Divergence, gene flow and the origin of leapfrog geographic distributions: The history of color pattern variation in *Phyllobates* poison-dart frogs

Running Head: Poison frog leapfrog distribution

Roberto Márquez[1,2,*], Tyler P. Linderoth[3,†], Daniel Mejía-Vargas[2], Rasmus Nielsen[3,4,5], Marcus R. Kronforst[1,‡], Adolfo Amézquita[2,‡]

[1]Department of Ecology and Evolution, University of Chicago. Chicago, IL. 60637, USA.
[2]Department of Biological Sciences, Universidad de los Andes. A.A. 4976, Bogotá, D.C., Colombia.
[3]Department of Integrative Biology and Museum of Vertebrate Zoology, University of California, Berkeley. Berkeley, CA. 94720, USA.
[4]Department of Statistics, University of California, Berkeley. Berkeley, CA. 94720, USA.
[5]Center for GeoGenetics, University of Copenhagen, Copenhagen 1350, Denmark.

[*]**Corresponding author.** Department of Ecology and Evolution, University of Chicago. 1101 East 57th St. Zoology 206. Chicago, IL. 60637. USA. Ph. 312-709-8658. Email: rmarquezp@uchicago.edu.
[†]Current address: Department of Genetics, University of Cambridge, Downing Street, Cambridge CB2 3EH, UK.
[‡] Joint senior authors.

Submitted for consideration as an Original Article.

Abstract

The geographic distribution of phenotypic variation among closely related populations is a valuable source of information about the evolutionary processes that generate and maintain biodiversity. Leapfrog distributions, in which phenotypically similar populations are disjunctly distributed and separated by phenotypically distinct populations, represent geographic replicates for the existence of a phenotype, and are therefore especially informative. These geographic patterns have mostly been studied from phylogenetic perspectives to understand how common ancestry and divergent evolution drive their formation. Other processes, such as gene flow between populations, have not received as much attention. Here we investigate the roles of divergence and gene flow between populations in the origin and maintenance of a leapfrog distribution in *Phyllobates* poison frogs. We found evidence for high levels of gene flow between neighboring populations but not over long distances, indicating that gene flow between central populations may have a homogenizing effect that maintains their similarity, and that introgression between "leapfroging" taxa has not played a prominent role as a driver of phenotypic diversity in *Phyllobates*. Although phylogenetic analyses suggest that the leapfrog distribution was formed through independent evolution of the peripheral (i.e. leapfrogging) populations, the elevated levels of gene flow between geographically close populations poise alternative scenarios, such as the history of phenotypic change becoming decoupled from genome-averaged patterns of divergence, which we cannot rule out. These results highlight the importance of incorporating gene flow between populations into the

study of geographic variation in phenotypes, both as a driver of phenotypic diversity and as a confounding factor of phylogeographic inferences.

Key Words: Phylogeography, spatial population genetics, convergent evolution, Dendrobatidae

Introduction

Geography has a strong influence on the diversification of closely related lineages, since it largely mediates the level of gene flow between them (Huxley, 1942; Mayr, 1942). Therefore, studying the geographic distribution of phenotypic and genetic variation among such lineages can generate valuable insights into the processes that generate biological diversity. An intriguing pattern of geographic variation is the "leapfrog" distribution, where phenotypically similar, closely related populations (of the same or recently diverged species) are disjunctly distributed and separated by phenotypically different lineages to which they are also closely related (Chapman, 1923; Remsen, 1984). Such patterns have been reported in multiple taxa, such as birds (e.g. Cadena, Cheviron, & Funk, 2010; Chapman, 1923; Norman, Christidis, Joseph, Slikas, & Alpers, 2002; Remsen, 1984), flowering plants (Shun-ichi Matsumura, Yokoyama, Tateishi, & Maki, 2006; Shun'Ichi Matsumura, Yokohama, Fukuda, & Maki, 2009), and butterflies (Brower, 1996; Emsley, 1965; Hovanitz, 1940; Sheppard, Turner, Brown, Benson, & Singer, 1985). Since leapfrog patterns represent repeated instances of similar phenotypes in space, they provide a rich opportunity to understand the processes generating geographic variation in phenotypes.

Two main hypotheses have been put forward to explain the origin of leapfrog distributions (Norman et al., 2002; Remsen, 1984): First, the phenotypically similar, geographically disjunct populations can owe their resemblance to recent common ancestry (i.e. they are descendants of an ancestral population with the same

phenotype), and the disjunct range of "leapfrogging" forms is due to biogeographic processes such as long-range migration or the extinction of geographically intermediate populations. Second, the distribution of phenotypes may be due to evolutionary convergence of populations with similar phenotypes, or divergence of the central (intervening) populations from the ancestral phenotype. Clear phylogenetic predictions can be drawn from these hypotheses: If phenotypic similarity among the leapfrogging populations is due solely to recent common ancestry, then such populations should be more closely related to one another than to geographically close populations that display the intervening phenotype. If the geographic distribution of phenotypes is due to convergent or divergent evolution then a correspondence between phylogeny and phenotypes is not expected. In this case, however, ancestral state reconstructions can identify whether the central or peripheral populations exhibit derived (i.e. divergent) phenotypes. Accordingly, efforts to elucidate the evolutionary mechanisms behind leapfrog distributions have mainly focused on inferring the phylogenetic relationships among populations and using them to reconstruct the evolution of the phenotype in question (e.g. Brower, 1996; Cadena et al., 2010; Shun'Ichi Matsumura et al., 2009; Norman et al., 2002; Quek et al., 2010).

Although a cladogenetic description of population history can reveal a great deal about the origin of leapfrog distributions, it is unable to capture some important aspects of the diversification process. Among them is the extent of historical and contemporary gene flow between populations (or its absence), which can play an important role in the formation of leapfrog distributions. For instance, reduced levels

of genetic exchange between populations with different phenotypes will promote the existence of such differences, while introgressive hybridization between populations can facilitate phenotypic convergence between them. Furthermore, if gene flow between geographically close populations with different phenotypes is pervasive, it can homogenize previous genetic divergence between these populations, decoupling the history of the phenotype from genome-wide patterns of divergence (Hines et al., 2011; James, Arenas-Castro, Groh, Engelstaedter, & Ortiz-Barrientos, 2020), which can complicate inferences related to the origin of leapfrog distributions.

Here we examine the processes driving the origin of a leapfrog distribution present in *Phyllobates* poison-dart frogs. This genus is found from Southern Nicaragua to Western Colombia, and is composed of five nominal species (Myers, Daly, & Malkin, 1978; Silverstone, 1976): *P. vittatus*, *P. lugubris*, and *P. aurotaenia,* which exhibit a bright dorsolateral stripe on a dark background, and *P. terribilis* and *P. bicolor*, which display solid bright-yellow dorsal coloration (Fig. 1A; note that at a handful of localities *P. terribilis* exhibit solid mint-green color patterns). The two latter species exhibit a leapfrog distribution in Western Colombia, separated by *P. aurotaenia*: *P. bicolor* occurs on the slopes of the Western Andes, in the upper San Juan river basin, *P. aurotaenia* in the lowlands along the San Juan and Atrato Drainages and onto the Pacific coast, and *P. terribilis* along the Pacific coast south of the San Juan's mouth (Fig. 1C).

Early systematic work grouped *P. terribilis* and *bicolor* as sister species based on morphological and ontogenetic characters (Maxson & Myers, 1985; Myers et al., 1978). Although an early mitochondrial phylogeny supported these relationships (Widmer, Lötters, & Jungfer, 2000), subsequent work has consistently recovered *P. terribilis* and *P. bicolor* as non-sister taxa (Grant et al., 2006, 2017; Márquez, Corredor, Galvis, Góez, & Amézquita, 2012; Santos et al., 2009), and even suggested that *P. aurotaenia* may actually represent two distinct lineages, one sister to *P. bicolor* and the other to *P. terribilis* (Grant et al., 2017; Santos et al., 2009). Although these studies only included 1-4 samples per *Phyllobates* species, and were based on DNA sequences from a small number of markers (1-7 loci), their results are compatible with convergent evolution giving rise to the leapfrog distribution.

In this study we aim to shed light on the demographic processes involved in the origin of the current geographic distribution of aposematic coloration in *Phyllobates* poison frogs. Based on substantially increased sampling across Colombian populations and thousands of genome-wide markers, we leverage phylogenetics and spatial population genetics to 1) elucidate the extent of genetic structure and evolutionary relationships among populations, and 2) evaluate the role of gene flow among populations in the formation of the leapfrog distribution.

Materials and Methods

To obtain a representative sample of Colombian *Phyllobates* populations, we conducted field expeditions to 23 localities throughout the genus's range (Fig. 2B), resulting in tissue (i.e. mouth swab, toe-clip, or liver) samples from 108 individuals (Table S1). In addition, we obtained eight samples of *P. vittatus* and *P. lugubris*, (four samples per species; Table S1) to serve as outgroups in our analyses. Both species are distributed in Central America, and have been consistently found to be the sister group of Colombian *Phyllobates* (Grant et al., 2006, 2017; Santos et al., 2009; Widmer et al., 2000).

mtDNA Sequencing and analysis

To gain initial insight into the levels of genetic variation and structure among populations we sequenced fragments of three mtDNA markers: 16S rRNA (16S; 569bp), Cytochrome Oxidase I (COI barcoding fragment; 658bp), and Cytochrome b (Cytb; 699bp) from 74 individuals. We extracted DNA using either Qiagen DNeasy spin columns or a salt precipitation protocol (Miller, Dykes, & Polesky, 1988), and used primers 16Sar and 16Sbr (Palumbi et al., 1991), Chmf4 and Chmr4 (Che et al., 2012), and CytbDen3-L and CytbDen1-H (Santos & Cannatella, 2011) to amplify the 16S, COI, and Cytb loci, respectively. Thermal cycling protocols consisted of 2 min at 95ºC, 30-35 cycles of 30 sec at 95ºC, 1 min at 45ºC and 1.5 min at 72ºC, and a final 5 min at 72ºC. PCR products were purified with ExoSAP (Affymetrix) and sequenced in both directions using an ABI 3500 Genetic Analyzer (Applied Biosystems).

Chromatograms were assembled and visually inspected in Geneious R9 (Kearse et al., 2012) to produce finalized consensus sequences.

We aligned our sequences and those available in GenBank (Table S1) using MUSCLE (Edgar, 2004), and built mtDNA trees with PhyML 3.3 (Guindon et al., 2010) and MrBayes 3.2.6 (Altekar, Dwarkadas, Huelsenbeck, & Ronquist, 2004; Ronquist et al., 2012). MrBayes analyses consisted of 10 million iterations (two runs with four chains each), sampling every 1,000 iterations, and discarding the first 2,500 trees (25%) as burnin. PhyML runs started from five different random trees, and used SPR moves to search the tree space. Nodal support was evaluated using aBayes scores (Anisimova, Gil, Dufayard, Dessimoz, & Gascuel, 2011). To obtain an estimate of divergence times between mtDNA haplotypes, we inferred a time-calibrated tree using BEAST v. 2.5.0. (Bouckaert et al., 2019). Based on results from previous work (Santos et al., 2014), we set a log-normal prior with mean 8.13 million years (MY) and standard deviation 1.2 MY (i.e. log(mean) = 2.12, log(s.d.) = 0.1) for the root age of *Phyllobates*. We used a Calibrated Yule tree prior, and set default priors for all other parameters, except for the clock rate mean and the Yule birth rate, which were set to *gamma*(0.01, 1000). We ran the MCMC sampler for 100 million iterations, sampling every 10,000, and generated a maximum clade credibility (MCC) tree using Tree Annotator (distributed with BEAST) after discarding the first 5% of trees as burnin. Mixing and stationarity of BEAST and MrBayes runs were evaluated visually and based on effective sample sizes (ESS) using Tracer v. 1.5 (Rambaut & Drummond, 2009). All mtDNA analyses were performed under partitioning schemes

and molecular evolution models chosen with PartitionFinder2 (Lanfear, Frandsen, Wright, Senfeld, & Calcott, 2017).

Transcriptome-enabled exon capture

Based on the results of mtDNA analyses we chose 63 samples (60 ingroup, 3 outgroup) from 17 localities (Table S1) representing the range of observed mtDNA variation among Colombian populations, and used them to perform transcriptome-enabled exon capture (Bi et al., 2012; Hodges et al., 2007). Briefly, we designed a set of DNA capture probes based on a transcriptome assembly and used them to enrich standard Illumina libraries for a subset of the genome.

