamma = 1$$ and nonlinear when $$\gamma \neq 1$$ (Shirk et al. 2010; Dudaniec et al. 2013, 2016). This approach has been shown to effectively identify IBR including linear and nonlinear relationships (Shirk et al. 2010; Dudaniec et al. 2013, 2016; Yadav et al. 2019). The equation expresses resistance as a function of the effect of landscape features. Based on previous information (Ferreira et al. 2018), we assume that the effects of land cover and temperature seasonality on resistance are negative and positive, respectively (Figure 3).

We used values of 0, 5, 10, 100, 1000 for intercept ($$\alpha$$), and values of 0.01, 0.1, 0.5, 1, 5, 100 for slope ($$\gamma$$) to create linear and nonlinear resistance surfaces. Models where $$\alpha$$ is equal to zero (7 models for each landscape feature) are identical regardless of $$\gamma$$ values, indicating no influence of resistance on genetic connectivity, which reduced the resistance surfaces for each dataset to 29 unique models. Values of $$\gamma < 1$$ represent resistance surfaces with increased sensitivity, $$\gamma = 1$$ represents a linear resistance relationship and $$\gamma > 1$$ are resistance surfaces with reduced sensitivity (Figure 2). We calculated pairwise resistance distance matrices for all landscape features using circuit theory (McRae et al. 2008; Hanks and Hooten 2013) as implemented in CIRCUITSCAPE v. 4.0.5 (McRae 2006). This approach identifies all possible pathways of movement between focal points across a given raster dataset and calculates average cumulative resistance between all pairwise sampling sites.

Landscape Genetic Resistance Modeling

To evaluate the contribution of landscape features to genetic differentiation, we fitted a maximum-likelihood population-effects (MLPE) mixed-effects model as implemented within the mlpe_rga function using the R package ResistanceGA v. 4.0-4 (Peterman 2018). This model uses individual pairwise metrics for genetic differentiation and landscape resistance, considering each pairwise data
point as an observation. The lack of independence is incorporated as a population-level factor which distinguishes between data points that share a common deme, and those that do not (Clarke et al. 2002; Row et al. 2017). Individual-based pairwise genetic distance was measured as $F_{ST}(1 - F_{IS})$ and used as the dependent variable, resistance distance as the independent variable, and population as the random variable. We fitted the mixed-effects models using parameterization to account for the nonindependence of values within pairwise distance matrices without restricting maximum likelihood (Clarke et al. 2002; Van Strien et al. 2012). Next, to identify which model best described genetic distance among sites, we performed a model selection approach using Akaike information criteria (AICc). We then calculated the difference between the AIC of each model and the minimum AIC value found (Burnham and Anderson 2002; Diniz-Filho et al. 2008) with the lowest change in AICc score ($\Delta$AICc = 0) and the largest AIC weight (wAICc) considered the most parsimonious model. These analyses were performed using the R package ResistanceGA v. 4.0-4 (Peterman 2018), with MLPE models fitted with mlpe_rga using the standard lm() v. 1.1–17 formula interface (Clarke et al. 2002; Bates et al. 2015), magrittr v. 1.5 (Bache and Wickham 2014), and dplyr v. 0.7.4 (Wickham et al. 2017).

**Effects of IBD and IBR on Genetic Differentiation**

We used a Mantel test (Mantel 1967) to estimate the significance of any relationship between pairwise $F_{ST}$ and geographic distance (km) using the function mantel.randtest implemented in the ade4 R package (Dray and Dufour 2007), with 10,000 permutations. The number of geographic distance classes was selected by the Strurges equation, Pearson correlation, and correction of F values through FDR in the R package vegan v. 2.5.1 (Oksanen et al. 2018).

The effect of IBR decoupled from IBD was calculated using distance-based redundancy analysis (dbRDA) using vegan v. 2.5.1 (Oksanen et al. 2018). dbRDA is a direct extension of a multiple regression to model multivariate response data (Legendre and Gallagher 2001; Benestan et al. 2016) and was used to quantify the correlation between the best MLPE model for each landscape variable and $F_{ST}/(1 - F_{ST})$, assuming models with genetic differentiation as the dependent variable and cost distances as independent variables, conditioned on IBD. We obtained statistical significance from each dbRDA model using analyses of variance (ANOVA; 1000 permutations).

