Molecular Phylogenetics and Evolution 188 (2023) 107907

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

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

Frog phylogeny: A time-calibrated, species-level tree based on hundreds of loci and 5,242 species

Daniel M. Portik [a,b,1], Jeffrey W. Streicher [c,1], John J. Wiens [a,*]

[a] Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, AZ 85721 USA
[b] California Academy of Sciences, San Francisco, CA 94118, USA
[c] Natural History Museum, London SW7 5BD, UK

ARTICLE INFO

Keywords:
Anurans
Divergence times
Missing data
Phylogenomics
Supermatrix

ABSTRACT

Large-scale, time-calibrated phylogenies from supermatrix studies have become crucial for evolutionary and ecological studies in many groups of organisms. However, in frogs (anuran amphibians), there is a serious problem with existing supermatrix estimates. Specifically, these trees are based on a limited number of loci (15 or fewer), and the higher-level relationships estimated are discordant with recent phylogenomic estimates based on much larger numbers of loci. Here, we attempted to rectify this problem by generating an expanded supermatrix and combining this with data from phylogenomic studies. To assist in aligning ribosomal sequences for this supermatrix, we developed a new program (TaxonomyAlign) to help perform taxonomy-guided alignments. The new combined matrix contained 5,242 anuran species with data from 307 markers, but with 95% missing data overall. This dataset represented a 71% increase in species sampled relative to the previous largest supermatrix analysis of anurans (adding 2,175 species). Maximum-likelihood analyses generated a tree in which higher-level relationships (and estimated clade ages) were generally concordant with those from phylogenomic analyses but were more discordant with the previous largest supermatrix analysis. We found few obvious problems arising from the extensive missing data in most species. We also generated a set of 100 time-calibrated trees for use in comparative analyses. Overall, we provide an improved estimate of anuran phylogeny based on the largest number of combined taxa and markers to date. More broadly, we demonstrate the potential to combine phylogenomic and supermatrix analyses in other groups of organisms.

1. Introduction

Large-scale, time-calibrated phylogenies with hundreds or thousands of species have become essential for many evolutionary and ecological studies. These trees are typically based on "supermatrix" analyses (i.e., datasets assembled primarily from sequence data on GenBank, usually with large numbers of taxa and limited numbers of genes). These large-scale trees have become crucial for comparative analyses in many groups, especially tetrapods (e.g., amphibians: Jetz and Pyron, 2018; birds; Jetz et al., 2012; mammals: Upham et al., 2019; squamate reptiles: Tonini et al., 2016; Zheng and Wiens, 2016). One such group is amphibians, especially frogs (anurans). There have been dozens of studies that have utilized large-scale phylogenies in frogs to address evolutionary and ecological questions, such as the origins of species richness patterns and the evolution of life-histories (e.g., Pyron and Wiens, 2013; De Lisle and Rowe, 2015; Moen and Wiens, 2017; Furness and Capellini, 2019; Furness et al., 2022; Liedtke et al., 2022).

At the same time, a serious problem has become apparent in these anuran supermatrix phylogenies. Specifically, these widely used, large-scale phylogenies include thousands of species, but are based on a relatively limited number of loci (e.g., Pyron and Wiens, 2011; Pyron, 2014; Jetz and Pyron, 2018). Importantly, the higher-level anuran phylogenies based on these datasets are often incongruent with those from more recent phylogenomic studies (i.e., those based on very large numbers of genes, usually from new data collection). The broad-scale anuran phylogenomic studies have included dozens (e.g., Feng et al., 2017), hundreds (Hime et al., 2021), or thousands of loci (e.g., Streicher et al., 2018; Portik et al., 2023). These phylogenomic analyses suggest that many aspects of the widely used supermatrix trees are not correct, particularly the higher-level relationships among families. For example,

* Corresponding author.
E-mail address: wiensj@arizona.edu (J.J. Wiens).
[1] These authors contributed equally.

https://doi.org/10.1016/j.ympev.2023.107907
Received 15 January 2023; Received in revised form 15 August 2023; Accepted 15 August 2023
Available online 25 August 2023

in previous supermatrix studies (e.g., Pyron and Wiens, 2011; Pyron, 2014; Jetz and Pyron, 2018) the hyloid clades Amazorana, Commutabirana, and Cornucopirana were not supported, although these clades were supported in analyses based on hundreds (Hime et al., 2021) and thousands of loci (Streicher et al., 2018; Portik et al., 2023). What is needed is a phylogeny that has the extensive species-level sampling of these supermatrix studies but that is consistent with the higher-level phylogenies based on large numbers of loci.

Here, we attempt to rectify this problem. We combine data from recent phylogenomic studies (with dozens or hundreds of nuclear loci) with a novel supermatrix to yield an overall matrix with 5,242 species. This represents the largest supermatrix analysis of frogs to date, in terms of numbers of taxa. We show that the higher-level relationships (e.g., among families) from this combined supermatrix tree largely resemble those from the phylogenomic studies, even though most species are included based on only a limited number of genes. We also demonstrate that individual species are generally placed in the families expected from previous taxonomy, and that estimated relationships do not appear to be misled by the extensive missing data in the majority of species.

2. Materials and methods

2.1. Sequence processing protocol

We used SuperCRUNCH v1.3.2 (Portik and Wiens, 2020) to process all sequence data. SuperCRUNCH is a bioinformatics toolkit for creating large phylogenetic datasets from GenBank data and/or local (i.e., newly generated) sequence data. The overall workflow involves: parsing starting sequences to create marker-specific fasta files, filtering and selecting sequences, performing alignment, and conducting various post-alignment tasks (i.e., relabeling, trimming, concatenation, and format conversion). Our complete analysis for the de novo anuran supermatrix is freely available on OSF: https://osf.io/3cbsh/ https://osf.io/3cbsh/?view_only. This resource includes all starting sequences, inputs and outputs from each step, final alignments, and complete instructions to replicate our results.

2.2. Taxonomic sampling and molecular data

There are currently ~ 7,600 described, extant anuran species distributed across 457 genera and 54 families (AmphibiaWeb, 2022), and our goal was to sample as many species as possible to create a new molecular supermatrix. We also included outgroup taxa. For outgroups we targeted most families from the two other amphibian orders: Gymnophiona (10/10 families, 27 species) and Caudata (10/10 families, 48 species). We selected representatives for these groups based on the species that contained the greatest amount of molecular data for the targeted markers (described below). These species were identified by initial surveys performed with SuperCRUNCH. For use as more distant outgroups, we also included representatives of Mammalia (*Lycaon pictus*, *Sarcophilus harrisii*), Crocodilia (*Alligator mississippiensis*), Aves (*Gallus gallus, Struthio camelus*), Testudines (*Podocnemis expansa*), Rhynchocephalia (*Sphenodon punctatus*), and Squamata (*Python molurus*, *Shinisaurus crocodilurus*). The final matrix included a total of 84 outgroup species.

We identified 26 molecular markers that have been widely used in amphibian phylogenetics (e.g., Wiens et al., 2005; Bossuyt et al., 2006; Frost et al., 2006; Roelants et al., 2007; Hedges et al., 2008; Van Bocxlaer et al., 2009; Pyron and Wiens, 2011; Hutter et al., 2013; Portik and Papenfuss, 2015; Hutter et al., 2017; Jetz and Pyron, 2018; Portik et al., 2019). These included 19 nuclear markers (*BDNF, BMP2, CMYC, CRYBA, CXCR4, H3A, MC1R, MYH, NCX1, NT3, POMC, RAG1, RAG2, RHO, SIA, SLC8A3, TNS3, TYR, ZFX*) and 7 mitochondrial markers (*12S, 16S, CO1, CYTB, ND1, ND2, ND4*). For both *RAG1* and *CO1,* multiple primer designs have led to distinct regions being sequenced. Full-length sequences for *RAG1* and *CO1* are generally ~ 3,000 bp and ~ 1,700 bp (respectively), but many studies have sequenced smaller, non-overlapping regions that fall within the first or second half of these full-length sequences. We therefore treated each of these markers as having two distinct regions, denoted as *RAG1_p1, RAG1_p2, CO1_p1,* and *CO1_p2*. We refer to this overall set of 28 markers as the "legacy" dataset.

