MS. DANIELLE RIVERA (Orcid ID : 0000-0002-9100-9945)

MR. IVAN PRATES (Orcid ID : 0000-0001-6314-8852)

DR. THOMAS J FIRNENO (Orcid ID : 0000-0002-4975-2794)

Article type : Original Article

Phylogenomics, introgression, and demographic history of South American true toads (*Rhinella*)

Danielle Rivera[1,2], Ivan Prates[3], Thomas J. Firneno Jr[1,2], Miguel Trefaut Rodrigues[4], Janalee P. Caldwell[5], Matthew K Fujita[1,2]

[1]*Department of Biology, University of Texas at Arlington, Arlington, TX, USA*

[2]*Amphibian and Reptile Diversity Research Center, University of Texas at Arlington, TX, USA.*

[3]*Department of Ecology and Evolutionary Biology and Museum of Zoology, University of Michigan, Ann Arbor, MI, USA.*

[4]*Departamento de Zoologia, Instituto de Biociências, Universidade de São Paulo, São Paulo, SP, Brazil.*

[5]*Sam Noble Museum & Department of Biology, University of Oklahoma, Norman, Oklahoma, 73072-7029, USA*

Correspondence

Danielle Rivera, Department of Biology, University of Texas at Arlington, Arlington, TX, USA.

Email: danielle.rivera@uta.edu

Running Head: Phylogenomics within *Rhinella marina*

Keywords: Introgression, Hybridization, *Rhinella*, Demographic Modeling, Phylogenomics

Abstract

The effects of genetic introgression on species boundaries and how they affect species' integrity and persistence over evolutionary time have received increased attention. The increasing availability of genomic data has revealed contrasting patterns of gene flow across genomic regions, which impose challenges to inferences of evolutionary relationships and of patterns of genetic admixture across lineages. By characterizing patterns of variation across thousands of genomic loci in a widespread complex of true toads (*Rhinella*), we assess the true extent of genetic introgression across species thought to hybridize to extreme degrees based on natural history observations and multi-locus analyses. Comprehensive geographic sampling of five large-ranged Neotropical taxa revealed multiple distinct evolutionary lineages that span large geographic areas and, at times, distinct biomes. The inferred major clades and genetic clusters largely correspond to currently recognized taxa; however, we also found evidence of cryptic diversity within taxa. While previous phylogenetic studies revealed extensive mito-nuclear discordance, our genetic clustering analyses uncovered several admixed individuals within major genetic groups. Accordingly, historical demographic analyses supported that the evolutionary history of these toads involved cross-taxon gene flow both at ancient and recent times. Lastly, ABBA-BABA tests revealed widespread allele sharing across species boundaries, a pattern that can be confidently attributed to genetic introgression as opposed to incomplete lineage sorting. These results confirm previous assertions that the evolutionary history of *Rhinella* was characterized by various levels of hybridization even across environmentally heterogeneous regions, posing exciting questions about what factors prevent complete fusion of diverging yet highly interdependent evolutionary lineages.

1 INTRODUCTION

How introgression affects reproductive isolation and speciation is an enduring question in evolutionary biology. Reproductive isolation has long been viewed as the primary factor behind lineage divergence and stable boundaries between closely related species (Avise et al., 1998; Mayr, 1963; Rabosky, 2016). When closely related populations come into contact, however, gene flow via hybridization can lead to the introgression of alleles (Mallet, 2005; O'Connell et al., 2021). Introgression levels can vary starkly across genome regions. In particular, in the presence of strong divergent selection, those loci underlying adaptive phenotypes can maintain marked differentiation even with extensive gene flow among closely related populations (Feder et al., 2012). Thus, these varying degrees of isolation across the genome may contribute to the maintenance of species boundaries despite the homogenizing effects of gene flow (Yeaman & Whitlock, 2011).

Differential introgression across genomic regions can lead to dramatic topological discordance between genealogies inferred from distinct genes, as illustrated by instances of mito-nuclear discordance (Bernardo et al., 2019; Bessa-Silva et al., 2020; Firneno et al., 2020). This gene-tree heterogeneity must be accounted for as it can make reconstructing evolutionary relationships and historical demography challenging (Carstens & Knowles, 2007; Firneno et al., 2020; Liu et al., 2010). The increasing availability of high-throughput sequencing datasets for non-model organisms has improved our ability to discern patterns of introgression in closely related species or populations (Firneno et al., 2020; Graham et al., 2018; Lavretsky et al., 2016) and thus clarify phylogenetic relationships and species limits. This is especially so in large, widely distributed species complexes with limited variation in external morphological traits and hybridization blurring species limits (Guo et al., 2016; Phuong et al., 2017; Potter et al., 2016).

The increasing availability of genome-scale datasets has also fostered the development of model-based approaches to infer historical demographic events such as population size shifts and pulses of gene flow (Portik et al., 2017; Prates, Xue et al., 2016). These approaches have transformed our understanding of how landscape and climate changes have contributed to the assembly of regional species pools, for instance by limiting dispersal, promoting speciation, or

77 leading to lineage fusion (Graham et al., 2018; Lavretsky et al., 2016; Leaché et al., 2019; Portik
78 et al., 2017). One flexible approach involves simulating population histories to compare the fit of
79 empirical genome-scale data to data simulated under alternative biogeographical scenarios (Dal
80 Vechio et al., 2019; Portik et al., 2017; Prates et al., 2016). This modeling framework can
81 facilitate hypothesis testing, such as how climate-driven habitat shifts may have led to migration,
82 introgression, or isolation across geographic regions. These approaches have been instrumental
83 to shed light on the historical factors behind present-day spatial biodiversity patterns in regions
84 that concentrate large proportions of biodiversity. This is the case of the Neotropics, where
85 demographic inference has supported that Late-Quaternary climate fluctuations and Neogene
86 geomorphological change have played a major role in shaping species range limits, genetic
87 diversity levels, and lineage divergence (Gehara et al., 2017; Pirani et al., 2020; Prates, Xue et
88 al., 2016). Nevertheless, biogeographic investigations in the Neotropics have often shown
89 geographic and taxonomic bias, which questions the generality of the mechanisms invoked to
90 explain species richness and distributions. For instance, taxa with wide ranges across South
91 America's open vegetation biomes – the dry and highly seasonal Cerrado, Caatinga, and Chaco –
92 have received relatively less attention than rainforest biotas (Fonseca et al., 2018; Gehara et al.,
93 2017; Werneck, 2011).

94 One example of a Neotropical clade whose biogeographic history remains poorly known
95 is the true South American toads, genus *Rhinella* (Bufonidae). Despite being the focus of a
96 handful of phylogeographic studies, the evolutionary relationships and species limits between
97 these toads remain elusive, perhaps due to wildly varying patterns of introgression and
98 hybridization across species (Maciel et al., 2010; Pereyra et al., 2016; Pereyra et al., 2021;
99 Sequeira et al., 2011; Vallinoto et al., 2009). As such, not only the evolutionary history of this
100 group is unclear, but so are the environmental and geographic factors that may have favored
101 introgression and its variation, or how hybridization may have contributed to lineage divergence
102 or fusion (Azevedo et al., 2003; Correa et al., 2012; Malone & Fontenot, 2008; Pereyra et al.,
103 2016; Sequeira et al., 2011). *Rhinella* is composed of multiple species complexes that are each
104 distributed across much of the Neotropics. These groups are known to harbor high levels of
105 cryptic lineage diversity, as revealed by single and multi-locus genetic analyses (Maciel et al.,

2010; Pereyra et al. 2016; Pereyra et al. 2021; Vallinoto et al., 2009). Among them is the *Rhinella marina* group, best known for the globally invasive species *R. marina*. Previous studies of this group have identified both mitochondrial and nuclear introgression across species (Azevedo et al., 2003; Maciel et al., 2010; Vallinoto et al., 2009). However, lack of data about persisting genetically admixed populations in the wild makes it difficult to assess the magnitude of presumed hybridization and how it affects species boundaries (Azevedo et al., 2003; Malone & Fontenot, 2008; Pereyra et al. 2021). Despite the ecological diversity seen in *Rhinella*, with taxa that span savannas, rainforests, and xeric shrublands, biogeographic analyses have largely focused on taxa occurring within a single biome (Sequeira et al., 2011; Thomé et al., 2010), which is also the case of other South American anuran clades (Fonseca et al., 2018; Gehara et al., 2017; Oliveira et al., 2018). As a result, how habitat transitions may contribute to patterns of gene flow and species range limits remains unclear.

In this investigation, we focus on the *R. marina* group to investigate evolutionary relationships, quantify the extent of hybridization, and examine whether landscape transitions among South America's biomes impose limits to gene flow and species ranges. For this purpose, we focus on *R. marina*, *R. poeppigii*, *R. horribilis*, *R. jimi*, and *R. schneideri* (also known as *R. diptycha*), which have established contact zones throughout the continent. We infer population structure, gene flow, and relationships based on geographically comprehensive sampling of genomic variation within each taxon. We then proceed to test alternative historical hypotheses to quantify plausible demographic events such as population size shifts and historical gene flow. With this approach, we seek to answer the following questions: what are the levels of genetic structure across and within each species? Do genomic data corroborate a pattern of widespread admixture or introgression across these species, as previously suggested based on only a few loci? Lastly, what historical demographic processes may explain species distributions and genetic diversity patterns within this clade?

2 MATERIAL AND METHODS

2.1 Sample collection

Our sampling included 191 individuals belonging to the *Rhinella marina* species group, as follows: 72 *R. marina*, 39 *R. schneideri*, 23 *R. horribilis*, 11 *R. jimi*, and nine *R. cf. poeppigii*, as well as four *R. veredas*, eight *R. rubescens*, and 25 *R. icterica* which were used in the mitochondrial analysis only. We also included samples from the *Rhinella granulosa* and *R. margaritifera* major clades within *Rhinella* as outgroups in the divergence time estimation analyses (see below). Within each species in the focal group consisting of *R. marina*, *R. horribilis*, *R. schneideri*, *R. jimi*, and *R. poeppigii*, we sample multiple individuals from each locality across their known ranges, with the exception of *R. cf. poeppigii*, which was identified as distinct from *R. marina a posteriori* based on the genetic data (see Results)(Acevedo, et al. 2016; Maciel, et al., 2010; Vallinoto, et al. 2010). Tissue samples were obtained from the MTR herpetological tissue collection hosted at Instituto de Biociências, University of São Paulo (IBUSP) with vouchers at Museum of Zoology, University of São Paulo, as well as from the Amphibian and Reptile Diversity Research Center (ARDRC), and the Louisiana State University Museum of Natural Science (LSUMNS)(Table S1).

2.2 DNA extraction, amplification, & sequencing

We extracted genomic DNA using a standard phenol-chloroform extraction protocol (Sambrook & Russell, 2006). Fragments of the mitochondrial 16S were amplified using 16Sar and 16Sbr primers and sequenced on an ABI 3730xL (Primer information and PCR conditions in the Supplementary Text S1). Sequences were edited and aligned in Geneious Prime 2020.0.4 (Identification and Accession numbers in Supplementary Table S1). We generated double-digest restriction-site associated DNA sequencing (ddRADseq) data following (Peterson, et al., 2012), with modifications as described in Streicher et al. (2014). Briefly, 200-500 ng of DNA were digested using the *Sbf*I (restriction site 5′-CCTGCAGG-3′) and *Msp*I (restriction site 5′-CCGG-3′) restriction enzymes in a single reaction using the manufacturer's recommended buffer (New England Biolabs) for 5 hr at 37°C. Digested DNA was bead-purified before ligating barcodes and index adaptors, then samples with the same index were pooled and size-selected

(415-515 bp) on a Blue Pippin Prep size selector (Sage Science). Final Library preparation was analyzed and quantified on a BioAnalyzer (Agilent) and Qubit Fluorometer 4 (Thermo Fisher Scientific). The resulting 100 bp single-end libraries were sequenced at MedGenome on an Illumina HiSeq2500.