To verify that our limited sample size did not affect the MLPE and dbRDA inferences, we subsampled our data with 3 random individuals for each module, recalculated $F_{ST}$ values, and correlated the complete and subsampled $F_{ST}$ matrices against each other. A correlation coefficient of 1.00 suggested that the sample sizes in the analyses provided reliable estimates.

**Results**

$F_{ST}$ Outlier Analysis and Environmental Association Analysis

Outlier analysis with BayeScan detected 174 SNPs with significantly high $F_{ST}$ (2.28%). The analysis with LFMM identified 1281, 912, 859, and 689 SNPs associated with land cover, the Walsh index, silt content, and temperature seasonality, respectively. Of these, 43 SNPs were associated with each of the 4 environmental variables (Supplementary Figure S1). Twenty-three outliers were in common for the BayeScan and LFMM analyses, none of which resulted in significant matches to either the *N. parkeri* genome or during BLAST searches using GenBank.

We removed the 23 loci in consensus between EAA and outlier approaches to produce an approximately neutral data set for population and landscape genetic analyses. Preliminary analyses indicated that inclusion or exclusion of these loci deviating from neutral expectations made no detectable difference to the results. Because of the strong genetic divergence of modules 1 and 2 from the remaining modules (see SNAPP analysis below), these 2 modules were excluded from the landscape genetic analyses to allow for subtle environmental influences on genetic structure to be detected. With the exclusion of the SNPs with signatures of selection and data from M1 and M2, a total of 7609 SNPs were available for analysis. Summary statistics for modules M6–M14 are provided in Table 1.

**Population Tree**

The population tree constructed with SNAPP showed that individuals from the northern modules M1 and M2 (yellow femoral spot) was measured as $ST/(1 − 2O)$, expected heterozygosity ($E_o$), inbreeding coefficient ($F_{IS}$) and their low and high values (95%), number of private alleles (PA), and probability of deviating from Hardy–Weinberg equilibrium (HWE) are provided.