We also targeted two sets of phylogenomic markers that were each available for smaller numbers of taxa. The first phylogenomic marker set included the 92 markers sequenced in Feng et al. (2017) and Tu et al. (2018). These markers are a subset of the PCR-based nuclear protein-coding loci (NPCL) described in Shen et al. (2013). The second phylogenomic marker set included 220 loci sequenced by Hime et al. (2021). These were conserved exon sequences obtained through anchored hybrid enrichment (AHE; Lemmon et al., 2012). We recognize that both the legacy dataset and AHE datasets contain nuclear protein-coding loci: these names (legacy, NPCL, AHE) are used simply as labels to indicate the origin of each dataset, rather than to provide a unique or complete description of their contents. The labeling of these markers revealed 34 gene names that were shared between the NPCL and AHE marker sets, but it was not clear if the same regions were sequenced for these genes (many genes are > 10 kb). We investigated potential sequence overlaps by attempting to align sequences from both datasets for each of these genes. Alignments for 26 of the 34 genes displayed partial to complete sequence overlap, indicating the NPCL and AHE sequences could be treated as the same marker for these respective genes (Supplementary Table S1). For this study, we considered the 24 shared markers as belonging to the NPCL set. Overall, we included 28 legacy markers and 286 phylogenomic markers, including 194 AHE markers and 92 NPCL markers (Supplementary Table S1).

Sequences for the legacy and NPCL markers were available on GenBank. To obtain all relevant GenBank sequences for Anura, we identified key taxonomic terms in the NCBI Taxonomy database that covered the phylogeny of Anura (Ascaphidae, Leiopelmatidae, Alytidae, Pelobatoidea, Bombinatoridae, Neobatrachia, Pipoidea). We performed individual searches for these terms on July 2, 2022. We then downloaded the sequence sets for each respective group, and concatenated the resulting fasta files into a single fasta file. We separately obtained all available GenBank sequences for caecilians and salamanders by searching for "Gymnophiona" and "Caudata", respectively. Finally, for the remaining outgroups we performed searches using individual species names and then concatenated the resulting fasta files. The sequence set for Anura contained a total of 730,873 sequences (4.6 Gb in size), Gymnophiona contained 2,381 sequences (3 Mb in size), and Caudata contained 7,100 sequences (1.5 Gb in size). The AHE sequences were obtained by downloading alignment files from the supplemental materials of Hime et al. (2021).

Hypothetically, we could also have included the datasets of Streicher et al. (2018) or Portik et al. (2023), which included > 2,000 nuclear loci. However, we found that combining this number of loci with a large-scale supermatrix yielded a matrix so large that it was not computationally tractable to analyze (see Portik et al., 2023).

2.3. Data pre-processing

SuperCRUNCH requires a list of taxon names to identify and select sequences for desired species. Sequence records without a matched taxon name are filtered out during analyses. Differences between the user-provided taxonomy and the taxonomy originally associated with a set of sequences can cause useful sequences to be inadvertently excluded. SuperCRUNCH provides straightforward steps for identifying such taxonomic conflicts. These conflicts can then be resolved by relabeling sequences with valid names.

To prepare the anuran sequences for the main analysis, we used SuperCRUNCH to examine the compatibility of the AmphibiaWeb (2022) taxonomy with the taxon names of the GenBank sequences. We downloaded the AmphibiaWeb taxonomy on July 2, 2022 and created a taxon list containing all recognized anuran species (n = 7,579). This taxon list was used to screen the sequences with the *Taxon_Assessment.py*

module in SuperCRUNCH, which detected 653,712 sequences with valid names and 77,161 sequences with invalid names. We inspected the unmatched names list to identify invalid names resulting from spelling errors or changes in taxonomy, that could easily be assigned to valid names. We identified 1,492 names which resulted from these error types and assigned each an updated valid name from the AmphibiaWeb taxonomy. We used *Rename_Merge.py* to relabel invalid sequences with the updated valid names, which relabeled 31,182 sequences. During this process, sequences that were successfully relabeled were merged with the sequences already containing valid names, and the remaining sequences with uncorrectable names were discarded. After the taxonomic analysis was completed, the starting sequence set for Anura contained 684,894 sequences that were fully compatible with the AmphibiaWeb taxonomy.

The AHE sequences were available as alignment files. To use sequences with SuperCRUNCH, they must be in fasta format with description lines that contain a unique identifier, taxon name, and locus abbreviation/description (similar to NCBI GenBank format).

To make the AHE data compatible with SuperCRUNCH, we stripped the alignments (i.e., removed all gaps), updated the taxon labels to be compatible with the taxonomy of AmphibiaWeb (2022), assigned each sequence a unique identifier (i.e., "accession" number), and added gene-description terms. The relabeled, SuperCRUNCH-compatible AHE sequences have been made available as a fasta file (https://osf.io/3cbsh/ https://osf.io/3cbsh/?view_only).

SuperCRUNCH requires a single input fasta file for the main analysis. We concatenated the fasta files for the anuran sequences (corrected for taxonomy), caecilians, salamanders, and the remaining outgroups. We also added the subset of the AHE sequences that matched the 26 shared NPCL markers to the larger combined sequence set. The final combined fasta file contained 988,872 total sequences and was 6.8 Gb in size, and is available from: https://osf.io/3cbsh/ https://osf.io/3cbsh/?view_only.

2.4. SuperCRUNCH analysis

In addition to the fasta file containing the set of starting sequences, SuperCRUNCH requires a list of taxa and marker search terms to assemble marker-specific fasta files. For the final taxon list, we added the relevant outgroup names (27 caecilian species, 48 salamanders, and 9 additional outgroup taxa) to the taxon list of 7,579 anuran species obtained from AmphibiaWeb (2022). We created two locus-search term files. One file contained search terms exclusive to the AHE loci and was designed to work with the relabeled AHE sequence set. The other search-term file was designed for the main analysis of GenBank sequences (and the overlapping subset of AHE sequences), and contained information for the legacy markers and NPCL markers.