We used the command line version of ipyrad v. 0.9.45 (Eaton & Overcast, 2020) (available at https://ipyrad.readthedocs.io) to de-multiplex and assign reads to individuals based on sequence barcodes (allowing no mismatches from individual barcodes), perform *reference* read assembly (minimum clustering similarity threshold = 0.90), align reads into loci, and call single nucleotide polymorphisms (SNPs). As a reference, we used the *Rhinella marina* genome (Edwards et al., 2018). A minimum Phred quality score (= 33), sequence coverage (= 6x), read length (= 35 bp), and maximum proportion of heterozygous sites per locus (= 0.5) were enforced, while ensuring that variable sites had no more than two alleles (i.e., a diploid genome). Following the initial assembly, we used Matrix Condenser (de Medeiros & Farrell, 2018) to assess levels of missing data across samples and then re-assembled our dataset to ensure a minimum sample coverage of less than 35% missing loci within each sample and at least 75% of samples at each locus. This strategy resulted in a final dataset composed of 49,376 SNPs at 3,318 RAD loci with less than 12% missing data. Additionally, Weir and Cockerham mean F_{ST} estimates for the ddRADseq dataset using VCFTools (Danecek et al., 2011) and Nei's G_{ST} for the mitochondrial dataset were calculated using the R package `mmod` (Winter, 2012).

2.3 Inferring population structure and genetic admixture

Based on the ddRAD data, we used a genetic clustering approach to estimate the number of demes and if admixture was present among them. We assembled a SNP dataset as described above but excluding outgroups and using only one SNP per RAD locus to maximize sampling of independent SNPs. This approach resulted in a dataset composed of 3,314 SNPs. Genetic clustering was performed using the maximum likelihood method `ADMIXTURE`, testing up to 15 populations with 20 replicates per K and a 10-fold cross-validation (Alexander, et al., 2009; Portik, 2016). The best K was determined by assessing the replicate with the lowest cross-

validation error. To further characterize population structure, we used the non-parametric method of discriminant analysis of principal components (DAPC), implemented in the R package `adegenet` (Jombart & Ahmed, 2011; Jombart, et al., 2010). The *find.clusters* function was used to test the fit of 1-15 clusters (K). The K with the lowest Bayesian information criterion (BIC) score was considered the best-fit number of demes. The resulting ancestry coefficient matrices (Q-matrices) were then imported into QGIS (QGIS Development Team 2020. QGIS Geographic Information System. Open Source Geospatial Foundation Project. http://qgis.osgeo.org) to make average-per-locality pie-charts indicating admixture levels at each sampled locality for each species.

2.4 Phylogenetic Analyses

We reconstructed maximum likelihood phylogenies for both the mitochondrial and the unlinked ddRADseq loci datasets using `IQTREE` v2.1.2, utilizing the built-in model selection tool ModelFinder Plus, implementing 1000 ultrafast bootstraps (Hoang et al., 2018; Kalyaanamoorthy et al., 2017; Nguyen et al., 2015). We specified that all partitions share the same branch lengths and selected the best-fit partitioning scheme by merging partitions (which implements the greedy algorithm of PartitionFinder), testing the “MrBayes" substitution model set and considering the top 10% partition schemes using the fast relaxed clustering algorithm from PartitionFinder2 to save computational time (Chernomor et al., 2016; Lanfear et al., 2012; Lanfear et al., 2014; Lanfear et al., 2017). In addition, we performed phylogenetic inference under a Bayesian framework for both datasets using MrBayes 3.2.6 (Ronquist et al. 2012), implementing three independent runs of four Markov chains of 10 million generations each and sampling every 1,000 generations with the first 25% generations discarded as burn-in. We used Tracer 1.7 (Rambaut et al. 2018) to assess whether Markov chain mixing was adequate (effective sample sizes > 200) and to visually assess model parameter stationarity and convergence between runs. We then summarized a 50% majority-rule consensus tree.

To estimate divergence dates and inform the delimitation of species boundaries, we conducted Bayesian divergence dating analyses based on the mtDNA dataset in `BEAST2` using

218 an HKY model of nucleotide substitution, a log-normal relaxed molecular clock, and a
219 coalescent constant population size prior. We follow Pramuk et al. (2008) by enforcing a
220 minimum age for the root node between the *Rhinella marina* and *R. granulosa* species
221 complexes based on a *Rhinella marina* fossil from the Clarendonian North American Stage of
222 the middle Miocene (ca. 11 mya), described by Sanchiz (1998), and employed a log-normally
223 distributed prior with a standard deviation of 0.5. We ran this analysis for 20 million generations
224 sampling every 1000 generations. Runs were assessed using TRACER v1.6 (Rambaut &
225 Drummond, 2009) to examine convergence. We then summarized a maximum clade credibility
226 tree using `TreeAnnotator` discarding the first 25% of trees as burn-in (Bouckaert et al., 2019;
227 Stamatakis, 2014). All phylogenetic tree-based methods were analyzed on Cipres (Miller et al.,
228 2010).

229

230 **2.5 Demographic modeling with *∂a∂i***

231 We used the diffusion-approximation method ∂a∂i (Gutenkunst, et al., 2009) to test alternative
232 hypotheses of population history within the *Rhinella marina* clade based on species ranges,
233 previously reported potential hybridization events, common biogeographic patterns exhibited by
234 amphibians across this region, as well as computational limitations. Using both two- and three-
235 dimensional joint site frequency spectra (2D- and 3D-JSFS), we divided the dataset into two
236 population subsets: one comprised of *R. marina, R. horribilis*, and *R. jimi*; and another
237 comprised of *R. schneideri* and *R. cf. poeppigii*. Folded-JSFS datasets were used in all ∂a∂i
238 analyses.

239 We filtered the ddRAD data for each subset to allow no more than 35% missing data
240 from any sample, removed singletons, and selected one SNP per locus using VCFtools (Danecek
241 et al., 2011; Gutenkunst et al., 2009; Portik et al., 2017). We then used the *stacks_pipeline*
242 Python script from Portik et al. (2017) to create the SNP input file for ∂a∂i. We used the python
243 script easySFS (https://github.com/isaacovercast/easySFS) to determine the projection size of
244 each population, which was determined by balancing a downscaled sample size that maximized
245 the number of segregating sites (Gutenkunst et al., 2009; Marth et al., 2004). In ∂a∂i, we then

246 tested a range of extrapolation grid sizes (40-100 in 10-unit increments, e.g., 50, 60, 70 to 100,
247 110, 120) in the divergence-with-no-migration model to determine the appropriate grid size by
248 selecting the model with the highest log-likelihood, implementing 4 rounds of optimization
249 totaling 100 replicates. Once an optimal grid size was determined, each tested model was run 3
250 times independently.

251 For the subset composed of *R. marina, R. horribilis*, and *R. jimi*, we used a 3D-JSFS to
252 test models incorporating gene flow at different times, including those accounting for ancient
253 migration, recent secondary contact, and past simultaneous divergence of all lineages (Fig. S5).
254 In addition to a model of 1) divergence with no migration, we tested the following models: 2)
255 divergence with continuous symmetric gene flow between all populations; 3) divergence with
256 continuous symmetric gene flow between geographically adjacent populations; 4) isolation
257 followed by secondary contact; 5) simultaneous divergence in isolation followed by more recent
258 secondary contact between adjacent populations; 6) simultaneous divergence with continuous
259 symmetric migration between adjacent populations; 7) ancient migration with very recent
260 isolation; 8) ancient migration with a longer period of recent isolation; 9) a short ancient period
261 of migration followed by a long period of isolation; and 10) ancient migration followed by
262 lineage isolation and population size change across two epochs (Barratt et al., 2018; Portik et al.,
263 2017).

264 For the subset composed of *R. schneideri* and *R. cf. poeppigii*, we tested 2D-JSFS models
265 incorporating differing migration levels at different time periods (Fig. S6). In addition to a model
266 of 1) divergence with no migration, we tested the following models: 2) divergence with
267 continuous symmetric migration; 3) divergence with continuous asymmetric migration; 4)
268 divergence with continuous symmetric migration and a varying rate of migration across two
269 epochs; 5) divergence with continuous asymmetric migration and a varying rate of migration
270 across two epochs; 6) divergence in isolation, followed by symmetric secondary contact; 7)
271 divergence in isolation, followed by asymmetric secondary contact; 8) ancient symmetric
272 migration then subsequent isolation; 9) ancient asymmetric migration then subsequent isolation;
273 10) divergence in isolation followed by symmetric secondary contact with subsequent isolation;

and 11) divergence in isolation followed by asymmetric secondary contact with subsequent isolation (Charles et al., 2018; Portik et al., 2017).

Best-fit models were chosen based on log-likelihood values, which we assumed to be the true likelihood (and not composite likelihood) given that we have kept only one SNP per RAD locus. Replicates with the consistently highest likelihood scores were used to calculate and compare models using the Akaike information criterion (AIC).

2.6 Inferring gene flow

To further explore potential hybridization between taxa, we inferred Patterson's D statistic, or the ABBA-BABA statistic, and the related admixture fraction estimates, or f_4-ratio statistics, based on the ddRAD data using `Dsuite` (Malinsky, et al., 2020; Patterson et al., 2012). Tests were designed with a 4-taxon fixed phylogeny (((P1,P2)P3)O), wherein a typical ancestral ("A") and derived ("B") allele pattern should follow BBAA. Under incomplete lineage sorting, conflicting ABBA and BABA patterns should occur in equal frequencies, resulting in a D statistic = 0. If, however, introgression between P3 and P1 or P2 has occurred, there should be an excess of these patterns and a D statistic significantly different from 0, with significance detected using a block-jackknifing approach (Durand, et al., 2011; Green et al., 2010; Malinsky et al., 2020; Patterson et al., 2012). We used the *f*-branch or f_b(C) metric to tease apart potentially correlated f_4-ratio statistics and estimate gene flow events between internal branches on the phylogeny (Malinsky et al., 2018; Martin et al., 2013). `Dsuite` uses a VCF file and a jackknifing approach to assess correlations in allele frequencies between closely-related species (Malinsky et al., 2020). Within `Dsuite`, we used the *Dtrios* and *Fbranch* programs to identify introgression between all combinations of species, as well as potential direction of gene flow, specifying *Rhinella veredas* as an outgroup and applying the Benjamini-Hochberg (BH) correction to control for the false discovery rate.