<table>
<thead>
<tr>
<th>Site</th>
<th>$N_{TOTAL}$</th>
<th>$H_o$</th>
<th>$H_e$</th>
<th>$F_{IS}$</th>
<th>$F_{IS}$ Low</th>
<th>$F_{IS}$ High</th>
<th>PA</th>
<th>HWE</th>
</tr>
</thead>
<tbody>
<tr>
<td>M6</td>
<td>5</td>
<td>0.08</td>
<td>0.09</td>
<td>0.0434</td>
<td>−0.0341</td>
<td>0.0827</td>
<td>234</td>
<td>1</td>
</tr>
<tr>
<td>M7</td>
<td>5</td>
<td>0.1</td>
<td>0.1</td>
<td>0.0126</td>
<td>−0.0756</td>
<td>0.0719</td>
<td>434</td>
<td>1</td>
</tr>
<tr>
<td>M8</td>
<td>4</td>
<td>0.12</td>
<td>0.12</td>
<td>0.0389</td>
<td>−0.1102</td>
<td>0.1448</td>
<td>170</td>
<td>1</td>
</tr>
<tr>
<td>B9_9</td>
<td>7</td>
<td>0.11</td>
<td>0.13</td>
<td>0.1672</td>
<td>0.0524</td>
<td>0.2390</td>
<td>258</td>
<td>1</td>
</tr>
<tr>
<td>M9</td>
<td>5</td>
<td>0.11</td>
<td>0.13</td>
<td>0.0023</td>
<td>−0.0517</td>
<td>0.2236</td>
<td>170</td>
<td>1</td>
</tr>
<tr>
<td>B9_10</td>
<td>3</td>
<td>0.11</td>
<td>0.11</td>
<td>0.0023</td>
<td>−0.0530</td>
<td>0.0985</td>
<td>200</td>
<td>1</td>
</tr>
<tr>
<td>M10</td>
<td>3</td>
<td>0.11</td>
<td>0.11</td>
<td>−0.0151</td>
<td>−0.0719</td>
<td>0.0770</td>
<td>189</td>
<td>1</td>
</tr>
<tr>
<td>M11</td>
<td>6</td>
<td>0.1</td>
<td>0.11</td>
<td>0.1244</td>
<td>−0.0025</td>
<td>0.1777</td>
<td>118</td>
<td>1</td>
</tr>
<tr>
<td>M12</td>
<td>6</td>
<td>0.09</td>
<td>0.1</td>
<td>0.0386</td>
<td>−0.0846</td>
<td>0.1041</td>
<td>119</td>
<td>1</td>
</tr>
<tr>
<td>M13</td>
<td>6</td>
<td>0.09</td>
<td>0.09</td>
<td>0.0090</td>
<td>−0.1280</td>
<td>0.0742</td>
<td>275</td>
<td>1</td>
</tr>
<tr>
<td>M14</td>
<td>4</td>
<td>0.08</td>
<td>0.08</td>
<td>−0.077</td>
<td>−0.2999</td>
<td>0.0515</td>
<td>416</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Site</th>
<th>$N_{TOTAL}$</th>
<th>$H_o$</th>
<th>$H_e$</th>
<th>$F_{IS}$</th>
<th>$F_{IS}$ Low</th>
<th>$F_{IS}$ High</th>
<th>PA</th>
<th>HWE</th>
</tr>
</thead>
<tbody>
<tr>
<td>M6</td>
<td>5</td>
<td>0.08</td>
<td>0.09</td>
<td>0.0434</td>
<td>−0.0341</td>
<td>0.0827</td>
<td>234</td>
<td>1</td>
</tr>
<tr>
<td>M7</td>
<td>5</td>
<td>0.1</td>
<td>0.1</td>
<td>0.0126</td>
<td>−0.0756</td>
<td>0.0719</td>
<td>434</td>
<td>1</td>
</tr>
<tr>
<td>M8</td>
<td>4</td>
<td>0.12</td>
<td>0.12</td>
<td>0.0389</td>
<td>−0.1102</td>
<td>0.1448</td>
<td>170</td>
<td>1</td>
</tr>
<tr>
<td>B9_9</td>
<td>7</td>
<td>0.11</td>
<td>0.13</td>
<td>0.1672</td>
<td>0.0524</td>
<td>0.2390</td>
<td>258</td>
<td>1</td>
</tr>
<tr>
<td>M9</td>
<td>5</td>
<td>0.11</td>
<td>0.13</td>
<td>0.0023</td>
<td>−0.0517</td>
<td>0.2236</td>
<td>170</td>
<td>1</td>
</tr>
<tr>
<td>B9_10</td>
<td>3</td>
<td>0.11</td>
<td>0.11</td>
<td>0.0023</td>
<td>−0.0530</td>
<td>0.0985</td>
<td>200</td>
<td>1</td>
</tr>
<tr>
<td>M10</td>
<td>3</td>
<td>0.11</td>
<td>0.11</td>
<td>−0.0151</td>
<td>−0.0719</td>
<td>0.0770</td>
<td>189</td>
<td>1</td>
</tr>
<tr>
<td>M11</td>
<td>6</td>
<td>0.1</td>
<td>0.11</td>
<td>0.1244</td>
<td>−0.0025</td>
<td>0.1777</td>
<td>118</td>
<td>1</td>
</tr>
<tr>
<td>M12</td>
<td>6</td>
<td>0.09</td>
<td>0.1</td>
<td>0.0386</td>
<td>−0.0846</td>
<td>0.1041</td>
<td>119</td>
<td>1</td>
</tr>
<tr>
<td>M13</td>
<td>6</td>
<td>0.09</td>
<td>0.09</td>
<td>0.0090</td>
<td>−0.1280</td>
<td>0.0742</td>
<td>275</td>
<td>1</td>
</tr>
<tr>
<td>M14</td>
<td>4</td>
<td>0.08</td>
<td>0.08</td>
<td>−0.077</td>
<td>−0.2999</td>
<td>0.0515</td>
<td>416</td>
<td>1</td>
</tr>
</tbody>
</table>
belong to a strongly divergent lineage (Figure 4, Supplementary Figure S2), consistent with the relatively high pairwise \( F_{ST} \) values found between M1 or M2 and the other localities (\( F_{ST} \) range 0.72–0.83). The remaining modules were split into 3 markedly shallower but distinct individual clades (posterior probability = 1.00 in all cases), with Cluster C formed by the most distal node (Figure 4).