Transcriptome sequencing. We generated a transcriptome assembly from liver, muscle, skin, and heart tissue of a single *P. bicolor* juvenile. RNA was extracted using Qiagen RNeasy spin columns, and pooled in equimolar ratios by tissue type to build a single cDNA library, which was sequenced on an Illumina HiSeq 2000. We filtered and trimmed reads using Trimmomatic v. 0.25 (Bolger, Lohse, & Usadel, 2014), and used Trinity (release 2013-02-25; Grabherr et al., 2011) to assemble them under default parameters, except for the minimum contig length, which was increased to 250bp. Finally we collapsed redundant contigs (e.g. alternative isoforoms) with CD-HIT-EST V.4.5.3 (Fu, Niu, Zhu, Wu, & Li, 2012).

Enrichment probe design. We annotated our transcriptome using BLASTX (Altschul, 1997) against *Xenopus tropicalis* proteins (JGI 4.2.72), and used Exonerate (Slater & Birney, 2005) to identify intron-exon boundaries in order to split transcripts into individual exons. We then chose a final set of exons to enrich in the following way: First we discarded those under 100bp, with GC content below 40% and above 70%, or which overlapped by more than 10bp based on Exonerate annotations. Next, we identified putatively repetitive elements and RNA-coding genes (e.g. rRNAs) in our transcriptome assembly with RepeatMasker v. 4.0 (Smith, Hubley, & Green, 2013) and BLASTn, respectively, and removed exons overlapping them. Finally, we blasted our exon set against itself with BLASTn under default parameters, and whenever two or more exons matched each other (e-value < 10^{-10}), we retained only one of them. This resulted in 38,888 exons (7.57Mb) that passed filters, which were used to design 1,943,120 100bp probes that were tiled at 3bp and printed on two Agilent SureSelect custom 1M-feature microarrays.

DNA library preparation, target enrichment, and sequencing. We extracted DNA as described above, and used a Diagenode Bioruptor to shear each extraction to a ~100-500bp fragment distribution by performing 3-4 rounds of sonication (7min of 30s on/off cycles per round). DNA libraries were built following Meyer & Kircher (2010), except for bead cleanups, which were done using a 1.6:1 ratio of beads to library (instead of the recommended 1.8:1) to obtain a slightly larger final fragment size distribution. Finished libraries were combined in equimolar ratios into two 22.5 µg pools (one per array) for target enrichment. Array hybridization was performed

largely following (Hodges et al., 2009) with minor modifications: Each library pool was mixed with xGen Universal P5 and P7 blocking oligonucleotides and a mixture of chicken, human, and mouse COT-1 DNA. The two capture eluates were amplified separately by 18 cycles of PCR. To reduce the propagation of PCR-induced errors, each eluate was amplified in four parallel reactions. PCR products were pooled so that both captures were equally represented, and sequenced on Ilumina HiSeq 2500 and 4000 machines.

Bioinformatic pipeline. De-multiplexed read files were filtered by 1) collapsing PCR-duplicate reads with SuperDeduper (Petersen, Streett, Gerritsen, Hunter, & Settles, 2015), 2) trimming low quality bases and removing adapter contamination with Trimmonatic (Bolger et al., 2014) and Skewer (Jiang, Lei, Ding, & Zhu, 2014) under default parameters, except for the minimum read length, which was increased to 36 bp, and 3) merging overlapping read pairs with FLASH (Magoč & Salzberg, 2011). To generate a reference for read mapping, we combined all cleaned reads from the ingroup species (i.e. *P. terribilis*, *aurotaenia*, and *bicolor*), and generated six *de novo* assemblies with different kmer sizes (k = 21, 31, 41, 51, 61, and 71) using ABySS (J. T. Simpson et al., 2009). We then merged the six assemblies using CD-HIT-EST and Cap3 (Huang & Madan, 1999). Finally, we identified contigs that matched our target exons using BLASTn, and retained only these for further analyses.

Reads from each sample were mapped to the reference using Bowtie2 v. 2.1.0 (Langmead & Salzberg, 2012), and outputs were sorted with Samtools v. 1.0 (Li et al.,

2009), de-duplicated with Picard v.1.8.4 (http://broadinstitute.github.io/picard), and re-aligned around indels with GATK v. 3.3.0 (McKenna et al., 2010). We filtered our data in the following ways: First, we performed a reciprocal blast using the methods described above and removed any contigs with more than one match (e-value<10^{-10}). Second, we used ngsParalog (https://github.com/tplinderoth/ngsParalog) to identify contigs with variants stemming from read mismapping due to paralogy and/or incorrect assembly. This program uses allele frequencies to calculate a likelihood ratio for whether the reads covering a site are derived from more than one locus in the genome, while incorporating the uncertainty inherent in NGS genotyping. We calculated p-values for these likelihood ratios based on a 50:50 mixed χ^2 distribution with one and zero degrees of freedom under the null, and removed any contigs with significantly paralogous sites after Bonferronni correction ($\alpha = 0.05$). Third, we restricted all analyses to contigs covered by at least one read in at least 20 individuals, bases with quality above 30, and read pairs mapping uniquely to the same contig (i.e. proper pairs) with mapping quality above 20. Finally, we removed samples with less than 2.5 million sites covered by at least one read after filtering. This resulted in a dataset of 32,516 contigs (12.95 Mb) and 57 samples, which were used in all downstream analyses.

Population Structure

To characterize genome-wide patterns of population differentiation we used our exon capture dataset to perform Principal Component Analysis (PCA) of genetic

288 covariances calculated in PCangsd v.0.94 (Meisner & Albrechtsen, 2018), to estimate
289 admixture proportions (k = 2-9) in ngsAdmix v.32 (Skotte, Korneliussen, &
290 Albrechtsen, 2013), and to build a minimum-evolution tree in FastME v.2.1.5
291 (Lefort, Desper, & Gascuel, 2015), using genetic distances estimated with ngsDist
292 (Vieira, Lassalle, Korneliussen, & Fumagalli, 2016). Nodal support for this tree was
293 evaluated using 500 bootstrapped distance matrices produced in ngsDist by sampling
294 blocks of 10 SNPs. These three analyses used genotype likelihoods (GL) as input,
295 which were estimated in Angsd v.0.9.18 (Korneliussen, Albrechtsen, & Nielsen,
296 2014) at sites covered by at least one read in at least 50% of the samples. PCA and
297 ngsAdmix analyses excluded sites with minor allele frequencies below 0.05, and
298 genetic distance estimation for the ME tree was restricted to variable sites (i.e. SNP
299 p-value < 0.05).
300
301 Finally, we reconstructed a population graph to evaluate the relationships among our
302 sampling localities using Treemix. We called genotypes using the HaplotypeCaller
303 and GenotypeGVCFs tools of GATK v.3.3.0 under default parameters, except for the
304 heterozygozity prior, minimum base quality, and minimum variant-calling confidence,
305 which were increased to 0.005, 30, and 20, respectively, to accommodate for the
306 multi-species nature of our dataset. We then obtained allele counts for biallelic SNPs
307 that were at least 1kb apart within each contig (usually resulting in a single SNP per
308 contig, since most contigs were under 1kb), and with at least 50% genotyping (20,275
309 SNPs), using Plink v.1.90 (Purcell et al., 2007). In two cases, two nearby populations
310 of the same color pattern (16.4 and 19.7 Km apart; Fig S1), which clustered closely in

all other genetic structure analyses, were merged into single demes for allele count estimation due to small sample sizes. In addition, since we only had exon capture data for one *P. vittatus* individual, only *P. lugubris* was used as outgroup in this analysis. We ran Treemix v.1.13 assuming m = 0–6 migration edges, and chose the optimal number of migration edges by performing likelihood ratio tests in which we compared each value of m to the one immediately smaller. P-values were calculated based on a χ^2 distribution with two degrees of freedom, since adding an extra edge adds two parameters (weight and direction of migration) to the model. This approach recovered m=2 as the most likely scenario (Table S2); results for m=0-6 are presented in Fig. S2.

Phylogenetic relationships between lineages

We reconstructed a species tree under the multispecies coalescent model, assuming independent sites, as implemented in SNAPP (Bryant, Bouckaert, Felsenstein, Rosenberg, & RoyChoudhury, 2012). SNAPP requires individuals to be assigned to operational taxonomic units (OTUs) *a priori.* Given our small sample sizes for some localities, as well as the evidence of gene flow between localities (see Results section), we took an *ad-hoc* approach and grouped our sampling localities into eight geographically and phenotypically coherent groups that showed evidence of being genetically distinct entities (see locality colors in Fig. 2B). Further details on our OTU selection criteria can be found in the online supplement.

For computational efficiency, SNAPP was run on a reduced version of the Treemix dataset described above, restricted to SNPs genotyped for at least 75% of individuals and at least one member of each OTU (5,938 SNPs). We ran the MCMC sampler under default priors for 1,000,000 iterations, sampling every 250, and discarded the first 150,000 as burnin. Stationarity and mixing were evaluated in Tracer (Rambaut & Drummond, 2009) as detailed above, and the posterior tree distribution was summarized as a maximum clade credibility (MCC) tree in TreeAnnotator. To obtain estimates of divergence times between OTUs, we assumed a mutation rate of $\mu = 1e^{-9}$ mutations per year (Crawford, 2003; Sun et al., 2015), and a generation time of one year (*Phyllobates* frogs are sexually mature at ~10-18 months after hatching; Myers et al., 1978; R. Márquez *pers obs.*), and converted branch lengths to time units as $T = (\tau g/\mu)$, where T is the divergence time in years, τ the branch length in coalescent units, g the generation time, and μ the mutation rate (Bryant et al., 2012).

Phylogenetic Comparative Analyses

We reconstructed ancestral color patterns along the SNAPP MCC tree using maximum parsimony (Fitch, 1971) in the R package *phangorn* (Schliep, 2011). Aposematic coloration has been shown to co-evolve with several other traits, such as body size, toxicity, and diet specialization in dendrobatid frogs (Pough & Taigen, 1990; Santos & Cannatella, 2011; Summers & Clough, 2001). Therefore, we investigated the extent of correlated evolution between color pattern, body size, and toxicity. We used the snout-to-vent length (SVL) as a proxy for body size, and the

356 average amount of batrachotoxin (BTX) in a frog's skin as a proxy for toxicity. BTX
357 is the most abundant and toxic alkaloid found in *Phyllobates* skins (Märki & Witkop,
358 1963; Myers et al., 1978). BTX levels were obtained from Table 2 of (Daly, Myers, &
359 Whittaker, 1987), and SVL was measured from specimens in natural history
360 collections (193 specimens; Table S3). We used mean SVL values for each lineage in
361 analyses, and log-transformed BTX levels to attain normality of residuals.
362 Correlations between traits were evaluated using phylogenetic generalized least
363 squares regression (pGLS; Grafen, 1989; Martins & Hansen, 1997) with either
364 Brownian motion (Felsenstein, 1985), Lambda (Pagel, 1999), or Ornstein–Uhlenbeck
365 (Martins & Hansen, 1997) correlation structures. The best correlation structure was
366 chosen by performing pGLS with the three correlation structures and comparing the
367 fit of each model based on the AIC. Correlation structures were generated using the R
368 package *ape* (Paradis, Claude, & Strimmer, 2004) and regressions were performed in
369 the *nlme* package (Pinheiro, Bates, DebRoy, & Sarkar, 2017). In addition to the
370 highest clade credibility tree, we also conducted tests of phylogenetic correlations on
371 1,000 randomly selected trees from the post-burnin SNAPP posterior distribution to
372 account for phylogenetic uncertainty.
373
374 *Spatial population genetics*
375
376 To investigate the extent of divergence and gene flow between species in an explicitly
377 spatial context we first we generated a geo-genetic map of the Colombian *Phyllobates*
378 populations in SpaceMix (Bradburd, Ralph, & Coop, 2016). This consists of a

bidimensional plot where the distance between two populations represents their expected geographic distance under stationary isolation by distance, accounting for the fact that a fraction of a population's alleles may have been acquired through migration from another region of the map, whose location is also estimated. As input we used allele counts generated as detailed above (see *Population Structure* section), for SNPs that were variable among Colombian individuals (8,093 sites). We then parameterized the full ("source_and target") SpaceMix model with an MCMC run comprised of 10 initial exploratory chains (500,000 giteration each), followed by a 500,000,000 iteration "long" run, which was sampled every 10,000 iterations. We used default prior settings, and centered spatial (i.e. location) priors for each population at their sampling location. For the two demes composed of individuals from nearby localities we used the midpoint of the segment connecting both localities (Fig. S1).

SpaceMix models long-distance gene flow between populations in terms of admixture proportions, which represent the probability that an allele in a given population migrated recently from a different location of the geo-genetic map. Since leapfrog distributions can arise through introgressive hybridization between disjunct populations, we evaluated the evidence for gene flow between distant populations in our data by comparing the fit of models where admixture proportions were either fixed to 0 or allowed to vary. We did so by estimating the Bayes Factor (Kass & Raftery, 1995) between both models using the Savage-Dickey density ratio (Dickey & Lientz, 1970), which approximates the Bayes Factor between nested models. Further

details on this estimator and our implementation for SpaceMix models can be found in the online supplement.

