Molecular Phylogenetics and Evolution 161 (2021) 107167

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

journal homepage: www.elsevier.com/locate/ympev

Phylogenomics and evolutionary history of *Oreobates* (Anura: Craugastoridae) Neotropical frogs along elevational gradients

Santiago Montero-Mendieta [a,1], Ignacio De la Riva [b], Iker Irisarri [b,2], Jennifer A. Leonard [a], Matthew T. Webster [c], Carles Vilà [a,*]

[a] Conservation and Evolutionary Genetics Group, Estación Biológica de Doñana (EBD-CSIC), Seville, Spain
[b] Department of Biodiversity and Evolutionary Biology, Museo Nacional de Ciencias Naturales (MNCN-CSIC), Madrid, Spain
[c] Department of Medical Biochemistry and Microbiology, Science for Life Laboratory, Uppsala University, Uppsala, Sweden

ARTICLE INFO

Keywords:
Amphibia
Diversification
Gene flow
Genetic differentiation
Montane habitats
Target enrichment

ABSTRACT

Mountain ranges offer opportunities for understanding how species evolved and diversified across different environmental conditions. Neotropical frogs of the genus *Oreobates* (Anura: Craugastoridae) are adapted to highland and lowland habitats along the Andes, but many aspects of their evolution remain unknown. We studied their evolutionary history using ~18,000 exons enriched by targeted sequence-capture. Since capture success was very variable across samples, we evaluated to what degree differing data filtering produced robust inferences. The inferred evolutionary framework evidenced phylogenetic discordances among lowland species that can be explained by taxonomic misidentification or admixture of ancestral lineages. Highland species showed smaller effective populations than lowland frogs, probably due to greater habitat fragmentation in montane environments. Stronger genetic drift likely decreased the power of purifying selection and led to an increased proportion of nonsynonymous mutations in highland populations that could play an important role in their adaptation. Overall, our work sheds light on the evolutionary history and diversification of this group of Neotropical frogs along elevational gradients in the Andes as well as on their patterns of intraspecific diversity.

1. Introduction

Geographical barriers (e.g., rivers, mountain ranges) play an important role in facilitating divergence among populations as they can hamper gene flow, which results in increased genetic drift and facilitates local adaptation (Slatkin, 1985, 1987). At the same time, gradual changes in habitat conditions offer differential selective forces that also promote disparity among populations (Coyne & Orr, 2004). Phenotypic and genetic differentiation along elevational gradients is widespread in animals, and such differences can increase with the topographical complexity of habitats (Guarnizo & Cannatella, 2013; Keller et al., 2013). In biodiversity hotspots, identifying spatial patterns of diversification along the speciation continuum not only improves our understanding of the evolution of endangered species, but also guides conservation priorities to protect them (Allendorf et al., 2010; Roux et al., 2016; Chan & Brown, 2020).

The American tropics (the Neotropics) are the richest biogeographic region in terms of overall vertebrate species on Earth (Antonelli & Sanmartín, 2011; Rahbek et al., 2019b). Amphibians inhabiting the Neotropical region comprise roughly 50% of the planet's total, and more than one-third of them are globally threatened (Stuart et al., 2008; Scheele et al., 2019). In South America, amphibian species richness and endemism are particularly high in mountain regions and in the transition zone between the Andes and the Amazon basin (Roberts et al., 2006; Hutter et al., 2017). The extreme habitat heterogeneity of these mountain gradients boosts diversification of Neotropical amphibians as it favours speciation, coexistence, and persistence of distinct evolutionary lineages (Rahbek et al., 2019a; Perrigo et al., 2020).

Multiple evolutionary mechanisms may contribute to the high rates of amphibian diversification in the Andes (González-Voyer et al., 2011). Orogenic and climatic processes facilitate the close proximity of very different environments (Rahbek et al., 2019b). Local adaptation to different selective pressures along the altitudinal cline can create a barrier to gene flow between populations and lead to speciation through

* Corresponding author.
E-mail address: carles.vila@ebd.csic.es (C. Vilà).
[1] Present address: Key Laboratory of Zoological Systematics and Evolution, Institute of Zoology, Chinese Academy of Sciences, Beijing, China.
[2] Present address: Department of Applied Bioinformatics, Institute for Microbiology and Genetics, University of Goettingen, Goettingen, Germany.

https://doi.org/10.1016/j.ympev.2021.107167
Received 17 June 2020; Received in revised form 11 March 2021; Accepted 25 March 2021
Available online 30 March 2021

climatic-niche divergence, as shown in lungless salamanders (Wiens et al., 2007; Kozak & Wiens, 2010). At the same time, in other amphibians, niche conservatism coupled with changes in the distribution of habitats during climatic oscillations and orogenies resulted in fragmentation of distribution ranges and allopatric speciation (Smith et al., 2007; Hutter et al., 2013, 2017). Studying phylogenetic relationships is key to understanding how amphibians diversified along environmental clines and across geographic barriers in the Neotropics (Fritz & Rahbek, 2012). In this regard, the application of genome-wide data for phylogenetic inference (i.e., phylogenomics) can provide basic knowledge, not only on the mechanisms leading to diversification, but also on the role of adaptation in species differentiation and on the effects on intraspecific diversity and evolutionary potential.

Here, we employed phylogenomics to assess the impact of elevation on the diversification and evolutionary history of the genus *Oreobates* Jiménez de la Espada, 1872 (Anura: Craugastoridae). This genus constitutes a small but interesting radiation of 25 direct-developing Neotropical frogs species (Terrarana) adapted to disparate habitats, including Amazonian lowlands, dry forests, inter-Andean dry valleys, humid forests, the yungas transitional zone, cloud forests, elfin forests, and puna grasslands of the Andes up to 3,800 m above sea level (m asl) (Padial et al., 2012). Previous studies have increased our knowledge on the biodiversity of this group of frogs (16 species have been described over the last two decades; see Frost, 2020). Yet, their contribution to the understanding of the processes associated with the diversification along the Andes has been limited; mainly because phylogenies in such studies were based on small regions of the genome (i.e., gene trees) that do not necessarily reflect the overall evolutionary history of the species (i.e. the species tree) (Mallo & Posada, 2016). Heterogeneous signals across different genomic regions might be an outcome of various evolutionary processes such as incomplete lineage sorting (ILS), hybridization, gene flow, and gene duplication or loss (Bravo et al., 2019).

Phylogenomics has resolved the evolutionary history of a variety of species and groups above the species level in the tree of life (e.g., Delsuc et al., 2005; McCormack et al., 2013; Irisarri & Meyer, 2016). Ideally, it involves comparing sequences of whole genomes; but in amphibians, this is still a major undertaking due to their large and highly repetitive genomes that make assembly challenging (Funk et al., 2018; Liedtke et al., 2018). One way to obtain large genomic datasets is with reduced-representation sequencing strategies such as targeted sequence-capture (or target enrichment), that selectively enrich for genomic regions of interest (Mamanova et al., 2010; Faircloth et al., 2012; Lemmon et al., 2012). This approach has proved useful for disentangling taxonomic issues among major frog lineages (e.g., Portik et al., 2016; Feng et al., 2017; Heinicke et al., 2018; Streicher et al., 2018). Considering the variability in the genome size of amphibians (Liedtke et al. 2018), the development of a sequence-capture approach based on sequences from the target species could potentially increase the proportion of homologous sequences captured.

In this study, we used the transcriptome of one species of *Oreobates* (*O. cruralis*; Montero-Mendieta et al., 2017) to develop an enrichment panel of ~18,000 exons and obtain homologous regions for most species in this genus and two outgroup species from closely-related genera. We applied a combination of different phylogenomic methods to explore the evolution of this group of Neotropical frogs along elevational gradients. Our aim was to test whether the evolutionary mechanisms that led to the diversification of these frogs differed between species living in highland and lowland habitats. We hypothesized that geographical isolation in combination with divergent selective pressures in highland and lowland habitats have impacted both the evolutionary and the demographic history of *Oreobates* frogs, with few transitions from one habitat to the other. Adaptation to highlands may have resulted in more fragmented habitats, smaller effective populations, and lower gene flow between incipient species, thus facilitating speciation.

2. Material and methods

2.1. Probe design for sequence-capture

We used the transcriptome of *O. cruralis* (NCBI accession number: PRJNA384528) to design enrichment probes to capture homologous sequences. We performed a BLAST homology search with NCBI-BLAST v2.4.0+ (Altschul et al., 1997) to the SwissProt database (Bairoch, 2000) and kept unigenes with at least one gene ontology (GO) term. This was done to restrict downstream analyses to unigenes with known functions. We refined the dataset following Portik et al. (2016) to achieve higher performance in the sequence-capture. In brief, we retained unigene sequences with a GC content of 40–60%, we selected those with a length of 500–850 base pairs (bp), and we trimmed to 850 bp those unigenes that were longer (see Portik et al., 2016). We used RepeatMasker v4.0.6 (http://www.repeatmasker.org/) to remove repetitive elements and low complexity regions with 'vertebrata metazoa' as reference database and the 'cross_match' search engine. All these steps produced a final set of 17,879 exonic sequences with a total length of 14,157,267 bp. We used these exons as a reference to design a myBaits® custom target capture kit (Arbor Biosciences, Ann Arbor, USA) with ~2x flexible tiling density and 120 nucleotides per probe to maximize the capture of divergent sequences. The selected bait kit was a MYbaits-11 with up to 220,000 probes (220 K). A relaxed filtering was applied by the vendor after screening candidate baits against the genomes of *Nanorana parkeri* and *Xenopus tropicalis* using proprietary software. As a result, the probe library for sequence-capture had 213,879 unique probes.

2.2. Sampling

All samples used in this study were provided by the Museo Nacional de Ciencias Naturales (MNCN-CSIC), Madrid, Spain. The specimens were collected between 1999 and 2016, and tissue samples were stored at the MNCN's frozen tissue and DNA collection. When available, we selected at least two samples from different localities for each species. We included more than two individuals for species with extensive distributions to ensure representative sampling. In total, our initial dataset included 65 specimens from 18 species of *Oreobates*: *O. amarakaeri* (1 sample), *O. ayacucho* (2), *O. barituensis* (2), *O. berdemenos* (1), *O. choristolemma* (2), *O. cruralis* (11), *O. discoidalis* (4), *O. gemcare* (2), *O. heterodactylus* (3), *O. lehri* (6), *O. machiguenga* (3), *O. madidi* (2), *O. pereger* (1), *O. quixensis* (13), *O. sanctaecrucis* (4), *O. saxatilis* (5), *O. yanucu* (1), and *O. zongoensis* (2). All these species inhabit the eastern slopes of the Andes or the Amazonian Piedmont, except *O. heterodactylus*, which occurs in semi-dry low mountain systems of the Cerrado formation in southeastern Bolivia. We did not include other species of *Oreobates* that occur further to the east and to the north in the Brazilian Cerrado (states of Mato Grosso, Goiás, and Minas Gerais; see Frost [2020]). Based on the evolutionary relationships among Terrarana frogs (Hedges et al., 2008), we also included three outgroup samples of two species from two closely related genera: *Phrynopus barthlenae* (2), and *Lynchius tabaconas* (1).

2.3. DNA extraction, genomic libraries, and target enrichment

We extracted genomic DNA from tissue samples using a phenol–chloroform protocol (Sambrook et al., 1989) and quantified its concentration with a Qubit 3.0 Fluorometer (Invitrogen, Carlsbad, USA). The resulting DNA was sheared for 40 s (s) to a target length of ~300 bp using the E220 focused-ultrasonicator (Covaris, Inc., Woburn, USA) and purified using the 'Genomic DNA Clean & Concentrator-10′ kit (Zymo Research, Irvine, USA). We prepared individual genomic libraries following Meyer & Kircher (2010), adding the following steps: (1) blunting and phosphorylation of DNA ends with 'Fast DNA End Repair' kit (ThermoFisher, Waltham, USA); (2) A-tailing with 'AmpliTaq Gold'

polymerase (Applied Biosystems, Foster City, USA) for 20 minutes (min) at 72 °C; (3) ligation of Y-shaped adapters (Illumina, Inc., San Diego, USA) with 'T4 DNA' ligase (ThermoFisher); (4) double-indexed PCRs with 'Kapa HiFi HS RM' polymerase (Roche Diagnostics, Risch-Rotkreuz, Switzerland). The PCR program included an initial step of 95 °C for 3 min, followed by 12 cycles of 98 °C for 20 s, 60 °C for 15 s, and 72 °C for 30 s, with a final step of 72 °C for 5 min. We purified the amplified DNA using 2X magnetic beads (Rohland & Reich, 2012).