Corresponding with the genetic lineages identified using SNAPP, the population genetic inferences with ADMIXTURE, sNMF and DAPC produced a congruent result of 3 inferred genetic clusters from Module 6 onward (Figure 5, see also Figure S3). The first Cluster A comprised 14 individuals with red femoral spots across modules M6–M14 in dense forest. It was distinct from a second Cluster B, which comprised 24 individuals from 5 populations (BM8_M9 - M11) across dense and open forest. This cluster largely comprised individuals with yellow femoral spots, with the exception of population BM8_9 with an intermediate (orange) coloration and evidence of genetic admixture (Figure 5). A third cluster (C, characterized by red femoral spots) was confined to 16 individuals from the eastern bank of the upper Madeira river (M12 to M14), an open forest area separated from the remainder of the transect by patches of savannah. Reducing the dataset to 3 individuals for all modules did not alter the genetic partitioning revealed by each of the 3 clustering methods, demonstrating that the sampling regime was sufficient to resolve genetic structure (Supplementary Figure S4).

**IBD and IBR**

Pairwise genetic distances \( (F_{ST}) \) across modules M6–M14 ranged from 0.020 (M13 and M14) to 0.207 (M6 and M14; Table 2), with a strong association between genetic and geographic distances and therefore IBD (Mantel test: \( P < 0.0001, r^2 = 0.96, \) Figure 6). The Mantel corrollograms calculated for 7 classes of geographic distance revealed spatial autocorrelation in 4 cases: positively at geographic distances to 60 km \( (r = 0.67, P < 0.001) \) and 143 km \( (r = 0.24, P = 0.02) \) and negatively at distances of 476 km \( (r = -0.61, P = 0.03) \) and 560 km \( (r = -0.61, P < 0.001; \) Supplementary Figure S5).

Our MLPE analysis showed that a land cover model with \( \alpha = 5 \) and \( \gamma = 10 \) explained 98% of the genetic variation (Table 3). The Walsh index explained 96% of the genetic variation at \( \alpha = 100 \) and \( \gamma = 5 \), and temperature seasonality and silt content explained 95% of the genetic variation each, at \( \alpha = 10 \) and 1000, and at \( \gamma = 5 \) and 1, respectively (Table 3). The \( \alpha \) values determine the maximum resistance of the variables (e.g., in the case of Walsh index, \( \alpha = 100 \) suggests that landscape resistance to gene flow is 100 times greater than zero), and the \( \gamma \) values indicate whether the variable influenced genetic connectivity linearly or non-linearly. Silt presented a value of \( \gamma = 1 \), suggesting a linear resistance relationship. All other confidence sets of resistance surfaces presented values \( \gamma > 1 \), supporting resistance surfaces with reduced resistance sensitivity. ΔAIC values were identical for the 4 landscape features (0.00), supporting the maximum-likelihood models. In the dbRDA models, the Walsh index captured 8.4% of the observed genetic variation \( (F_{ST} = 41.72, P = 0.001) \), followed by land cover (6.4%; \( F_{ST} = 26.85, P = 0.001 \)), temperature seasonality (5.3%; \( F_{ST} = 20.54, P = 0.001 \)) and silt content (3.5%; \( F_{ST} = 11.79, P = 0.001; \) Table 3; Supplementary Figure S6).