Next, we used EEMS (Petkova, Novembre, & Stephens, 2015) to identify areas of the landscape where gene flow between populations is especially prevalent or reduced. Briefly, this algorithm estimates the rate at which genetic similarity decays with distance (i.e. the effective migration). Regions where this decay is quick or slow can be interpreted as barriers or corridors of migration, respectively. We estimated mean squared genetic differences between samples from genotype likelihoods in ATLAS (Link et al., 2017), and used them as input for EEMS. We set the number of demes to 500, and averaged across 10 independent 10,000,000-step MCMC runs logged every 1,000 steps (20% burnin). Since our genetic dissimilarity matrix was inferred from genotype likelihoods, specifying the number of SNPs used to compute the matrix (required by EEMS) was not straightforward. We used the number of sites with SNP p-value below 0.05, as calculated with Angsd (221,825 sites).

Finally, we assessed how attributes of the landscape influence genetic divergence between populations. Based on the results of EEMS and SpaceMix analyses, we evaluated the effect of three landscape features on genetic divergence: geographic distance, differences in elevation, and the presence of the San Juan River as a potential corridor of gene flow. To do so, we used the multiple matrix regression with randomization (MMRR) approach proposed by Wang (2013), which is an extension of multiple linear regression for distance matrices.

Our regression model consisted of genetic distance as a response variable and geographic distance, difference in elevation, and the effect of the San Juan river as a dispersal corridor as explanatory variables. As a proxy for genetic distance, we used the linearized genome-wide weighted F_{ST} ($F_{ST}/[1-F_{ST}]$; J. Reynolds, Weir, & Cockerham, 1983; Weir & Cockerham, 1984), estimated using Angsd based on 2D-site frequency spectra (SFS). To maximize the number of sites used to estimate each SFS, we included contigs with data for less than 20 individuals (but that passed all other filters) in this analysis. We estimated geodesic distances among populations based on GPS coordinates taken in the field using the pointDistance() function of the raster R package (Hijmans, 2017), and calculated elevation differences based on measurements taken in the field or extracted from Google Earth. To generate a proxy for the San Juan river as a dispersal corridor we built a resistance layer where every pixel overlapping the San Juan river had a value of 1 and all others had values of 100, and used this layer to calculate least cost distances between populations with the costDistance() function of the gdistance R package (Dijkstra, 1959; van Etten, 2017). Finally, we regressed the least cost distance against the geodesic distance, and saved the model residuals as a measure of the component of the resistance distance not explained by geographic distance. These residuals were used as an explanatory variable in our model. The MMRR analysis was run using the script archived by Wang (2013; https://doi.org/10.5061/dryad.kt71r) with 10,000 permutations to estimate p-values.

Results

Population structure among Colombian Phyllobates

As expected from a multi-species dataset, we found multiple genetically structured clusters of individuals, which were largely concordant across our exon-enrichment and mtDNA analyses (Fig. 2, Fig. S2-S3). However, these clusters align much more closely with geography than either coloration or the current taxonomy: All populations of *P. terribilis* grouped with the southern populations of *P. aurotaenia*, while the northeastern populations of *P. aurotaenia* clustered closely with the northern populations of *P. bicolor*. The southern *P. bicolor* and the *P. aurotaenia* populations east and west of the Baudó mountains also formed independent clusters, but their relationship to other lineages was less clear. Finally, two sequences from captive-bred *P. aurotaenia* of unknown origin (sequenced by Grant et al. [2006] and Santos et al. [2009]) were sister to those from the southern populations of *P. bicolor* in our mtDNA genealogy (Fig. 2A). These results highlight the existence of several previously unrecognized (i.e. cryptic) lineages. Notably, they reveal the existence three independent solid-yellow lineages, instead of two as previously thought, since the populations currently classified as *P. bicolor* clustered as two clearly separate and independent lineages. This points to an even greater discordance between coloration phenotypes and genetic similarity than previously thought.

Phylogenetic relationships and divergence times.

The inferred species tree was generally consistent with our genetic structure results, since tree topologies largely mirrored geography (Fig. 2-3). The three yellow lineages were recovered as sister to a striped lineage. The topology of the SNAPP tree was largely concordant with those obtained in mtDNA and Treemix analyses. We only found inconsistencies in the placement of the *P. aurotaenia* populations from the eastern and western flanks of the Serranía del Baudó: Mitochondrial haplotypes from these two populations were part of a closely-related clade that also included all *P. aurotaenia* sequences from the Atrato river. This clade was sister to another one containing sequences from the northern *P. bicolor* and the (Fig. 2A). Treemix also recovered the eastern and western Atrato populations as sister taxa, but these two populations were sister to the rest of the Colombian populations (Fig. 2D and S2). Finally SNAPP recovered only the western Baudó *P. aurotaenia* as sister to all other Colombian populations, while the eastern Baudó *P. aurotaenia* was sister to the southern *P. bicolor* (Fig. 3). Treemix inferred a migration edge from the base of the clade containing the populations of *P. bicolor* and *P. aurotaenia* from the San Juan and Atrato drainages into the eastern Baudó *P. aurotaenia* (Fig. 2D). Since Treemix reconciles instances where a bifurcating tree model, such as the one used by SNAPP, does not fit the data well by incorporating migration edges between branches of the tree, this result suggests that these differences may be due to gene flow among populations.

Divergence time estimation based on the SNAPP tree revealed a Plio-Pleistocene diversification of *Phyllobates*, and were generally concordant with previous estimates (Santos et al., 2009, 2014), indicating that our mutation rate and generation time assumptions are reasonable. The most recent common ancestor (MRCA) of *Phyllobates* was placed at 5.1 million years ago (MYA), with subsequent cladogenesis events from the late Pliocene to the Pleistocene (2.9-0.6 MYA; Fig. 3). These divergence times were slightly older but within the 95% HPD intervals of those estimated from mtDNA sequences (Fig. S4).

Comparative analyses

Ancestral state reconstructions found the striped phenotype to be ancestral to solid-yellow (Fig. 4A). Phylogenetic regressions revealed a strong relationship between color pattern and size, with solid-yellow lineages being significantly larger than striped ones (Brownian Motion: $\beta = 11.95$, $t = 9.92$, $df = 10$, $p = 9.03e\text{-}6$; Fig. 4A), but a much weaker relationship between coloration and toxicity (Ornstein–Uhlenbeck: $\beta = 1.19$, $t = 2.48$, $df = 5$, $p = 0.089$; Fig. 4B). These results, suggest that at least two co-evolving traits (solid yellow coloration and larger size), possibly related to predator avoidance, are distributed in a leapfrog fashion in *Phyllobates*. Regressions performed over a set of posterior trees instead of the summary tree resulted in effect sizes and p-values centered around and qualitatively equivalent to those estimated using the summary tree, showing that the above conclusions are robust to the phylogenetic uncertainty present in our species tree reconstruction (Fig. S5).

517

518 *Spatial Population Genetics*

519

520 The effective migration surface estimated by EEMS revealed a clear corridor of
521 migration that matches the course of the San Juan river to a remarkable degree,
522 considering that this method is completely agnostic to the topography of the landscape
523 (Fig. 6A). This corridor connects most of the sampled *P. aurotaenia* populations and
524 the northern *P. bicolor*, and could explain the discordance between mtDNA and exon
525 capture datasets in the phylogenetic placement of *P. aurotaenia* populations from the
526 Eastern and Western Baudó mountains. Concordantly, SpaceMix estimated geo-
527 genetic locations of populations along the San Juan corridor that were much closer to
528 one another than their actual geographic positions: *P. aurotaenia* populations from the
529 upper San Juan and Atrato drainages and the northern *P. bicolor* converged to very
530 close locations in the upper/mid San Juan, overlapping considerably. The Baudó (east
531 and west) and southern populations of *P. aurotaenia* were also shifted towards this
532 area, but to a lesser extent (Fig. 6B).

533

534 In addition EEMS estimated very low levels of migration in the area enclosing the
535 two southern *P. bicolor* populations, suggesting the existence of barriers to gene flow
536 around these populations. Interestingly, the geo-genetic location of these populations
537 was inferred north of its geographic location, past the mid San Juan cluster, and only
538 slightly overlapping with other populations. The estimated long-distance admixture
539 proportions were minimal for all populations (Fig. 6C), and the model with these

proportions fixed at 0 was overwhelmingly supported over one where they were allowed to vary (Bayes factor = 1748).

In agreement with EEMS and SpaceMix results, the MMRR analysis found significant effects of geographic distance, elevation differences, and the San Juan as a migration corridor on genetic divergence between localities (Table 1). Geographic distance was, expectedly, the strongest predictor ($\beta = 0.64$, $t = 10.05$, $p < 0.00001$), but elevation differences ($\beta = 0.29$, $t = 4.60$, $p = 0.003$), and the San Juan as a barrier ($\beta = 0.24$, $t = 4.01$, $p = 0.013$) still had appreciable effects on genetic divergence.

Discussion

Our main goal for this study was to understand the evolutionary and biogeoographic processes that have shaped the leapfrog distribution of color pattern among *Phyllobates* populations, focusing on the roles of genetic divergence and gene flow. We found patterns of genetic structure and phylogenetic affinity between populations that closely match geography, and a strong signal of short-range gene flow, especially along the San Juan river, together with compelling evidence against gene flow between distant populations.

These results provide strong evidence against the hypothesis that introgression of color pattern alleles between disjunct populations has played a role in generating the geographic distribution of this trait. Instead, they suggest an important role for short-range gene flow between neighboring populations. The high level of migration among the central striped populations along the San Juan river suggests that allele movement between these populations can have a homogenizing effect that maintains their phenotypic similarity. In addition, we find evidence for a barrier to gene flow that encloses the two sampled populations of the southern *P. bicolor* lineage, probably associated with differences in elevation, which could be helping maintain the phenotypic distinctiveness of this population. Conversely, the northern *P. bicolor* populations showed a strong signature of gene flow with their neighboring striped populations, suggesting that other forces, possibly selection, are maintaining the phenotypic differences between these populations in the face of recurrent gene flow. Nevertheless, to fully reject or accept these hypotheses, the history the alleles underlying color pattern differences must me taken into account (Hines et al., 2011).

Our phylogenetic reconstructions are consistent with a scenario in which the disjunct solid-yellow populations evolved their color patterns independently. However, the high levels of gene flow between geographically proximal populations and the close correspondence between phylogeny and geography lead us to suspect that, at least to an extent, the recovered phylogenetic relationships are a product of prevalent gene flow between neighboring populations, and therefore do not reflect the history of color pattern. A recent simulation study showed that even moderate levels of gene

585 flow between geographic neighbors can confound phylogenetic analyses (James et al.,
586 2020). This scenario seems especially likely in the case of the northern populations of
587 *P. bicolor*, given the signature of gene flow with their nearby striped populations (e.g.
588 the San Juan and Atrato *P. aurotaenia*), but less so in the case of the southern *P.*
589 *bicolor*, considering the strong barriers to gene flow inferred around the populations
590 of this lineage. For *P. terribilis* we cannot favor either scenario, since we did not find
591 strong evidence for or against gene flow with its sister *P. aurotaenia* populations.
592
593 It is therefore plausible that convergent evolution and common ancestry have both
594 played a role in the origin of this leapfrog distribution. Teasing apart these two
595 scenarios is challenging if gene flow between neighboring populations is prevalent,
596 since high levels of genetic exchange between geographically close populations can
597 erode existing differentiation between them, leading to patterns of
598 genetic/phylogenetic affinity across the genome that mirror geography, regardless of
599 their previous history. In the case of leapfrog distributions, this means that, in the face
600 of persistent gene flow, peripheral populations will be closest to their phenotypically
601 distinct neighbors, even if their phenotypic similarity stems from common ancestry.
602 However, admixture between lineages is seldom uniform across the genome, since
603 selection (see below) can restrict gene flow at certain genomic regions (J. R. Turner,
604 Johnson, & Eanes, 1979; T. L. Turner, Hahn, & Nuzhdin, 2005; Wu, 2001). Such
605 regions can therefore preserve historic signatures that have been erased by gene flow
606 elsewhere in the genome. This is likely to be the case for loci underlying color pattern
607 variation in *Phyllobates*, especially in cases such as the Northern *P. bicolor*, where