Libraries were enriched in pools of up to six individuals, grouped by relatedness as suggested by the phylogenetic relationships inferred by Köhler & Padial (2016). For each pool, we performed capture reactions following the manufacturer's protocol (MYbaits® kit manual version 3). Due to the limited amount of initial tissue available for DNA extraction (just about 0.05 gr per sample), we used half of the recommended volume of DNA and reagents/probes. We purified individual capture reactions using 1.3X SPRI (Solid Phase Reversible Immobilization) magnetic beads and amplified post-capture products using two separate PCR reactions of 15 cycles each. The resulting PCR products were purified, combined, and quantified with QuantiFluor-ST (Promega, Madison, USA) assays. Post-capture libraries were paired-end sequenced by John Hopkins Genomics (Baltimore, USA) on a HiSeq 2500 instrument (Illumina Inc.) using the v4 chemistry according to the manufacturer's protocols.

2.4. Data processing, assembly, and ortholog detection

To extract and examine the exons captured for each sample we used the HybPhyloMaker pipeline v1.6 (https://github.com/tomas-fer/HybPhyloMaker) (Fér & Schmickl, 2018). This pipeline consists of a series of bash scripts that perform multiple analyses from raw enriched data to the species tree.

Probe sequences were concatenated and separated with a chain of 400 Ns each using the script 'HybPhyloMaker0b', which produced a pseudo-reference sequence for *O. cruralis* consisting of 21,309,667 bp. For each sample, raw FASTQ reads were processed by the script 'HybPhyloMaker1′. This script uses Bowtie2 (Langmead & Salzberg, 2012) and SAMtools v1.5 (Li et al., 2009) to remove Illumina's PhiX reads, Trimmomatic v0.36 (Bolger et al., 2014) for adapter trimming and quality filtering, and FastUniq v1.1 (Xu et al., 2012) to remove duplicate reads. Filtered reads were mapped to the pseudo-reference by the script 'HybPhyloMaker2′ using both the Burrows-Wheeler Aligner (BWA-MEM) v0.7.5 (Li & Durbin, 2009) and Bowtie2. Consensus sequences (contigs) were obtained with Kindel v0.2 (Constantinides & Robertson, 2017) by considering bases present in more than half of the reads (minimum site coverage for SNP calling = 2). Contigs that matched target sequences were identified with BLAT v35 (Kent, 2002) using the script 'HybPhyloMaker3′ (minimum sequence identity between probe and sample = 90%).

As we did not previously filter our probe sequences to select for single-copy exons, high enrichment of paralogs was expected. To find putative orthologs, HybPhyloMaker applies a majority rule consensus, i.e., the most abundant sequence of a locus is taken as the ortholog, as paralogs are usually less prevalent in the mapping compared to orthologs due to a higher sequence dissimilarity to the probe sequences (Fér & Schmickl, 2018). Additionally, we used the script 'HybPhyloMaker4a2′ to filter out exons with more than four heterozygous sites (default setting) because that could be an indication of paralogous sequences mapping to the same locus.

2.5. Phylogenetic inference and missing data evaluation

We aligned orthologous exons with MAFFT v7.313 (Katoh & Toh, 2008) using the script 'HybPhyloMaker4a' and adjusted them to the correct reading frame with the script 'HybPhyloMaker4b' via EMBOSS v6.5.7 (Rice et al., 2000). This script removes exons with ambiguous reading frames and more than five stop codons per alignment, as well as incomplete triplets at the beginning or at the end of the alignments (Carlsen et al., 2018). The remaining stop codons were converted to Ns in both nucleotide and amino acid alignments, as they are considered likely errors induced via sequencing or read mapping (Fér & Schmickl, 2018). The program AMAS v0.98 (Borowiec, 2016) was used to join exons from the same gene into a single gene alignment.

To keep as much DNA sequence information as possible, we specified a relaxed missing data filtering in the script 'HybPhyloMaker5′ (MISSINGPERCENT = 99 in the HybPhyloMaker settings file, i.e., include exons with up to 99% of Ns per alignment, and SPECIESPRESENCE = 1, i.e., include exons that appear in at least 1% of the samples). However, we also evaluated the effect of different filtering parameters (see next paragraph). The script 'HybPhyloMaker6′ was used to infer gene trees with RAxML v8.2.10 (Stamatakis, 2014) from alignments applying the GTRGAMMA model with rapid bootstrapping and 100 bootstrap pseudo-replicates. The resulting gene trees were used as input of 'HybPhyloMaker7′ and 'HybPhyloMaker8′ to reconstruct the tree depicting the relationship between samples (i.e., the 'species trees', but we avoid using this term here because we included several individual samples per species, and it could lead to misinterpretation; hereafter 'sample tree') on the basis of the following different methods with default parameters: ASTRAL v5.6.1 (Mirarab et al., 2014), ASTRID v1.4 (Vachaspati & Warnow, 2015; both algorithmic approximations based on coalescence), and MRL (Nguyen et al., 2012; supertree approach that uses maximum parsimony heuristics). We concatenated the alignments with AMAS to infer a maximum likelihood tree using IQ-TREE v2.0.3 and 1000 bootstrap replicates (Minh et al., 2020a; hereafter 'ML concatenated'). To do that, the data were partitioned by gene and the GTRGAMMA model was applied to all partitions. Also, we used the concatenated alignment to infer a tree of samples under the coalescent scheme with SVDquartets (SVDq) (Chifman & Kubatko, 2014) implemented in PAUP* v4a164 (Swofford, 2002). All possible quartets were evaluated and clade support was assessed by running 100 bootstrap replicates (svdq evalq = all bootstrap). For comparative purposes, we also ran the same analyses using frame-uncorrected alignments (i.e. not adjusted to the correct reading frame).

To evaluate the impact of missing data on sample tree estimation, we applied a series of different filtering parameters via the script 'HybPhyloMaker5′. In particular, we removed exons that had >99%, 75%, 50%, 25%, and 1% of Ns in the alignment (i.e., 'MISSINGPERCENT'). For each of these datasets, we also required that the minimum number of taxa with data for any given exon was at least 1%, 25%, 50%, 75%, and 100% of the total (i.e., 'SPECIESPRESENCE'). This produced 25 datasets with varying levels of missing data (labeled as 'maxmissing99_minsp1′ 'maxmissing99_minsp25′, 'maxmissing99_minsp50′ … 'maxmissing1_minsp100′) that were used as input for ASTRAL, ASTRID, MRL, ML concatenated, and SVDq. Thus, we inferred a total of 125 sample trees (25 trees for each of the 5 methods). For the ML concatenated analyses, we used PartitionFinder (Lanfear et al., 2017) to find the best partition scheme in datasets ≤ 200 exons (datasets >200 exons were partitioned by gene as they were computationally impractical). We rooted each tree with *P. barthlenae* using the function 'nw_reroot' from the package Newick Utilities v1.6 (Junier & Zdobnov, 2010). For each method, we calculated the normalized Robinson–Foulds (nRF) symmetric distance (Robinson & Foulds, 1981) between the tree inferred with the most complete dataset ('maxmissing99_minsp1′) and those resulting from all other datasets, using the function 'compare' in ETE3 (Huerta-Cepas et al., 2016).

In addition to HybPhyloMaker, we also ran an alternative data analysis pipeline, HybPiper v1.2 (https://github.com/mossmatters/HybPiper/) (Johnson et al., 2016) to recover and examine enriched gene regions of interest. Unlike HybPhyloMaker, this pipeline implements a *de-novo* strategy for contig assembly. However, it does not correct alignments by reading frame nor remove paralogs. We retained potential orthologous sequences by removing exons in which the number of mapped reads was excessively elevated compared to other exons within each individual. The resulting sequences were used to infer gene trees with RAxML and a

sample tree with ASTRAL. This tree was compared to the ASTRAL tree inferred using HybPhyloMaker ('maxmissing99_minsp1′). Further details can be found in Appendix 1.

2.6. Pairwise species similarity and population structure

To quantify the degree of differentiation among species, we used a custom Python script to calculate pairwise uncorrected *p*-distances (i.e., the proportion of nucleotides sequenced in both taxa that are different after removing Ns) between all pairs of individuals and species on the 'maxmissing99_minsp1′ data. For three pairs of species that showed low interspecific genetic variability and had more than one individual per species (*O. quixensis* and *O. saxatilis*; *O. barituensis* and *O. discoidalis*; *O. machiguenga* and *O. lehri*), we assessed potential admixture as follows. First, we used SAMtools v1.8 utility 'mpileup' (Li et al., 2009) with the options "-I -C 50 -t DP,SP -v -A -E" on the BAM files resulting from read mapping with BWA-MEM (see subsection 2.4) to call SNPs as genotype probabilities in VCF format. Next, we converted SNPs from genotype probabilities to genotype calls with BCFtools v1.9 (Li, 2011). Then, we used VCFtools v0.1.16 (Danecek et al., 2011) to filter SNPs with the following cutoffs "–maf 0.1 –minDP 10 –max-missing 0.25" and reformatted the data from VCF to PED using PLINK v1.90b4 (Purcell et al., 2007). Lastly, we ran two different clustering programs to analyse population structure, sNMF v1.2 (Frichot et al., 2014) and STRUCTURE v2.3.4 (Pritchard et al., 2000) setting the flag "-K" (where K is the number of ancestral populations) from $K = 1$ to $K =$ the total number of individuals included in the analysis. These programs use different algorithms to cluster genetically similar individuals together. For each K, we carried out 10 independent replicate runs and combined them using CLUMPP v1.1.2 (Jakobsson & Rosenberg, 2007) with default settings. To determine the most likely number of K, we compared the values of the cross-entropy criterion resulting from sNMF's analyses at each K (option "-c").

2.7. Individual heterozygosity and rates of evolution

We calculated the proportion of heterozygous positions per individual sample via a folded Site Frequency Spectrum (SFS) as implemented in ANGSD v0.929–25-g458c20d (Korneliussen et al., 2014) using the BAM files mentioned above. This estimation was based on individual genotype likelihoods assuming Hardy-Weinberg equilibrium (option: "-gl 1"). As ancestral sequences were not available, we used the *O. cruralis* pseudo-reference to polarize our data. The worst reads and bases were removed by adding the parameters "-C 50 -ref ref.fa -minQ 20 -minmapq 30".

To compute the ratio of nonsynonymous to synonymous substitution rates, i.e., d_N/d_S or 'ω', we used PAML v4 (Yang, 2007) as implemented in ETE3′s function 'evol'. We ran three different branch model tests including the null or 'M0′, which assumes that all tree branches evolve at the same ω rate (Yang et al., 2000); the 'b_free', which considers that branches of interest, usually called 'foreground', evolve at a different ω rate than the rest of the tree, termed 'background'; and the 'b_neut', where the ω of foreground branches is fixed to one (Yang & Nielsen, 2002). These models were compared with a likelihood-ratio test (LRT). The comparison between 'b_free' and 'M0′ models tests if foreground branches have a ω significantly different from the rest of the tree, while the comparison between 'b_free' and 'b_neut' models tells if the ω of foreground branches is significantly higher than 1. We restricted our analyses to a single individual per species (the individual with the lowest amount of missing data) and used as input the concatenated alignment corresponding to the 'maxmissing75_minsp75′ dataset with reading frame correction (hereafter, 'species tree' refers to trees inferred with the frame-corrected 'maxmissing75_minsp75′ dataset and with one sample per species). We chose this dataset as it resulted in similar phylogenetic relationships to those obtained with the most complete dataset, 'maxmissing99_minsp1′, despite including far fewer loci (see Results). We performed two PAML analyses on the basis of the species tree obtained with ASTRAL: one marking as foreground all branches of the highland clade consisting of *O. ayacucho*, *O. gemcare*, *O. lehri*, *O. machiguenga*, and *O. pereger*; and another marking as foreground all branches of a lowland clade consisting of *O. barituensis*, *O. cruralis*, *O. discoidalis*, *O. heterodactylus*, and *O. madidi*.