We used the GenBank marker search terms with the *Parse_Loci.py* module to create marker-specific fasta files. We then performed sequence-similarity filtering steps to ensure that the identity of the sequences was consistent with their marker labels. We used *Cluster_Blast_Extract.py* to automate similarity-filtering for markers composed of "simple" records (those generally containing a single gene region with limited length variation), which included 109 markers. For each marker, this method: (i) creates sequence clusters using CD-HIT-EST (Li and Godzik, 2006) at the 80% sequence identity threshold, (ii) performs BLASTn searches (dc-megablast) for all sequences using the largest sequence cluster as the reference database, and (iii) outputs sequences that are trimmed based on their merged BLAST hit coordinates (Portik and Wiens, 2020). However, this method is expected to fail for "complex" records, which contain the target region plus non-target sequence. This category often includes mtDNA markers (which are sequenced for short gene fragments, full genes, or complete mitogenomes) and genes sequenced in different regions that have little or no overlap. In our dataset, markers containing "complex" records included all mtDNA markers (*12S, 16S, CO1_p1, CO1_p2, CYTB, ND1, ND2, ND4*), *RAG1_p1, RAG1_p2,* and the NPCL marker *SACS*. For these 11 markers, we used *Reference_Blast_Extract.py* for similarity-filtering. This method is highly similar to *Cluster_Blast_Extract.py*, but differs by requiring a user-supplied set of reference sequences to use as the BLAST reference database (as opposed to a cluster generated from the input sequences). This strategy offers finer control over which target regions are included. For example, it can be used to extract a single mtDNA gene region from a record set containing a mix of whole mitochondrial genomes, long multi-gene mtDNA sequences, and shorter target sequences (Portik and Wiens, 2020). For reference sequences, we used the anuran mtDNA and vertebate *RAG1* reference sets included with SuperCRUNCH. For *SACS,* we used the unaligned sequences from Feng et al. (2017) as references. Both of the similarity-filtering methods were run using the discontiguous-megablast algorithm option (-b dc-megablast) and the defaults for merging BLAST coordinates (-m span –bp_bridge 100). As a final quality check for the mtDNA sequences, we screened the sequences against human mtDNA sequences using the *Contamination_Filter.py* module, which removed one 12S sequence (GU191110.1) and two 16S sequences (HM921175.1, HM921174.1) that were unambiguously identified as human (but labeled as anurans on GenBank).

After similarity filtering, we selected a representative sequence for each species for each of the 120 markers using *Filter_Seqs_and_Species.py*. We selected one sequence per species per marker (-s oneseq) based on the longest available sequence (-f length), with a minimum sequence-length requirement of 300 bp (-m 300). Additionally, we used the –accessions_exclude option to exclude 384 specific accessions from being selected. This list of excluded accessions was based on the identification of several problematic sequences during preliminary analyses (e.g., incorrect taxonomic identification; see Section 2.8 below for a description of these analyses). The sequence-selection step eliminated the legacy marker *CRYBA*, for which all sequences were less than 300 bp (after similarity-filtering). We note that given the number of sequences per species per marker, the total number of possible supermatrix combinations for this dataset is $1.24 \times 10^{12,465}$ (assessed using the *Infer_Supermatrix_Combinations.py* module). After the selection of representative sequences, we screened the remaining 119 markers to ensure that they contained data for at least 75 species using *Fasta_Filter_by_Min_Seqs.py*. We found 113 markers that passed this filter, and six markers that did not and were removed (*B3GALT1, FUT9, MC1R, MYH, ZBED4, ZFX*). The cutoff at 75 species is somewhat arbitrary, but we found that three markers had 4–5 species, three others had 55–60 species, whereas all other markers had 84 species or more (mean = 395 species). To prepare for multiple sequence alignment, we used *Adjust_Direction.py* to ensure all sequences were in the correct orientation.

To perform sequence alignment for protein-coding genes, we used the MACSE v2 translation aligner (Ranwez et al., 2018). MACSE can align coding sequences with respect to their translation, while accounting for multiple frameshifts or stop codons. MACSE can be run using a set of reliable sequences (e.g., those without translation errors), but it also has a feature allowing simultaneous alignment of a set of reliable sequences and a set of unreliable sequences (e.g., those with translation errors). The latter feature is particularly useful for GenBank sequences, which vary in quality. To prepare sequences for alignment with MACSE, we used *Coding_Translation_Tests.py* to identify the correct reading frame of sequences, adjust them to the first codon position, and ensure completion of the final codon. We performed this step for the 105 nuclear markers (using standard code for translation), and for 6 of the 8 mtDNA markers (using the vertebrate mitochondrial code; see below regarding alignment for the remaining two mitochondrial markers, *12S* and *16S rRNA*). For each marker, the outputs include a file of adjusted sequences that passed translation, and a file of sequences that failed translation (if any). The single or paired outputs were then used to run *Align.py* with the MACSE alignment option, which auto-detects whether a single set of reliable sequences is used or if the simultaneous alignment should be performed. This strategy worked well for all markers except *TTN*, which contained a mix of many long and several shorter sequence

fragments. We used the MAFFT (Katoh et al., 2002; Katoh and Standley, 2013) option in *Align.py* to align *TTN* instead, using the automatic alignment algorithm selection option.

The alignment of the *12S* and *16S rRNA* markers was more challenging, given that they are not protein-coding and may have insertions and deletions that modify their lengths. Therefore, we developed here a new program to help perform taxonomy-guided alignments called TaxonomyAlign. This method relies on a mapping file that defines taxonomic groups for all sequences present (e.g., genera, families). TaxonomyAlign begins by dividing the starting sequences into subsets based on the defined groups, and builds sub-alignments for each group using MAFFT (with multiple alignment algorithm options available). Upon completion of sub-alignments, the –merge option in MAFFT is used to merge the set of sub-alignments while preserving each sub-alignment. The merging step is followed by an optional polishing step, which uses iterative refinement to produce a final refined alignment. In cases where a grouping is represented by a single species, that sequence is aligned during the merging step. TaxonomyAlign is open-source and freely available on github: https://github.com/dportik/TaxonomyAlign. We used TaxonomyAlign to align *12S* and *16S* sequences based on the membership of species in genera and families. To benchmark the performance of TaxonomyAlign, we also used MAFFT to generate parallel alignments for these markers. The alignments from MAFFT and TaxonomyAlign were used to estimate gene trees, which allowed a comparison of the number of monophyletic groupings occurring across taxonomic ranks (described in Section 2.6 below).

Given that the AHE sequences were already validated by Hime et al. (2021), we only performed a subset of SuperCRUNCH steps for these markers. Our goal was to create alignments using MACSE, to be consistent with our approach for the legacy and NPCL markers. We used the AHE marker search terms to create marker-specific fasta files from the AHE sequence set, but skipped similarity-filtering. We performed sequence selection as described above, adjusted sequence directions, adjusted reading frames, and used MACSE to perform alignments.

We then used SuperCRUNCH to perform the final steps to create the full concatenated dataset. After multiple sequence alignment, we relabeled sequences to include only species names (vs. full description lines) using *Fasta_Relabel_Seqs.py*. Finally, we concatenated the final alignments using *Concatenation.py*. We produced concatenated alignments for the AHE markers, NPCL markers, legacy markers, and a combination of the AHE, NPCL, and legacy markers. We used the concatenated alignments of each marker set to compare various metrics (number of taxa, sequences, informative sites, etc.). The combined marker set was used in the phylogenetic analyses.

We used the *Alignment_Assessment* tool from Portik et al. (2016) to generate summary statistics for all individual and concatenated alignments. These statistics included alignment length, number of informative sites (defined here as sites containing at least two different nucleotides that are each present in at least two sequences, synonymous with parsimony-informative sites), and percent missing data (relative frequency of cells with missing data in the alignment or concatenated matrix). We give alignment lengths in base pairs (bp) but note that these lengths can also include inferred insertions/gap positions.

Finally, for each marker we evaluated sequence-length heterogeneity using *Sequence_Length_Heterogeneity.py*. This method calculates the coefficient of variation (CV; the ratio of the standard deviation to the mean) from the set of aligned sequence lengths for each marker (ASL-CV; Portik and Wiens, 2021), and indicates how much variation in sequence length is present.