3 RESULTS

3.1 Phylogenetic relationships

The 16S phylogeny (160 individuals; 481 base pairs) suggested little phylogenetic structure within the *Rhinella marina* complex. One clade included most of the *R. horribilis* samples, while individuals from the remaining taxa formed a polytomy (Fig. S1). Maximum likelihood and Bayesian phylogenies based on the ddRADseq dataset (128 individuals) resulted in fully concordant phylogenies (Fig. 1). These analyses inferred six highly supported clades, two corresponding to *R. marina* and the other four corresponding to *Rhinella schneideri*, *R. horribilis*, *R. jimi*, and a clade tentatively assigned to *R. cf. poeppigii* (BS = 100; PP = 1.0; Fig. 1-3). These putative *R. poeppigii* samples were originally identified as *R. marina*, which would render *R. marina* to be paraphyletic; however, after re-examining these specimens morphologically, we were able to positively identify samples from western Amazonia in Brazil's state of Acre as *R. poeppigii*, while closely related samples from eastern localities in the state of Pará were morphologically more similar to *R. marina* (Fig. S7). Pairwise Nei's G_{ST} estimates for the 16S data were much lower than the Weir and Cockerham weighted F_{ST} estimates for the ddRADseq data. Across all taxa, the average pairwise G_{ST} for the mitochondrial data was 0.117 (0.025-0.228) while the average pairwise F_{ST} for the nuclear data was 0.506 (0.379-0.843) (Table S2).

The time-calibrated phylogeny based on the 16S mitochondrial data dated the root of *Rhinella marina* at 5.58 mya (95% HPD: 2.75–9.40 million years ago; Fig 4). Though many relationships had poor support due to lack of variability within the locus, some clades showed high support, including a clade with most of the *R. horribilis* samples, which was dated at 1.03 mya (95% HPD: 0.35–2.04 mya). Two samples not included in this clade were samples distributed in the northern Andes, which clustered with other *R. marina* samples (Fig 4). One highly supported clade consisted of two clades with a divergence date of 0.67 mya (95% HPD: 0.18–1.48 mya): one clade consisted of *Rhinella cf. poeppigii* samples from eastern Amazonia, and other consisted of the western Amazonia *R. poeppigii* sample as well as *R. marina* from southern Amazonia (Fig. 4). Additionally, *R. granulosa* is estimated to be sister to the *R. marina*

329 complex, with *R. margaritifera* more distantly related. Due to the lack of variation within the *R.*
330 *marina* group, we interpret dates within this complex with caution.

331

332 **3.2 Population structure**

333 Despite the high posterior probabilities of each clade in our ddRAD tree, the ADMIXTURE
334 results supported genetic admixture both within and across multiple taxa within the *Rhinella*
335 *marina* complex (Fig. 1), with a best-fit K of 7. Each ddRAD clade corresponded to a cluster,
336 except for the *Rhinella schneideri* clade which consisted of two non-reciprocally monophyletic
337 units. *Rhinella horribilis* (blue, Fig. 1-2) showed admixture from the northern cluster of *R.*
338 *marina* into one northern Andes locality. One cluster of *Rhinella marina* was relegated to
339 northern Amazonia (light green, Fig. 1-2), while the other cluster showed a cline of admixture
340 across its western and southern Amazonia clades (light green to purple, Fig. 1-2) and admixture
341 from *R. jimi* and *R. cf. poeppigii* (dark green and orange, Fig. 1). The two genetic clusters within
342 *R. schneideri* (pink and yellow, Fig. 1,3) followed an east-west admixture gradient across the
343 Cerrado to the northern Atlantic Forest, as well as intermediate ecotones. *Rhinella jimi* occurs
344 mostly in the semi-arid Caatinga shrublands of northeastern Brazil, but also in the adjacent
345 coastal rainforest (dark green, Fig. 1-2). The DAPC analysis supported this clustering scheme as
346 well; however, BIC scores suggested similar support for six to eight clusters (Fig. S4). The seven
347 clusters recovered were concordant with phylogenetic structure (Fig. 1).

348

349 **3.3 Demographic inference**

350 For the subset composed of *R. marina, R. horribilis*, and *R. jimi*, the best 3D-JFSF model was
351 one that incorporated ancient migration with a short period of recent isolation since divergence,
352 with a log-likelihood of -1572.69 and AIC of 3165.38 (Fig. 5, Table S3). This model included an
353 ancient period of migration between all lineages (mA, Fig. 5), then another period of migration
354 between geographically adjacent species after the divergence between *R. marina* and *R. jimi*, and
355 then subsequent lineage isolation. Parameter estimates indicated a much longer ancient period of

migration between all lineages with smaller migration rates (T1 = 10.82; mA = 0.05) compared to the shorter time of adjacent-species migration with higher rates of migration (T2 = 0.12; m1 = 1.36; m2 = 0.85) and the shortest period of isolation (T3 = 0.10) (Table S3).

For the subset composed of *R. schneideri* and *R. cf. poeppigii*, the best 2D-JFSF model incorporated divergence in isolation followed by secondary contact with asymmetric gene flow, with a log-likelihood of -539.27 and AIC of 1090.54 (Fig. 5, Table S3). Parameter estimates inferred a period of divergence in isolation (T1 = 0.07) with a shorter period of secondary contact (T2 = 0.01) and a much higher rate of migration from *R. cf. poeppigii* into *R. schneideri* (m12 = 15.5) than from *R. schneideri* into *R. cf. poeppigii* (m21 = 1.82) (Table S3).

3.4 D-statistics

Nearly all topological trios tested (((P1,P2)P3)O) had significant *D*-statistics (Table S4), indicating that the majority of gene flow within this group is not due to incomplete lineage sorting. The *R. jimi-marina-horribilis* trio was not significant ($p > 0.05$), indicating that we could not reject the null hypothesis of no gene flow, with ABBA-BABA patterns arising solely due to incomplete lineage sorting (Malinsky et al., 2020). *D*-statistics for all significant trios ranged from 0.12 to 0.49 (Table S4). The highest *D*-statistics were for *R. horribilis-jimi-schneideri* (0.49), *R. marina-jimi-schneideri* (0.37), and *R. schneideri-poeppigii-marina* (0.30). The f_b(C) statistic is a summary of f_4 admixture ratios and shows excess allele sharing between the branch on the y-axis and the sample on the x-axis (Malinsky et al. 2018). The f_b(C) statistics indicated the highest percentages of gene flow between *R. cf. poeppigii* and *R. marina* (11%), between *R. cf. poeppigii* and *R. horribilis* (8%), and between *R. jimi* and *R. schneideri* (7%) (Fig. 6, Table S5).

4 DISCUSSION

Based on comprehensive geographic and genomic sampling within a clade of South American toads, this investigation found evidence of multiple distinct evolutionary lineages that span large

383 geographic areas and, at times, distinct biomes. The inferred major clades and genetic clusters
384 largely correspond to currently recognized taxa within *Rhinella*; however, we also found
385 evidence of potentially cryptic diversity within *R. marina*, *R. schneideri*, and potentially *R.*
386 *poeppigii*. Genetic clustering analyses suggested that many of the inferred groups include
387 admixed individuals. Accordingly, demographic analyses supported that the evolutionary history
388 of these toads involved cross-taxon gene flow both at ancient (in the case of *R. marina, R.*
389 *horribilis*, and *R. jimi*) and recent (in the case of *R. schneideri* and *R. cf. poeppigii*) times. Both
390 demographic inference and ABBA-BABA tests inferred patterns of genetic introgression across
391 species, supporting previous assertions that the evolutionary history of *Rhinella* was
392 characterized by various levels of hybridization (Pereyra et al. 2016; Sequeira et al., 2011).

393

394 **4.1 Phylogenetic patterns and species boundaries**

395 The phylogenetic findings of this study improve our knowledge about species diversity and
396 distributions in South America. Our sampling validates previous reports of *Rhinella poeppigii*
397 present in western Amazonia (Venâncio et al., 2017). *Rhinella poeppigii* has a history of both
398 taxonomic uncertainty and misidentification, due to its similarity to *R. marina* (De la Riva, 2002;
399 Venâncio et al., 2017; Venegas & Ron, 2014). After the first individuals were identified and
400 collected in Ecuador, subsequent specimens collected in the region that were previously
401 misidentified were discovered at Museo de Zoología, Pontificia Universidad Católica del
402 Ecuador (QCAZ) (Venegas & Ron, 2014). In this study we included another individual from
403 Porto Walter, Acre, Brazil, which further corroborates *R. poeppigii* extending into Brazil.
404 Furthermore, we uncovered a group of *R. cf. poeppigii* specimens in eastern Amazonia near the
405 Belo Monte Hydroelectric dam on the Xingu River (Fig. 3). These samples, however, do not
406 display distinct *R. poeppigii* morphology, and in fact are more similar morphologically to *R.*
407 *marina*, to which they were originally assigned (Fig. S7). Unfortunately, as sampling of this
408 clade was initially unintentional, we sampled only a few individuals that could confidently be
409 called *R. poeppigii* that did not occur across the range of *R. poeppigii*, which may be
410 misrepresenting the genetic admixture visualized within this clade (Fig. 1,3). Given this

411 restricted sampling and the more than 2,000 km distance in sampled individuals, it may be that
412 eastern *R. cf. poeppigii* is actually a yet undescribed cryptic species within the *Rhinella marina*
413 complex.

414 The mitochondrial 16S rRNA marker has been used extensively for identification and
415 barcoding of amphibians (Maya-Soriano et al., 2012; Rockney et al., 2015; Vences et al., 2005).
416 Despite this marker being extremely useful in taxonomic identification for a number of closely
417 related species (Firneno & Townsend, 2019), even within the *Rhinella* genus (Pereyra et al.
418 2016), there is an inherent lack of diversity recovered across all focal species within the *R.*
419 *marina* complex (Fig. S1-S2). It is possible that purifying selection has acted on this region of
420 the mitochondrial genome, thereby greatly reducing genetic diversity across the complex
421 (Charlesworth et al., 1995; Cvijović et al., 2018). Considering that processes like purifying
422 selection can also reduce genetic diversity at linked neutral sites, previous estimates of potential
423 introgression within *Rhinella* species using mitochondrial data may be similarly affected
424 (Cvijović et al., 2018). This phenomenon could have resulted in an overestimation of shared loci
425 by any other means, such as hybridization, as opposed to a constraint on particular loci. With the
426 16S fragment sequenced being relatively short (~480 bp), an analysis of the entire 16S rRNA
427 gene or even the whole-mitochondrial genome in this group could prove useful in disentangling
428 the reasons for such low genetic diversity seen here.

429 By contrast, despite evidence of admixture both within and between species, nuclear data
430 estimated a phylogeny with substantial structure and support (Fig. 1). When compared to other
431 phylogenies generated with single or multi-locus datasets, high-throughput sequencing of the
432 *Rhinella marina* complex has revealed a surprising amount of genetic complexity, introgression,
433 and interspecific resolution (Bessa-Silva et al., 2020; Maciel et al., 2010; Vallinoto et al., 2009).
434 These patterns suggest that in groups with such complex demographic histories, and especially
435 those with a likelihood of hybridization between divergent populations or species, large-scale
436 genetic data can be very useful in disentangling relationships and histories. These types of data
437 could be utilized in future studies to identify potentially adaptive regions of the genome that
438 correlate with phenotypic or ecological differences between populations.