To avoid biases when detecting selection that could be caused by replacing stop codons by Ns or due to genes' and species' histories being different, we repeated PAML analyses using an ASTRAL tree inferred from a concatenated alignment in which for each locus we deleted taxa that had at least one stop codon (rather than removing loci with more than five stop codons). Additionally, we used as input for PAML analyses the gene trees derived from this new data, which were inferred with RAxML as described in Section 2.5. To avoid biases due to missing taxa, we not only excluded exons in which outgroup species were not present but also exons containing less than three highland or lowland species.

To identify possible correlations between d_S, d_N or ω, and six environmental variables (midpoint elevation, mean annual temperature, mean annual precipitation, mean annual relative humidity, mean net primary productivity, and mean annual UVB radiation) for each species, we used the program Coevol v1.4 (Lartillot & Poujol, 2011). Both the collection of the environmental variables and the analysis of correlations were carried out as described in Dugo-Cota et al. (2015) with the ML concatenated species tree and its corresponding alignment (frame-corrected 'maxmissing75_minsp75' dataset) as input. We obtained temperature and precipitation data from the WorldClim climatic maps (Hijmans et al., 2005), relative humidity from the Atlas of the Biosphere (New et al., 1999), net primary productivity data from the HANNP datasets (Imhoff et al., 2004; Imhoff & Bounoua, 2006), and UVB radiation data from the glUV dataset (Beckmann et al., 2014).

2.8. Conflicting phylogenetic signals and networks

To determine whether gene trees were concordant or conflicting with the overall sample tree, we ran PhyParts (Smith et al., 2015) using as input the species tree obtained with ASTRAL and its corresponding gene trees (frame-corrected 'maxmissing75_minsp75′ dataset). As PhyParts requires rooted trees, we used 'nw_reroot' to root trees with *P. barthlenae*. If this species was not present, we rooted with *L. tabaconas*. We discarded gene trees in which outgroups were not present. Pie charts representing gene tree and species tree discordance at each node were obtained with the script 'phypartspiecharts.py' (available at https://github.com/mossmatters/MJPythonNotebooks). Additionally, we used the ML concatenated species tree and its gene trees to assess the degree of conflict at internal branches using gene concordance factors (gCF) and gene internode certainty (gIC) as implemented in IQ-TREE 2 (Minh et al., 2020b) (further details are available at http://www.robertlanfear.com/blog/files/concordance_factors.html).

We reconstructed species networks to assess the presence of recent and/or historical gene flow using the above-mentioned rooted gene trees as input for the pseudo-maximum likelihood phylogenetic network approach (Yu & Nakhleh, 2015) implemented in PhyloNet v3.7.1 (Than et al., 2008). We allowed for 0 to 4 reticulations and compared them using the Akaike information criterion (AIC) (Akaike, 1974).

3. Results

3.1. Efficiency of the sequence-capture approach

We successfully enriched target sequences in 40 (37 *Oreobates* + 3 outgroups) out of the 68 collected samples corresponding to 19 (17 *Oreobates* + 2 outgroups) out of 20 species (Table S1). The only species that could not be enriched was *O. amarakaeri*. The concentration of the extracted DNA varied between 0.03 and 36.85 ng/μL. The enrichment protocol failed in 28 samples that initially had very low DNA concentration (<1 ng/μL) regardless of the species. The number of raw reads

obtained differed greatly between individuals (1,566,470–49,786,466 reads) and a large percentage of the reads were duplicates (57–87%). After removing these duplicates, approximately 13–70% of the reads mapped to the targets (depth of coverage: 2–53X), recovering a total of 2,640–14,821 exons per sample. With HybPiper, we retrieved 139–12,464 exons per sample (see Table S2).

Although the probes were based on the transcriptome of *O. cruralis*, the species with the largest proportion of mapped reads were *O. zongoensis, O. sanctaecrucis,* and *O. choristolemma* (Fig. 1). The capture efficiency per species of our probes, measured as the percentage of the target bases that were recovered, varied between 5.37% and 36.90%, with an average of about 19%. We retained a total of 3,810 exons after removing loci with ambiguous reading frames or more than five stop codons per alignment and incomplete triplets at the beginning or at the end of the alignments. The characteristics of these exons in terms of percent of missing data, the number of taxa in which they have been sequenced, the proportion of parsimony-informative and variable sites are shown in Fig. 2 and further details are available in Table S3.

3.2. Phylogenetic relationships of Oreobates

The genus *Oreobates* was monophyletic regardless of the pipeline used to examine the data (HybPhyloMaker and HybPiper) or the method employed to infer the sample tree (ASTRAL, ASTRID, ML concatenated, MRL, and SVDq). In Fig. 3a, we show the relationships recovered by the ML concatenated tree based on the 3,810 exons obtained with HybPhyloMaker. However, other methods supported similar tree topologies (see below).

All methods recovered two major lineages within *Oreobates*: one encompassing species distributed from Colombia to Peru (*O. ayacucho*, *O. gemcare*, *O. lehri*, *O. machiguenga*, *O. pereger*, *O. quixensis*, and O. *saxatilis*) and another that included species from Argentina and Bolivia (*O. barituensis*, *O. berdemenos*, *O. choristolemma*, *O. cruralis*, *O. discoidalis*, *O. heterodactylus*, *O. madidi*, *O. sanctaecrucis*, *O. yanucu*, and *O. zongoensis*) (Fig. 3b). In the first lineage, two divergent clades were recovered. One clade is composed of *O. quixensis* and *O. saxatilis*; the former species includes populations distributed along the Amazonian lowlands <1,000 m asl and the latter is restricted to the lower hills of the Andes <1,200 m asl (Padial et al., 2008, 2012; Köhler & Padial, 2016). These two species were recovered as polyphyletic, with one individual from each species consistently found more closely related to the other species (*O. saxatilis* RS66 within *O. quixensis*, and *O. quixensis* RS48 within *O. saxatilis*). The other clade inhabits highland habitats and consists of three Andean species from cloud forests found > 1,500 m asl (*O. gemcare*, *O. lehri*, *O. machiguenga*) and two species from higher elevations, occurring in elfin forests up to 2,900 m asl or puna grasslands up to 3,850 m asl (*O. pereger* and *O. ayacucho*, respectively). In the second lineage, species from the southern yungas forests of Argentina and Bolivia between 1,600–1,700 m asl (*O. barituensis* and *O. discoidalis*) were sister to two species that occur throughout Peruvian and Bolivian humid forests t<1,300 m asl (*O. cruralis* and *O. madidi*). The sister taxa to these two clades, *O. heterodactylus*, is the species located farthest to the east in this study (at the Cerrado dry forests of southeastern Bolivia).

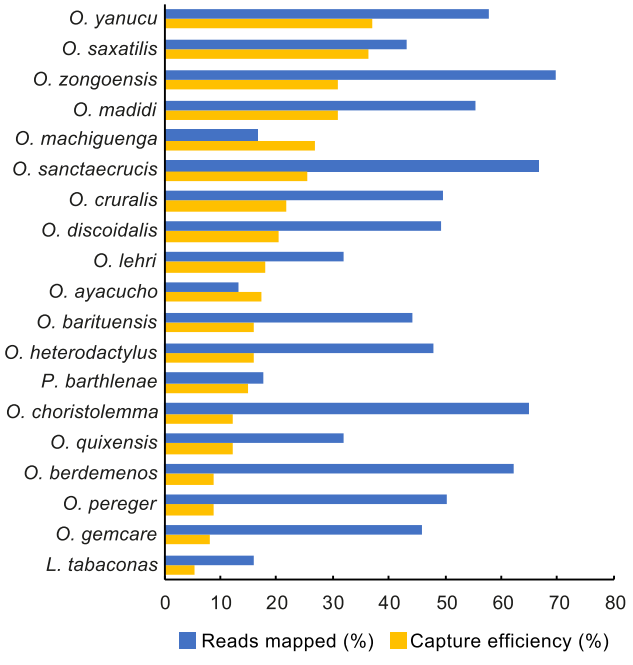

Fig. 1. Percent of mapped reads (blue) and capture efficiency (yellow) per species as determined by the HybPhyloMaker pipeline. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

The topology of the trees recovered by ASTRAL using HybPiper and HybPhyloMaker pipelines were similar and almost all differences corresponded to intraspecific relationships, which are not expected to be well portrayed in bifurcating trees. Most nodes had lower support using HybPiper than with HybPhyloMaker, and the species *O. quixensis* and *O. saxatilis* were still recovered as polyphyletic in both trees (Figure S1). The ASTRAL-HybPiper tree inferred *O. barituensis* and *O. discoidalis* to be polyphyletic as well (Figure S1a). Unexpectedly, these species were not monophyletic in ASTRAL, MRL, and SVDq trees inferred with HybPhyloMaker (Figure S1b; Figure S2b,c) but were monophyletic with the ML concatenated and ASTRID approaches (Fig. 3; Figure S2a). The ASTRAL-HybPiper tree did not group *O. berdemenos* and *O. yanucu* together. In addition, with this pipeline, the species *O. machiguenga* was not sister to *O. lehri*, but to *O. gemcare*.

3.3. Influence of reading frame correction and missing data

To test whether correcting the alignment by reading frame could affect the resulting phylogenies, we repeated the inferences for each sample tree method without adjusting to the correct reading frame. In this case, HybPhyloMaker retrieved 10,190 exons (Table S4). We did not find large differences in the proportion of variable sites, parsimony informative sites, or missing sites between frame-corrected and frame-uncorrected data (0.120, 0.049 and 0.68 in the frame-corrected; 0.124, 0.050 and 0.72 in the frame-uncorrected, respectively). However, there was a greater proportion of possibly saturated exons ("Satur_slope" >1) in the data without correction by frame (2,877 exons; 28.23%) compared to the data with frame correction (207 exons; 5.43%). Nonetheless, for each tree method, we found that both frame-corrected (3,810 exons) and frame-uncorrected data (10,190 exons) reconstructed similar topologies (Figure S3). At the interspecific level, only the method ASTRID inferred different topologies when using frame-corrected and frame-uncorrected alignments: the tree with frame-corrected data supported the monophyly of *O. barituensis* (Figure S3c) but the one with frame-uncorrected data showed paraphyly (Figure S3d). At the intraspecific level, different topologies either within *O. cruralis*, *O. lehri*, or *O. quixensis* were found between trees inferred with frame-corrected and frame-uncorrected data across different methods.