2.5. Partitioning

Given the large number of markers, the full matrix contained many potential partitions (i.e., sets of characters with distinct rates and other properties). In preliminary analyses with reduced numbers of taxa, we used PartitionFinder2 (Lanfear et al., 2016) to identify the best-fitting set of partitions. We performed two analyses. For the first, the input partitions consisted of separate partitions for each codon position in each gene (with a single partition for non-protein coding genes, like *12S* and *16S*). For the second, there was a single partition per gene. The best-fit partitioning schemes for these two analyses had 916 and 225 partitions, respectively. However, running these partitioned datasets was computationally intractable on the Cyberinfrastructure for Phylogenetic Research (CIPRES) Science Gateway (Miller et al., 2010) that was used for our likelihood tree searches. Indeed, the analyses failed in the first few minutes because of memory limitations. We therefore used a partitioning scheme with only seven partitions: a partition for each of the three codon positions in the nuclear protein-coding genes, a partition for each of the three codon positions in the mitochondrial protein-coding genes, and a single partition for mitochondrial ribosomal RNA (including *12S* and *16S*). This partitioning scheme is based on the idea that the most important partitions will be different codon positions within protein-coding genes, and between nuclear and mitochondrial protein-coding genes. This limited number of partitions proved to be computationally tractable.

2.6. Phylogenetic analyses

We performed concatenated maximum-likelihood analyses using RAxML v8.2 (Stamatakis, 2014) on the Cyberinfrastructure for Phylogenetic Research (CIPRES) Science Gateway (Miller et al., 2010). We chose RAxML for all ML analyses because RaxML was found to yield better results than IQ-TREE (Nguyen et al., 2015) for phylogenomic datasets containing $>$ 200 taxa (Zhou et al., 2018). We conducted 25 alternate runs on distinct starting trees to find the best-scoring ML tree, using the GTR + CAT model (general-time-reversible model with the CAT approximation of the gamma distribution of rates among sites). To complete the analysis within CIPRES run-time allotments, we performed five separate analyses involving 5 alternate runs on distinct starting trees, with a final branch-length optimization on the best tree found in each analysis.

We performed partitioned rapid bootstrapping analyses to obtain 100 bootstrap trees with branch lengths (option -k) using RAxML v8.2. This type of bootstrapping analysis requires a parameter optimization under GAMMA at the end of each replicate, which increases run times. However, the estimated branch lengths were necessary to generate a set of 100 time-calibrated trees for use in comparative analyses (see below). To complete the analysis within CIPRES run-time allotments, we performed four separate analyses with 25 bootstrap replicates each (and specifying a unique seed number for each analysis with the -b option). The 100 bootstrap trees allowed us to assess branch support and to generate confidence intervals for our divergence time estimates (see below).

We acknowledge that our search for the best tree was somewhat limited, with only 25 replicate searches. Nevertheless, given the very large number of taxa and markers, these searches took almost 2 months of computational time to complete (see Results). It is possible that some aspects of the topology of the optimal likelihood tree reported here reflect the failure to find the tree with the maximum likelihood for these data. However, it seems very unlikely that the same random tree-search artefact would be repeated dozens of times across the multiple searches of the 100 bootstrap replicates and thus appear as a strongly supported branch. Therefore, weakly supported branches of the tree may also reflect failure to find the optimal tree and should be treated with the usual caution. Thus, it would be problematic to dismiss our overall results simply because the best-fit tree found might not be truly optimal.

To evaluate the match of the inferred phylogenies with taxonomy, we used the program MonoPhylo (Portik and Wiens, 2021). MonoPhylo assesses the status (monophyletic, paraphyletic, polyphyletic) of user-defined groupings (genus, subfamily, family, etc.) for a given phylogenetic tree. We assessed the status of the 439 anuran genera and 54 families that were included in the tree. These taxa were defined based on

the AmphibiaWeb taxonomy (rechecked November 2022). We used MonoPhylo to summarize whether these taxa were recovered as monophyletic or not, and to obtain support values for monophyletic groups.

The alignments from TaxonomyAlign and MAFFT alone for *12S* and *16S* were used to estimate gene trees for each of these markers. For each alignment, we conducted 10 runs on distinct starting trees to find the best-scoring ML tree, using the GTR + CAT model (general-time-reversible model with the CAT approximation of the gamma distribution of rates among sites). We used MonoPhylo to quantify the number of monophyletic genera, subfamilies, and families for the tree from each marker and alignment method. We also compared the alignment length, number of sites with gaps, and the number of informative sites for the same marker from each alignment method.

2.7. Data checking

After conducting our initial concatenated analyses, we found some unusual phylogenetic results in some parts of tree, especially within Bufonidae (e.g., congeneric species placed in different parts of the tree). We suspected that these might be related to misidentified specimens or to mislabeled sequences on GenBank. Therefore, we performed a set of data-checking analyses. Specifically, we analyzed each marker separately using maximum likelihood, as described above. We looked for mismatches between these gene trees and taxonomy that were not present across genes (i.e., a species from genus A placed inside genus B for gene 1, when the species is placed in genus A for gene 2). We also looked for cases in which pairs of species had nearly identical sequences for a given gene but were distinct in other genes. The accession numbers of potentially problematic sequences were added to the list of accessions to exclude during sequence selection. We then re-ran *Filter_Seqs_and_Species.py* using the –accessions_exclude feature with this updated list. This had the effect of replacing the potentially problematic sequences with different sequences from GenBank for the same marker and species. If a suitable replacement was not found (e.g., there was only a single sequence for that species for that marker), the sequence was simply deleted for that species and marker. All subsequent steps in SuperCRUNCH were run as described above to obtain alignments for the next round of gene tree analyses. We repeated this overall process twice, until all clearly problematic sequences were eliminated. We then re-concatenated the data and performed the final analyses.

2.8. Divergence-time estimation

We conducted divergence-time estimation using penalized likelihood (Sanderson, 2001) implemented with treePL (Smith and O'Meara, 2012). This method is widely used for dating phylogenies that have very large numbers of species. We used 33 calibration points based on fossil evidence. We started from a review of fossil calibration points used in other recent amphibian studies (Feng et al., 2017; Jetz and Pyron, 2018; Hime et al., 2021). The details of the final set of calibration points used and their selection are given in Supplementary File S1.

Using the best maximum-likelihood tree obtained from RAxML, we conducted a thorough analysis in treePL during the two-round optimization phase. A "thorough" analysis iterates until convergence of log likelihoods is observed. We tested the following standard smoothing values in treePL: 0.1, 1, 10, 100, 1,000, and 10,000. The best smoothing value was found to be 1,000 by a cross-validation analysis, based on having the lowest chi-square value in the "cvoutfile" file. We used this analysis to time-calibrate branches in our primary analysis. We then inferred branch lengths for 100 bootstrap trees in RAxML and then ran these trees iteratively through treePL to obtain a distribution of 100 time-calibrated trees. For this latter analysis, we assumed that the optimal smoothing parameter for the original data also applied to the bootstrapped trees. Hypothetically, we could have determined the optimal smoothing parameter separately for each replicate, but the cross-validation analysis was extremely time intensive. Furthermore, the optimal smoothing parameter for the observed data should be optimal for datasets based on resampling the observed data.

We used the 100 time-calibrated bootstrap trees to generate confidence intervals for the divergence times of all nodes in the best tree. Divergence-time confidence intervals (95%) were generated using TreeAnnotator 1.10.4 (Drummond and Rambaut, 2007). We visualized variation in trees and divergence times among families across these 100 trees using the R package *phangorn* 2.2 (Schliep, 2011), with the "densiTree" function.