4.2 Biogeographic drivers of species range limits

Inferred species range limits can be attributed to both present-day spatial environmental gradients and the history of topographic change in South America, as suggested for a number of other South American taxa (Carnaval et al., 2009; Fonseca et al., 2018; Gehara et al., 2014; Prates, Rivera et al., 2016). Mitochondrial divergence time analyses are consistent with the idea that the Andean uplift contributed to divergence between *R. marina* and *R. horribilis* (Fig. 4); pronounced genetic divergence between populations on each side of the Andean chain supports the recent recognition of *R. horribilis* as a taxon distinct from *R. marina* (Vallinoto et al., 2009). While the Andes likely limits contemporary gene flow between these two taxa, our finding of admixture between them suggests that the northern Andes may be a semi-permeable barrier (Fig. 2), in agreement with patterns seen in other organisms (Acevedo et al., 2016; Bessa-Silva et al., 2020; Maciel et al., 2010). Additionally, like other amphibians (Noonan & Wray, 2006) and reptiles (Gamble et al., 2008), the extensive fluvial network formed in western Amazonia by periodic Miocene flooding, known as the Pebas formation, may have contributed to divergence not only between *R. horribilis* and *R. marina*, but also between the northeast and south-southwestern Amazonian clades within *R. marina* (Vallinoto et al., 2009; Wesselingh & Salo, 2006). *Rhinella marina*, which is comprised of two well-supported clades, is distributed across Amazonian climates, which are known to have asynchronous historical eastern-western climatic cycles and have had an effect on species composition and genetic diversity within the biome (Cheng et al., 2013; Prates, Rivera et al., 2016). Considering that the distinct clades have a northern-southern distribution, as opposed to an eastern-western distribution, however, it may be more plausible that geographic barriers, such as fluctuating fluvial networks from the Miocene through the Pleistocene (Cooke et al., 2012; Lundberg et al., 1998), have had a higher impact in promoting divergence between these clades within *R. marina*.

Similar to what is observed within *R. marina*, we see patterns of species distributed across environmental gradients repeated across the phylogeny; *R. schneideri* is distributed across the Cerrado, through Cerrado-Caatinga-Atlantic Forest ecotones, and into the northern Atlantic

Forest, with an east-west gradient of admixture (Fig. 1,3). The Seasonally Dry Tropical Forests and savannas of South America have been known to harbor complex and cryptic genetic diversity and have been especially affected by Quaternary climate fluctuations (Bandeira et al., 2021; Fonseca et al., 2018; Gehara et al., 2017; Prado et al., 2012; Vasconcellos et al., 2019; Werneck et al., 2015). Considering the phylogenetic pattern that we see within *R. schneideri*, we can posit that this species expanded eastward during Plio-Pleistocene climate change (Bandeira et al., 2021; Lisiecki & Raymo, 2007). Paleoclimatic modeling of the biogeographic history and niche of *R. schneideri* on a finer scale is recommended to validate this hypothesis.

A puzzling biogeographic pattern that emerged from our results is the extremely disjunct distribution between *R. poeppigii* in western Amazonia and its sister clade, *R. cf. poeppigii*, from eastern Brazilian Amazonia, more than 2,000 km apart. This mysterious pattern has also been reported for other herpetofaunal species, including the lizards *Anolis* trachyderma (Ribeiro-Júnior, 2015) and *Potamites ecpleopus* (Ribeiro-Júnior & Amaral, 2017) and the horned treefrog *Hemiphractus scutatus* (de Lima Moraes & Pavan, 2018). Despite this large geographic distance, as well as the effects of contrasting climatic seasonality between the eastern and western localities in this region on other herpetofauna (Cheng et al., 2013; Prates, Rivera et al., 2016; Wang et al., 2017), this and other studies indicate limited genetic divergence across disjunct regions (de Lima Moraes & Pavan, 2018). A comprehensive analysis of museum specimens and available tissues from these areas, in conjunction with a more thorough sampling of *R. poeppigii* across its range, will be required to confirm this unexpected pattern of genetic divergence within this group.

4.3 Hybridization and introgression

The interspecific relationships inferred with historical demographic modeling suggest extremely varied patterns of migration and hybridization through time within the *Rhinella marina* complex. Our study indicates that species within this group have diverged across multiple biomes and amassed significant genetic differentiation despite continuous gene flow among species (Fig. 5). Many of the species within the *Rhinella marina* complex also have a shared introgressive history

(Fig. 1,6; Table S4-S5). Hypothesis testing of demographic models suggests that the *R. marina-horribilis-jimi* clade continued to exchange genes throughout its dispersion across the continent, and species within this clade exchanged genes with other species within the complex (Fig. 1,5-6, Table S5). *Rhinella* species have long been observed to have overlapping ranges with the potential for hybridization, especially given the propensity for these toads to participate in "mating balls", wherein different toad species – and sometimes from different genera – seemingly mate with one another (Fontenot et al., 2011; Maciel et al., 2010; Pereyra et al. 2016; Sequeira et al., 2011; Thomé et al., 2012). There is a relative lack of data, however, surrounding behavioral aspects of intraspecific interactions among species outside of these mating balls. Despite evidence of gene flow between species, there was no evidence of population-wide hybridization or the presence of hybrid species within our sampling. Potential proposed hybridization events have been reported within or between *Rhinella* species groups, such as within the *R. granulosa* complex (Guerra et al., 2011; Pereyra et al. 2016) and the *R. crucifer* complex (Júnior et al., 2004; Thomé et al., 2012), where either instances of morphologically intermediate individuals or hybrid populations have been reported. Much of the speculation surrounding hybridization in Neotropical toads has been accompanied by a lack of data from natural populations to assess the biological reality of presumed hybrid species (Fontenot et al., 2011; Malone & Fontenot, 2008; Thomé et al., 2012). Within the *R. marina* group, however, we found that recurrent gene flow between species at low levels is much more prevalent than the persistence of repeated gene flow in massive multi-species mating events.

ACKNOWLEDGEMENTS

We thank all the Brazilian collectors who made this study possible and particularly students in MTR's laboratory for field collection and support. Brazil's Instituto Chico Mendes de Conservação da Biodiversidade issued collecting permits (SISBIO 36753-1, 36753-4, and 27290-3). This work was co-funded by Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP; BIOTA 2013/50297-0), the National Science Foundation (DEB 1343578), and the National Aeronautics and Space Administration through the Dimensions of Biodiversity

Program. MTR acknowledges additional funding from FAPESP grants 2003/10335-8, 2011/50146-6, 2011/50206-9, 2012/15754-8, and 2017/08357-6. DR was funded by NSF GRFP. IP was funded by NSF grant DEB 1754398.

DATA AVAILABILITY STATEMENT

16S sequences and demultiplexed genetic data are available through NCBI (BioProject PRJNA772164; Table S1) and associated scripts and input files are available through dryad: https://doi.org/10.5061/dryad.7pvmcvdtp.

AUTHOR CONTRIBUTIONS

DR and MKF conceived and designed the study. DR, MKF, MTR and IP acquired funding. MTR, IP, and JPC provided samples. DR, IP, and TJF wrote software used for analysis. DR and TJF analyzed the data. DR, IP, and MKF wrote the manuscript with input from all authors.

REFERENCES

Acevedo, A. A., Lampo, M., & Cipriani, R. (2016). The cane or marine toad, *Rhinella marina* (Anura, Bufonidae): two genetically and morphologically distinct species. *Zootaxa*, *4103*(6), 574–586.

Alexander, D. H., Novembre, J., & Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, *19*(9), 1655–1664.

Avise, J. C., Walker, D., & Johns, G. C. (1998). Speciation durations and Pleistocene effects on vertebrate phylogeography. *Proceedings. Biological Sciences / The Royal Society*, *265*(1407), 1707–1712.

Azevedo, M. F. C., Foresti, F., Ramos, P. R. R., & Jim, J. (2003). Comparative cytogenetic studies of *Bufo ictericus*, *B. paracnemis* (Amphibia, Anura) and an intermediate form in sympatry. *Genetics and Molecular Biology*, *26*(3), 289–294.

Bandeira, L. N., Villalobos, F., Werneck, F. P., Peterson, A. T., & Anciães, M. (2021). Different elevational environments dictate contrasting patterns of niche evolution in Neotropical *Pithecopus* treefrog species. *Biotropica*, (btp.12929). doi: 10.1111/btp.12929

Barratt, C. D., Bwong, B. A., Jehle, R., Liedtke, H. C., Nagel, P., Onstein, R. E., … Loader, S. P. (2018). Vanishing refuge? Testing the forest refuge hypothesis in coastal East Africa using genome-wide sequence data for seven amphibians. *Molecular Ecology*, *27*(21), 4289–4308.

Bernardo, P. H., Sánchez-Ramírez, S., Sánchez-Pacheco, S. J., Álvarez-Castañeda, S. T., Aguilera-Miller, E. F., Mendez-de la Cruz, F. R., & Murphy, R. W. (2019). Extreme mito-nuclear discordance in a

peninsular lizard: the role of drift, selection, and climate. *Heredity*, *123*(3), 359–370.

Bessa-Silva, A., Vallinoto, M., Sampaio, I., Flores-Villela, O. A., Smith, E. N., & Sequeira, F. (2020). The roles of vicariance and dispersal in the differentiation of two species of the *Rhinella marina* species complex. *Molecular Phylogenetics and Evolution*, *145*, 106723.

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

Carnaval, A. C., Hickerson, M. J., Haddad, C. F. B., Rodrigues, M. T., & Moritz, C. (2009). Stability predicts genetic diversity in the Brazilian Atlantic forest hotspot. *Science*, *323*(5915), 785–789.

Carstens, B. C., & Knowles, L. L. (2007). Estimating species phylogeny from gene-tree probabilities despite incomplete lineage sorting: an example from *Melanoplus* grasshoppers. *Systematic Biology*, *56*(3), 400–411.

Charles, K. L., Bell, R. C., Blackburn, D. C., Burger, M., Fujita, M. K., Gvoždík, V., … Portik, D. M. (2018). Sky, sea, and forest islands: diversification in the African leaf-folding frog *Afrixalus paradorsalis* (Anura: Hyperoliidae) of the Lower Guineo-Congolian rain forest. *Journal of Biogeography*, *45*(8), 1781–1794.

Charlesworth, D., Charlesworth, B., & Morgan, M. T. (1995). The pattern of neutral molecular variation under the background selection model. *Genetics*, *141*(4), 1619–1632.

Cheng, H., Sinha, A., Cruz, F. W., Wang, X., Edwards, R. L., d'Horta, F. M., … Auler, A. S. (2013). Climate change patterns in Amazonia and biodiversity. *Nature Communications*, *4*, 1411.

Chernomor, O., von Haeseler, A., & Minh, B. Q. (2016). Terrace aware data structure for phylogenomic inference from supermatrices. *Systematic Biology*, *65*(6), 997–1008.

Cooke, G. M., Chao, N. L., & Beheregaray, L. B. (2012). Marine incursions, cryptic species and ecological diversification in Amazonia: the biogeographic history of the croaker genus *Plagioscion* (Sciaenidae). *Journal of Biogeography*, *39*(4), 724–738.

Correa, C. L., Méndez, M. A., Veloso, A., & Sallaberry, M. (2012). Genetic and reproductive evidence of natural hybridization between the sister species *Rhinella atacamensis* and *Rhinella arunco* (Anura, Bufonidae). *Journal of Herpetology*, Vol. 46, pp. 568–577. doi: 10.1670/10-266

Cvijović, I., Good, B. H., & Desai, M. M. (2018). The effect of strong purifying selection on genetic diversity. *Genetics*, *209*(4), 1235–1278.

Dal Vechio, F., Prates, I., Grazziotin, F. G., Zaher, H., Grabosky, R., & Rodrigues, M. T. (2019). Rain forest shifts through time and riverine barriers shaped the diversification of South American

terrestrial pit vipers (*Bothrops jararacussu* species group). *Journal of Biogeography*, *47*(2), 516–526.