The different data filtering parameters that we applied on the frame-corrected data produced 25 datasets with variable levels of missing data that ranged from 3,810 exons (dataset 'maxmissing99_minsp1′: a 99% maximum of Ns and at least 1% of samples in the alignment) to 1 exon (dataset 'maxmissing1_minsp100′: 1% maximum of Ns and 100% of samples). Nine of these datasets had >1,000 exons, 4 datasets had 100–1,000 exons, and 12 datasets had <100 exons (Fig. 4a). In general, we observed an increase in the nRF distance between trees and the reference (i.e., 'maxmissing99_minsp1′) when fewer data were used. However, in most cases, datasets for which we required at least 75% of samples per alignment (i.e., 'maxmissing99_minsp75′,

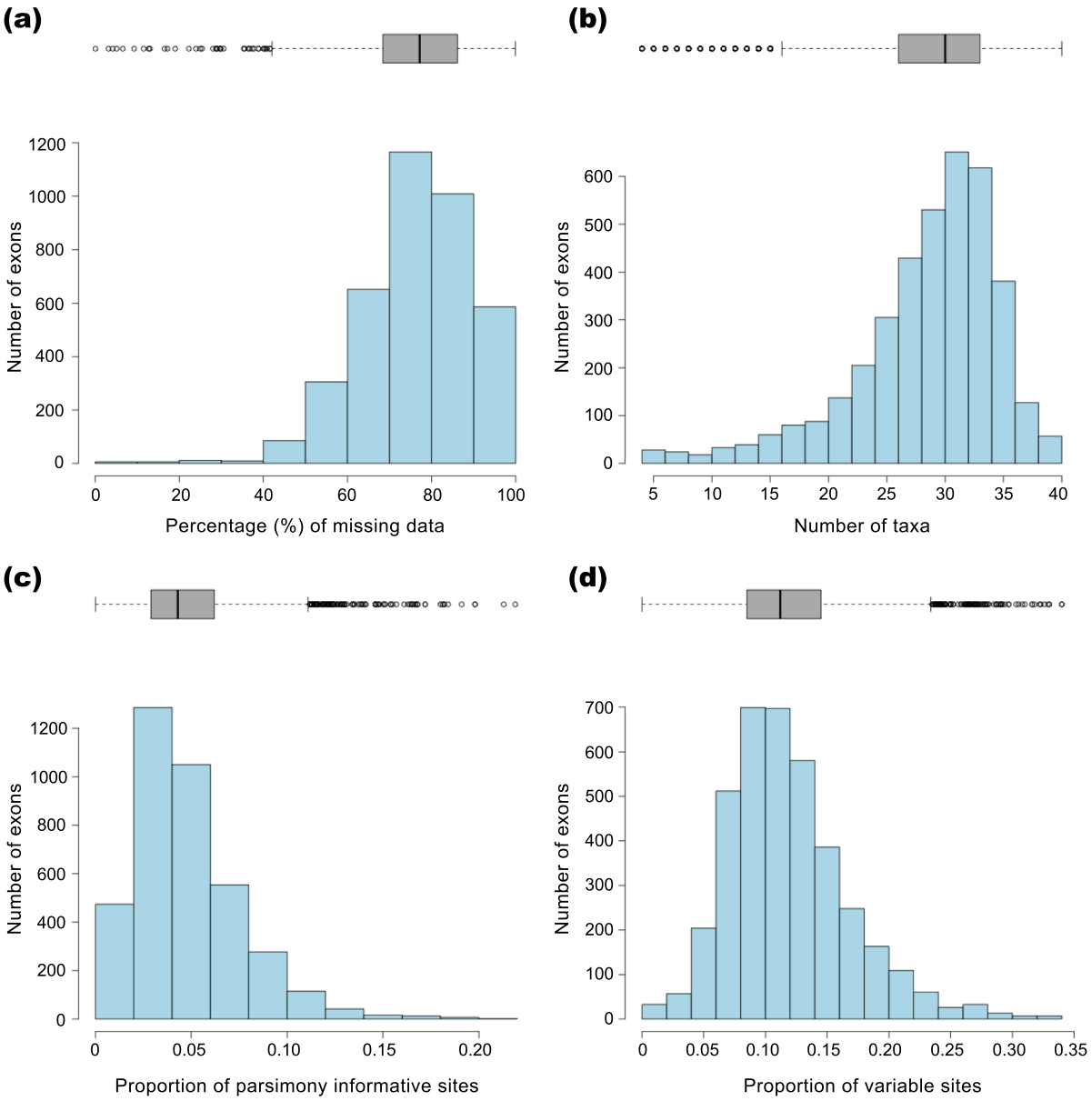

Fig. 2. Characteristics of the 3,810 frame-corrected exons obtained with the HybPhyloMaker pipeline in terms of (a) percentage missing data, (b) number of taxa, (c) proportion of parsimony informative, and (d) proportion of variable sites per individual exon.

'maxmissing75_minsp75′, 'maxmissing50_minsp75′ and 'maxmissing25_minsp75′) led to topologies with lower nRF distances to the reference than other datasets with a similar number of exons (Fig. 4b). We observed high concordance between sample trees inferred with the dataset 'maxmissing99_minsp1′ and with the dataset 'maxmissing75_minsp75′ across all methods (average nRF = 0.238). These two datasets had a substantially different number of exons (3,810 exons and 3,463,041 bp vs. 159 exons and 117,579 bp, respectively). In contrast, the average nRF distance across the 25 datasets and all five sample tree methods (15,625 comparisons) was much higher (0.639). Therefore, we used the frame-corrected 'maxmissing75_minsp75′ dataset for computationally-demanding analyses that would otherwise not be feasible with the 'maxmissing99_minsp1′ data (see subsections 3.5 and 3.6). Trees generated with this dataset were concordant with the complete 'maxmissing99_minsp1′ dataset in all interspecific relationships except for those relationships that changed depending on the method used and described earlier.

3.4. Pairwise species similarity and population structure

The average uncorrected genetic *p*-distance across species was 0.043 after excluding the outgroups (*P. barthlenae* and *L. tabaconas*). The lowest interspecific genetic differentiation was found between *O. quixensis* and *O. saxatilis* (0.018), followed by *O. barituensis* and *O. discoidalis* (0.022), and *O. choristolemma* and *O. sanctaecrucis* (0.023) (Figure S4). We examined shared ancestry among the species pairs in which at least two individuals were available per species (i.e., *O. quixensis* and *O. saxatilis*; *O. barituensis* and *O. discoidalis*; and *O. machiguenga* and *O. lehri*) (Fig. 5). Across all individuals in each of these pairs, we detected 3,055 SNPs, 1,599 SNPs, and 1,655 SNPs, respectively. Population structure analyses showed that the data for the samples from *O. quixensis* and *O. saxatilis* was most consistent with the presence of two clusters ($K = 2$), as inferred by the mean cross-entropy criterion across 10 replicate runs (Figures 5, S5 and S6). One cluster contained seven individuals of *O. quixensis* and one *O. saxatilis* (RS66), while the other cluster consisted of the two other *O. saxatilis* and one

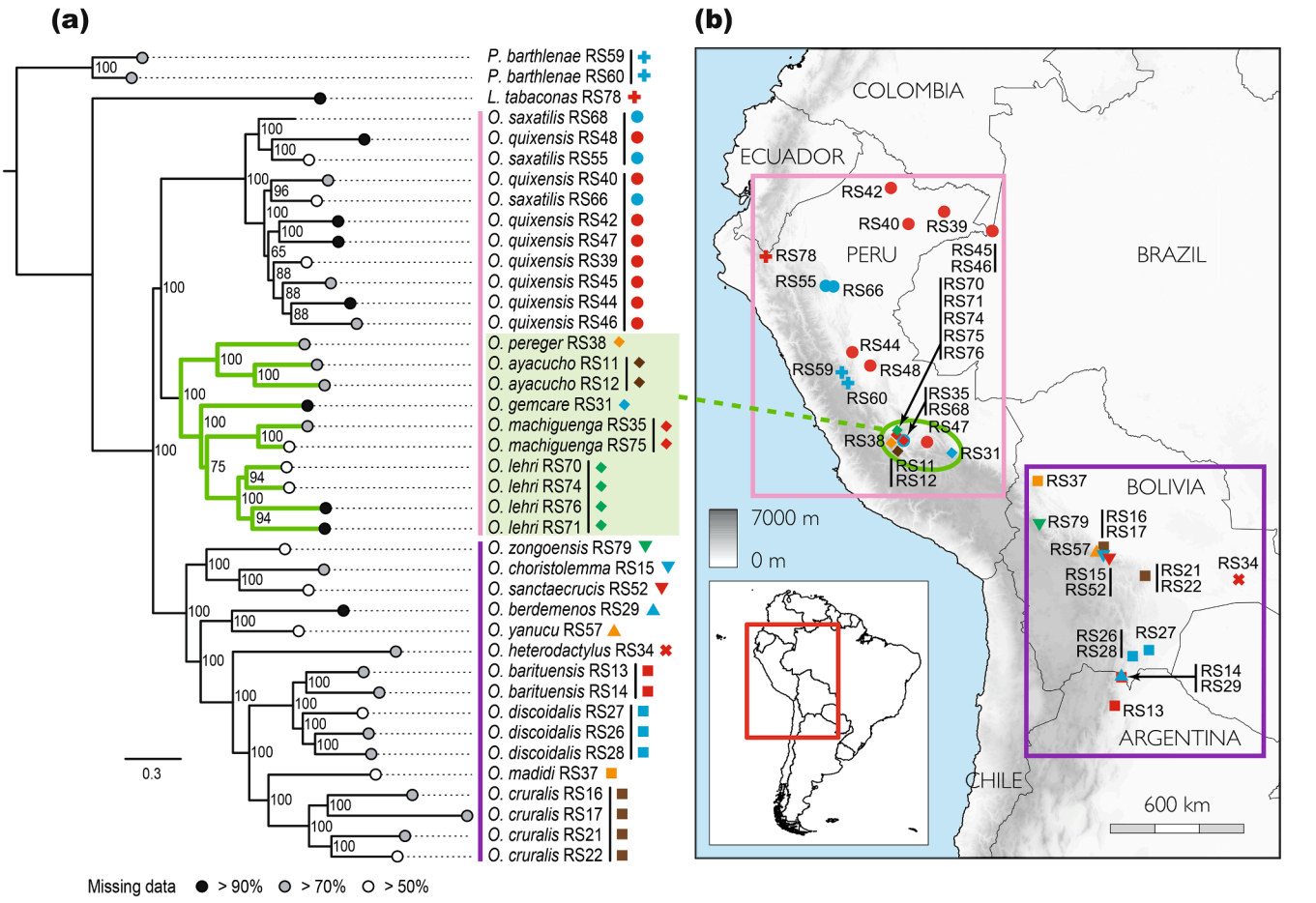

Fig. 3. Phylogenetic relationships of *Oreobates* using genomic data. (a) Concatenated maximum likelihood sample tree based on the 3,810 frame-corrected exons obtained with the HybPhyloMaker pipeline. Circles at the branch tips indicate percent of missing data: black, >90%; grey, >70%; white, >50%; no circle, <50%. Branches in green indicate highland species. (b) Sampling localities in this study. Coloured lines and rectangles indicate individuals distributed from Colombia to Peru (pink) and from Bolivia to Argentina (purple). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

O. quixensis (RS48). The data for the samples from *O. machiguenga* and *O. lehri* showed alike results, but in this case, the two species appeared well-defined at $K = 2$. In contrast, for *O. barituensis* and *O. discoidalis*, population structure analyses supported the presence of one cluster ($K = 1$) and they were unable to separate the individuals into different genetic groups despite using >1000 polymorphic positions.

3.5. Conflicting phylogenetic signals and gene flow

Inspection of gene tree/species tree discordance using PhyParts (Fig. 6a) showed that the split of *Oreobates* into two lineages (separating northern and southern species; see subsection 3.2) was supported by a large percentage of concordant gene trees (~78%; other trees supported alternative topologies or were uninformative for this split). One lineage was supported by ~16% gene trees, whereas the other was supported by ~42% gene trees. In the first lineage, the highland clade (*O. ayacucho*, *O. gemcare*, *O. lehri*, *O. machiguenga*, and *O. pereger*) was supported by ~26% of the gene trees. As expected, the presence of a larger proportion of exons supporting internal nodes is associated with longer branches. Similarly, short internal nodes tend to have a smaller proportion of supporting trees, in addition to lower values of gCF and gIC (Figure S7). Thus, trees involving fewer exons would be less likely to provide support for those internal nodes.

We used a PhyloNet network to assess the possible presence of ancient gene flow. The AIC was lowest for the presence of two reticulations (Table S5). Within the southern clade, the ancestor of the clade of *O. barituensis* and *O. discoidalis* could have had gene flow at different time-periods with members of the lineage represented by *O. heterodactylus*, *O. cruralis*, and *O. madidi* (Fig. 6b).

3.6. Individual heterozygosity and rates of evolution

Species of the highland clade showed a significant trend of reduced individual heterozygosity compared to species of the lowland clade consisting of *O. barituensis*, *O. cruralis*, *O. discoidalis*, *O. heterodactylus*, and *O. madidi* (*t* test comparison of highland vs. lowland clades, $p = 0.002$; Fig. 6b,c). The outgroup species, which are also found in Andean highland habitats, had low heterozygosity as well. However, a species found ≤ 1,500 m asl (*O. saxatilis*) had the smallest proportion of heterozygous positions per individual and its sister taxa (*O. quixensis*) also exhibited low heterozygosity compared to other lowland species. Surprisingly, STRUCTURE and sNMF analyses for these two species showed that lineages may not be separated according to elevation as anticipated and additional fragmentation could be present (see results for $K = 3$ in Figure S5).