3. Results

3.1. Supermatrix parameters

The supermatrix created with SuperCRUNCH contained 5,242 anuran species representing 439 genera and 54 families. It also included 84 outgroup species, including 27 caecilians and 48 salamanders. The matrix included 5,326 species overall. The AmphibiaWeb taxonomy used for this analysis (July 2022) contained 7,579 described anuran species, 457 genera, and 54 families. Therefore, the matrix included 69% of described anuran species, 96% of the genera, and 100% of the families.

There were 307 molecular markers (24/28 legacy markers, 194/194 AHE, and 89/92 NPCL markers) and 90,196 total sequences included. The concatenated alignment was 381,201 bp in length with 95% missing data overall (Table 1; Supplementary Table S1). Characteristics of the concatenated alignments for each marker category (i.e., legacy, NPCL, AHE) are shown in Table 1, and characteristics for all markers are provided in Supplementary Table S1. A table of GenBank accession numbers including all sequences used is provided as Supplementary Table S2.

In the legacy dataset, different genes had data for different numbers of species (Fig. 1; Supplementary Table S1). The markers with the greatest numbers of species were *16S* ($n = 4{,}926$) and *12S* ($n = 3{,}743$), followed by *RAG1_p2* ($n = 2{,}363$), *CO1_p1* ($n = 2{,}227$), *CYTB* ($n = 2{,}118$), *TYR* ($n = 1{,}659$), *ND1* ($n = 1{,}328$), *RHO* ($n = 1{,}257$), and *POMC* ($n = 1{,}241$).

The basic properties of these three categories of markers (legacy, NPCL, AHE) are illustrated in Fig. 2. All three were similar in length, percentage of informative sites, and percentage of missing data. However, the legacy markers tended to be more variable for all three properties, and the NPCL and AHE datasets tended to have less missing data.

The total run time for the SuperCRUNCH analysis was ~ 88 h, of which 17 h were used for similarity-filtering and 68 h were used for alignment (65 with MACSE and 3 for TaxonomyAlign; Supplementary Table S3). However, this estimate assumes that the similarity-filtering and alignment steps were run for all files sequentially. In practice, subsets of files can be run in parallel at these steps to reduce run time. Using this strategy, the complete analysis to construct the supermatrix required just under 24 h of run time.

We also compared the alignments of *12S* and *16S* from TaxonomyAlign (which uses MAFFT) and from MAFFT alone. TaxonomyAlign produced a slightly longer alignment (3%) than MAFFT for *12S* (2,234 bp vs. 2,177 bp) but a shorter alignment (5%) for *16S* (3,775 bp vs. 3,972 bp). For both markers and methods, all sites had one or more gaps. TaxonomyAlign yielded a lower percentage of informative sites relative to MAFFT, for both *12S* (57% vs. 64%) and *16S* (59% vs. 63%). Nevertheless, maximum likelihood analyses of the alignments of *12S* and *16S* from TaxonomyAlign consistently supported more monophyletic genera, subfamilies, and families than those from MAFFT. For *12S*, TaxonomyAlign supported the monophyly of 225 genera, 25 subfamilies, and 38 families, versus 214, 24, and 30 (respectively) for MAFFT. For *16S*, TaxonomyAlign supported the monophyly of 245 genera, 28 subfamilies, and 45 families, whereas MAFFT supported 228, 26, and 39, respectively. Importantly, the largest proportional

Table 1
Information for concatenated alignments of different marker categories, including the final supermatrix (which contains all three categories). For all marker categories, the number of species includes outgroups: the overall matrix includes 84 outgroup taxa and 5,242 ingroup species. Alignment length is the overall length of the concatenated alignment of all markers in that category, given in number of base pairs (bp).

Marker category	Species	Number of markers	Number of sequences	Alignment length (bp)	Informative sites (%)	Missing data (%)
AHE, NPCL, Legacy	5,326	307	90,196	381,201	55.4	95.2
AHE	230	194	43,282	250,299	56.0	5.6
NPCL	659	89	18,232	102,045	53.9	70.1
Legacy	5,324	24	28,682	28,857	56.3	85.6
Legacy (mitochondrial)	5,270	8	16,628	13,329	62.9	79.9
Legacy (nuclear)	3,554	16	12,054	15,528	50.6	85.6

AHE markers are nuclear exon sequences obtained from anchored hybrid enrichment. NPCL refers to the set of nuclear protein coding loci from Feng et al. (2017). Legacy markers are a set of mitochondrial and nuclear genes that are widely used in anuran phylogenetics, including large-scale supermatrix studies (e.g., Jetz and Pyron, 2018). Informative sites are defined here as sites containing at least two different nucleotides that are each present in at least two sequences, equivalent to parsimony-informative sites.

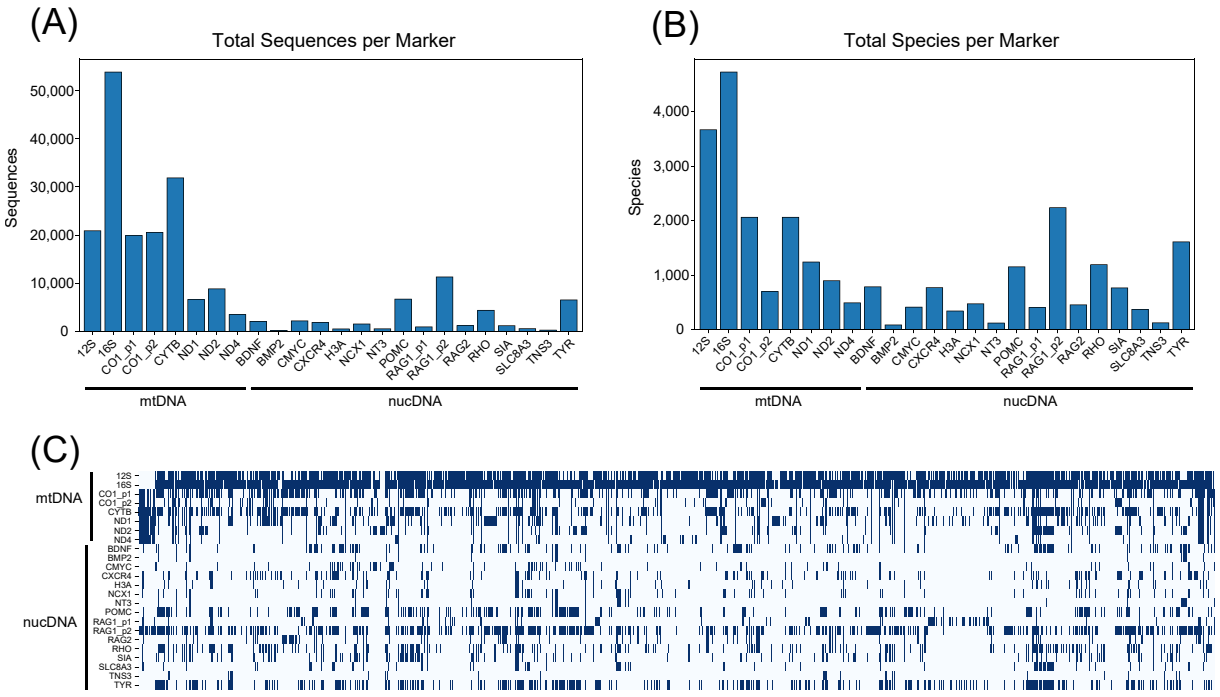

Fig. 1. (A) Total number of amphibian sequences available for 24 legacy markers according to a GenBank search in July 2022. (B) Total number of species for each locus according to a GenBank search in July 2022. (C) Distribution of legacy markers across the 5,242 amphibian species included in this study.