608 phenotypic differences persist in spite of gene flow. Hence, the history of alleles at
609 these loci should provide unique insights into the history of this phenotype. Future
610 studies to identify such loci and understand their evolutionary history in relation to
611 our current results will be instrumental to uncover the demographic processes leading
612 to the current geographic and phylogenetic distribution of solid-yellow color pattern
613 in *Phyllobates*, since they will allow for much more explicit tests of the hypotheses
614 presented here.
615
616 Regardless of whether common ancestry or convergent evolution are at play in this
617 system, it seems clear that differential selective pressures on the striped and solid-
618 yellow populations have been involved in the origin and/or maintenance of the
619 geographic distribution of color patterns. Independent evolution of similar phenotypes
620 is many times promoted by similar changes in selective regimes (Darwin, 1859; Mayr,
621 1963; G. G. Simpson, 1953), and selection is required to maintain phenotypic
622 differences between populations in the face of gene flow (Endler, 1977). Two of the
623 three solid-yellow lineages (the northern and southern *P. bicolor*) occur at higher
624 elevations (~600-1000 m.a.s.l) than the rest of the genus (~0-500 m.a.s.l). It is
625 therefore possible that these mid-elevation habitats pose selective pressures (e.g.
626 predator communities or light environments) different from those of lowland forests,
627 which favor solid-yellow patterns over striped ones. The many known examples of
628 variation in coloration across altitudinal gradients lend support to this idea (e.g.
629 Köhler, Samietz, & Schielzeth, 2017; Rebelo & Siegfried, 1985; Reguera, Zamora-
630 Camacho, & Moreno-Rueda, 2014; Richmond & Reeder, 2002; Rios & Álvarez-

631 Castañeda, 2007) A similar situation could also be the case with *P. terribilis*, given its
632 distribution at the southern edge of the genus's range, where it may also experience
633 different selective pressures from those faced by its closely related striped lineages.
634
635 The nature of color pattern variation in *Phyllobates* (i.e. solid vs striped) suggests that
636 differential predation pressures may be important for the origin/maintenance of solid-
637 yellow patterns. In aposematic species, advertisement signals with complex pattern
638 elements, such as stripes, have been shown to serve a distance-dependent purpose,
639 acting as conspicuous signals at short distances, while providing camouflage at long
640 distances (Barnett et al., 2017; Barnett & Cuthill, 2014; Tullberg, Merilaita, &
641 Wiklund, 2005). In contrast, bright, solid-colored signals remain conspicuous over a
642 much wider range of distances. For example, a recent study focusing on the poison
643 frog *Dendrobates tinctorius* found that striped patterns of this species are highly
644 detectable at close range, but become camouflaged when observed from further away.
645 Solid-yellow patterns, on the other hand, remained easily detectable over the whole
646 range of distances tested (Barnett, Michalis, Scott-Samuel, & Cuthill, 2018).
647 Therefore, it is likely that the striped and solid color patterns represent alternative
648 aposematic strategies that are advantageous under different environments and/or
649 predator communities.
650
651 The fact that we find a signature of correlated evolution between size and color
652 pattern is compatible with this idea, since larger aposematic signals have been shown
653 to be more detectable and memorable for predators (Forsman & Merilaita, 1999;

Gamberale & Tullberg, 1996). Accordingly, size and conspicuousness are positively correlated among Dendrobatid poison frog species (Hagman & Forsman, 2003; Santos & Cannatella, 2011). However, we do not find a comparable pattern for toxicity, which has also been shown to co-vary with conspicuousness in poison frogs (Santos & Cannatella, 2011; Summers & Clough, 2001). This could be an artifact of low statistical power, since data are available only for one striped and two plain yellow Colombian lineages, but we cannot rule out the possibility that toxicity is indeed comparable between solid and striped populations. Furthermore, considering that aposematism relies on avoidance learning, it is possible that, despite similar levels of BTX, solid and striped populations differ in levels of palatability to predators. In any case, a scenario where solid and striped populations are similarly toxic and/or palatable is still compatible with predation pressures driving evolutionary convergence, since all species are, in any case, considerably toxic (Daly et al., 1987; Myers et al., 1978). However, other explanations, such as geographic variation in mate preference (R. G. Reynolds & Fitzpatrick, 2007; Summers, Symula, Clough, & Cronin, 1999; Twomey, Vestergaard, & Summers, 2014; Yang, Richards-Zawacki, Devar, & Dugas, 2016), could also explain our results and cannot be ruled out.

It is worth noting, however, that the correlated evolution of body size and color pattern could also be due to ontogenetic integration (Olson & Miller, 1958). Tadpoles of all *Phyllobates* species are dark grey, and all of them develop a dorsolateral stripe shortly before metamorphosis, which remains unchanged until adulthood in striped lineages. Solid-yellow frogs, on the other hand, gradually lose dark pigmentation,

677 until the solid adult pattern is attained a few months after metamorphosis (Myers et
678 al., 1978). Therefore it is possible that, for example, the evolution of an extended
679 growth period could generate changes in both body size and color pattern. If this is the
680 case, then the concerted evolution of advertisement signal and body size would not
681 necessarily be evidence of striped and solid patterns representing alternative predator
682 avoidance strategies.
683
684 Our divergence time estimates indicate that the diversification of *Phyllobates* has
685 followed the Plio-Pleistocene history of the Central American and the Chocó
686 bioregions. The first cladogenesis event in our tree, which divides the Central
687 American and Chocoan taxa was reconstructed between 4.5-5.9 MYA, which
688 coincides with previously identified increases in faunal migration between Central
689 and South America at ~6 MYA (Bacon et al., 2015; Santos et al., 2009). Further
690 branching within South American lineages occurred later than 3 MYA, after both the
691 Atrato (Duque-Caro, 1990b, 1990a) and Tumaco (Borrero et al., 2012) basins
692 emerged above sea level to form the current Chocoan landscape. The Pleistocene was
693 characterized by recurrent climatic and environmental fluctuations, which have been
694 proposed as major drivers of neotropical rainforest biodiversity (Baker et al., 2020;
695 Haffer, 1969; Hooghiemstra & Van Der Hammen, 1998; Vanzolini & Williams,
696 1970). Although the central Chocó has traditionally been regarded as a relatively
697 stable Pleistocene forest refuge throughout the Quaternary (Gentry, 1982; Haffer,
698 1967; Hooghiemstra & Van Der Hammen, 1998), a notion supported by multiple
699 palynological studies (Behling, Hooghiemstra, & Negret, 1998; Berrío, Behling, &

Hooghiemstra, 2000; González, Urrego, & Martínez, 2006; Jaramillo & Bayona, 2000; Ramírez & Urrego, 2002; Urrego, Molina, Urrego, & Ramírez, 2006), there is some evidence of fluctuations in humidity, temperature, fluvial discharge, and sea level throughout the Quaternary in this region (González et al., 2006; Urrego et al., 2006). Despite being less dramatic than those experienced by other tropical forests (e.g. Amazonia), these fluctuations may have promoted periodic retractions of *Phyllobates* populations towards the San Juan, perhaps resulting in increased rates of gene flow among them. Future work to understand how climatic fluctuations over the Quaternary have shaped the distribution of suitable habitat for *Phyllobates* frogs should shed further light on the biogeographic history of this genus in Northern South America.

Finally, our findings have broad implications for the systematics of *Phyllobates*. First and foremost, this study provides definitive evidence that the populations currently grouped under *P. aurotaenia* represent multiple independently-evolving lineages, many of which have probably been reproductively isolated for enough time to warrant recognition as separate species. Furthermore, we find that *P. bicolor* is comprised of two well-structured lineages that may have evolved similar phenotypes independently. Third, we find highly variable levels of mtDNA divergence within *P. lugubris* (0-4% 16S, 0.2-8% COI, and 0-5.7% Cytb uncorrected p-distances), which could also be due to the existence of cryptic species.

It is, therefore, evident that a thorough revision of *Phyllobates* systematics is warranted. At this point, however, we refrain from modifying the group's current taxonomy for several reasons. First, our results clearly illustrate that, as previously suggested (Grant et al., 2006; Myers & Daly, 1976), coloration alone is not effective for diagnosing species of Dendrobatid frogs. Therefore, integrating multiple lines of evidence (e.g. coloration, genetic variation, alkaloid profiles, bioacoustic data, larval and adult morphology) is needed to disentangle species limits. Second, although our study represents a substantial increase in geographic sampling, there are still considerable gaps, such as the lower San Juan drainage, or the mid-elevation forests south of the distribution of *P. bicolor,* that need to be considered. The fact that the two captive-bred *P. aurotaenia* included in mtDNA analyses are sister to the southern *P. bicolor* (Fig. 2A) suggests that we have not yet sampled the full diversity of P*hyllobates* lineages in Colombia. Finally the holotype of *P. aurotaenia* was collected in Condoto, Chocó, which is considerably distant from any of our sampling localities (Fig. S6), and the type locality of *P. bicolor* is unknown (Myers et al., 1978). This situation poses nomenclature issues, since, even if a robust species delimitation were available, naming these species would not be straightforward until the type specimens of *P. bicolor* and *P. aurotaenia* can be confidently assigned to one of them. Further work with increased sampling, including type specimens, and integrating multiple lines of evidence is therefore still needed to generate a taxonomy for *Phyllobates* that more accurately represents the genus's evolutionary history.

Concluding remarks

Leapfrog distributions constitute geographic replicates for the occurrence of a phenotype, and therefore provide important information about the origins of phenotypic diversity among closely related lineages. Here we show that, despite marked genetic structure and differentiation, there is considerable gene flow between phenotypically similar populations at the center of a poison-dart frog leapfrog distribution. This has probably been important for the origin and maintenance of the geographic distribution of color patterns in this group. Furthermore, we found instances of both reduced and increased levels of gene flow between neighboring populations with different phenotypes, suggesting that in some cases reduced gene exchange can contribute to the maintenance of phenotypic differences between populations in a leapfrog distribution, while in others these differences actually persist in the face of gene flow, probably due to local adaptation of different forms.

However, we are unable to answer a commonly addressed question about leapfrog distributions: whether phenotypic differences between populations stem from common ancestry or independent evolution. Even though our phylogenetic reconstructions unambiguously suggest the latter on their own, our finding of extensive gene flow among neighboring populations casts doubt on this conclusion. Several other studies on the history of leapfrog distributions have obtained similar phylogenies that align with geography instead of phenotypic similarity (Cadena et al., 2010; Garcia-Moreno & Fjeldså, 1999; Norman et al., 2002; Quek et al., 2010; Toon,

768 Austin, Dolman, Pedler, & Joseph, 2012), leading to the view that leapfrog
769 distributions are often due to independent evolution. Our results, therefore, add to the
770 notion that alternative explanations such as pervasive gene flow (Hines et al.,
771 2011) or incomplete lineage sorting (Avise & Robinson, 2008) can decouple the
772 history of phenotypic change at a given trait from genome-wide patterns of
773 divergence, possibly leading to erroneous inferences of convergent evolution (Hahn &
774 Nakhleh, 2015).
775
776 **Data accessibility**
777
778 mtDNA sequences were uploaded to GenBank under accessions XXX-YYY (To be
779 added upon acceptance), and raw Illumina reads were uploaded to the NCBI SRA
780 under accession zzzz. The assemblies, bam, and vcf files, and body size data, as well
781 as the code used for analyses are available on the DRYAD repository (Citation to be
782 added upon acceptance) or as supplementary material.
783
784 **Acknowledgements**
785
786 We thank Pablo Palacios-Rodríguez, José Alfredo Hernández, Carolina Esquivel,
787 Diana Galindo, Mabel González, and Fernando Vargas-Salinas for assistance in the
788 field, Lydia Smith, Valeria Ramírez-Castañeda, Alvaro Hernández, and Ke Bi for help
789 with molecular and bioinformatic procedures, Andrea Paz for advice on MMRR
790 analyses, and Alan Resetar (FMNH), Andrew Crawford and Alberto Farfán

(ANDES), Rayna Bell and Addison Wynn (USNM), Andrés Acosta and Carlos Montaña (IAvH), John Taylor Rengifo (UTCh), Greg Schneider (UMMZ), and David Kizirian (AMNH) for facilitating access to preserved specimens. Comments and suggestions from Trevor Price, John Novembre, Valentina Gómez-Bahamon, John Bates, Daniel Matute, the Bates/Hackett lab, the Kronforst lab, and two anonymous reviewers greatly improved this paper. We sincerely thank Lina M. Arenas for allowing us to use her beautiful frog illustrations for our figures. This work was funded by a Basic Sciences Grant from the Vice Chancellor of Research at Universidad de los Andes, a Colombia Biodiversa Scholarship from the Alejandro Angel Escobar foundation, a Pew Biomedical Scholarship, Neubauer Family funds and a Steiner Award from the Univeristy of Chicago, and NSF grant DEB-1655336. RM was partially supported by a Fellowship for Young Researchers and Innovators (Otto de Greiff) from COLCIENCIAS. Computations were performed on the University of Chicago's Gardner HPC cluster, funded by NIH grant TR000430. Tissue collections were authorized by permits No. 2194 and 1380 from the Colombian Ministry of Environment and Authority for Environmental Licenses (ANLA).