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., … 1000 Genomes Project Analysis Group. (2011). The variant call format and VCFtools. *Bioinformatics* , *27*(15), 2156–2158.

De la Riva, I. (2002). Taxonomy and distribution of the South American toad *Bufo poeppigii* Tschudi, 1845 (Amphibia, Anura, Bufonidae). *Graellsia / Editada Por El Instituto Espanol de Entomologia Del Consejo Superior de Investigaciones Cientificas*, *58*(1), 49–57.

de Lima Moraes, L. J. C., & Pavan, D. (2018). On the occurrence of *Hemiphractus scutatus* (Spix, 1824) (Anura: Hemiphractidae) in eastern Amazonia. *Amphibian & Reptile Conservation*, *12*(1), 5–14 (e151).

de Medeiros, B. A. S., & Farrell, B. D. (2018). Whole-genome amplification in double-digest RADseq results in adequate libraries but fewer sequenced loci. *PeerJ*, *6*, e5089.

Durand, E. Y., Patterson, N., Reich, D., & Slatkin, M. (2011). Testing for ancient admixture between closely related populations. *Molecular Biology and Evolution*, *28*(8), 2239–2252.

Eaton, D. A. R., & Overcast, I. (2020). ipyrad: Interactive assembly and analysis of RADseq datasets. *Bioinformatics* , *36*(8), 2592–2594.

Edwards, R. J., Tuipulotu, D. E., Amos, T. G., O'Meally, D., Richardson, M. F., Russell, T. L., … White, P. A. (2018). Draft genome assembly of the invasive cane toad, *Rhinella marina*. *GigaScience*, Vol. 7. doi: 10.1093/gigascience/giy095

Feder, J. L., Egan, S. P., & Nosil, P. (2012). The genomics of speciation-with-gene-flow. *Trends in Genetics: TIG*, *28*(7), 342–350.

Firneno, T. J., Jr, O'Neill, J. R., Portik, D. M., Emery, A. H., Townsend, J. H., & Fujita, M. K. (2020). Finding complexity in complexes: Assessing the causes of mitonuclear discordance in a problematic species complex of Mesoamerican toads. *Molecular Ecology*, *29*(18), 3543–3559.

Firneno, T. J., Jr., & Townsend, J. H. (2019). Evaluation of species boundaries in sympatric and parapatric populations of Mesoamerican toads. *Zoologica Scripta*, *48*(4), 454–465.

Fonseca, E. M., Gehara, M., Werneck, F. P., Lanna, F. M., Colli, G. R., Sites, J. W., Jr, … Garda, A. A. (2018). Diversification with gene flow and niche divergence in a lizard species along the South American "diagonal of open formations." *Journal of Biogeography*, *45*(7), 1688–1700.

Fontenot, B. E., Makowsky, R., & Chippindale, P. T. (2011). Nuclear–mitochondrial discordance and gene flow in a recent radiation of toads. *Molecular Phylogenetics and Evolution*, *59*(1), 66–80.

Gamble, T., Simons, A. M., Colli, G. R., & Vitt, L. J. (2008). Tertiary climate change and the diversification of the Amazonian gecko genus *Gonatodes* (Sphaerodactylidae, Squamata). *Molecular Phylogenetics and Evolution*, *46*(1), 269–277.

Gehara, M., Crawford, A. J., Orrico, V. G. D., Rodríguez, A., Lötters, S., Fouquet, A., … Köhler, J. (2014). High levels of diversity uncovered in a widespread nominal taxon: continental phylogeography of the neotropical tree frog *Dendropsophus minutus*. *PloS One*, *9*(9), e103958.

Gehara, M., Garda, A. A., Werneck, F. P., Oliveira, E. F., da Fonseca, E. M., Camurugi, F., … Burbrink, F. T. (2017). Estimating synchronous demographic changes across populations using hABC and its application for a herpetological community from northeastern Brazil. *Molecular Ecology*, *26*(18), 4756–4771.

Graham, A. M., Lavretsky, P., Muñoz-Fuentes, V., Green, A. J., Wilson, R. E., & McCracken, K. G. (2018). Migration-selection balance drives genetic differentiation in genes associated with high-altitude function in the speckled teal (*Anas flavirostris*) in the Andes. *Genome Biology and Evolution*, *10*(1), 14–32.

Green, R. E., Krause, J., Briggs, A. W., Maricic, T., Stenzel, U., Kircher, M., … Pääbo, S. (2010). A draft sequence of the Neandertal genome. *Science*, *328*(5979), 710–722.

Guerra, C., Baldo, D., Rosset, S., Borteiro, C., & Kolenc, F. (2011). Advertisement and release calls in Neotropical toads of the *Rhinella granulosa* group and evidence of natural hybridization between *R. bergi* and *R. major* (Anura: Bufonidae). *Zootaxa*, *3092*, 26–42.

Guo, B., Lu, D., Liao, W. B., & Merilä, J. (2016). Genomewide scan for adaptive differentiation along altitudinal gradient in the Andrew's toad *Bufo andrewsi*. *Molecular Ecology*, *25*(16), 3884–3900.

Gutenkunst, R. N., Hernandez, R. D., Williamson, S. H., & Bustamante, C. D. (2009). Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLoS Genetics*, *5*(10), e1000695.

Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B. Q., & Vinh, L. S. (2018). UFBoot2: Improving the Ultrafast Bootstrap Approximation. *Molecular Biology and Evolution*, *35*(2), 518–522.

Jombart, T., & Ahmed, I. (2011). adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. *Bioinformatics* , *27*(21), 3070–3071.

Jombart, T., Devillard, S., & Balloux, F. (2010). Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genetics*, *11*, 94.

Júnior, F. A. B., Caramaschi, U., & Haddad, C. F. B. (2004). Review of the *Bufo crucifer* species group, with descriptions of two new related species (Amphibia, Anura, Bufonidae). *Arquivos Do Museu*

Nacional. Museu Nacional , *62*(3), 255–282.

Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A., & Jermiin, L. S. (2017). ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature Methods*, *14*(6), 587–589.

Lanfear, R., Calcott, B., Ho, S. Y. W., & Guindon, S. (2012). Partitionfinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Molecular Biology and Evolution*, *29*(6), 1695–1701.

Lanfear, R., Calcott, B., Kainer, D., Mayer, C., & Stamatakis, A. (2014). Selecting optimal partitioning schemes for phylogenomic datasets. *BMC Evolutionary Biology*, *14*, 82.

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.

Lavretsky, P., Peters, J. L., Winker, K., Bahn, V., Kulikova, I., Zhuravlev, Y. N., … McCracken, K. G. (2016). Becoming pure: identifying generational classes of admixed individuals within lesser and greater scaup populations. *Molecular Ecology*, *25*(3), 661–674.

Leaché, A. D., Portik, D. M., Rivera, D., Rödel, M., Penner, J., Gvoždík, V., … Fujita, M. K. (2019). Exploring rain forest diversification using demographic model testing in the African foam-nest treefrog *Chiromantis rufescens*. *Journal of Biogeography*, *46*(12), 2706–2721.

Lisiecki, L. E., & Raymo, M. E. (2007). Plio–Pleistocene climate evolution: trends and transitions in glacial cycle dynamics. *Quaternary Science Reviews*, *26*(1), 56–69.

Liu, K., Wang, F., Chen, W., Tu, L., Min, M.-S., Bi, K., & Fu, J. (2010). Rampant historical mitochondrial genome introgression between two species of green pond frogs, *Pelophylax nigromaculatus* and *P. plancyi*. *BMC Evolutionary Biology*, *10*, 201.

Lundberg, J. G., Marshall, L. G., Guerrero, J., Horton, B., Malabarba, M., & Wesselingh, F. (1998). The stage for Neotropical fish diversification: a history of tropical South American rivers. *Phylogeny and Classification of Neotropical Fishes*, *27*, 13–48.

Maciel, N. M., Collevatti, R. G., Colli, G. R., & Schwartz, E. F. (2010). Late Miocene diversification and phylogenetic relationships of the huge toads in the *Rhinella marina* (Linnaeus, 1758) species group (Anura: Bufonidae). *Molecular Phylogenetics and Evolution*, *57*(2), 787–797.

Malinsky, M., Matschiner, M., & Svardal, H. (2020). Dsuite - Fast D-statistics and related admixture evidence from VCF files. *Molecular Ecology Resources*. doi: 10.1111/1755-0998.13265

Malinsky, M., Svardal, H., Tyers, A. M., Miska, E. A., Genner, M. J., Turner, G. F., & Durbin, R. (2018).

680 Whole-genome sequences of Malawi cichlids reveal multiple radiations interconnected by gene flow.
681 *Nature Ecology & Evolution*, *2*(12), 1940–1955.
682 Mallet, J. (2005). Hybridization as an invasion of the genome. *Trends in Ecology & Evolution*, *20*(5),
683 229–237.
684 Malone, J. H., & Fontenot, B. E. (2008). Patterns of reproductive isolation in toads. *PloS One*, *3*(12),
685 e3900.
686 Marth, G. T., Czabarka, E., Murvai, J., & Sherry, S. T. (2004). The allele frequency spectrum in genome-
687 wide human variation data reveals signals of differential demographic history in three large world
688 populations. *Genetics*, *166*(1), 351–372.
689 Martin, S. H., Dasmahapatra, K. K., Nadeau, N. J., Salazar, C., Walters, J. R., Simpson, F., … Jiggins, C.
690 D. (2013). Genome-wide evidence for speciation with gene flow in *Heliconius* butterflies. *Genome*
691 *Research*, *23*(11), 1817–1828.
692 Maya-Soriano, M. J., Holt, W. V., & Lloyd, R. E. (2012). Biobanked amphibian samples confirmed to
693 species level using 16S rRNA DNA barcodes. *Biopreservation and Biobanking*, *10*(1), 22–28.
694 Mayr, E. (1963). *Animal Species and Evolution*. Harvard University Press.
695 Miller, M. A., Pfeiffer, W., & Schwartz, T. (2010). Creating the CIPRES Science Gateway for inference
696 of large phylogenetic trees. *2010 Gateway Computing Environments Workshop (GCE)*. doi:
697 10.1109/gce.2010.5676129
698 Nguyen, L.-T., Schmidt, H. A., von Haeseler, A., & Minh, B. Q. (2015). IQ-TREE: a fast and effective
699 stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and*
700 *Evolution*, *32*(1), 268–274.
701 Noonan, B. P., & Wray, K. P. (2006). Neotropical diversification: the effects of a complex history on
702 diversity within the poison frog genus *Dendrobates*. *Journal of Biogeography*, *33*(6), 1007–1020.
703 O'Connell, K. A., Prates, I., Scheinberg, L. A., Mulder, K. P., & Bell, R. C. (2021). Speciation and
704 secondary contact in a fossorial island endemic, the São Tomé caecilian. *Molecular Ecology*, *30*(12),
705 2859–2871.
706 Oliveira, E. F., Gehara, M., São-Pedro, V. A., Costa, G. C., Burbrink, F. T., Colli, G. R., … Garda, A. A.
707 (2018). Phylogeography of Muller's termite frog suggests the vicariant role of the Central Brazilian
708 Plateau. *Journal of Biogeography*, *45*(11), 2508–2519.
709 Patterson, N., Moorjani, P., Luo, Y., Mallick, S., Rohland, N., Zhan, Y., … Reich, D. (2012). Ancient
710 admixture in human history. *Genetics*, *192*(3), 1065–1093.
711 Pereyra, M. O., Baldo, D., Blotto, B. L., Iglesias, P. P., Thome, M. T. C., Haddad, C. F. B., … Faivovich,

J. (2016). Phylogenetic relationships of toads of the *Rhinella granulosa* group (Anura: Bufonidae): a molecular perspective with comments on hybridization and introgression. *Cladistics*, *32*, 36–53.