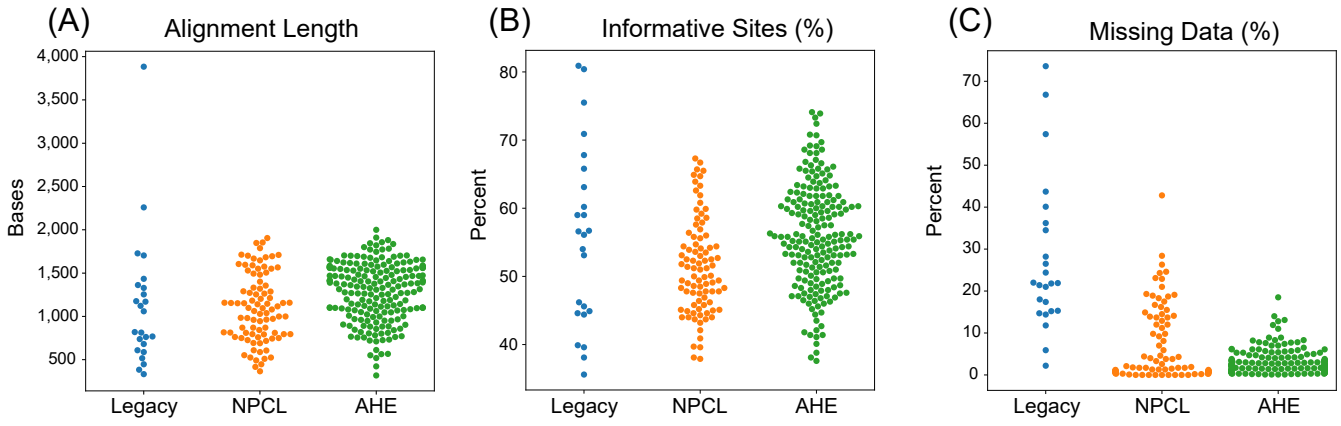

Fig. 2. Qualities of amphibian DNA sequence alignments used in this study from 24 GenBank markers (Legacy; mixed sources), 89 nuclear protein coding loci (NPCL; Feng et al., 2017) and 194 anchored hybrid enrichment markers (AHE; Hime et al., 2021). (A) Alignment length (in number of bases) for legacy, NPCL, and AHE markers. (B) Percent informative sites (%) for legacy, NPCL, and AHE markers. (C) Percent missing data (%) for legacy, NPCL, and AHE markers.

differences are among families, where TaxonomyAlign yields a 27% (*12S*) to 15% (*16S*) increase in the number of families supported as monophyletic. Furthermore, the monophyly of most anuran families is independently supported by phylogenomic analyses (e.g., Feng et al., 2017; Streicher et al., 2018; Hime et al., 2021; Portik et al., 2023), suggesting that the trees from TaxonomyAlign might better reflect the true relationships.

3.2. Phylogenetic analyses and divergence dating

The final ln-likelihood of the best RAxML tree was −18,637,415. The best tree search analyses (using 25 alternate runs on distinct starting trees) required a minimum of 56 Gb of memory, used 64 threads, and ran for a combined total of 1,323 h (55 days) on CIPRES. The bootstrapping analysis (including estimation of branch lengths for each of the 100 replicates) also used 64 threads and ran for a combined total of 491 h (20 days), taking roughly 5 h per bootstrap replicate.

We examined whether genera, subfamilies, and families of anurans were recovered as monophyletic using MonoPhylo (Table 2; Supplementary Table S4). Tests were performed only when a defined group was represented by two or more species. Groups represented by a single species are the result of a monotypic taxon or limited sampling, which we do not distinguish here. At the genus level, we found 259 of 345 testable genera were recovered as monophyletic (75%). We found that 34 of 43 (79%) testable subfamilies were monophyletic. Among families, 51 of 52 (98.0%) testable families were monophyletic. The single non-monophyletic family was Strabomantidae, the content of which is currently controversial (Heinicke et al., 2018; Barrientos et al., 2021). We address non-monophyletic taxa within families in Section 3.4 below.

Higher-level relationships from the likelihood analysis are summarized in Fig. 3, and the time-calibrated higher-level tree is in Fig. 4. An overall summary of the time-calibrated species-level tree is provided in Fig. 5, and detailed subtrees are given in Figs. 6–60. The full maximum-likelihood tree with branch lengths and support values is available as Supplementary File S2. The time-calibrated version of this tree is available as Supplementary File S3. The 100 time-calibrated bootstrap trees are available as Supplementary File S4. Confidence intervals on the estimated ages for each node are given in Supplementary File S5. Variation in topologies and divergence times among these 100 replicates are illustrated in Supplementary File S6, based on the "densiTree" function in *phangorn*, with each thin purple line showing an individual topology and thicker and darker lines showing the overlap of multiple trees.

3.3. Higher-level relationships

Here we compare the family-level relationships found in our supermatrix tree (Fig. 3) to those from other recent, large-scale, molecular analyses. Specifically, we compare our results to recent phylogenomic analyses (e.g., Feng et al., 2017; Streicher et al., 2018; Hime et al., 2021) and the supermatrix study by Jetz and Pyron (2018). We also compare our trees to those of Portik et al. (2023), including those from their phylogenomic analyses (UCE; ultraconserved element; 3,784 markers, most frog families) and gigamatrix analyses (UCE + legacy + AHE + NPCL; 4,091 markers, most frog genera). Our goal here is not to review all recent analyses of anuran phylogeny and taxonomy, but rather to highlight similarities and differences with these key studies. We also mention support values, but note that these were not provided directly by Feng et al. (2017) or Jetz and Pyron (2018).

Table 2
Assessment of monophyletic groups across ranks for anurans. Testable groups include those represented by two or more sampled species.

Rank	Total groups	Testable groups	Number monophyletic	Percent monophyletic
Genus	439	345	259	75%
Subfamily	46	43	34	79%
Family	54	52	51	98%

We first provide a quantitative summary of the congruence among studies, followed by a more narrative summary. We compared the relationships among families across these studies (Supplementary Table S5), given that many of the studies compared here were not comprehensive in terms of sampling of genera. Furthermore, some studies lacked data for one or more families. The higher-level (among-family) relationships found here were most similar to those of the phylogenomic study of Hime et al. (2021): among 49 comparable nodes, 94% were congruent. Conversely, these relationships were least congruent with those of Jetz and Pyron (2018), with 65% of 52 comparable nodes congruent. Our tree was also highly congruent with the phylogenomic tree of Feng et al. (2017; their Fig. 1) with 93% of 43 nodes congruent. However, Feng et al. (2017) also presented a more complete tree, with more families added with more limited data. This tree was relatively incongruent (65% of 52 nodes congruent), but due almost entirely to the unusual placement of these added families (e.g., Cycloramphidae, Micrixalidae). Finally, the tree was relatively congruent with the phylogenomic (UCE) tree of Portik et al. (2023), with 83% of 43 comparable nodes congruent, and with the gigamatrix tree of that study (87% of 52 nodes congruent).

Our results agree with these other recent studies regarding the major clades of frogs (Fig. 3). Specifically, these studies agree with ours on the relationships among Leiopelmatoidea (Ascaphidae, Leiopelmatidae), Discoglossoidea (Alytidae, Bombinatoridae), Pipoidea (Rhinophrynidae, Pipidae), Pelobatoidea (Scaphiopodidae, (Pelodytidae, (Pelobatidae, Megophryidae))), and Neobatrachia. These relationships were strongly supported in our tree (Fig. 3).