Author contributions

R.M., A.A., and M.R.K. conceived the project, R.M., T.P.L, R.N., M.R.K, and A.A. designed the research, R.M., A.A., R.N. and M.R.K. acquired funding, A.A., D.M-V., and R.M. collected samples, R.M. and T.P.L. generated the data, and R.M. analyzed the data and wrote the paper with input from M.K. and edits from all authors.

Supplementary information

Supplementary note. Criteria used to select Operational Taxonomic Units (OTUs) for phylogenetic inference with SNAPP.

Table S1. Information and accession numbers for samples used in this study. Locality data is limited to prevent illegal traffic. Further details are available from the authors upon request.

Table S2. Results of likelihood ratio tests performed on Treemix analyses run with m = 0-6 migration edges.

Table S3. Information and snout-to-vent lengths of museum specimens measured for comparative analyses.

Figure S1. Localities joined into a single deme for locality-level analyses.

Figure S2. Results of Treemix analyses run with m=0-6 migration edges.

Figure S3. Minimum-evolution tree based on genetic distances.

Figure S4. Mitochondrial DNA time tree inferred using BEAST 2.

Figure S5. Results of phylogenetic comparative analyses run on 1000 randomly-drawn trees from the SNAPP posterior tree distribution.

Figure S6. Map of the type locality of *P. aurotaenia*.

Figure S7-S8 and associated text. Details on our implementation of the Savage-Dickey ratio to estimate Bayes Factors between nested SpaceMix models.

837 **Tables**

838

839 Table 1. Results from the multiple matrix regression with randomization (MMRR)

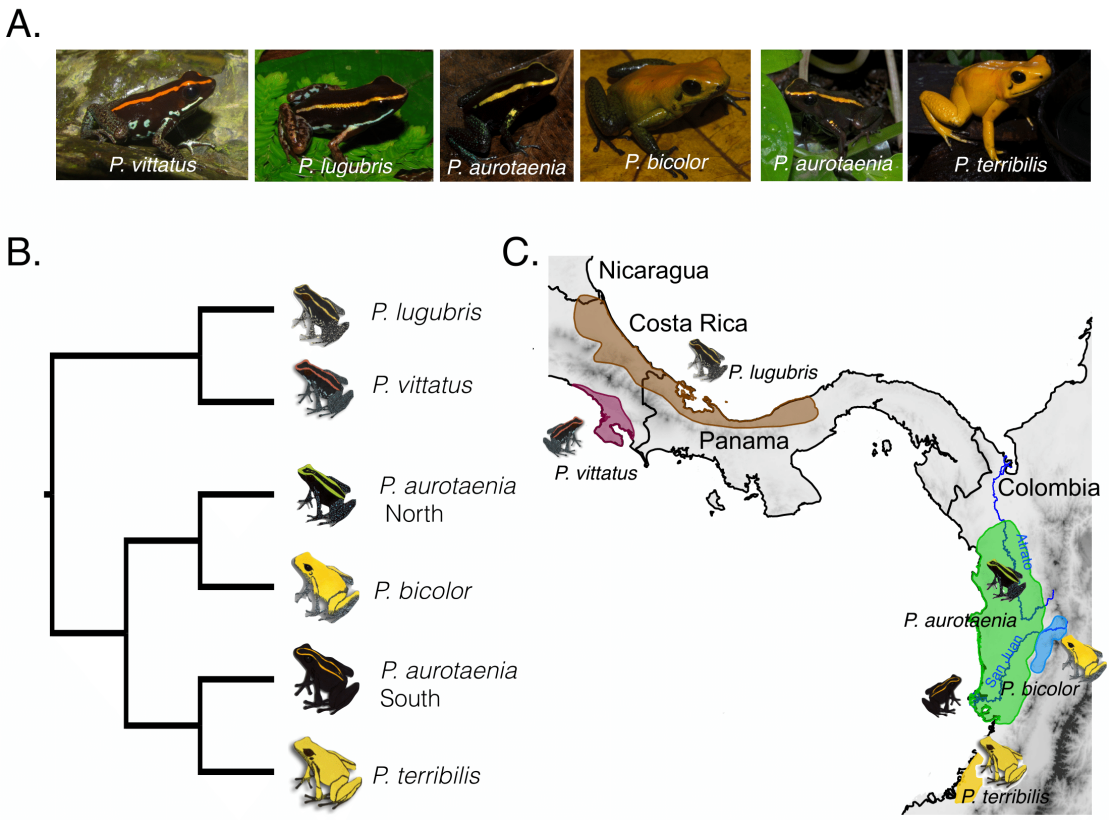
840 analysis. P-values were estimated using 10,000 permutations.

Predictor	Coefficient	t-statistic	p-value
Intercept	0.0325	0.5286	1.00000
Geodesic Distance	0.6419	10.0547	< 0.00001
San Juan LC Distance	0.2386	4.0094	0.01370
Elevation Difference	0.2886	4.5994	0.00320
$r^2 = 0.62$, $F = 48.13$, $p < 0.00001$			

841

842 **Figures**

843



845

846 **Fig 1.** A) Color pattern diversity, B) geographic distribution, and C) currently
847 accepted phylogenetic relationships of *Phyllobates* poison frogs (Grant et al., 2017).
848 Species distribution polygons were obtained from the IUCN red list of threatened
849 species website (https://www.iucnredlist.org/) and modified to fit natural history
850 collection records and our own observations.

851

852

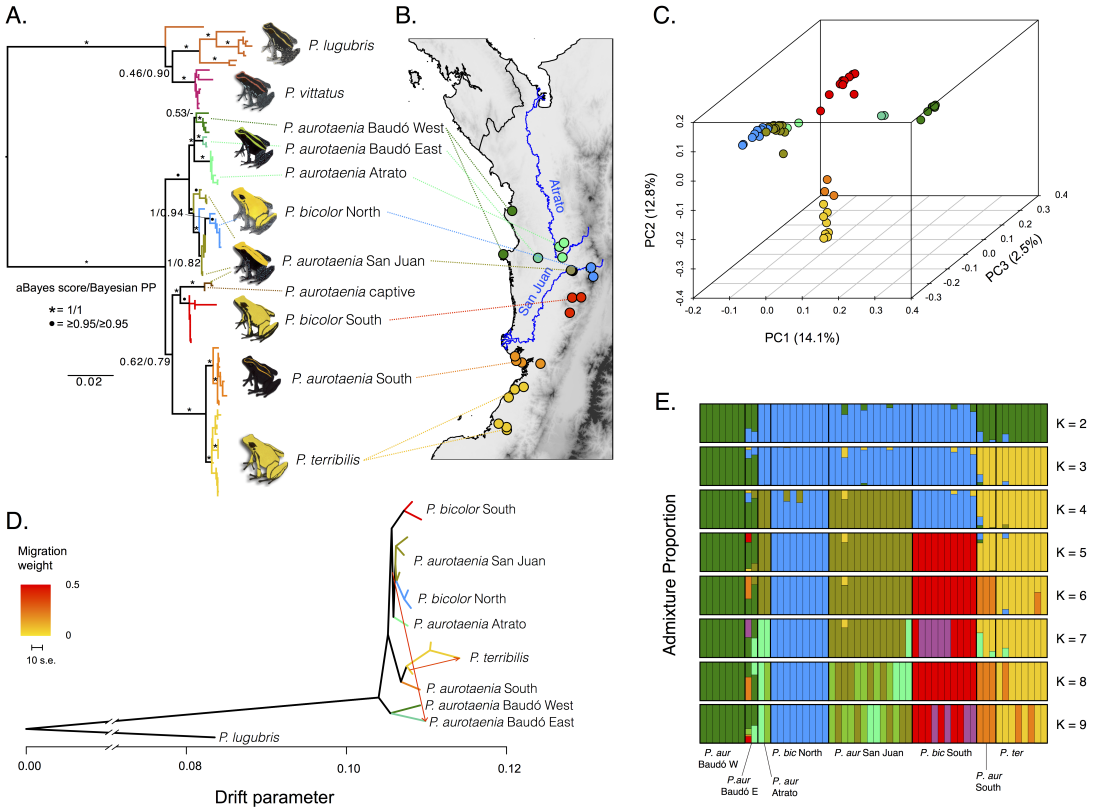
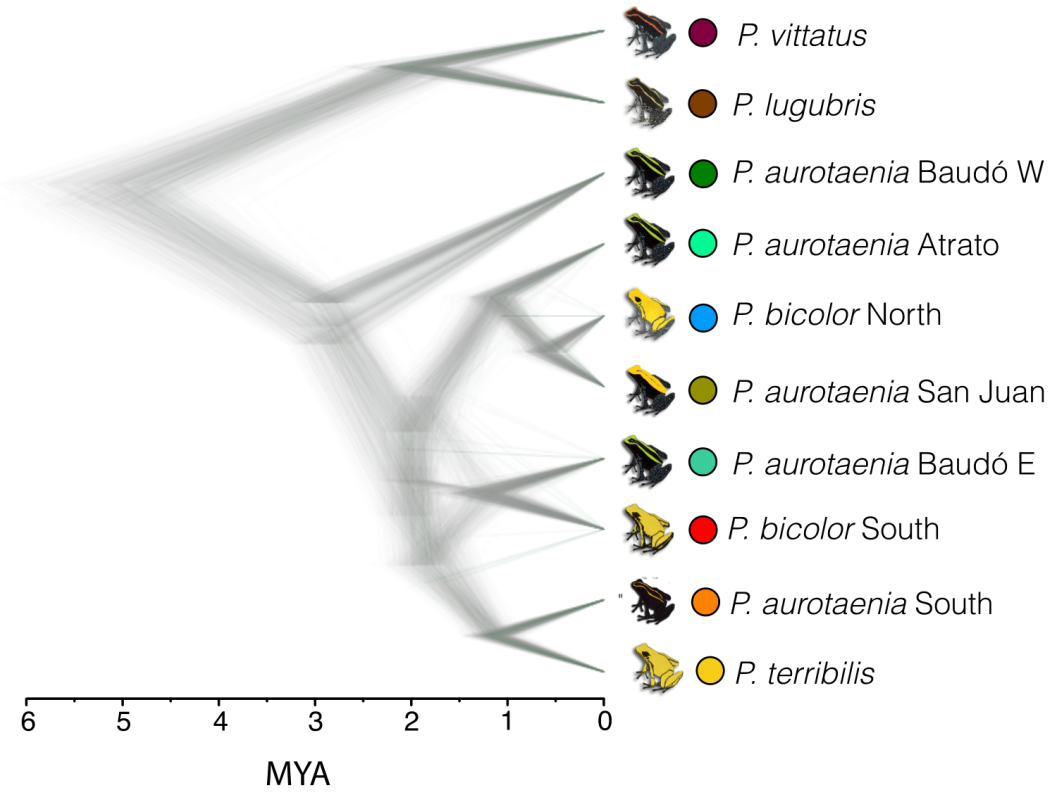
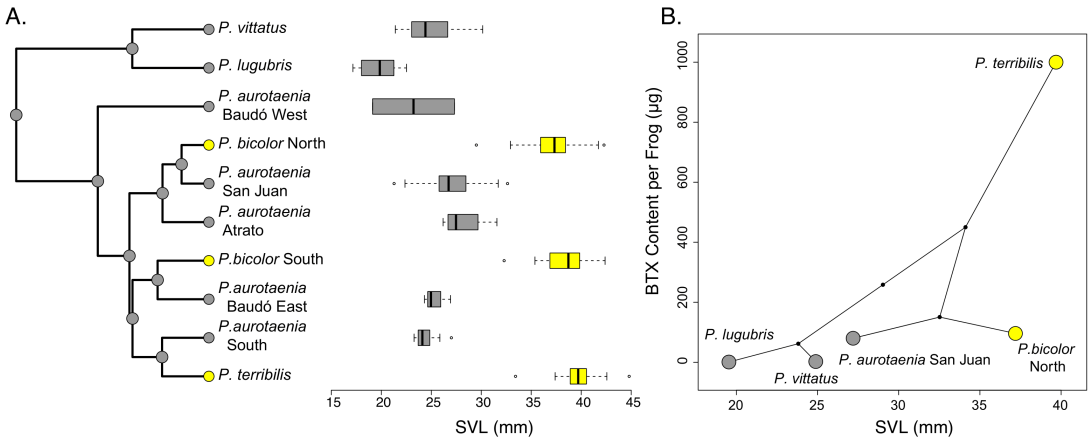


Fig 2. Genetic structure among *Phyllobates* populations in western Colombia. A) Maximum likelihood mtDNA genealogy inferred from 1926bp. B) Sampling localities for this study. C) Principal component analysis plot based on the first three components accounting for 29.4% of the variance. D) Treemix population graph assuming 2 migration edges. E) Individual admixture proportions assuming 2-9 ancestral populations. Colors in A-D correspond to the operational taxonomic units (OTU) used for phylogenetic analyses. Colors in E were chosen to loosely represent these clusters.