**Discussion**

We characterized patterns of genetic structure and femoral spot coloration for the brilliant-thighed poison frog *A. femoralis* that was sampled along an 880 km transect through continuous rainforest in a major Amazonian interflue. We revealed 4 genetically distinct clusters, one derived from a deep lineage divergence, and each cluster corresponding with a consistent femoral spot coloration which differed from the color possessed by individuals from adjacent clusters. Transitions between major forest types were not consistently associated with the boundaries of genetic clusters. Genetic variation was characterized by a pattern of IBD across hundreds of kilometers, and subtle but significant effects of contemporary landscape features on the distribution of individual measures of genetic variation.

Under a pronounced pattern of IBD, as is the case for our study system, genetic clustering algorithms can overestimate the number of partitions or lead to misleading admixture inferences (Frantz et al. 2009; Garcia-Erill and Albrechtsen 2020). We nevertheless argue that the clusters identified along our *A. femoralis* transect represent biologically meaningful entities, as they were identified through 4 independent approaches and conform to variation observed with a phenotypic trait (femoral spot coloration). Although precise time calibrations are beyond the scope of the present study, the phenotypic differences suggest that the clusters have arisen from past rather than contemporary phenomena, reflecting the "time lag problem" of landscape genetic inferences (see, e.g., Epps and Keyghobadi 2015). That the DAPC approach failed to identify the zone of admixture is...
expected because it does not assess differential ancestry proportions for each individual (see also Miller et al. 2020).

Possible taxonomic implications of the deeply diverged population of *A. femoralis* from the northeast of the PMI (localities M1 and M2) will require further work. Timing the divergence is needed to evaluate the role of historical processes in isolating these localities from the remainder of the PMI. The northeast of the PMI is well drained, of young sedimentary origin (Late Pleistocene-Early Holocene, see, e.g., Sombroek 2001) and due to the proximity to the Amazon river subject to rapid changes in topography and hydrology that might have resulted in periods of isolation (Hoorn et al. 2010; Latrubesse et al. 2010; Pupim et al. 2019). At present, the populations from M1 and M2 are separated from the remainder of the transect by approximately 150 km of lowland dense forest unoccupied by *A. femoralis* (Ferreira et al. 2018). Isolation by unsuitable habitat is also suggested for cluster C (M12–M14, red femoral spots), which is separated from the remainder of the modules by secondary vegetation, including intervening savannah over about 150 km, an ecological barrier that is likely to have been further strengthened during the glacial periods in the late Pleistocene (Cohen et al. 2014).

In contrast to the association of Clusters A and C, the area of contact between Clusters A and B (M8–M9) does not occur at the location of a current ecotone. This implies that the divergence of Clusters A and B might be linked to a barrier, which is no longer present. Our finding for *A. femoralis* contrasts with recent data on the genetic structure of a treefrog (Ortiz et al. 2018) and with plumage coloration in birds (De Abreu et al. 2018) along the same transect, which both reveal a zone of divergence spatially matching with the ecotone between open and closed forest (M10 and M11). For these species, it was concluded that present day environmental differences were responsible for the genetic partitioning.

Individuals in Cluster A possess different femoral spot coloration (red) from those in Cluster B (yellow), except in a relatively narrow (~100 km) zone of admixture where individuals possess orange femoral spots. This color transition mirrors a well-studied model hybrid zone system between the European red (fire)-bellied toad *Bombina bombina* and the yellow-bellied toad *B. variegata*, that form orange-bellied hybrids in parapatry (e.g., Szymura and Barton 1986, 1991). In this system, spatial separation through differential habitat preferences leads to a narrow zone of admixture despite the lack of pronounced postzygotic mating barriers (Vines et al. 2003). Another mechanism that can lead to narrow zones of admixture is sexual selection, and assortative mating in accordance with red or yellow femoral spot coloration has been demonstrated with *A. femoralis*. 