Global ω (d_N/d_S) across all branches in the species tree was 0.273 as estimated by the null (M0) model using the frame-corrected 'maxmissing75_minsp75' dataset. When applying the 'b_free' branch model, ω for the highland clade was 0.302 (significantly different from M0 model, p-value <0.05), whereas the ω for the lowland clade was 0.269 (not significantly different from M0 model, p-value = 0.54). Similar values of ω were obtained using the dataset in which taxa with

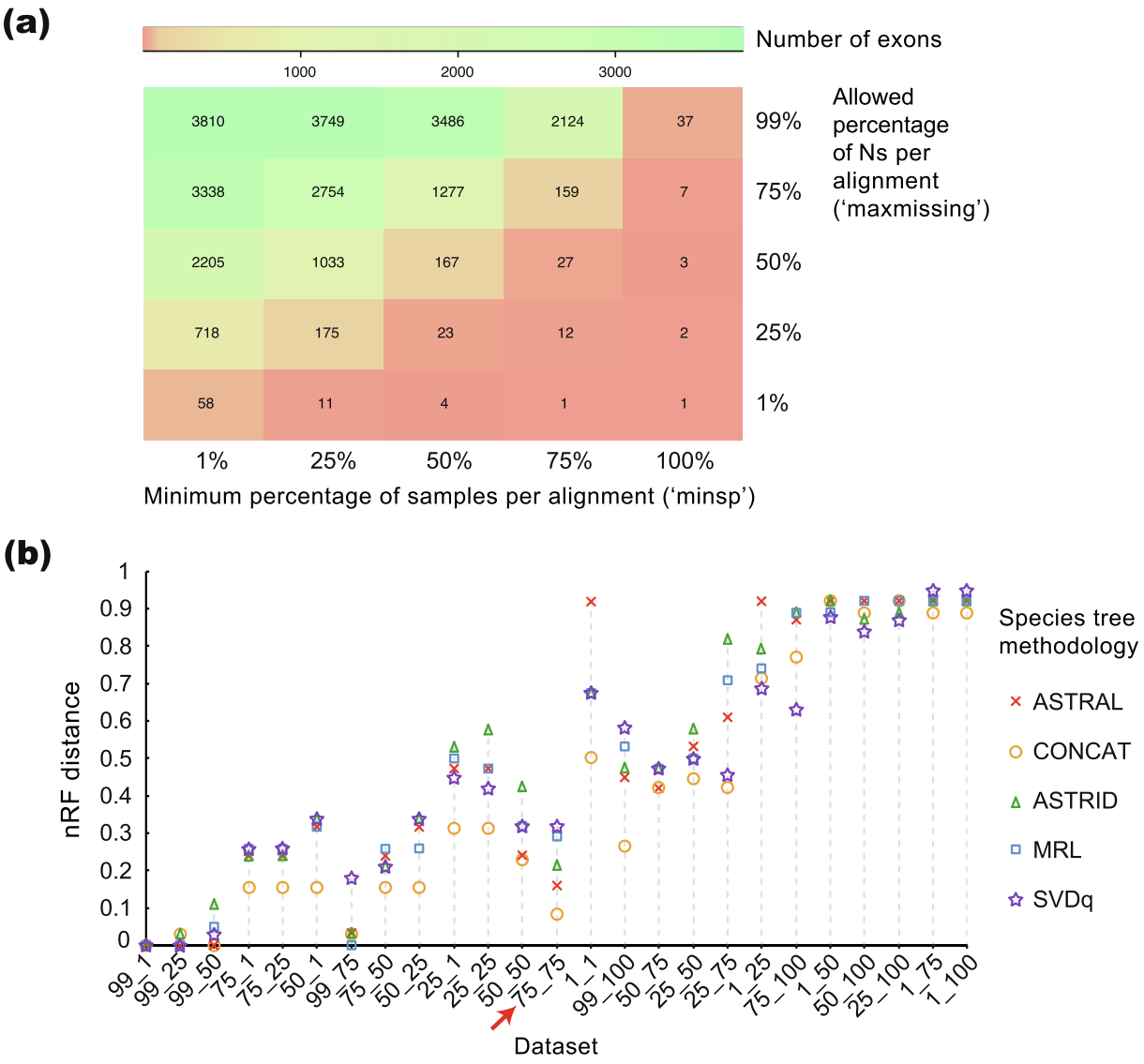

Fig. 4. Evaluation of the effect of missing data using the HybPhyloMaker pipeline: (a) Number of exons for each of the 25 datasets. Two different data filtering thresholds were applied based on: the maximum percentage of allowed Ns per alignment ('maxmissing'); the minimum percentage of samples present per alignment ('minsp'). (b) Normalized RF (nRF) distances between each of the 25 datasets' sample trees (source trees) inferred with five methodologies (ASTRAL, ASTRID, ML CONCAT, MRL, and SVDq) and their corresponding reference tree ('maxmissing99_ minsp1′). Datasets were sorted by decreasing number of exons. The red arrow highlights the similar phylogenetic relationships obtained between sample trees inferred with the 'maxmissing75_ minsp75′ and the 'maxmissing99_minsp1′ datasets (see text). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

stop codons had been deleted (Table S6). Selection analyses based on gene trees did not qualitatively change the results obtained with the species tree (Table S7). Thus, ω is lower in lowland than in highland species. Nonetheless, we did not find a significant correlation between d_S, d_N or ω and sampling elevation nor any of the environmental variables tested (temperature, precipitation, relative humidity, net primary productivity, and UVB radiation).

4. Discussion

4.1. Evolutionary history of Oreobates frogs

Our phylogenomic analyses clarify the biogeographic history of *Oreobates*. Due to a lack of calibration points within our phylogeny, we have not attempted to date it. Recent phylogenomic studies (Feng et al., 2017; Hime et al., 2020) suggest that the Terrarana group appeared 50–60 Myr and the clade including our ingroup plus outgroup may have diverged very early during the central Andean orogeny (Armijo et al., 2015), although high elevations were not attained until much later (Gregory-Wodzicki, 2000). The first diversification event in our phylogeny involved the divergence of two major clades, separating species currently distributed in Argentina and Bolivia from species found in Colombia and Peru. The frogs of the Argentinian-Bolivian lineage developed smaller and more slender bodies as compared to the members of the Colombian-Peruvian lineage, which evolved more robust bodies and with marked granular skin (Padial et al., 2008). The second diversification event took place within the latter lineage and led to the separation of a highland and a lowland clade. The observation that highland and lowland species tend to be separated in the phylogeny shows that the colonization of lowland habitats from highlands or the opposite are rare occurrences. This is surprising considering that lowland *Oreobates* occupy extensive areas along the Andean slopes and have apparently had ample opportunity to colonize the highlands. This emphasizes that the switch between highland and lowland habitats may not be easy and could involve many genomic changes.

We observed that individuals of *O. quixensis* and *O. saxatilis* were

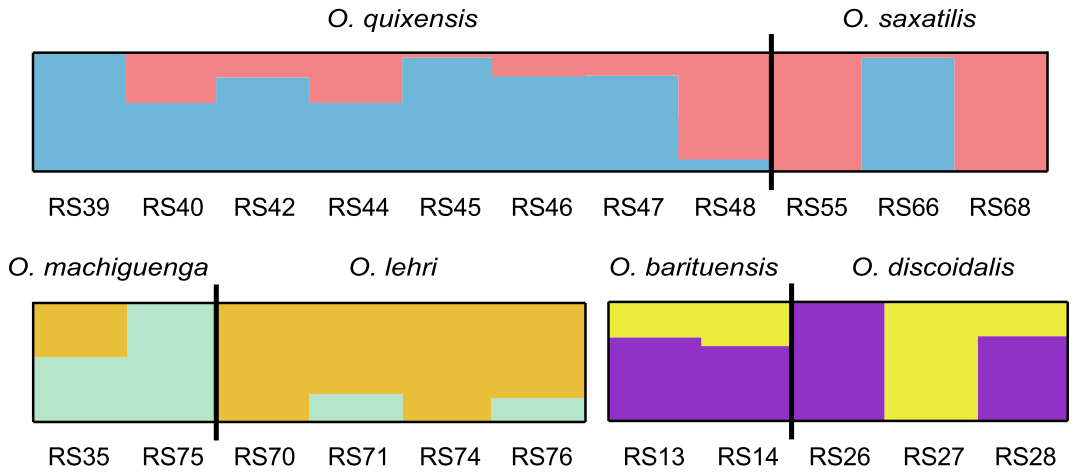

Fig. 5. Population structure for $K = 2$ of three different species pairs using the program sNMF: *O. quixensis* and *O. saxatilis*; *O. machiguenga* and *O. lehri*; *O. barituensis* and *O. discoidalis*.

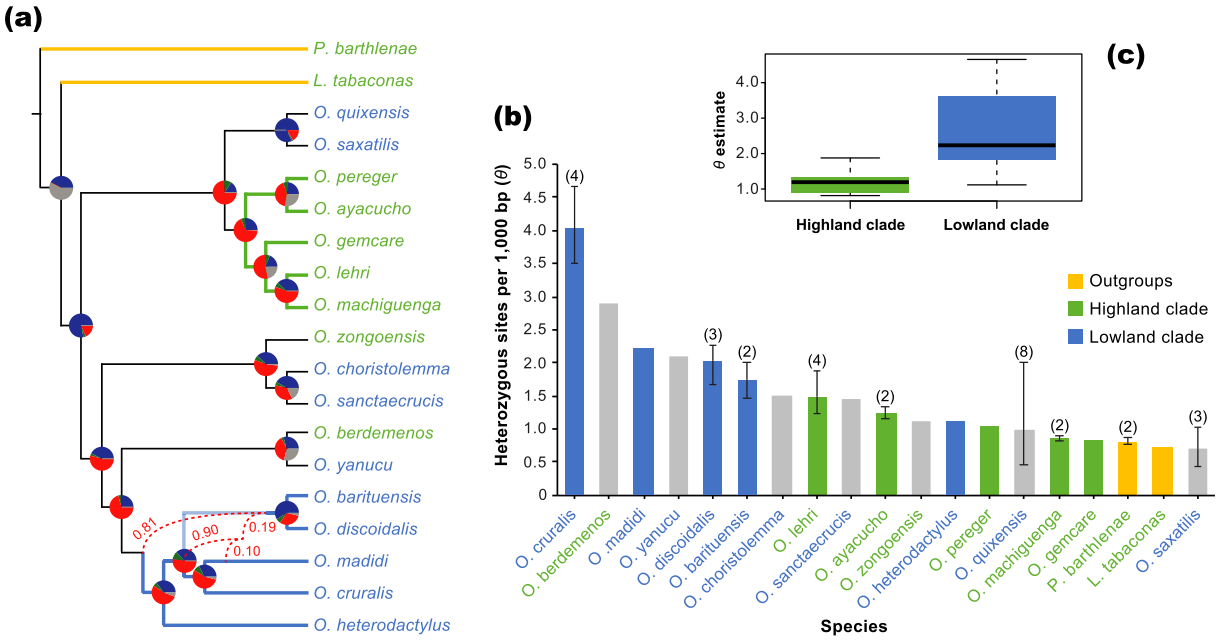

Fig. 6. Genetic diversity and gene flow in *Oreobates*: (a) Gene flow probability estimates between lineages according to the PhyloNet network. Pie charts show gene tree conflict at each node relative to the ASTRAL species tree as estimated by phyparts (blue, proportion concordant with the shown topology; green, proportion that support the dominant alternative topology; red, proportion that support remaining alternatives; grey: unsupported (<50% bootstrap support, BS). Branches coloured in green and blue indicate species of the highland and lowland clade, respectively. Yellow branches are the outgroup taxa. Names in green colour refer to species found >1,500 m asl, whereas names in blue colour refer to species found ≤ 1,500 m asl. (b) Average individual heterozygosity (θ) for each species as calculated by ANGSD. The total number of individuals for species with more than one sample is shown in brackets. Whiskers represent the minimum and maximum values. Grey bars represent species that occupy wide elevational ranges. (c) Heterozygosity estimates of species corresponding to highland vs. lowland clades. $p = 0.002$; significantly different means. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

consistently paraphyletic in all our phylogenetic analyses. Although STRUCTURE and sNMF suggested the existence of two clusters, these groups did not correspond to the two species. Interestingly, this species pair had already been recovered as paraphyletic in mitochondrial genealogies based on a fragment of the 16S rRNA by Padial et al. (2012). While two samples of *O. saxatilis* were grouped within *O. quixensis* in Padial et al. (2012), here we found that one individual of each species was genetically more similar to the other species. A possible explanation for this could be that some museum specimens were incorrectly identified. Even though adults of the two species are reciprocally diagnosable (Padial et al., 2012), their immature and juvenile stages share similar morphological traits, and errors in the species identification are not unlikely. Genomic analyses also suggest gene flow between the two lineages. An alternative explanation would be incomplete lineage sorting (ILS). However, the fact that both concatenation and summary coalescent methods converged on very similar topologies when using many loci, suggests that there is no ILS or that the effect of ILS is overwhelmed by genuine phylogenetic signal (Irisarri et al., 2018). To fully disentangle the relationship between these two species, a more complete analysis including genetic and phenotypic data of as many individuals as possible from across their distribution ranges would be needed and topotypes from both species would need to be sequenced as well. This could allow assessing possible misidentifications and population structure.