865 **Fig. 3** Phylogenetic relationships and divergence times among *Phyllobates* lineages
866 inferred using SNAPP. Divergence times assume a mutation rate of 10^{-9} mutations per
867 year and a generation time of one year. Each individual tree represents a sample from
868 the posterior distribution, and clades present in more samples have higher posterior
869 probabilities. A maximum clade credibility tree derived from this distribution is
870 displayed on Fig. 4A. Colors are as in Fig. 2.
871

872



874 **Fig. 4** Evolutionary patterns of body size, color pattern, and toxicity in *Phyllobates*.
875 A) Maximum clade credibility tree derived from the SNAPP posterior distribution
876 and phylogenetic distribution of color pattern and snout-to-vent length (SVL) values
877 among lineages. B) Phylogenetic biplot depicting the relationship between mean SVL
878 and mean batrachotoxin concentration. Grey boxes/points represent striped lineages
879 while yellow ones represent solid-yellow lineages.

880

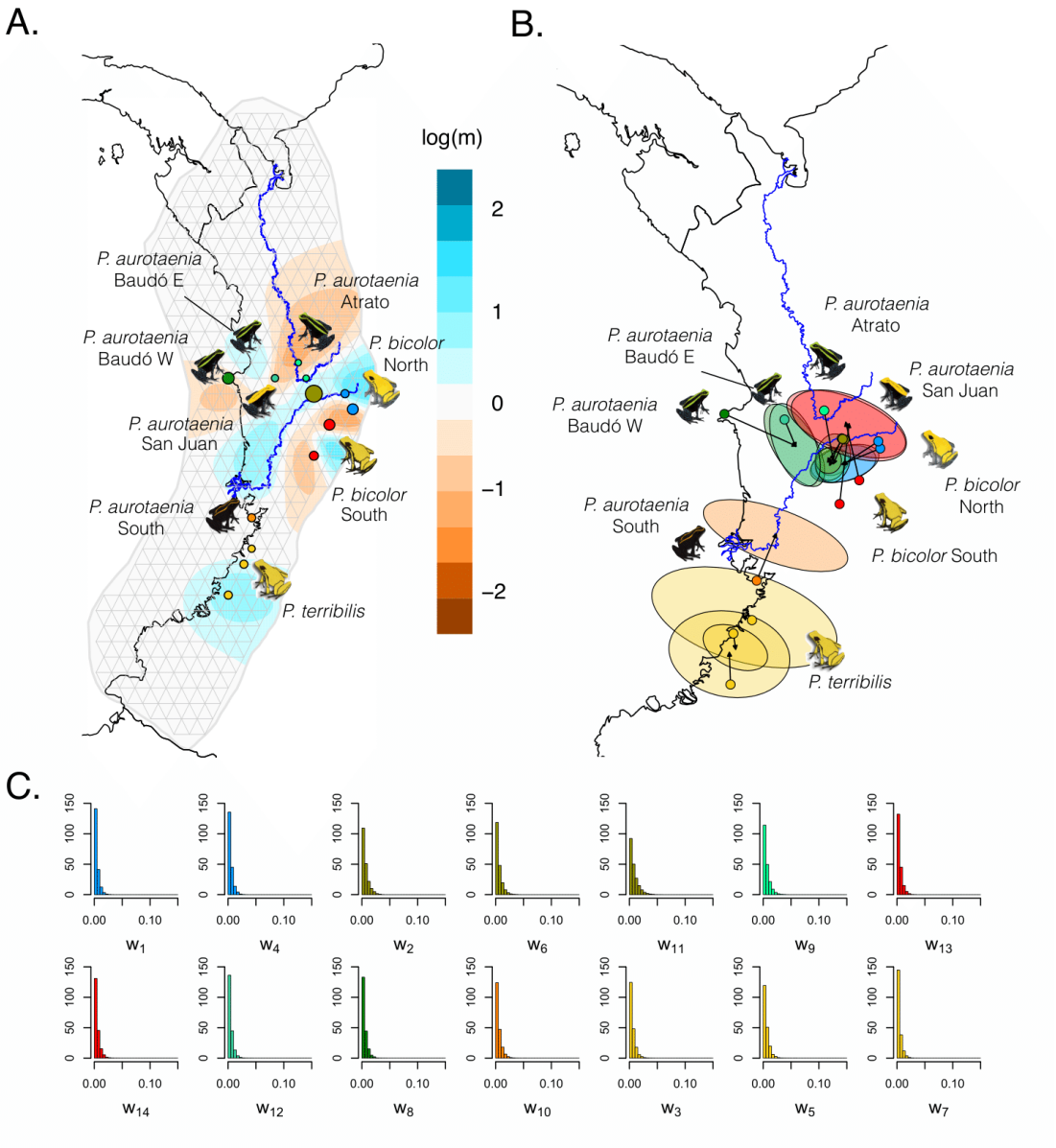
881

882

883

884

885



887 **Fig. 5** Gene flow among *Phyllobates* species across Western Colombia. A) Effective
888 migration surface estimated using EEMS. Cyan and brown areas of the map are those
889 where migration between demes is higher (cyan) or lower (brown) than expected
890 under isolation by distance. Grey lines depict the population grid and habitat outline
891 used by EEMS. B) Geogenetic map inferred with SpaceMix. Ellipses represent the

892 95% Bayesian credible intervals around each population's geogenetic location, and
893 colored dots represent actual sampling locations. Arrows connect sampling and
894 geogenetic locations. C) Posterior distributions of the admixture proportion
895 parameters from the SpaceMix model for each population. Histograms, points, and
896 ellipses are colored by OTU as in Fig. 2.
897
898
899

References

Altekar, G., Dwarkadas, S., Huelsenbeck, J. P., & Ronquist, F. (2004). Parallel Metropolis coupled Markov chain Monte Carlo for Bayesian phylogenetic inference. *Bioinformatics*, *20*(3), 407–415. doi:10.1093/bioinformatics/btg427

Altschul, S. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research*, *25*(17), 3389–3402. doi:10.1093/nar/25.17.3389

Anisimova, M., Gil, M., Dufayard, J. F., Dessimoz, C., & Gascuel, O. (2011). Survey of branch support methods demonstrates accuracy, power, and robustness of fast likelihood-based approximation schemes. *Systematic Biology*, *60*(5), 685–699. doi:10.1093/sysbio/syr041

Avise, J. C., & Robinson, T. J. (2008). Hemiplasy: A new term in the lexicon of phylogenetics. *Systematic Biology*, *57*(3), 503–507. doi:10.1080/10635150802164587

Bacon, C. D., Silvestro, D., Jaramillo, C., Smith, B. T., Chakrabarty, P., & Antonelli, A. (2015). Biological evidence supports an early and complex emergence of the Isthmus of Panama. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(19), 6110–6115. doi:10.1073/pnas.1423853112

Baker, P. A., Fritz, S. C., Battisti, D. S., Dick, C. W., Vargas, O. M., Asner, G. P., … Prates, I. (2020). Beyond Refugia: New insights on Quaternary climate variation and the evolution of biotic diversity in tropical South America. In V. Rull & A. C. Carnaval (Eds.), *Neotropical Diversification: Patterns and Processes*. Berlin: Springer. doi:10.1007/978-3-030-31167-4

Barnett, J. B., & Cuthill, I. C. (2014). Distance-dependent defensive coloration. *Current Biology*, *24*(24), R1157–R1158. doi:10.1016/j.cub.2014.11.015

Barnett, J. B., Michalis, C., Scott-Samuel, N. E., & Cuthill, I. C. (2018). Distance-dependent defensive coloration in the poison frog *Dendrobates tinctorius*, Dendrobatidae. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(25), 6416–6421. doi:10.1073/pnas.1800826115

Barnett, J. B., Redfern, A. S., Bhattacharyya-Dickson, R., Clifton, O., Courty, T., Ho, T., … Cuthill, I. C. (2017). Stripes for warning and stripes for hiding: Spatial frequency and detection distance. *Behavioral Ecology*, *28*(2), 373–381. doi:10.1093/beheco/arw168

933 Behling, H., Hooghiemstra, H., & Negret, A. J. (1998). Holocene history of the Chocó
934 rain forest from Laguna Piusbi, southern Pacific lowlands of Colombia.
935 *Quaternary Research*, *50*(3), 300–308. doi:10.1006/qres.1998.1998

936 Berrío, J. C., Behling, H., & Hooghiemstra, H. (2000). Tropical rain-forest history
937 from the Colombian Pacific area: A 4200-year pollen record from Laguna
938 Jotaordó. *Holocene*, *10*(6), 749–756. doi:10.1191/09596830094999

939 Bi, K., Vanderpool, D., Singhal, S., Linderoth, T., Moritz, C., & Good, J. M. (2012).
940 Transcriptome-based exon capture enables highly cost-effective comparative
941 genomic data collection at moderate evolutionary scales. *BMC Genomics*, *13*(1),
942 403. doi:10.1186/1471-2164-13-403

943 Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for
944 Illumina sequence data. *Bioinformatics*, *30*(15), 2114–2120.
945 doi:10.1093/bioinformatics/btu170

946 Borrero, C., Pardo, A., Jaramillo, C. M., Osorio, J. A., Cardona, A., Flores, A., …
947 Castillo, H. (2012). Tectonostratigraphy of the Cenozoic Tumaco forearc basin
948 (Colombian Pacific) and its relationship with the northern Andes orogenic build
949 up. *Journal of South American Earth Sciences*, *39*, 75–92.
950 doi:10.1016/j.jsames.2012.04.004

951 Bouckaert, R., Vaughan, T. G., Barido-Sottani, J., Duchêne, S., Fourment, M.,
952 Gavryushkina, A., … Drummond, A. J. (2019). BEAST 2.5: An advanced
953 software platform for Bayesian evolutionary analysis. *PLoS Computational*
954 *Biology*, *15*(4), e1006650. doi:10.1371/journal.pcbi.1006650

955 Bradburd, G. S., Ralph, P. L., & Coop, G. M. (2016). A spatial framework for
956 understanding population structure and admixture. *PLoS Genetics*, *12*(1), 1–38.
957 doi:10.1371/journal.pgen.1005703

958 Brower, A. V. Z. (1996). Parallel race formation and the evolution of mimicry in
959 *Heliconius* butterflies : A phylogenetic hypothesis from mitochondrial DNA
960 sequences. *Evolution*, *50*(1), 195–221. doi:10.2307/2410794

961 Bryant, D., Bouckaert, R., Felsenstein, J., Rosenberg, N. A., & RoyChoudhury, A.
962 (2012). Inferring species trees directly from biallelic genetic markers: Bypassing
963 gene trees in a full coalescent analysis. *Molecular Biology and Evolution*, *29*(8),
964 1917–1932. doi:10.1093/molbev/mss086

965 Cadena, C. D., Cheviron, Z. A., & Funk, W. C. (2010). Testing the molecular and
966 evolutionary causes of a ‘leapfrog’ pattern of geographical variation in

coloration. *Journal of Evolutionary Biology*, *24*(2), 402–414. doi:10.1111/j.1420-9101.2010.02175.x

Chapman, F. M. (1923). Mutation among birds on the genus *Buarremon*. *Bulletin of the American Museum of Natural History*, *48*, 243–278.

Che, J., Chen, H.-M., Yang, J.-Xi., Jin, J.-Q., Jiang, K., Yuan, Z.-Y., … Zhang, Y.-P. (2012). Universal COI primers for DNA barcoding amphibians. *Molecular Ecology Resources*, *12*(2), 247–258. doi:10.1111/j.1755-0998.2011.03090.x

Crawford, A. J. (2003). Relative rates of nucleotide substitution in frogs. *Journal of Molecular Evolution*, *57*(6), 636–641. doi:10.1007/s00239-003-2513-7

Daly, J. W., Myers, C. W., & Whittaker, N. (1987). Further classification of skin alkaloids from neotropical poison frogs (Dendrobatidae), with a general survey of toxic/noxious substances in the Amphibia. *Toxicon*, *25*(10), 1023–1095. doi:10.1016/0041-0101(87)90265-0

Darwin, C. (1859). *On the origin of species, by means of natural selection; or, The preservation of favoured races in the struggle for life*. London: John Murray.