Most frog families and species belong to Neobatrachia. Among the major clades of Neobatrachia, our results agree with these other recent studies in the following points (Fig. 3). First, we show Heleophrynidae as the sister taxon to all other neobatrachians. Second, we show the clade Sooglossidae + Nasikabatrachidae as the sister taxon of Ranoidea. Third, we show the clade Myobatrachidae + Calyptocephalellidae as the sister group of Hyloidea. All of these relationships were strongly supported in our tree (Fig. 3).

Within Hyloidea, our tree is largely congruent with recent phylogenomic studies (Feng et al., 2017; Streicher et al., 2018; Hime et al., 2021; Portik et al., 2023), but more discordant with the supermatrix tree of Jetz and Pyron (2018). These phylogenomic analyses agree with our tree in placing Rhinodermatidae as the sister taxon to all other Hyloidea, Neoaustrarana (Alsodidae, Batrachylidae, Cycloramphidae, Hylodidae) as the sister group to the remaining Hyloidea (Cornucopirana), and placing Telmatobiidae as the sister taxon to other members of Cornucopirana. The remaining hyloids belong to Amazorana (Hylidae (Ceratophryidae, Hemiphractidae)) and Commutabirana (Terraranae (Dendrobatidae ((Bufonidae, Odontophrynidae), (Leptodactylidae (Allophrynidae, Centrolenidae)))). The support here for many of these clades in Hyloidea is limited (Cornucopirana = 77%; Neoaustrarana = 81%; Amazorana = 87%; Commutabirana = 73%), but these clades are strongly supported by Streicher et al. (2018), Hime et al. (2021), and the phylogenomic tree of Portik et al. (2023). Note that the phylogenomic analyses of Feng et al. (2017) did not include all families, and when they included Cycloramphidae based on more limited data, it was placed in Commutabirana and not Neoaustrarana.

By contrast, the tree of Jetz and Pyron (2018) was broadly discordant with our tree and others within Hyloidea. For example, Jetz and Pyron (2018) placed Dendrobatidae as the sister group to all other Hyloidea, and Hemiphractidae + Terranae as the sister group to all remaining hyloids. They inferred a clade that consisted of Ceratophryidae, Rhinodermatidae, Telmatobiidae, Alsodidae, Batrachylidae, Cycloramphidae, and Hylodidae. Hylidae was the sister taxon to a clade that contained Allophrynidae, Centrolenidae, Bufonidae, Leptodactylidae, and Odontophrynidae. In short, the tree of Jetz and Pyron (2018)

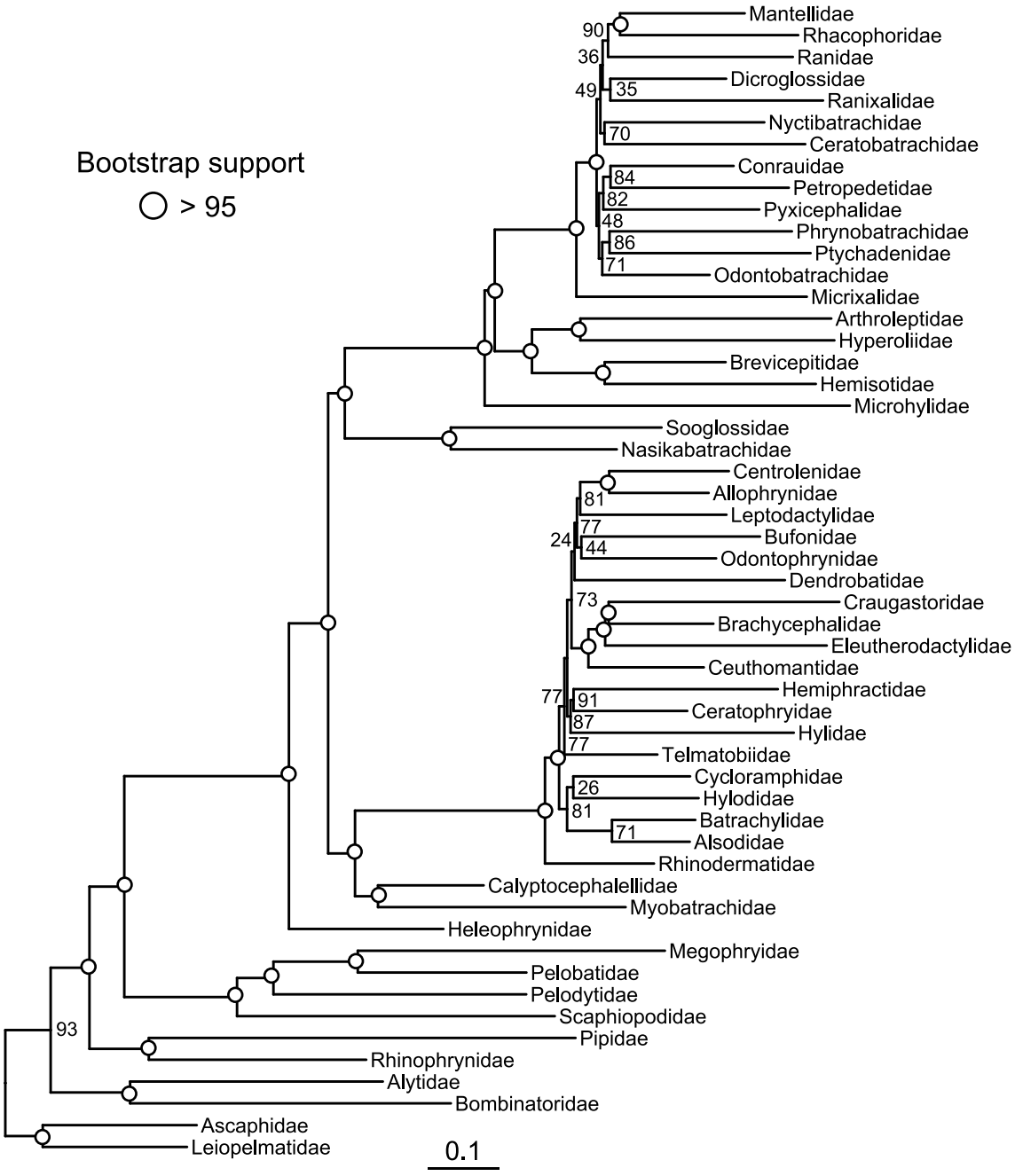

Fig. 3. Maximum-likelihood phylogeny reduced to one tip per anuran family. Bootstrap support values are indicated when less than 96%. Family-level taxonomy follows AmphibiaWeb (2022) except for Strabomantidae, which was not monophyletic in our analyses and was included within Craugastoridae.

conflicts with our tree and those of recent phylogenomic analyses in rejecting the major clades of Hyloidea, including Cornucopirana, Amazorana, and Commutabirana. There were particularly striking disagreements regarding the placements of Ceratophryidae, Dendrobatidae, Hemiphractidae, Rhinodermatidae, and Telmatobiidae. However, there were some points of agreement, including monophyly of Neoaustrarana, Terraranae, and the clade consisting of Allophrynidae, Centrolenidae, Bufonidae, Leptodactylidae, and Odontophrynidae.

There were also disagreements with previous analyses regarding relationships within many of these major hyloid clades. Furthermore, many relationships in our tree were only weakly supported relative to the phylogenomic studies. We note several of the disagreements here. First, within Neoaustrarana, we found the weakly supported relationships: ((Cycloramphidae, Hylodidae), (Alsodidae, Batrachylidae)), as did Portik et al. (2023) in their gigamatrix tree. Feng et al. (2017) did not support monophyly of Neoaustrarana. Streicher et al. (2018) strongly supported the relationships: ((Alsodidae, Hylodidae), (Cycloramphidae, Batrachylidae)). Hime et al. (2021) found the relationships: (Cycloramphidae (Hylodidae, (Alsodidae, Batrachylidae))). The clade of Alsodidae + Batrachylidae was strongly supported by them and congruent with our results here, whereas the placement of Hylodidae had moderate support in their tree. Jetz and Pyron (2018) found: (Cycloramphidae, (Batrachylidae, (Alsodidae, Hylodidae))).