We found that individuals of two lowland species (*O. barituensis* and *O. discoidalis*) were not genetically separated into two distinct clusters.

The near-threatened *O. barituensis* (according to IUCN criteria) was described on the basis of morphometrics and bioacoustics and occurs in a small geographical area that partially overlaps with the broad distribution range of *O. discoidalis* (Vaira & Ferrari, 2008; Lavilla et al., 2010; IUCN, 2019). While some phylogenetic approaches have suggested reciprocal monophyly for these species, other analyses suggested paraphyly. Our clustering genomic analyses indicate that the two species do not represent separate gene pools and should be considered a single species (*O. discoidalis*). Interestingly, PhyloNet analyses suggest that the lineage *O. barituensis* + *O. discoidalis* derived from the admixture of two or three ancestral lineages, as suggested for other amphibians (Rancilhac et al., 2020). It is possible that this diverse origin contributed to the phenotypic variability that led to the claim of the existence of two species. This admixture at different points in the phylogeny implies interspecific gene flow after a secondary contact. Potentially, physical barriers are less likely to maintain lowland species spatially segregated over evolutionary time.

Highland habitats are characterized by higher fragmentation, lower temperatures, higher ultraviolet radiation, reduced levels of oxygen, lower water availability (remaining frozen during part of the year), and different flora and fauna compared to lowland areas (Montero-Mendieta et al., 2019). As a result, highland and lowland species are subjected to distinct selective pressures that shape their evolution and diversification. We found that species of *Oreobates* occurring in montane areas (cloud forests, elfin forests, and puna grasslands) have in general lower heterozygosity (H_e). Neutral theory predicts that for a given mutation rate (μ), H_e is positively correlated with the effective population size (N_e) (Kimura, 1983; Ohta, 1992). At the same time, small N_e causes purifying selection to be less efficient at removing slightly deleterious mutations due to strong genetic drift, resulting in an increased ratio of nonsynonymous to synonymous substitution rates (ω, dN/dS; Mohlhenrich & Mueller, 2016; Lanfear et al., 2010). This is what we observe in our case, with higher ω in montane species compared to species inhabiting Amazonian lowlands, dry forests, inter-Andean dry valleys, humid forests, and the yungas <1,700 m. The observation that H_e is lower in highland species suggests that their N_e is smaller compared to lowland species, and have larger ω, as expected if montane habitats are more fragmented (Montero-Mendieta et al., 2019). On the other hand, and in contrast to findings by Dugo-Cota et al. (2015) and Lin et al. (2019) for another group of Neotropical frogs and bumblebees, we did not observe an acceleration of d_S (synonymous substitution rates) in lowland species. Therefore, our data suggest that the fragmentation in mountain habitats may have led to species with reduced distribution ranges and lower N_e. As a result, purifying selection was weaker and promoted the relatively faster accumulation of nonsynonymous mutations in highland species of *Oreobates*.

Until now, some aspects of the evolutionary history of the frogs of the genus *Oreobates* remained fairly unknown. This was partly because this group includes species that are extremely difficult to find, such as *O. yanucu* and *O. zongoensis*, which are known only from one or a few individuals from a single locality (Reichle & Köhler, 1997; Köhler & Padial, 2016). Here, we have shown the utility of genomic data to understand the evolution of *Oreobates* across a wide elevational range. However, some questions remain unsolved. Further work combining genomic, ecological, behavioural, and morphological data is needed to fully address these questions and to better understand the role of mountain systems in the evolutionary history of amphibians.

4.2. Target enrichment phylogenomics

Our custom probes designed from the transcriptome of one species of *Oreobates* captured a set of homologous loci across most species in this genus and two other closely-related genera. During probe design, we did not filter for single-copy genes as our aim was to keep as many sequences as possible (~18,000 exons). A potential pitfall of this procedure is that paralogous loci could have been recovered in our data. However, the consistency of our inferences is confirmed by the overall congruence observed between phylogenies reconstructed with the two pipelines (HybPhyloMaker and HybPiper) despite one filtering out potential paralogs but not the other. We observed variation in the number of loci captured per species by HybPhyloMaker and HybPiper, which is likely due to the use of different methodologies for the assembly and identification of contigs that match the target sequences (Fér & Schmickl, 2018). Putative intronic flanking regions were not explicitly removed because HybPhyloMaker is not capable of extracting them, and they can be valuable for reconstructing shallow phylogenies (Forcina et al., 2021). Analyses using HybPhyloMaker consumed fewer computing resources and were faster than with HybPiper. There were minor topological differences between the trees inferred by the two pipelines that may reflect taxonomic problems for some species (discussed in Section 4.1), but we used the data obtained with HybPhyloMaker for subsequent analyses because the inferences by HybPiper had lower support values overall. Thus, HybPhyloMaker recovered data of higher phylogenetic informativeness compared to HybPiper.

In terms of the percentage of target bases that were retrieved, our probes delivered higher average capture success than the ones designed by Heinicke et al. (2018) to enrich conserved exons across Terrarana frogs (19% vs. 10%). This shows that capture probes based on the transcriptome of one or few species can be useful to obtain data for phylogenomic and evolutionary history reconstructions among closely-related taxa (such as the genus *Oreobates* compared to the entire Terrarana group) when no other genomic resources are available. However, the lack of phylogenetic diversity in our probe design implies that this approach may not be suitable for species with deep genetic divergences. In such a case, the use of the recently available FrogCap marker set (Hutter et al., 2019) may be more appropriate. This marker set effectively captures a large array of markers across variable timescales in the amphibian tree of life. Ultimately, to choose one method or the other, one must consider the availability, sensitivity and specificity of initial genomic resources for the ingroup taxa, as well as the cost of the project.

An important consideration before reconstructing phylogenies with target enrichment data is to minimize “noise” that may mislead the phylogenetic inference process (Herrando-Moraira et al., 2018). We explored the impact of two types of dataset noise filtering: removing loci with incorrect reading frames, and selecting different thresholds for the inclusion of loci and taxa with missing data. The observation that the data without frame correction consists of many more exons compared to the data with frame correction (10,190 and 3,810 exons, respectively) means that ~63% of the loci had ambiguous reading frames or more than five stop codons. This may have been induced by sequencing errors that insert or delete a nucleotide and produce reading frameshifts, but more likely causes could be the criteria for probe designing and read mapping being not restrictive enough, which can lead to ambiguous sequence calls. The latter represents a trade-off between the depth of coverage of protein-coding genes and the accuracy of the genotype calls (Miya et al., 2015). We found that the removal of loci with incorrect reading frames decreased the proportion of exons with high saturation levels. This implies that the reading frame correction produced data with a stronger phylogenetic signal than the data without this correction, and therefore, more accurate inferences. Homoplasy in the frame-uncorrected data may have contributed to the differences observed between trees reconstructed with frame-corrected and frame-uncorrected data for two of the inference methods. Accordingly, we focused on frame-corrected data to examine the evolutionary history of *Oreobates*.

Regardless of reading frame correction, the average percentage of missing data was ~ 70% in both datasets. In addition to missing nucleotides resulting from biological processes such as insertions and deletions, target enrichment data increased noise due to: (1) inconsistent performance of the sequence-capture among samples leading to missing data; and (2) the stochasticity inherent in collecting data, resulting in variable taxon coverage across loci (Hosner et al., 2016). As a result, the inferred phylogenies may exhibit very long branches (Darriba et al.,

2016). In the concatenated tree of *Oreobates* inferred with the 3,810 frame-corrected exons, samples with high missing data have longer branches than others (see Fig. 3a). By comparing trees inferred from subsets of frame-corrected loci with different amounts of missing data, we found that retaining exons with >75% taxa per loci, each with <75% of ambiguous positions (N) solves that issue. The dataset consisting of 159 exons had considerably less missing data (~35%) and shorter branches than the complete frame-corrected data but resulted in similar tree topologies with all inference methods. Because downstream analyses were based on this reduced data, and we only retained one individual per species (i.e., the one with less missing data), the inferred evolutionary framework is unlikely affected by branch length biases.

The species relationships recovered among *Oreobates* with these data are congruent to those inferred in previous studies with short fragments of the 12S, 16S, RAG1, and TYR genes, at least for branches receiving high bootstrap values or posterior probabilities (Padial et al., 2008, 2012, 2014; Hedges et al., 2008; Pyron & Wiens, 2011; Pereyra et al., 2014; Köhler & Padial, 2016; Jetz & Pyron, 2018; Vaz-Silva et al., 2018; Pansonato et al., 2020). There are several valuable implications that can be extracted from these results. First, compared to phylogenomic inferences, phylogenetic trees of *Oreobates* based on a few mitochondrial and nuclear gene markers can be considered reliable in most cases, as noticed by Heinicke et al. (2018) for other amphibians. Second, the use of the different phylogenomic pipelines to extract target sequences and inference programs does not have a strong impact on the inferred evolutionary relationships between species. On the other hand, the application of different filtering parameters is a critical step to phylogenetic reconstructions. Third, reading frame correction is recommended to obtain data with higher phylogenetic signal, but allowing alignments with a few stop codons can help recover valid exons with errors induced via sequencing or read mapping. Finally, as shown in previous studies, denser taxon coverage (~75% taxa per loci, in this case) seems to be more important than sequencing thousands of genes to resolve species relationships with phylogenomic data (Betancur-R et al., 2019; Montero-Mendieta & Dheer, 2019). In contrast, having high percentages of ambiguous positions in the alignments (75–99% of Ns) does not have a clear effect in the inference process.

5. Conclusions

We examined the evolutionary history of *Oreobates*, a genus of frogs that inhabits low and high elevation environments in the Andes. Using sequence-capture, we obtained the sequence ofnumerous independent loci for phylogenetic inference and evolutionary analyses. We tested two alternative pipelines for assembling phylogenomic datasets from capture data and compared the effect of key methodological decisions. Robust evolutionary relationships were inferred, supported by both concatenation and summary coalescent methods. The inferred evolutionary framework showed different selective regimes between highland and lowland species, as well as conflicting phylogenetic relationships among the latter. Taken together, our results not only improve our understanding of the evolution of *Oreobates* frogs but also provide insights into the use of target enrichment data for phylogenomic studies.

CRediT authorship contribution statement

Santiago Montero-Mendieta: Methodology, Software, Investigation, Formal analysis, Data curation, Writing - original draft, Writing - review & editing, Visualization. **Ignacio De Riva:** Resources, Writing - review & editing, Funding acquisition. **Iker Irisarri:** Writing - review & editing. **Jennifer A. Leonard:** Methodology, Investigation, Resources, Writing - review & editing. **Matthew T. Webster:** Supervision, Writing - review & editing. **Carles Vilà:** Conceptualization, Supervision, Methodology, Writing - original draft, Writing - review & editing, Project administration, Funding acquisition.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The tissue samples used for this study were provided by the frozen tissue collection of the Museo Nacional de Ciencias Naturales (MNCN-CSIC) in Madrid, Spain (MNCN/ADN collection; we thank Isabel Rey and Beatriz Dorda for their help). We would like to acknowledge Doñana's Singular Scientific-Technical Infrastructure (ICTS-RBD) for informatic resources and Hanyi Li for computational assistance. We thank Irene Quintanilla and the Laboratorio de Ecologia Molecular (LEM-EBD) for laboratory support. We also thank Rosa Fernández, Lisa Pokorny, and Tamara Villaverde for their discussions and feedback on early drafts of the manuscript. This research was supported by grants from the Spanish Government (Ministerio de Economía y Competitividad) to C.V. (CGL2013-47547-P and CGL2016-75227-P), I.D.l.R (CGL2011-30393), and an FPI (Formación de Personal Investigador) fellowship (BES-2014-069006) as well as three travel grants (EEBB-I-16-10576, EEBB-I-17-12168, and EEBB-I-18-12878) to S.M.-M. The funders had no role in study design, data collection, analysis, decision to publish, or preparation of the manuscript.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ympev.2021.107167.