Dickey, J. M., & Lientz, B. P. (1970). The weighted likelihood ratio, sharp hypotheses about chances, the order of a Markov chain. *The Annals of Mathematical Statistics*, *41*(1), 214–226. doi:10.1214/aoms/1177697203

Dijkstra, E. W. (1959). A note on two problems in connexion with graphs. *Numerische Mathematik*, *1*(1), 269–271. doi:10.1007/BF01386390

Duque-Caro, J. H. (1990a). Neogene stratigraphy, paleoceanography and paleobiogeography in northwest South America and the evolution of the Panama seaway. *Palaeogeography, Palaeoclimatology, Palaeoecology*, *77*(3–4), 203–234. doi:10.1016/0031-0182(90)90178-A

Duque-Caro, J. H. (1990b). The Chocó block in the northwestern corner of South America: Structural, tectonostratigraphic, and paleogeographic implications. *Journal of South American Earth Sciences*, *3*(1), 71–84. doi:10.1016/0895-9811(90)90019-W

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

Emsley, M. G. (1965). The geographical distribution of the color-pattern components of *Heloconius erato* and *Heliconius melpomene* with genetical evidence for the systematic relationship between the two species. *Zoologica*, *49*(15), 245–286.

1001 Endler, J. A. (1977). *Geographic Variation, Speciation, and Clines. Monographs In*
1002 *Population Biology*. Princeton: Princeton University Press.

1003 Fitch, W. M. (1971). Toward defining the course of evolution: Minimum change for a
1004 specific tree topology. *Systematic Zoology*, *20*(4), 406. doi:10.2307/2412116

1005 Forsman, A., & Merilaita, S. (1999). Fearful symmetry: pattern size and asymmetry
1006 affects aposematic signal efficacy. *Evolutionary Ecology*, *13*(2), 131–140.
1007 doi:10.1023/A:1006630911975

1008 Fu, L., Niu, B., Zhu, Z., Wu, S., & Li, W. (2012). CD-HIT: Accelerated for clustering
1009 the next-generation sequencing data. *Bioinformatics*, *28*(23), 3150–3152.
1010 doi:10.1093/bioinformatics/bts565

1011 Gamberale, G., & Tullberg, B. S. (1996). Evidence for a Peak-Shift in Predator
1012 Generalization among Aposematic Prey. *Proceedings of the Royal Society B:*
1013 *Biological Sciences*, *263*(1375), 1329–1334. doi:10.1098/rspb.1996.0195

1014 Garcia-Moreno, J., & Fjeldså, J. (1999). Re-evaluation of species limits in the genus
1015 *Atlapetes* based on mtDNA sequence data. *Ibis*, *141*(2), 199–207.
1016 doi:10.1111/j.1474-919X.1999.tb07542.x

1017 Gentry, A. H. (1982). Phytogeographic patterns as evidence for a Chocó refuge. In G.
1018 T. Prance (Ed.), *Biological Diversification in the Tropics* (pp. 112–136). New
1019 York: Plenum Press.

1020 González, C., Urrego, L. E., & Martínez, J. I. (2006). Late Quaternary vegetation and
1021 climate change in the Panama Basin: Palynological evidence from marine cores
1022 ODP 677B and TR 163-38. *Palaeogeography, Palaeoclimatology,*
1023 *Palaeoecology*, *234*(1), 62–80. doi:10.1016/j.palaeo.2005.10.019

1024 Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. a, Amit, I., …
1025 Regev, A. (2011). Full-length transcriptome assembly from RNA-Seq data
1026 without a reference genome. *Nature Biotechnology*, *29*(7), 644–652.
1027 doi:10.1038/nbt.1883

1028 Grafen, A. (1989). The Phylogenetic Regression. *Philosophical Transactions of the*
1029 *Royal Society B: Biological Sciences*, *326*(1233), 119–157.
1030 doi:10.1098/rstb.1989.0106

1031 Grant, T., Frost, D. R., Caldwell, J. P., Gagliardo, R., Haddad, C. F. B., Kok, P. J. R.,
1032 … Wheeler, W. C. (2006). Phylogenetic systematics of dart-poison frogs and
1033 their relatives (Amphibia: Athesphatanura: Dendrobatidae). *Bulletin of the*

American Museum of Natural History, *299*, 1–262. doi:10.1206/0003-0090(2006)299[1:PSODFA]2.0.CO;2

Grant, T., Rada, M., Anganoy-Criollo, M., Batista, A., Dias, P. H., Jeckel, A. M., … Rueda-Almonacid, J. V. (2017). Phylogenetic systematics of Dart-Poison frogs and their relatives revisited (Anura: Dendrobatoidea). *South American Journal of Herpetology*, *12*(s1), S1–S90.

Guindon, S., Dufayard, J. F., Lefort, V., Anisimova, M., Hordijk, W., & Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. *Systematic Biology*, *59*(3), 307–321. doi:10.1093/sysbio/syq010

Haffer, J. (1967). Speciation in Colombian forest birds west of the Andes. *American Museum Novitates*, *2294*, 1–57.

Haffer, J. (1969). Speciation in Amazonian Forest Birds. *Science*, *165*(3889), 131–137. doi:10.1126/science.165.3889.131

Hagman, M., & Forsman, A. (2003). Correlated evolution of conspicuous coloration and body size in poison frogs (Dendrobatidae). *Evolution*, *57*(12), 2904–2910. doi:10.1111/j.0014-3820.2003.tb01531.x

Hahn, M. W., & Nakhleh, L. (2015). Irrational exuberance for resolved species trees. *Evolution*, *70*(1), 7–17. doi:10.1111/evo.12832

Hijmans, R. J. (2017). raster: Geographic Data Analysis and Modeling. *R Package Version 2.6-7*.

Hines, H. M., Counterman, B. A., Papa, R., Albuquerque de Moura, P., Cardoso, M. Z., Linares, M., … McMillan, W. O. (2011). Wing patterning gene redefines the mimetic history of *Heliconius* butterflies. *Proceedings of the National Academy of Sciences of the United States of America*. doi:10.1073/pnas.1110096108

Hodges, E., Rooks, M., Xuan, Z., Bhattacharjee, A., Benjamin Gordon, D., Brizuela, L., … Hannon, G. J. (2009). Hybrid selection of discrete genomic intervals on custom-designed microarrays for massively parallel sequencing. *Nature Protocols*, *4*(6), 960–974. doi:10.1038/nprot.2009.68

Hodges, E., Xuan, Z., Balija, V., Kramer, M., Molla, M. N., Smith, S. W., … McCombie, W. R. (2007). Genome-wide in situ exon capture for selective resequencing. *Nature Genetics*, *39*(12), 1522–1527. doi:10.1038/ng.2007.42

Hooghiemstra, H., & Van Der Hammen, T. (1998). Neogene and Quaternary development of the neotropical rain forest: The forest refugia hypothesis, and a

literature overview. *Earth Science Reviews*, *44*(3–4), 147–183. doi:10.1016/S0012-8252(98)00027-0

Hovanitz, W. (1940). Ecological color variation in a butterfly and the problem of "protective coloration." *Ecology*, *21*(3), 371–380. doi:10.2307/1930846

Huang, X. Q., & Madan, A. (1999). CAP3: A DNA sequence assembly program. *Genome Research*, *9*(9), 868–877. Retrieved from http://sequoia.ucmerced.edu/SymBioSys/index.php

Huxley, J. S. (1942). *Evolution The Modern Synthesis*. New York City: Harper and Brothers.

James, M. E., Arenas-Castro, H., Groh, J. S., Engelstaedter, J., & Ortiz-Barrientos, D. (2020). Highly replicated evolution of parapatric ecotypes. *BioRxiv*, 2020.02.05.936401. doi:10.1101/2020.02.05.936401

Jaramillo, C., & Bayona, G. (2000). Mangrove distribution during the Holocene in Tribugá Gulf, Colombia. *Biotropica*, *32*(1), 14. doi:10.1646/0006-3606(2000)032[0014:mddthi]2.0.co;2

Jiang, H., Lei, R., Ding, S. W., & Zhu, S. (2014). Skewer: A fast and accurate adapter trimmer for next-generation sequencing paired-end reads. *BMC Bioinformatics*, *15*(1), 1–12. doi:10.1186/1471-2105-15-182

Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, *90*(430), 773–795. doi:10.1080/01621459.1995.10476572

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., … Drummond, A. (2012). Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, *28*(12), 1647–1649. doi:10.1093/bioinformatics/bts199

Köhler, G., Samietz, J., & Schielzeth, H. (2017). Morphological and colour morph clines along an altitudinal gradient in the meadow grasshopper *Pseudochorthippus parallelus*. *PLoS ONE*, *12*(12), 1–13. doi:10.1371/journal.pone.0189815

Korneliussen, T. S., Albrechtsen, A., & Nielsen, R. (2014). ANGSD: Analysis of next generation sequencing data. *BMC Bioinformatics*, *15*(1), 1–13. doi:10.1186/s12859-014-0356-4

Lanfear, R., Frandsen, P. B., Wright, A. M., Senfeld, T., & Calcott, B. (2017). Partitionfinder 2: New methods for selecting partitioned models of evolution for

molecular and morphological phylogenetic analyses. *Molecular Biology and Evolution*, *34*(3), 772–773. doi:10.1093/molbev/msw260

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, *9*(4), 357–359. doi:10.1038/nmeth.1923

Lefort, V., Desper, R., & Gascuel, O. (2015). FastME 2.0: A comprehensive, accurate, and fast distance-based phylogeny inference program. *Molecular Biology and Evolution*, *32*(10), 2798–2800. doi:10.1093/molbev/msv150

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., … Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, *25*(16), 2078–2079. doi:10.1093/bioinformatics/btp352

Link, V., Kousathanas, A., Veeramah, K., Sell, C., Scheu, A., & Wegmann, D. (2017). ATLAS: Analysis Tools for Low-depth and Ancient Samples. *BioRxiv*, 105346. doi:10.1101/105346

Magoč, T., & Salzberg, S. L. (2011). FLASH: Fast length adjustment of short reads to improve genome assemblies. *Bioinformatics*, *27*(21), 2957–2963. doi:10.1093/bioinformatics/btr507

Märki, F., & Witkop, B. (1963). The venom of the Colombian arrow poison frog *Phyllobates bicolor*. *Experientia*, *19*(8), 329–338. doi:10.1007/bf02152303

Márquez, R., Corredor, G., Galvis, C., Góez, D., & Amézquita, A. (2012). Range extension of the critically endangered true poison-dart frog, *Phyllobates terribilis* (Anura: Dendrobatidae), in western Colombia. *Acta Herpetologica*, *7*(2), 341–345. doi:10.13128/Acta_Herpetol-11387

Martins, E. P., & Hansen, T. F. (1997). Phylogenies and the comparative method: A general approach to incorporating phylogenetic information into the analysis of interspecific data. *The American Naturalist*, *149*(4), 646–667. doi:10.1086/286013

Matsumura, Shun-ichi, Yokoyama, J., Tateishi, Y., & Maki, M. (2006). Intraspecific variation of flower colour and its distribution within a sea lavender, *Limonium wrightii* (Plumbaginaceae), in the northwestern Pacific Islands. *Journal of Plant Research*, *119*(6), 625–632. doi:10.1007/s10265-006-0022-7

Matsumura, Shun'Ichi, Yokohama, J., Fukuda, T., & Maki, M. (2009). Origin of the disjunct distribution of flower colour polymorphism within *Limonium wrightii* (Plumbaginaceae) in the Ryukyu Archipelago. *Biological Journal of the Linnean Society*, *97*(4), 709–717. doi:10.1111/j.1095-8312.2009.01253.x

1136 Maxson, L. R., & Myers, C. W. (1985). Albumin evolution in tropical poison frogs
1137 (Dendrobatidae): A preliminary report. *Biotropica*, *17*(1), 50–56.
1138 doi:10.2307/2388378

1139 Mayr, E. (1942). *Systematics and the Origin of Species from, the Viewpoint of a*
1140 *Zoologist*. New York: Columbia University Press.

1141 Mayr, E. (1963). *Animal Species and Evolution*. Cambridge: Harvard University
1142 Press.

1143 McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., …
1144 DePristo, M. A. (2010). The Genome Analysis Toolkit: A MapReduce
1145 framework for analyzing next-generation DNA sequencing data. *Genome*
1146 *Research*, *20*(9), 1297–1303. doi:10.1101/gr.107524.110

1147 Meisner, J., & Albrechtsen, A. (2018). Inferring population structure and admixture
1148 proportions in low-depth NGS data. *Genetics*, *210*(2), 719–731.
1149 doi:10.1534/genetics.118.301336

1150 Meyer, M., & Kircher, M. (2010). Illumina sequencing library preparation for highly
1151 multiplexed target capture and sequencing. *Cold Spring Harbor Protocols*,
1152 *2010*(6), pdb.prot5448-pdb.prot5448. doi:10.1101/pdb.prot5448

1153 Miller, S. A., Dykes, D. D., & Polesky, H. F. (1988). A simple salting out procedure
1154 for extracting DNA from human nucleated cells. *Nucleic Acids Research*, *16*(3),
1155 1215. doi:10.1093/nar/16.3.1215

1156 Myers, C. W., & Daly, J. W. (1976). Preliminary evaluation of skin toxins and
1157 vocalizations in taxonomic and evolutionary studies of poison-dart frogs
1158 (Dendrobatidae). *Bulletin of the American Museum of Natural History*, *157*(3),
1159 173–262.