Figure 4. A population tree generated using SNAPP, and a histogram showing individual ancestry proportions, estimated using ADMIXTURE. The location of the collection modules are color coded to reflect the color assigned to each genetic cluster in the ADMIXTURE plot (the white circles for M3–M5 indicate the absence of *Allobates femoralis*). Posterior probabilities obtained at each node are shown on the tree. Cluster 1 corresponds to individuals with yellow femoral spots, Cluster A corresponds to individuals with red femoral spots, Cluster B corresponds to individuals with yellow femoral spots, with a zone of admixture between Cluster A–B (BM8-9) with an intermediate color phenotype (orange), and Cluster C corresponds to individuals with red femoral spot. See online version for full colors.
mate choice experiments (Ferreira et al., unpublished data). In addition, femoral spot coloration in *A. femoralis* spatially varies in association with mimicry with syntopic toxic species (Amézquita et al. 2017), and evaluating locally co-occurring taxa to investigate such relationships may help shed light on the mechanisms underpinning the distribution of color variation at this locality.

Spatially structured transitions of coloration across an area of genetic admixture could serve as a mechanism to generate new phenotypes (Stelkens and Seehausen 2009; Sefc et al. 2017). In other poison dart frogs, hybridization has been shown to result in independent aposematic lineages and novel color morphs (Medina et al. 2013; Ebersbach et al. 2020). Examining the evolutionary history of admixed individuals with color variation across the wider distribution of *A. femoralis* in the Amazon basin will help establish the role of hybridization in generating this polymorphism. In addition, testing for assortative mating particularly for individuals possessing the orange phenotype and conditions allowing disassortative mating (e.g., low mate availability; Medina et al. 2013) will contribute toward a better understanding of the isolating processes involved.

Although contemporary environmental variation was not consistently associated with the 4 distinct genetic clusters we have described, genetic connectivity still varies with environmental...
conditions. Environmental variables have been shown to influence gene flow in other anurans. For example, IBR contributed an additional 10–20% in variation to models governed by IBD for the European common frog *Rana temporaria* (Van Buskirk and Jansen van Rensburg 2020). Given that this study was conducted in rugged, alpine terrain, the magnitude of these values are consistent with the environmental influence that we measured in a more gradually varying environment. We found that the influence of land cover was strongly supported by our MLPE models, confirming previous evidence that dense forest flooded by streams and overflowing rivers are not favorable habitats for *A. femoralis* (Ferreira et al. 2018).

Our dbRDA results showed that the Walsh index was also associated with less connectivity. A possible explanation is that rainfall strongly determines the existence and persistence of water-filled ditches on the forest floor, a requirement for reproduction for many amphibians including *A. femoralis* (Menin et al. 2011; Ringler et al. 2015). Rainfall gradients and the 2 dominant forest phytophysiognomies in the PMI are autocorrelated, which likely explains the inconsistency with the highest ranking variable resolved with the MLPE and dbRDA results (forest cover vs. Walsh index, respectively). Open forests in the drier, southwestern areas of the PMI are more seasonal and have lower stem densities and higher tree mass compared with wetter, dense forest at northeastern parts (Sombroek 2001; Cintra et al. 2013; Schietti et al. 2016).

Environmental variation also appears to impose different selective pressures along the PMI, with environmental association analyses showing the largest number of SNP loci associated with the Walsh index and forest cover. Further work with greater SNP densities and a reference genome will contribute toward the identification of genes under selection. Nonetheless, our existing results suggest that both the levels of connectivity and differences in fitness associated with environmental variation may contribute to the observed fine-scale patterns of genetic variation. We reduced the risk of false positives in such inferences (see Hoban et al. 2016; Ahrens et al. 2018) by considering only those loci which were identified by both BayeScan and LFMM.

Although strong IBD and environmental-based selection are conditions that may lead to divergence in accordance with the gradient diversification hypothesis (Endler 1977), our data also suggest a role of historical processes in the generation of the patterns of genetic divergence we describe for *A. femoralis*. In particular, the relatively rapid restructuring of the Amazon region may give rise to conditions where historical isolation and processes associated with secondary contact reduce the potential for environmental gradients to strongly influence genetic and phenotypic variation. For example, reinforcement by the development of reproductive character displacement could potentially be a stronger influence on gene flow than the effects of environmental gradients (Rojas et al. 2019). Accumulating genetic data from additional species using the standardized sampling system along the PMI provides a unique opportunity to look for traits (e.g., variation in mating cues) that predict whether current environmental transitions or mechanisms associated with past landscapes generate diversity in areas of continuous habitat.