Pereyra, M. O., Blotto, B. L., Baldo, D., Chaparro, J. C., Ron, S. R., Elias-Costa, A. J., … Faivovich, J. (2021). Evolution in the Genus *Rhinella*: A total evidence phylogenetic analysis of Neotropical true toads (Anura: Bufonidae). *Bulletin of the American Museum of Natural History*, Vol. 447. doi: 10.1206/0003-0090.447.1.1

Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S., & Hoekstra, H. E. (2012). Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PloS One*, *7*(5). Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/pmc3365034/

Phuong, M. A., Bi, K., & Moritz, C. (2017). Range instability leads to cytonuclear discordance in a morphologically cryptic ground squirrel species complex. *Molecular Ecology*, *26*(18), 4743–4755.

Pirani, R. M., Peloso, P. L. V., Prado, J. R., Polo, É. M., Knowles, L. L., Ron, S. R., … Werneck, F. P. (2020). Diversification history of clown tree frogs in Neotropical rainforests (Anura, Hylidae, *Dendropsophus leucophyllatus* group). *Molecular Phylogenetics and Evolution*, *150*, 106877.

Portik, D. (2016). *Phylo_Wrapper_Scripts*. Retrieved from https://github.com/dportik/Phylo_Wrapper_Scripts

Portik, D. M., Leaché, A. D., Rivera, D., Barej, M. F., Burger, M., Hirschfeld, M., … Fujita, M. K. (2017). Evaluating mechanisms of diversification in a Guineo-Congolian tropical forest frog using demographic model selection. *Molecular Ecology*, *26*(19), 5245–5263.

Potter, S., Bragg, J. G., Peter, B. M., Bi, K., & Moritz, C. (2016). Phylogenomics at the tips: inferring lineages and their demographic history in a tropical lizard, *Carlia amax*. *Molecular Ecology*, *25*(6), 1367–1380.

Prado, C.P.A., Haddad, C.F.B., Zamudio, K.R. (2012). Cryptic lineages and Pleistocene population expansion in a Brazilian Cerrado frog. *Molecular Ecology, 21*, 921–941.

Prates, I., Rivera, D., Rodrigues, M. T., & Carnaval, A. C. (2016). A mid-Pleistocene rainforest corridor enabled synchronous invasions of the Atlantic Forest by Amazonian anole lizards. *Molecular Ecology*, *25*(20), 5174–5186.

Prates, I., Xue, A. T., Brown, J. L., Alvarado-Serrano, D. F., Rodrigues, M. T., Hickerson, M. J., & Carnaval, A. C. (2016). Inferring responses to climate dynamics from historical demography in neotropical forest lizards. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(29), 7978–7985.

744 Rabosky, D. L. (2016). Reproductive isolation and the causes of speciation rate variation in nature.
745 *Biological Journal of the Linnean Society. Linnean Society of London*, *118*(1), 13–25.
746 Ribeiro-Júnior, M. A. (2015). Catalogue of distribution of lizards (Reptilia: Squamata) from the Brazilian
747 Amazonia. I. Dactyloidae, Hoplocercidae, Iguanidae, Leiosauridae, Polychrotidae, Tropiduridae.
748 *Zootaxa*, *3983*(1), 1–110.
749 Ribeiro-Júnior, M. A., & Amaral, S. (2017). Catalogue of distribution of lizards (Reptilia: Squamata)
750 from the Brazilian Amazonia. IV. Alopoglossidae, Gymnophthalmidae. *Zootaxa*, *4269*(2), 151–196.
751 Rockney, H. J., Ofori-Boateng, C., Porcino, N., & Leaché, A. D. (2015). A comparison of DNA
752 barcoding markers in West African frogs. *African Journal of Herpetology: Journal of the*
753 *Herpetological Association of Africa*, *64*(2), 135–147.
754 Sambrook, J., & Russell, D. W. (2006). Purification of nucleic acids by extraction with
755 phenol:chloroform. *CSH Protocols*, *2006*(1). doi: 10.1101/pdb.prot4455
756 Sequeira, F., Sodré, D., Ferrand, N., Bernardi, J. A. R., Sampaio, I., Schneider, H., & Vallinoto, M.
757 (2011). Hybridization and massive mtDNA unidirectional introgression between the closely related
758 Neotropical toads *Rhinella marina* and *R. schneideri* inferred from mtDNA and nuclear markers.
759 *Biomedical Chromatography: BMC*, *11*(264).
760 Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large
761 phylogenies. *Bioinformatics* , *30*(9), 1312–1313.
762 Streicher, J.W., Devitt, T.J., Goldberg, C.S., Malone, J.H., Blackmon, H., Fujita M.K. (2014).
763 Diversification and asymmetrical gene flow across time and space: lineage sorting and hybridization
764 in polytypic barking frogs. *Molecular Ecology, 23*(13), 3273-3291.
765 Thomé, M. T. C., Zamudio, K. R., Giovanelli, J. G. R., Haddad, C. F. B., Baldissera, F. A., Jr, &
766 Alexandrino, J. (2010). Phylogeography of endemic toads and post-Pliocene persistence of the
767 Brazilian Atlantic Forest. *Molecular Phylogenetics and Evolution*, *55*(3), 1018–1031.
768 Thomé, M. T. C., Zamudio, K. R., Haddad, C. F. B., & Alexandrino, J. (2012). Delimiting genetic units in
769 Neotropical toads under incomplete lineage sorting and hybridization. *BMC Evolutionary Biology*,
770 *12*, 242.
771 Vallinoto, M., Sequeira, F., Sodre, D., Bernardi, J. A. R., Sampaio, I., & Schneider, H. (2009). Phylogeny
772 and biogeography of the *Rhinella marina* species complex (Amphibia, Bufonidae) revisited:
773 implications for Neotropical diversification hypotheses. *Zoologica Scripta*, *39*(2), 128–140.
774 Vasconcellos, M.M., Colli, G.R., Weber, J.N., Ortiz, E.M., Rodrigues, M.T., & Cannatella, D.C. (2019).
775 Isolation by instability: Historical climate change shapes population structure and genomic

776 divergence of treefrogs in the Neotropical Cerrado savanna. *Molecular Ecology, 28*(7), 1748–1764.
777 Venâncio, N. M., Antonio de Freitas, M., Abegg, A. D., & Kokubum, M. N. de C. (2017). First record of
778 *Rhinella poeppigii* (Tschudi, 1845) in Brazil (Anura, Bufonidae). *Check List*, *13*(6), 747–750.
779 Vences, M., Thomas, M., van der Meijden, A., Chiari, Y., & Vieites, D. R. (2005). Comparative
780 performance of the 16S rRNA gene in DNA barcoding of amphibians. *Frontiers in Zoology*, *2*(1), 5.
781 Venegas, P. J., & Ron, S. R. (2014). First records of *Rhinella poeppigii* (Tschudi 1845) from Ecuador,
782 with a distribution map (Anura: Bufonidae). *Herpetology Notes*, *7*, 713–716.
783 Wang, X., Lawrence Edwards, R., Auler, A. S., Cheng, H., Kong, X., Wang, Y., … Chiang, H.-W.
784 (2017). Hydroclimate changes across the Amazon lowlands over the past 45,000 years. *Nature*,
785 *541*(7636), 204–207.
786 Werneck, F. P. (2011). The diversification of eastern South American open vegetation biomes: Historical
787 biogeography and perspectives. *Quaternary Science Reviews*, *30*(13), 1630–1648.
788 Werneck, F. P., Leite, R. N., Geurgas, S. R., & Rodrigues, M. T. (2015). Biogeographic history and
789 cryptic diversity of saxicolous Tropiduridae lizards endemic to the semiarid Caatinga. *BMC*
790 *Evolutionary Biology*, *15*, 94.
791 Wesselingh, F. P., & Salo, J. A. (2006). A Miocene perspective on the evolution of the Amazonian biota.
792 *Scripta Geologica*, *133*, 439–458.
793 Winter, D. J. (2012). MMOD: an R library for the calculation of population differentiation statistics.
794 *Molecular Ecology Resources*, *12*(6), 1158–1160.
795 Yeaman, S., & Whitlock, M. C. (2011). The genetic architecture of adaptation under migration-selection
796 balance: The genetic architecture of local adaptation. *Evolution; International Journal of Organic*
797 *Evolution*, *65*(7), 1897–1911.

798

799 **TABLES AND FIGURES**

800 **FIGURE 1** (A) Maximum likelihood phylogeny of *Rhinella marina* complex focal species using
801 ddRADseq data and corresponding `ADMIXTURE` plot (K=7). Black circles on the phylogeny
802 denote ML bootstrap support (BS) > 95 and Bayesian posterior probability (PP) > 0.95. (B)
803 DAPC plot (K=7).

FIGURE 2 Locality map for the subset depicting average ADMIXTURE cluster assignments per locality (K=7) for *Rhinella horribilis*, *R. marina*, and *R. jimi*. Colors correspond to Figure 1. Map partitioned into biomes (Central America, Northern Andes, Northern Amazonia, Western Amazonia, Eastern Amazonia, Southern Amazonia, Pantanal, Chaco, Cerrado, Caatinga, Northern Atlantic Forest, Southern Atlantic Forest).

FIGURE 3 Locality map for the subset depicting average ADMIXTURE cluster assignments per locality (K=7) for *Rhinella poeppigii* and *R. schneideri*. Colors correspond to Figure 1. Map partitioned into biomes (Central America, Northern Andes, Northern Amazonia, Western Amazonia, Eastern Amazonia, Southern Amazonia, Pantanal, Chaco, Cerrado, Caatinga, Northern Atlantic Forest, Southern Atlantic Forest). *R. poeppigii* range adapted from (De la Riva, 2002; Venâncio et al., 2017; Venegas & Ron, 2014).

FIGURE 4 Time calibrated phylogeny based on mitochondrial 16S data. Black circles indicate PP > 0.95, * indicates the fossil calibration, and the bars represent the 95% HPD, which are also in parentheses. Colors correspond to the phylogeny in Fig. 1.