1160 Myers, C. W., Daly, J. W., & Malkin, B. (1978). A dangerously toxic new frog
1161 (*Phyllobates*) used by Emberá Indians of western Colombia, with discussion of
1162 blowgun fabrication and dart poisoning. *Bulletin of the American Museum of*
1163 *Natural History*, *161*, 307–366.

1164 Norman, J. A., Christidis, L., Joseph, L., Slikas, B., & Alpers, D. (2002). Unravelling
1165 a biogeographical knot: origin of the “leapfrog” distribution pattern of Australo-
1166 Papuan sooty owls (Strigiformes) and logrunners (Passeriformes). *Proceedings*
1167 *of the Royal Society B: Biological Sciences*, *269*(1505), 2127–2133.
1168 doi:10.1098/rspb.2002.2136

Olson, E. C., & Miller, R. L. (1958). *Morphological integration*. Chicago: University of Chicago Press.

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

Palumbi, S. R., Martin, A., Romano, S., McMillan, W. O., Stice, L., & Grabowski, G. (1991). *The simple fool's guide to PCR*. Honolulu: Department of Zoology, University of Hawaii.

Paradis, E., Claude, J., & Strimmer, K. (2004). APE: Analyses of phylogenetics and evolution in R language. *Bioinformatics*, *20*(2), 289–290. doi:10.1093/bioinformatics/btg412

Petersen, K. R., Streett, D. A., Gerritsen, A. T., Hunter, S. S., & Settles, M. L. (2015). Super deduper, fast PCR duplicate detection in fastq files. In *Proceedings of the 6th ACM Conference on Bioinformatics, Computational Biology and Health Informatics - BCB '15* (pp. 491–492). New York: ACM Press. doi:10.1145/2808719.2811568

Petkova, D., Novembre, J., & Stephens, M. (2015). Visualizing spatial population structure with estimated effective migration surfaces. *Nature Genetics*, *48*(1), 94–100. doi:10.1038/ng.3464

Pinheiro, J., Bates, D., DebRoy, S., & Sarkar, D. (2017). nlme: Linear and Nonlinear Mixed Effects Models. *R Package Version 3.1-131*.

Pough, F. H., & Taigen, T. L. (1990). Metabolic correlates of the foraging and social behaviour of dart-poison frogs. *Animal Behaviour*, *39*(1), 145–155. doi:10.1016/S0003-3472(05)80734-1

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., … Sham, P. C. (2007). PLINK: A tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics*, *81*(3), 559–575. doi:10.1086/519795

Quek, S.-P., Counterman, B. A., Albuquerque de Moura, P., Cardoso, M. Z., Marshall, C. R., McMillan, W. O., & Kronforst, M. R. (2010). Dissecting comimetic radiations in *Heliconius* reveals divergent histories of convergent butterflies. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(16), 7365–7370. doi:10.1073/pnas.0911572107

Rambaut, A., & Drummond, A. J. (2009). Tracer V1.5. *Available from http://beast.bio.ed.ac.uk/tracer*.

1203 Ramírez, L. F., & Urrego, L. E. (2002). Vegetación holocénica en el delta del río San
1204 Juan, Pacífico colombiano. In I. D. Correa & J. D. Restrepo (Eds.), *Geologia y*
1205 *Oceanografia del Dela del Río San Juan. Litoral Pacífico Colombiano.* (1st ed.,
1206 pp. 151–169). Medellín: Editorial Universidad EAFIT.

1207 Rebelo, A. G., & Siegfried, W. R. (1985). Colour and size of flowers in relation to
1208 pollination of *Erica* species. *Oecologia*, *65*(4), 584–590.
1209 doi:10.1007/BF00379677

1210 Reguera, S., Zamora-Camacho, F. J., & Moreno-Rueda, G. (2014). The lizard
1211 *Psammodromus algirus* (Squamata: Lacertidae) is darker at high altitudes.
1212 *Biological Journal of the Linnean Society*, *112*(1), 132–141.
1213 doi:10.1111/bij.12250

1214 Remsen, J. V. (1984). High incidence of "leapfrog" pattern of geographic variation in
1215 andean birds: implications for the speciation process. *Science*, *224*(4645), 171–
1216 173. doi:10.1126/science.224.4645.171

1217 Reynolds, J., Weir, B. S., & Cockerham, C. C. (1983). Estimation of the coancestry
1218 coefficient: Basis for a short-term genetic distance. *Genetics*, *105*(3), 767–779.
1219 doi:10.1534/genetics.103.025387

1220 Reynolds, R. G., & Fitzpatrick, B. M. (2007). Assortative mating in poison-dart frogs
1221 based on an ecologically important trait. *Evolution*, *61*(9), 2253–2259.
1222 doi:10.1111/j.1558-5646.2007.00174.x

1223 Richmond, J. Q., & Reeder, T. W. (2002). Evidence for parallel ecological speciation
1224 in sincid lizards of the Eumeces skiltonianus species group (Squamata:
1225 Scincidae). *Evolution*, *56*(7), 1498–1513. doi:10.1111/j.0014-
1226 3820.2002.tb01461.x

1227 Rios, E., & Álvarez-Castañeda, S. T. (2007). Environmental responses to altitudinal
1228 gradients and subspecific validity in pocket gophers (*Thomomys bottae*) from
1229 Baja California Sur, Mexico. *Journal of Mammalogy*, *88*(4), 926–934.
1230 doi:10.1644/06-mamm-a-226r1.1

1231 Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D. L., Darling, A., Hohna, S., …
1232 Huelsenbeck, J. P. (2012). MrBayes 3.2: Efficient Bayesian phylogenetic
1233 inference and model choice across a large model space. *Systematic Biology*,
1234 *61*(3), 539–542. doi:10.1093/sysbio/sys029

1235 Santos, J. C., Baquero, M., Barrio-Amorós, C., Coloma, L. A., Erdtmann, L. K.,
1236 Lima, A. P., & Cannatella, D. C. (2014). Aposematism increases acoustic

diversification and speciation in poison frogs. *Proceedings of the Royal Society B: Biological Sciences*, *281*(1796), 20141761. doi:10.1098/rspb.2014.1761

Santos, J. C., & Cannatella, D. C. (2011). Phenotypic integration emerges from aposematism and scale in poison frogs. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(15), 6175–6180. doi:10.1073/pnas.1010952108

Santos, J. C., Coloma, L. A., Summers, K., Caldwell, J. P., Ree, R., & Cannatella, D. C. (2009). Amazonian amphibian diversity is primarily derived from late miocene andean lineages. *PLoS Biology*, *7*(3), e56. doi:10.1371/journal.pbio.1000056.st014

Schliep, K. P. (2011). phangorn: phylogenetic analysis in R. *Bioinformatics*, *27*(4), 592–593. doi:10.1093/bioinformatics/btq706

Sheppard, P. M., Turner, J. R. G., Brown, K. S., Benson, W. W., & Singer, M. C. (1985). Genetics and the Evolution of Muellerian Mimicry in *Heliconius* Butterflies. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *308*(1137), 433–610. doi:10.1098/rstb.1985.0066

Silverstone, P. A. (1976). A revision of the poison-arrow frogs of the genus *Phyllobates* Bibron in Sagra (Family Dendrobatidae). *Natural History Museum of Los Angeles County, Science Bulletin*, *27*, 1–53.

Simpson, G. G. (1953). *The major features of evolution*. New York: Columbia University Press.

Simpson, J. T., Wong, K., Jackman, S. D., Schein, J. E., Jones, S. J. M., & Birol, I. (2009). ABySS: A parallel assembler for short read sequence data. *Genome Research*, *19*(6), 1117–1123. doi:10.1101/gr.089532.108

Skotte, L., Korneliussen, T. S., & Albrechtsen, A. (2013). Estimating individual admixture proportions from next generation sequencing data. *Genetics*, *195*(3), 693–702. doi:10.1534/genetics.113.154138

Slater, G. S. C., & Birney, E. (2005). Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics*, *6*, 31. doi:10.1186/1471-2105-6-31

Smith, A. F. ., Hubley, R., & Green, P. (2013). RepeatMasker Open-4.0. Available at http://www.repeatmasker.org

Summers, K., & Clough, M. E. (2001). The evolution of coloration and toxicity in the poison frog family (Dendrobatidae). *Proceedings of the National Academy of*

Sciences of the United States of America, *98*(11), 6227–6232. doi:10.1073/pnas.101134898

Summers, K., Symula, R., Clough, M., & Cronin, T. (1999). Visual mate choice in poison frogs. *Proceedings of the Royal Society B: Biological Sciences*, *266*(1434), 2141–2145. doi:10.1098/rspb.1999.0900

Sun, Y.-B., Xiong, Z.-J., Xiang, X.-Y., Liu, S.-P., Zhou, W.-W., Tu, X.-L., … Zhang, Y.-P. (2015). Whole-genome sequence of the Tibetan frog *Nanorana parkeri* and the comparative evolution of tetrapod genomes. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(11), E1257–E1262. doi:10.1073/pnas.1501764112

Toon, A., Austin, J. J., Dolman, G., Pedler, L., & Joseph, L. (2012). Evolution of arid zone birds in Australia: Leapfrog distribution patterns and mesic-arid connections in quail-thrush (*Cinclosoma*, Cinclosomatidae). *Molecular Phylogenetics and Evolution*, *62*(1), 286–295. doi:10.1016/j.ympev.2011.09.026

Tullberg, B. S., Merilaita, S., & Wiklund, C. (2005). Aposematism and crypsis combined as a result of distance dependence: Functional versatility of the colour pattern in the swallowtail butterfly larva. *Proceedings of the Royal Society B: Biological Sciences*, *272*(1570), 1315–1321. doi:10.1098/rspb.2005.3079

Turner, J. R., Johnson, M. S., & Eanes, W. F. (1979). Contrasted modes of evolution in the same genome: allozymes and adaptive change in *Heliconius*. *Proceedings of the National Academy of Sciences of the United States of America*, *76*(4), 1924–1928. doi:10.1073/pnas.76.4.1924

Turner, T. L., Hahn, M. W., & Nuzhdin, S. V. (2005). Genomic Islands of Speciation in *Anopheles gambiae*. *PLoS Biology*, *3*(9), e285. doi:10.1371/journal.pbio.0030285.st001

Twomey, E., Vestergaard, J. S., & Summers, K. (2014). Reproductive isolation related to mimetic divergence in the poison frog *Ranitomeya imitator*. *Nature Communications*, *5*, 4749. doi:10.1038/ncomms5749

Urrego, L. E., Molina, L. A., Urrego, D. H., & Ramírez, L. F. (2006). Holocene space–time succession of the Middle Atrato wetlands, Chocó biogeographic region, Colombia. *Palaeogeography, Palaeoclimatology, Palaeoecology*, *234*(1), 45–61. doi:10.1016/j.palaeo.2005.10.018

van Etten, J. (2017). R Package gdistance : Distances and Routes on Geographical Grids. *Journal of Statistical Software*, *76*(13), 1–21. doi:10.18637/jss.v076.i13

Vanzolini, P. E., & Williams, E. E. (1970). South american anoles: the geographic differentiation and evolution of the *Anolis chrysolepis* species group (Sauria, Iguanidae). *Arquivos de Zoologia*, *19*(1–2), 1–258. doi:10.11606/issn.2176-7793.v19i3-4p125-298

Vieira, F. G., Lassalle, F., Korneliussen, T. S., & Fumagalli, M. (2016). Improving the estimation of genetic distances from next-generation sequencing data. *Biological Journal of Linnean Society*, *117*(1), 139–149. doi:10.1111/bij.12511

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

Weir, B. S., & Cockerham, C. C. (1984). Estimating F-statistics for the analysis of population structure. *Evolution*, *38*(6), 1358. doi:10.2307/2408641

Widmer, A., Lötters, S., & Jungfer, K. H. (2000). A molecular phylogenetic analysis of the neotropical dart-poison frog genus *Phyllobates* (Amphibia: Dendrobatidae). *Die Naturwissenschaften*, *87*(12), 559–562. doi:10.1007/s001140050

Wu, C.-I. (2001). The genic view of the process of speciation. *Journal of Evolutionary Biology*, *14*(6), 851–865. doi:10.1046/j.1420-9101.2001.00335.x

Yang, Y., Richards-Zawacki, C. L., Devar, A., & Dugas, M. B. (2016). Poison frog color morphs express assortative mate preferences in allopatry but not sympatry. *Evolution*, *70*(12), 2778–2788. doi:10.1111/evo.13079