**Table 3.** Summary of model selection using MLPE and dbRDA that evaluated the effects of IBR on genetic distance (log(*F*ₐₛ/1 − *F*ₛₐ))

<table>
<thead>
<tr>
<th>Variables</th>
<th>α</th>
<th>γ</th>
<th>AICc</th>
<th>ΔAIC</th>
<th>r²</th>
<th>SE</th>
<th>r-value</th>
<th>r²</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Land cover</td>
<td>5</td>
<td>10</td>
<td>36.66</td>
<td>0.00</td>
<td>0.98</td>
<td>0.0800</td>
<td>21.223</td>
<td>0.064</td>
<td>26.85</td>
<td>0.001</td>
</tr>
<tr>
<td>Walsh index</td>
<td>100</td>
<td>5</td>
<td>51.81</td>
<td>0.00</td>
<td>0.96</td>
<td>0.0707</td>
<td>17.025</td>
<td>0.084</td>
<td>41.72</td>
<td>0.001</td>
</tr>
<tr>
<td>Temperature seasonality</td>
<td>10</td>
<td>5</td>
<td>52.73</td>
<td>0.00</td>
<td>0.95</td>
<td>0.0903</td>
<td>16.053</td>
<td>0.053</td>
<td>20.54</td>
<td>0.001</td>
</tr>
<tr>
<td>Silt content</td>
<td>1000</td>
<td>1</td>
<td>50.69</td>
<td>0.00</td>
<td>0.95</td>
<td>0.0887</td>
<td>16.557</td>
<td>0.035</td>
<td>11.79</td>
<td>0.001</td>
</tr>
</tbody>
</table>

For MLPE, the AIC, r² value, standard error (SE), and the parameter combination (α and γ) are given for the best models for each landscape variable. For dbRDA, the magnitude of difference is given by the r-value, and the F and P values were obtained by ANOVA. Bolded P values show significant effects of IBR on genetic distance.
Supplementary Material

Supplementary data are available at Journal of Heredity online.

Funding

This study was funded by the Brazilian Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) and Fundação de Amparo à Pesquisa do Estado do Amazonas (FAPEAM) under grants conceded to A.P.L. (CNPq; Programa Ciência sem Fronteiras process 401327/2012-4; FAPEAM/CNPq/PRONEX process 586/10, Edital 003/2009—number 137) and A.S.F. received a PhD scholarship by Fôs-Doutorado Júnior (PDJ) (number 161883/2014-1) and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) through Programa de Doutorado-Sanduíche no Exterior (PDSE) (Finance Code 001, number 88881.133683/2016-01).

Acknowledgments

Specimens were collected under permit numbers 13.777 and 7836-1 (A.P.L.) issued by the Instituto Chico Mendes de Conservação da Biodiversidade—ICMBio of the Ministry of Environment, Government of Brazil. We thank the Instituto Nacional de Pesquisas da Amazônia (INPA), the Programa de Pesquisas em Biodiversidade (PPBio/RAPELD), the Centro Integrado de Estudos da Biodiversidade Amazônica (INCT—CENBAM), and Santo Antônio Energia S.A. for logistical and institutional support. We thank Chris Barratt for providing R scripts and Sonu Yadav and Alex Carey for helping with R scripts. We are grateful to all members of the Conservation Genetics Lab of Macquarie University for valuable discussion during early stages of this study. We are very grateful to the field assistants for their help in collecting *Allobates femoralis* over the years and to Bill Magnussone and Antoine Fouquet for providing comments on the manuscript.

Data Availability

In accordance with the *Journal of Heredity* data archiving policy, we have submitted all the data and R scripts to Dryad: https://doi.org/10.5061/dryad.5hqbzkhh4.4.

References