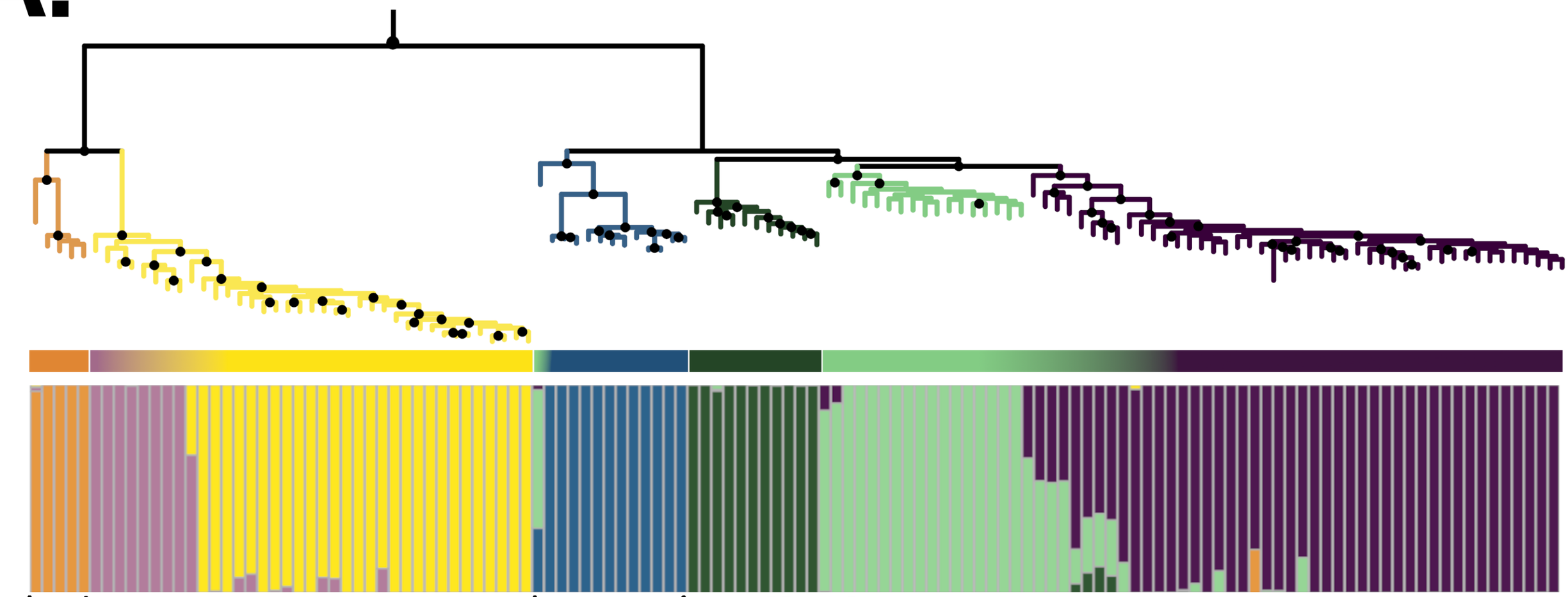
FIGURE 5 Optimal demographic models and residual plots for the (A) 3D-JSFS analysis of *Rhinella horribilis*, *R. marina*, and *R. jimi*, and (B) 2D-JSFS analysis of *R. poeppigii* and *R. schneideri*.

FIGURE 6 The f_b statistic (summary of f_4 admixture ratios). Grey color corresponds to tests that are not possible because of constraints on the phylogeny. * indicates a significant result.

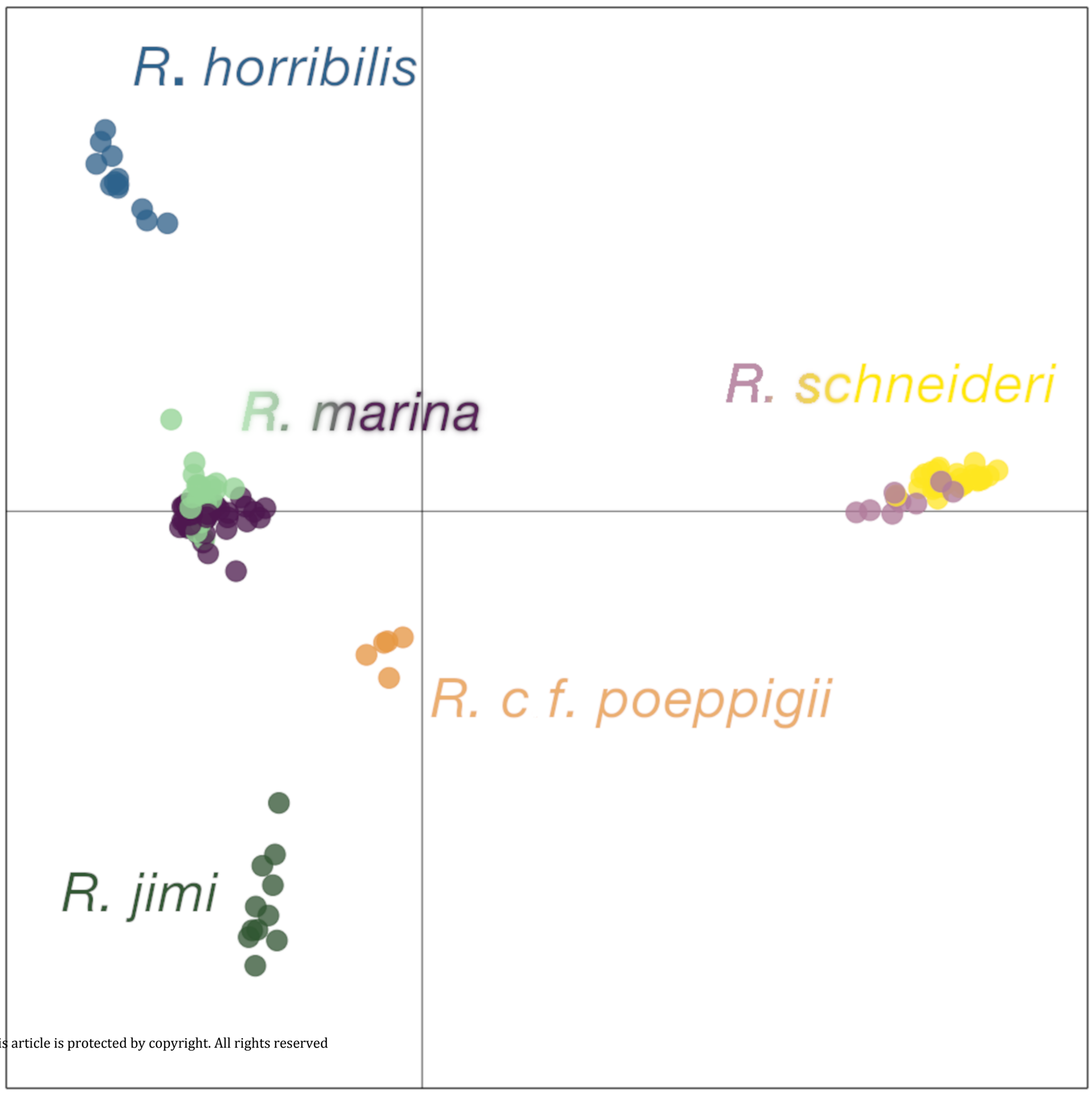
TABLE 1 Optimal demographic models and estimated parameters. Abbr: LL = log-likelihood; θ ($4N_{ref}\mu L$) = the effective mutation rate of the reference population (ancestral population); nu1, nu2 = effective population sizes under the constant population size model; nuA = effective population sizes of the ancestral population; mA, m1, m2 = migration rates between the ancestral (A), first (1) or second (2) population; m12 = migration rate from population two to population one; m21 = migration rate from population one to population two; T1, T2, T3 = unscaled time between demographic events.

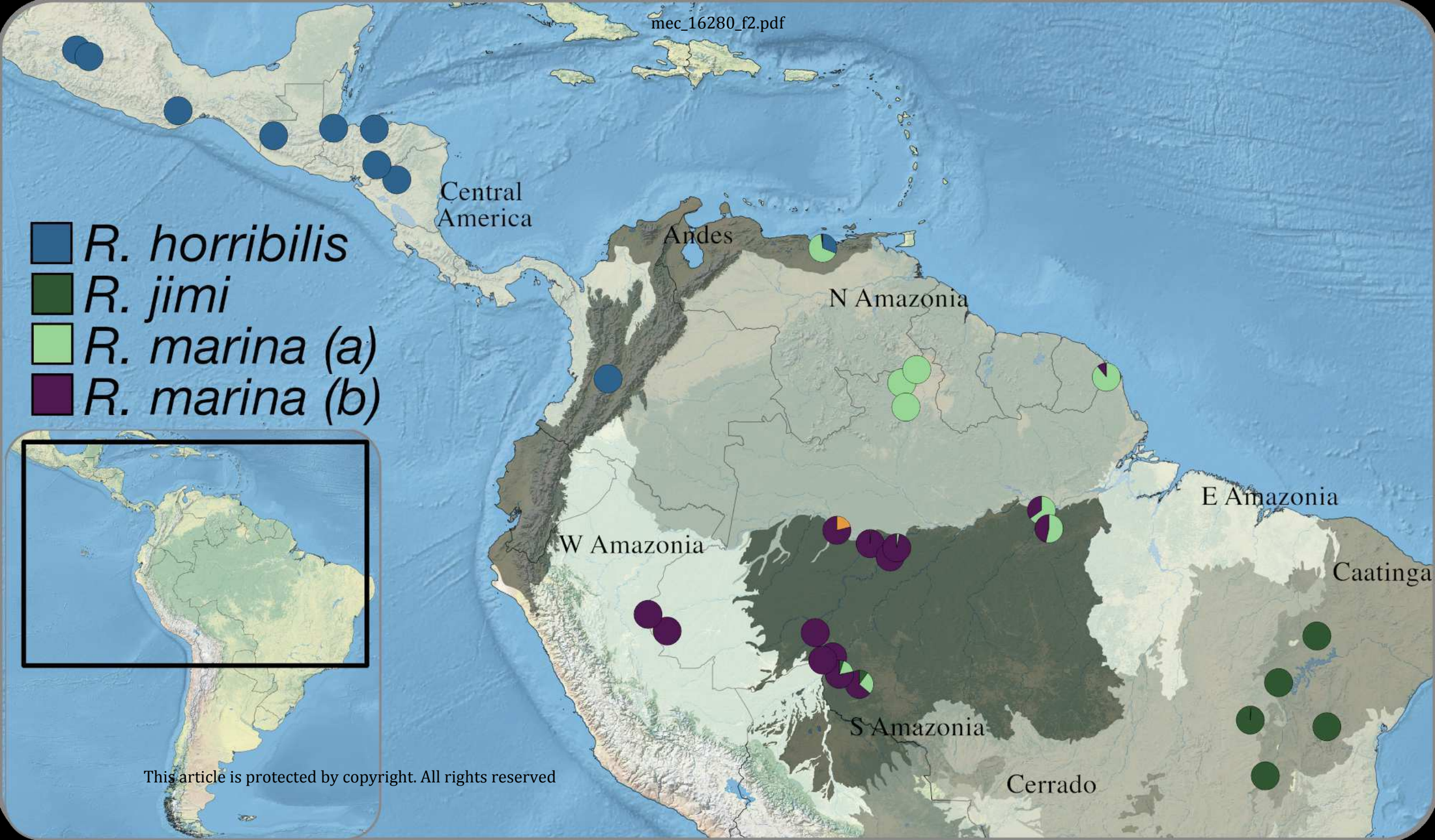
Model									
2D	LL	AIC	θ	nu1	nu2		m12	m21	
divergence in isolation with continuous asymmetric secondary contact	-539.27	1090.5	1424.5	0.01	0.05		15.5	1.82	
3D	LL	AIC	θ	nu1	nu2	nu3	nuA	mA	m1
ancient migration with shortest isolation	-1572.7	3165.4	95.33	0.57	0.15	1.31	6.15	0.05	1.36