References

Akaike, H., 1974. A new look at the statistical model identification. IEEE Trans. Autom. Control 19, 716–723.
Allendorf, F.W., Hohenlohe, P.A., Luikart, G., 2010. Genomics and the future of conservation genetics. Nat. Rev. Genet. 11, 697–709.
Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Lipman, D.J., 1997. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. Nucleic Acids Res. 25, 3389–3402.
Antonelli, A., Sanmartín, I., 2011. Why are there so many plant species in the Neotropics? Taxon 60, 403–414.
Armijo, R., Lacassin, R., Coudurier-Curveur, A., Carrizo, D., 2015. Coupled tectonic evolution of Andean orogeny and global climate. Earth Sci. Rev. 143, 1–35.
Bairoch, A., 2000. The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. Nucleic Acids Res. 28, 45–48.
Beckmann, M., Václavík, T., Manceur, A.M., Šprtová, L., von Wehrden, H., Welk, E., Cord, A.F., 2014. glUV: a global UV-B radiation data set for macroecological studies. Methods Ecol. Evol. 5, 372–383.
Betancur-R, R., Arcila, D., Vari, R.P., Hughes, L.C., Oliveira, C., Ortí, G., 2019. Phylogenomic incongruence, hypothesis testing, and taxonomic sampling: The monophyly of characiform fishes. Evolution 73, 329–345.
Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30, 2114–2120.
Borowiec, M.L., 2016. AMAS: a fast tool for alignment manipulation and computing of summary statistics. PeerJ 4, e1660.
Bravo, G.A., Antonelli, A., Bacon, C.D., Bartoszek, K., Blom, M.P.K., Edwards, S.V., 2019. Embracing heterogeneity: coalescing the Tree of Life and the future of phylogenomics. PeerJ 7, e6399.
Carlsen, M.M., Fér, T., Schmickl, R., Leong-Škorničková, J., Newman, M., Kress, W.J., 2018. Resolving the rapid plant radiation of early diverging lineages in the tropical Zingiberales: Pushing the limits of genomic data. Mol. Phylogenet. Evol. 128, 55–68.
Chan, K.O., Brown, R.M., 2020. Elucidating the drivers of genetic differentiation in Malaysian torrent frogs (Anura: Ranidae: Amolops): a landscape genomics approach. Zool. J. Linn. Soc. 190, 65–78.
Chifman, J., Kubatko, L., 2014. Quartet inference from SNP data under the coalescent model. Bioinformatics 30, 3317–3324.
Constantinides, B., Robertson, D.L., 2017. Kindel: indel-aware consensus for nucleotide sequence alignments. Journal of Open Source Software 2, 282.
Coyne, J., Orr, H., 2004. Speciation. Sinauer Associates Inc., Massachusetts. USA.
Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., Durbin, R., 2011. The variant call format and VCFtools. Bioinformatics 27, 2156–2158.
Darriba, D., Weiß, M., Stamatakis, A., 2016. Prediction of missing sequences and branch lengths in phylogenomic data. Bioinformatics 32, 1331–1337.
Delsuc, F., Brinkmann, H., Philippe, H., 2005. Phylogenomics and the reconstruction of the tree of life. Nat. Rev. Genet. 6, 361–375.

Dugo-Cota, Á., Castroviejo-Fisher, S., Vilà, C., González-Voyer, A., 2015. A test of the integrated evolutionary speed hypothesis in a Neotropical amphibian radiation. Glob. Ecol. Biogeogr. 19, 733–740.
Faircloth, B.C., McCormack, J.E., Crawford, N.G., Harvey, M.G., Brumfield, R.T., Glenn, T.C., 2012. Ultraconserved elements anchor thousands of genetic markers spanning multiple evolutionary timescales. Syst. Biol. 61, 717–726.
Feng, Y.J., Blackburn, D.C., Liang, D., Hillis, D.M., Wake, D.B., Zhang, P., 2017. Phylogenomics reveals rapid, simultaneous diversification of three major clades of Gondwanan frogs at the Cretaceous-Paleogene boundary. Proceedings of the National Academy of Sciences USA 114, E5864–E5870.
Fér, T., Schmickl, R.E., 2018. Hybphylomaker: Target enrichment data analysis from raw reads to species trees. Evolutionary Bioinformatics 14, 1–9.
Forcina, G., Camacho-Sánchez, M., Tuh, F.Y.Y., Moreno, S., Leonard, J.A., 2021. Markers for genetic change. Heliyon 7, e05583.
Frichot, E., Mathieu, F., Trouillon, T., Bouchard, G., François, O., 2014. Fast and efficient estimation of individual ancestry co-efficients. Genetics 196, 973–983.
Fritz, S.A., Rahbek, C., 2012. Global patterns of amphibian phylogenetic diversity. J. Biogeogr. 39, 1373–1382.
Frost, D.R., 2020. Amphibian Species of the World: an online reference. Version 6.1 Accessed on 22–04–2020. Electronic database accessible at. American Museum of Natural History, New York, USA.
Funk, W.C., Zamudio, K.R., Crawford, A.J., 2018. Advancing understanding of amphibian evolution, ecology, behavior, and conservation with massively parallel sequencing. Population Genomics. Springer, Cham, pp. 1–44.
González-Voyer, A., Padial, J.M., Castroviejo-Fisher, S., De la Riva, I., Vilà, C., 2011. Correlates of species richness in the largest Neotropical amphibian radiation. J. Evol. Biol. 24, 931–942.
Gregory-Wodzicki, K., 2000. Uplift history of the Central and Northern Andes: a review. Geol. Soc. Am. Bull. 112, 1091–1105.
Guarnizo, C.E., Cannatella, D.C., 2013. Genetic divergence within frog species is greater in topographically more complex regions. J. Zool. Syst. Evol. Res. 51, 333–340.
Hedges, S.B., Duellman, W.E., Heinicke, M.P., 2008. New World direct-developing frogs (Anura: Terrarana): Molecular phylogeny, classification, biogeography, and conservation. Zootaxa 1737, 1–182.
Heinicke, M.P., Lemmon, A.R., Lemmon, E.M., McGrath, K., Hedges, S.B., 2018. Phylogenomic support for evolutionary relationships of New World direct-developing frogs (Anura: Terraranae). Mol. Phylogenet. Evol. 118, 145–155.
Herrando-Moraira, S., Calleja, J.A., Carnicero, P., Fujikawa, K., Galbany-Casals, M., Vilatersana, R., 2018. Exploring data processing strategies in NGS target enrichment to disentangle radiations in the tribe Cardueae (Compositae). Mol. Phylogenet. Evol. 128, 69–87.
Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G., Jarvis, A., 2005. Very high resolution interpolated climate surfaces for global land areas. Int. J. Climatol. 25, 1965–1978.
Hime, P.M., Lemmon, A.R., Lemmon, E.C.M., Prendini, E., Brown, J.M., Weisrock, D.W., 2020. Phylogenomics reveals ancient gene tree discordance in the amphibian tree of life. Syst. Biol. 70, 49–66.
Hosner, P.A., Faircloth, B.C., Glenn, T.C., Braun, E.L., Kimball, R.T., 2016. Avoiding missing data biases in phylogenomic inference: an empirical study in the landfowl (Aves: Galliformes). Mol. Biol. Evol. 33, 1110–1125.
Huerta-Cepas, J., Serra, F., Bork, P., 2016. ETE 3: Reconstruction, analysis, and visualization of phylogenomic data. Mol. Biol. Evol. 33, 1635–1638.
Hutter, C.R., Cobb, K.A., Portik, D.M., Travers, S.L., Wood Jr., P.L., Brown, R.M., 2019. FrogCap: A modular sequence capture probe set for phylogenomics and population genetics for all frogs, assessed across multiple phylogenetic scales. bioRxiv, 825307. https://doi.org/10.1101/825307.
Hutter, C.R., Guayasamin, J.M., Wiens, J.J., 2013. Explaining Andean megadiversity: the evolutionary and ecological causes of glassfrog elevational richness patterns. Ecol. Lett. 16, 1135–1144.
Hutter, C.R., Lambert, S.M., Wiens, J.J., 2017. Rapid diversification and time explain amphibian richness at different scales in the tropical Andes, Earth's most biodiverse hotspot. Am. Nat. 190, 828–843.
Imhoff, M.L., Bounoua, L., 2006. Exploring global patterns of net primary production carbon supply and demand using satellite observations and statistical data. J. Geophys. Res. 111, D22S12.
Imhoff, M.L., Bounoua, L., Ricketts, T., Loucks, C., Harriss, R., Lawrence, W.T., 2004. Global patterns in human consumption of net primary production. Nature 429, 870–873.
Irisarri, I., Meyer, A., 2016. The identification of the closest living relative(s) of tetrapods: phylogenomic lessons for resolving short ancient internodes. Syst. Biol. 65, 1057–1075.
Irisarri, I., Singh, P., Koblmüller, S., Torres-Dowdall, J., Henning, F., Meyer, A., 2018. Phylogenomics uncovers early hybridization and adaptive loci shaping the radiation of Lake Tanganyika cichlid fishes. Nat. Commun. 9, 3159.
IUCN (2019). IUCN SSC Amphibian Specialist Group 2019. Oreobates barituensis. The IUCN Red List of Threatened Species 2019: e.T18435652A20174921. https://doi.org/10.2305/IUCN.UK.2019-1.RLTS.T18435652A20174921.en. Downloaded on 23–08–2019.
Jakobsson, M., Rosenberg, N.A., 2007. CLUMPP: A cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. Bioinformatics 23, 1801–1806.
Jetz, W., Pyron, R.A., 2018. The interplay of past diversification and evolutionary isolation with present imperilment across the amphibian tree of life. Nat. Ecol. Evol. 2, 850–858.
Johnson, M.G., Gardner, E.M., Liu, Y., Medina, R.B., Goffinet, B., Wickett, N.J., 2016. HybPiper: extracting coding sequence and introns for phylogenetics from high-throughput sequencing reads using target enrichment. Appl. Plant Sci. 4, 1600016.
Junier, T., Zdobnov, E.M., 2010. The Newick utilities: high-throughput phylogenetic tree processing in the UNIX shell. Bioinformatics 26, 1669–1670.
Katoh, K., Toh, H., 2008. Recent developments in the MAFFT multiple sequence alignment program. Briefings Bioinf. 9, 286–298.
Keller, I., Alexander, J.M., Holderegger, R., Edwards, P.J., 2013. Widespread phenotypic and genetic divergence along altitudinal gradients in animals. J. Evol. Biol. 26, 2527–2543.
Kent, W.J., 2002. BLAT - The BLAST-like alignment tool. Genome Res. 12, 656–664.
Kimura, M., 1983. The Neutral Theory of Molecular Evolution. Cambridge University Press, Cambridge.
Köhler, J., Padial, J.M., 2016. Description and phylogenetic position of a new (singleton) species of *Oreobates* Jiménez de la Espada, 1872 (Anura: Craugastoridae) from the yungas of Cochabamba, Bolivia. Annals of Carnegie Museum 84, 23–38.
Korneliussen, T.S., Albrechtsen, A., Nielsen, R., 2014. ANGSD: analysis of next generation sequencing data. BMC Bioinf. 15, 356.
Kozak, K.H., Wiens, J.J., 2010. Accelerated rates of climatic-niche evolution underlie rapid species diversification. Ecol. Lett. 13, 1378–1389.
Lanfear, R., Ho, S.Y.W., Love, D., Bromham, L., 2010. Mutation rate is linked to diversification in birds. Proceedings of the National Academy of Sciences USA 107, 20423–20428.
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. Mol. Biol. Evol. 34, 772–773.
Langmead, B., Salzberg, S.L., 2012. Fast gapped-read alignment with Bowtie 2. Nat. Methods 9, 357–359.
Lartillot, N., Poujol, R., 2011. A phylogenetic model for investigating correlated evolution of substitution rates and continuous phenotypic characters. Mol. Biol. Evol. 28, 729–744.
Lavilla, E., Cortez, C., Reichle, S., De la Riva, I., Köhler, J. (2010). Oreobates discoidalis. The IUCN Red List of Threatened Species 2010: e.T56564A11484610. https://doi.org/10.2305/IUCN.UK.2010-2.RLTS.T56564A11484610.en. Downloaded on 23–08–2019.
Lemmon, A.R., Emme, S.A., Lemmon, E.M., 2012. Anchored hybrid enrichment for massively high-throughput phylogenomics. Syst. Biol. 61, 727–744.
Li, H., 2011. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics 27, 2987–2993.
Li, H., Durbin, R., 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25, 1754–1760.
Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Durbin, R., 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25, 2078–2079.
Lin, G., Huang, Z., Wang, L., Chen, Z., Zhang, T., Zhao, F., 2019. Evolutionary rates of bumblebee genomes are faster at lower elevations. Mol. Biol. Evol. 36, 1215–1219.
Liedtke, H.C., Gower, D.J., Wilkinson, M., Gómez-Mestre, I., 2018. Macroevolutionary shift in the size of amphibian genomes and the role of life history and climate. Nat. Ecol. Evol. 2, 1792–1799.
Mallo, D., Posada, D., 2016. Multilocus inference of species trees and DNA barcoding. Philosophical Transactions of the Royal Society B: Biological Sciences 371, 20150335.
Mamanova, L., Coffey, A.J., Scott, C.E., Kozarewa, I., Emily, H., Turner, D.J., 2010. Target-enrichment strategies for next-generation sequencing. Nat. Methods 7, 111–118.
McCormack, J.E., Hird, S.M., Zellmer, A.J., Carstens, B.C., Brumfield, R.T., 2013. Applications of next-generation sequencing to phylogeography and phylogenetics. Mol. Phylogenet. Evol. 66, 526–538.
Meyer, M. & Kircher, M. (2010). Illumina sequencing library preparation for highly multiplexed target capture and sequencing. Cold Spring Harbor Protocols, 2010, prot5448.
Minh, B.Q., Schmidt, H.A., Chernomor, O., Schrempf, D., Woodhams, M.D., Lanfear, R., 2020a. IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. Mol. Biol. Evol. 37, 1530–1534.
Minh, B.Q., Hahn, M.W., Lanfear, R., 2020b. New methods to calculate concordance factors for phylogenomic datasets. Mol. Biol. Evol. 37, 2727–2733.
Mirarab, S., Reaz, R., Bayzid, M.S., Zimmermann, T., Swenson, M.S., Warnow, T., 2014. ASTRAL: genome-scale coalescent-based species tree estimation. Bioinformatics 30, 541–548.
Miya, F., Kato, M., Shiohama, T., Okamoto, N., Saitoh, S., Tsunoda, T., 2015. A combination of targeted enrichment methodologies for whole-exome sequencing reveals novel pathogenic mutations. Sci. Rep. 5, 9331.
Mohlhenrich, E.R., Mueller, R.L., 2016. Genetic drift and mutational hazard in the evolution of salamander genomic gigantism. Evolution 70, 2865–2878.
Montero-Mendieta, S., Grabherr, M., Lantz, H., De la Riva, I., Leonard, J.A., Vilà, C., 2017. A practical guide to build de-novo assemblies for single tissues of non-model organisms: The example of a Neotropical frog. PeerJ 5, e3702.
Montero-Mendieta, S., Tan, K., Christmas, M.J., Olsson, A., Vilà, C., Webster, M.T., 2019. The genomic basis of adaptation to high-altitude habitats in the eastern honey bee (*Apis cerana*). Mol. Ecol. 28, 746–760.
Montero-Mendieta, S., Dheer, A., 2019. Digest: Resolving phylogenomic conflicts in characiform fishes†. Evolution 73, 416–418.
New, M., Hulme, M., Jones, P., 1999. Representing 20th century space-time climate variability. I: development of a 1961–1990 mean monthly terrestrial climatology. J. Clim. 12, 829–856.