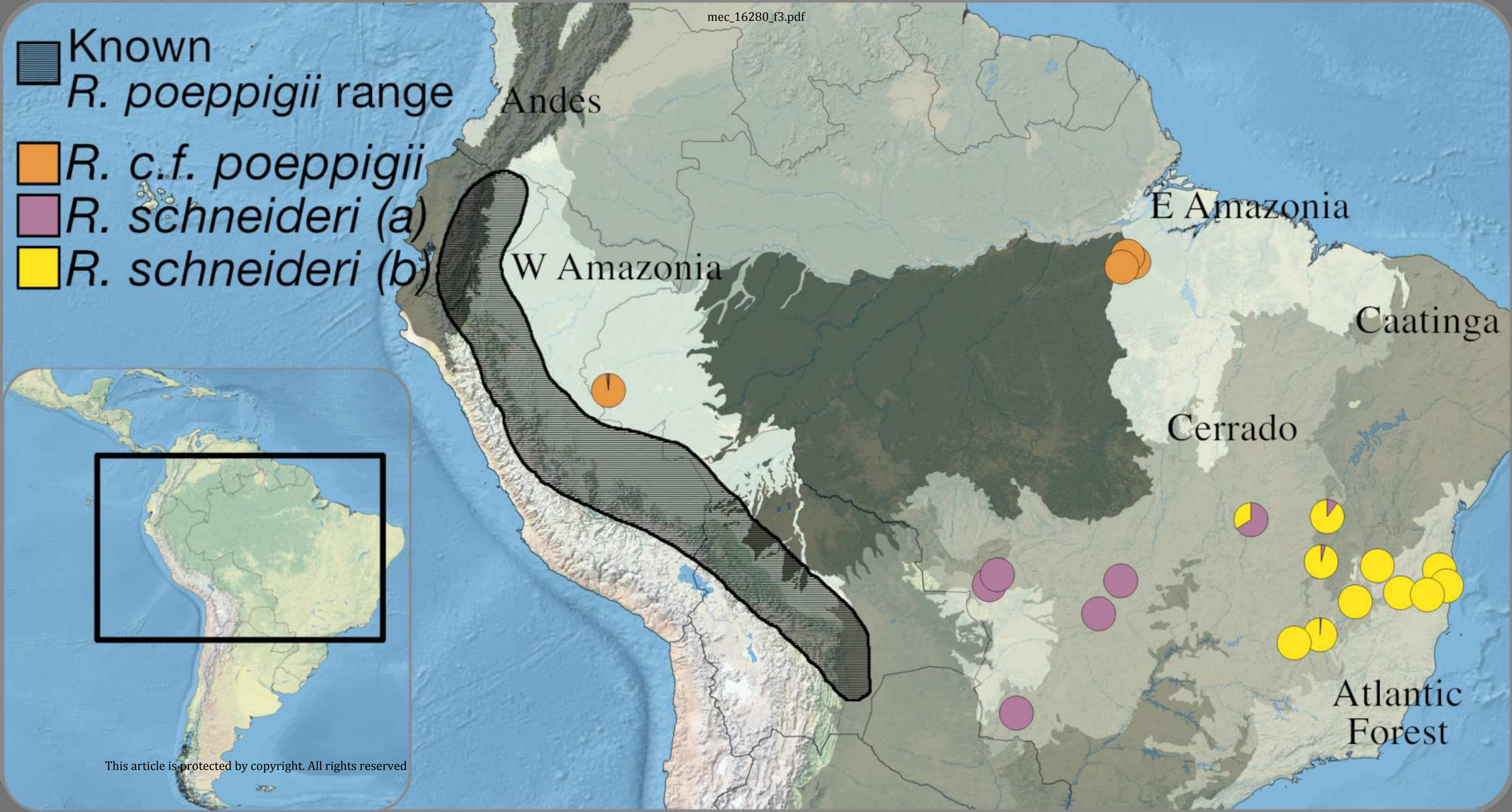
A.

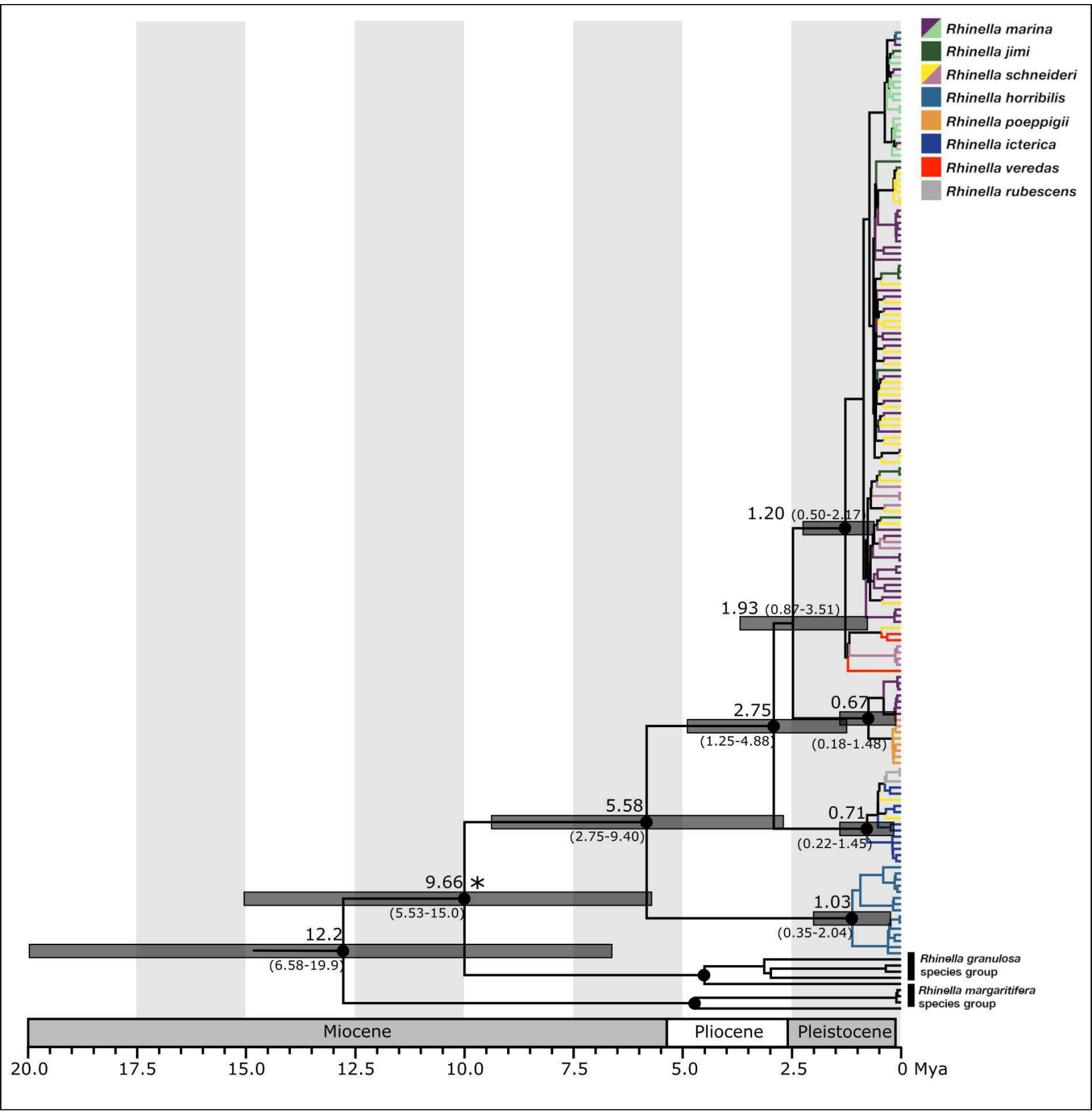


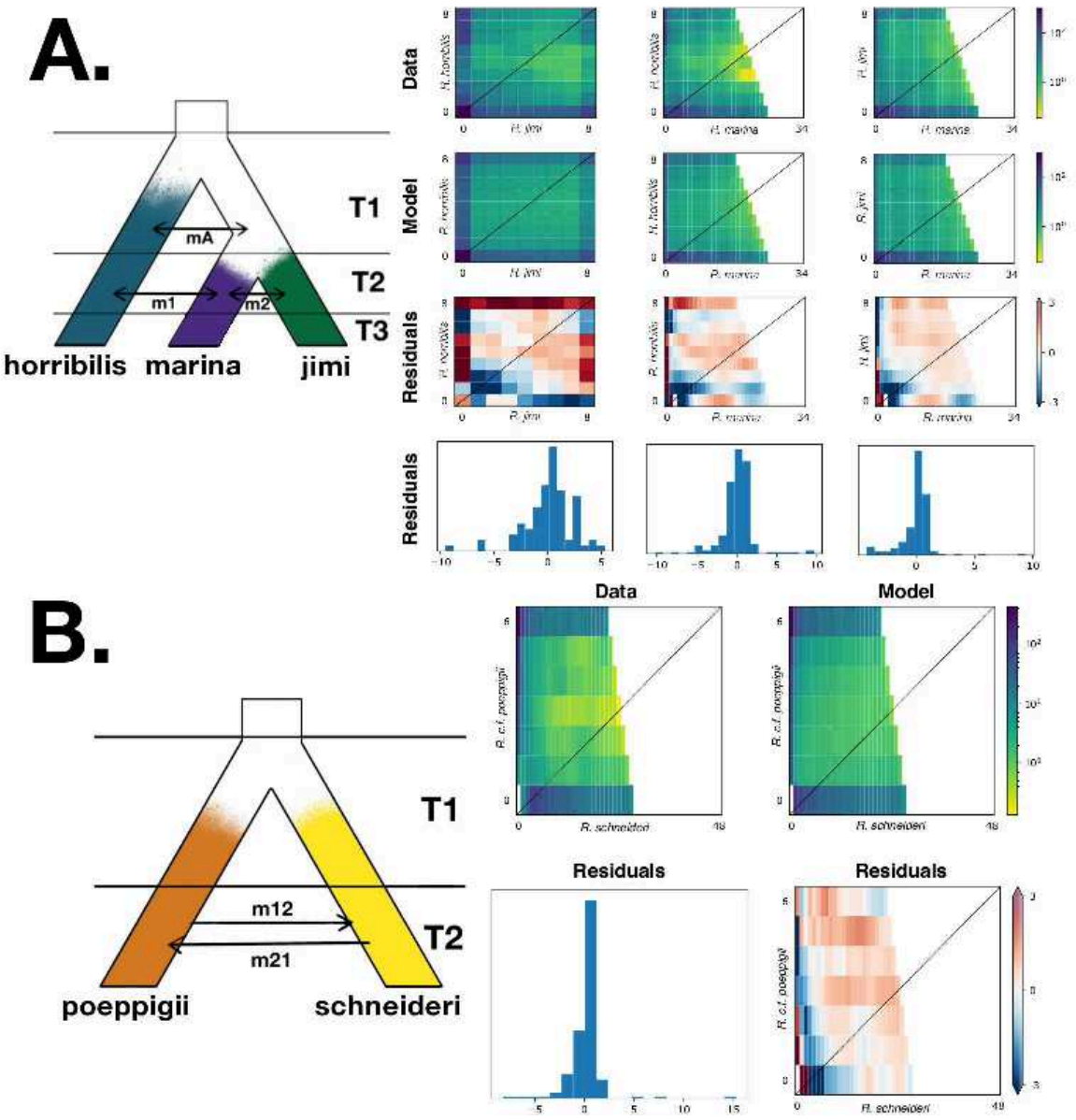
B.











mec_16280_f6.pdf