Nguyen, N., Mirarab, S., Warnow, T., 2012. MRL and SuperFine+MRL: new supertree methods. Algorithms Mol. Biol. 7, 3.
Ohta, T., 1992. The nearly neutral theory of molecular evolution. Annu. Rev. Ecol. Syst. 23, 263–286.
Padial, J.M., Chaparro, J.C., Castroviejo-Fisher, S., Guayasamin, J.M., Lehr, E., De la Riva, I., 2012. A revision of species diversity in the Neotropical genus *Oreobates* (Anura: Strabomantidae), with the description of three new species from the Amazonian slopes of the Andes. Am. Mus. Novit. 3752, 1–55.
Padial, J.M., Chaparro, J.C., De la Riva, I., 2008. Systematics of *Oreobates* and the *Eleutherodactylus discoidalis* species group (Amphibia, Anura), based on two mitochondrial DNA genes and external morphology. Zool. J. Linn. Soc. 152, 737–773.
Padial, J.M., Grant, T., Frost, D.R., 2014. Molecular systematics of terraranas (Anura: Brachycephaloidea) with an assessment of the effects of alignment and optimality criteria. Zootaxa 3825, 1–132.
Pansonato, A., Motta, A., Cacciali, P., Haddad, C., Strüssmann, C., Jansen, M., 2020. On the identity of species of *Oreobates* (Anura: Craugastoridae) from Central South America, with the description of a new species from Bolivia. Journal of Herpetology 54, 393–412.
Pereyra, M.M.O., Cardozo, D.E.D.E., Baldo, J., Baldo, D., 2014. Description and phylogenetic position of a new species of *Oreobates* (Anura: Craugastoridae) from Northwestern Argentina. Herpetologica 70, 211–227.
Perrigo, A., Hoorn, C., Antonelli, A., 2020. Why mountains matter for biodiversity. J. Biogeogr. 47, 315–325.
Portik, D.M., Smith, L.L., Bi, K., 2016. An evaluation of transcriptome-based exon capture for frog phylogenomics across multiple scales of divergence (Class: Amphibia, Order: Anura). Mol. Ecol. Resour. 16, 1069–1083.
Pritchard, J.K., Stephens, M., Donnelly, P., 2000. Inference of population structure using multilocus genotype data. Genetics 155, 945–959.
Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Sham, P.C., 2007. PLINK: A tool set for whole-genome association and population-based linkage analyses. The American Journal of Human Genetics 81, 559–575.
Pyron, R.A., Wiens, J.J., 2011. A large-scale phylogeny of Amphibia including over 2800 species, and a revised classification of extant frogs, salamanders, and caecilians. Mol. Phylogenet. Evol. 61, 543–583.
Rahbek, C., Borregaard, M.K., Antonelli, A., Colwell, R.K., Holt, B.G., Fjeldså, J., 2019a. Building mountain biodiversity: geological and evolutionary processes. Science 365, 1114–1119.
Rahbek, C., Borregaard, M.K., Colwell, R.K., Dalsgaard, B., Holt, B.G., Fjeldså, J., 2019b. Humboldt's enigma: What causes global patterns of mountain biodiversity? Science 365, 1108–1113.
Rancilhac, L., Irisarri, I., Angelinic, C., Arntzen, J.W., Babik, W., Vences, M., 2020. Phylotranscriptomic evidence for pervasive ancient hybridization among Old World salamanders. Mol. Phylogenet. Evol. in press.
Reichle, S., Köhler, J., 1997. A new species of *Eleutherodactylus* (Anura: Leptodactylidae) from the Andean slopes of Bolivia. Amphibia Reptilia 18, 333–337.
Rice, P., Longden, L., Bleasby, A., 2000. EMBOSS: The European Molecular Biology Open Software Suite. Trends Genet. 16, 276–277.
Roberts, J.L., Brown, J.L., May, R., Arizábal, W., Schulte, R., Summers, K., 2006. Genetic divergence and speciation in lowland and montane peruvian poison frogs. Mol. Phylogenet. Evol. 41, 149–164.
Robinson, D.F., Foulds, L.R., 1981. Comparison of phylogenetic trees. Math. Biosci. 53, 131–147.
Rohland, N., Reich, D., 2012. Cost-effective, high-throughput DNA sequencing libraries for multiplexed target capture. Genome Res. 22, 939–946.
Roux, C., Fraïsse, C., Romiguier, J., Anciaux, Y., Galtier, N., Bierne, N., 2016. Shedding light on the grey zone of speciation along a continuum of genomic divergence. PLoS Biol. 14 (10), 1371.
Sambrook, J., Fritsch, E.F., Maniatis, T., 1989. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory Press.
Scheele, B.C., Pasmans, F., Skerratt, L.F., Berger, L., Martel, A., Canessa, S., 2019. Amphibian fungal panzootic causes catastrophic and ongoing loss of biodiversity. Science 363, 1459–1463.
Slatkin, M., 1985. Gene flow in natural populations. Annu. Rev. Ecol. Syst. 16, 393–430.
Slatkin, M., 1987. Gene flow and the geographic structure of natural populations. Science 236, 787–792.
Smith, S.A., Montes, N., de Oca, A., Reeder, T.W., Wiens, J.J., 2007. A phylogenetic perspective on elevational species richness patterns in Middle American treefrogs: why so few species in low- land tropical rainforests? Evolution 61, 1188–1207.
Smith, S.A., Moore, M.J., Brown, J.W., Yang, Y., 2015. Analysis of phylogenomic datasets reveals conflict, concordance, and gene duplications with examples from animals and plants. BMC Evol. Biol. 15, 150.
Stamatakis, A., 2014. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30, 1312–1313.
Streicher, J.W., Miller, E.C., Guerrero, P.C., Correa, C., Ortiz, J.C., Wiens, J.J., 2018. Evaluating methods for phylogenomic analyses, and a new phylogeny for a major frog clade (Hyloidea) based on 2214 loci. Mol. Phylogenet. Evol. 119, 128–143.
Stuart, S., Hoffmann, M., Chanson, J.S., Cox, N.A., Berridge, R.J., ... Young, B.E. (2008). Threatened Amphibians of the World (Lynx Ediciones, Ed,). IUCN, Gland, Switzerland; and Conservation International, Arlington, Virginia, USA., Barcelona, Spain.
Swofford, D.L. (2002). PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4. Sinauer Associates.
Than, C., Ruths, D., Nakhleh, L., 2008. PhyloNet: A software package for analysing and reconstructing reticulate evolutionary relationships. BMC Bioinf. 9, 322.
Vachaspati, P., Warnow, T., 2015. ASTRID: Accurate species trees from internode distances. BMC Genomics 16, S3.
Vaira, M., Ferrari, L., 2008. A new species of *Oreobates* (Anura: Strabomantidae) from the Southern Andean Yungas of Argentina. Herpetozoa 1908, 41–50.
Vaz-Silva, W., Maciel, N.M., De Andrade, S.P., Amaro, R.C., 2018. A new cryptic species of *Oreobates* (Anura: Craugastoridae) from the seasonally dry tropical forest of central Brazil. Zootaxa 4441, 089–108.
Wiens, J.J., Parra-Olea, G., Garcia-Paris, M., Wake, D.B., 2007. Phylogenetic history underlies elevational patterns of biodiversity in tropical salamanders. Proceedings of the Royal Society B 274, 919–928.
Xu, H., Luo, X., Qian, J., Pang, X., Song, J., Chen, S., 2012. FastUniq: a fast de novo duplicates removal tool for paired short reads. PLoS ONE 7, e52249.
Yang, Z., 2007. PAML 4: Phylogenetic analysis by maximum likelihood. Mol. Biol. Evol. 24, 1586–1591.
Yang, Z., Nielsen, R., 2002. Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. Mol. Biol. Evol. 19, 908–917.
Yang, Z., Nielsen, R., Goldman, N., Pedersen, A.M.K., 2000. Codon-substitution models for heterogeneous selection pressure at amino acid sites. Genetics 155, 431–449.
Yu, Y., Nakhleh, L., 2015. A maximum pseudo-likelihood approach for phylogenetic networks. BMC Genomics 16, S10.