Within Amazorana, we placed the clade Ceratophryidae + Hemiphractidae as the sister group to Hylidae with relatively strong support (bootstrap = 91%), as did Portik et al. (2023) in their phylogenomic and gigamatrix trees. Feng et al. (2017) inferred that Ceratophryidae was the sister taxon to Hylidae + Hemiphractidae. Streicher et al. (2018) and Hime et al. (2021) placed Hemiphractidae as the sister taxon to Hylidae + Ceratophryidae (with strong and weak support, respectively). Jetz and

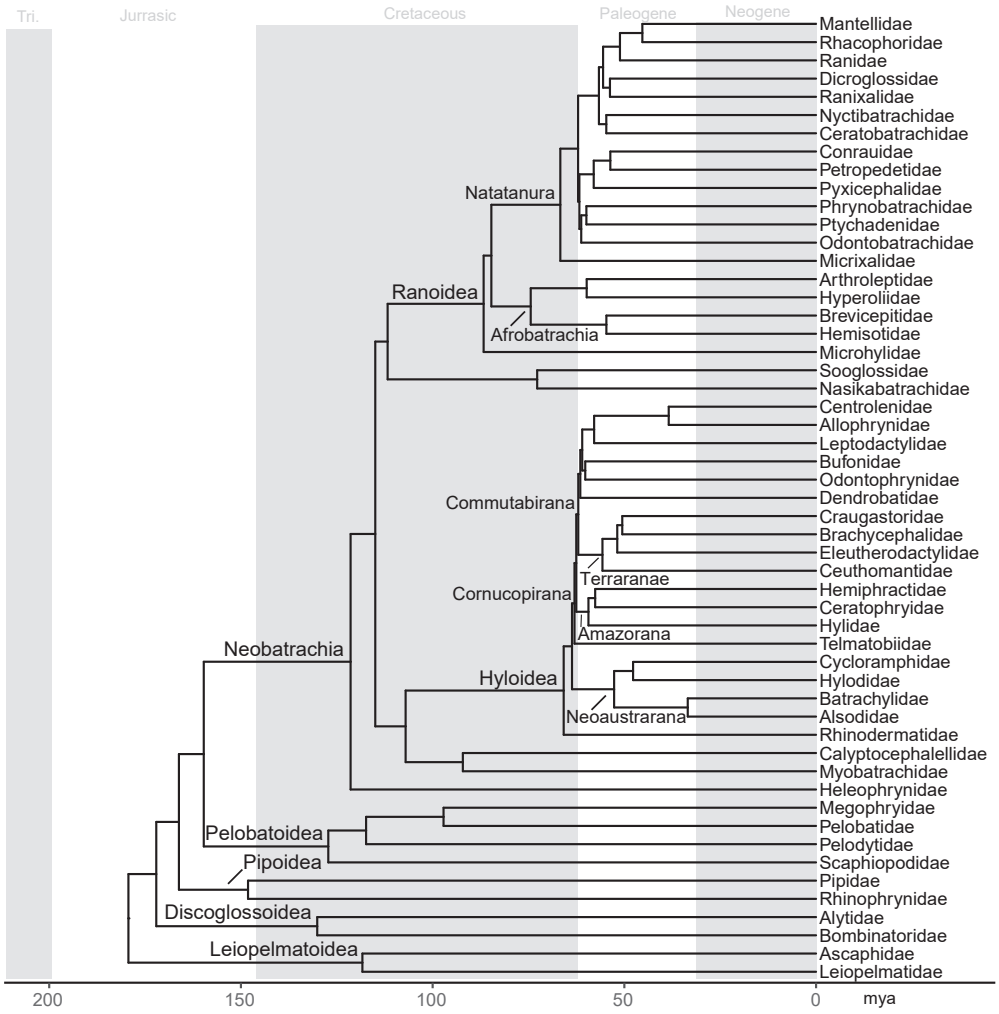

Fig. 4. Time-calibrated phylogeny based on partitioned maximum-likelihood analysis reduced to one tip per family. Clade names discussed in text are indicated on relevant branches. Family-level taxonomy follows AmphibiaWeb (2022) except for Strabomantidae, which was not monophyletic in our analysis and was included within Craugastoridae.

Pyron (2018) did not place these three families together.

Within Terraranae, most recent studies placed Ceuthomantidae as the sister taxon to the other families (as does our tree here), but there is otherwise considerable disagreement among studies. We strongly placed Eleutherodactylidae as the sister taxon to Brachycephalidae + Craugastoridae (with the latter including Strabomantidae). This is in agreement with Hedges et al. (2008), Heinicke et al. (2018), and Hime et al. (2021). However, Feng et al. (2017), Streicher et al. (2018), and Portik et al. (2023) supported: (Brachycephalidae (Eleutherodactylidae, Craugastoridae)). Other authors (e.g., Pyron, 2014; Hutter et al., 2017; Jetz and Pyron, 2018) have instead supported: (Craugastoridae (Eleutherodactylidae, Brachycephalidae)).

Within Commutabirana, we found the relationships (Fig. 3): (Terraranae (Dendrobatidae ((Bufonidae, Odontophrynidae) (Leptodactylidae (Allophrynidae, Centrolenidae))))). Yet the relationships among most of these families were not strongly supported. Hime et al. (2021) found similar relationships to those found here (but again with limited support for many relationships). However, they lacked data for Allophrynidae. Portik et al. (2023) found similar relationships (with strong phylogenomic support), but switching the placements of Dendrobatidae and Terraranae. Streicher et al. (2018) also found similar relationships, but placed Dendrobatidae with Leptodactylidae as the sister group to Allophrynidae + Centrolenidae (with only limited support for this clade of four families). Feng et al. (2017), in their Fig. 2, included Cycloramphidae in this clade and found quite different relationships among these families: (Dendrobatidae (Terraranae (Bufonidae (Allophrynidae, Centrolenidae) (Leptodactylidae (Odontophrynidae, Cycloramphidae))))). Finally, Jetz and Pyron (2018) also recovered the relationships: (Bufonidae, Odontophrynidae) (Leptodactylidae (Allophrynidae, Centrolenidae)). However, they placed Dendrobatidae and Terraranae at the base of Hyloidea and did not support monophyly of Commutabirana.

Besides Hyloidea, the other major, species-rich clade of Neobatrachia is Ranoidea (Fig. 3). Within Ranoidea, our tree strongly supported Microhylidae as the sister taxon to Afrobatrachia and Natatanura. This arrangement was also supported by the phylogenomic analyses of Feng et al. (2017), Hime et al. (2021), and Portik et al. (2023). However, Jetz and Pyron (2018) placed Microhylidae as the sister taxon to Afrobatrachia.

Within Afrobatrachia, our results supported the relationships: ((Arthroleptidae, Hyperoliidae) (Brevicepitidae, Hemisotidae)). This is in agreement with the phylogenomic studies (Feng et al., 2017; Hime et al., 2021; Portik et al., 2023) and Jetz and Pyron (2018).

Within Natatanura, there was considerably more conflict. First, our tree placed Micrixalidae as the sister taxon to all other members of Natatanura (with strong support). Portik et al. (2023) also placed Micrixalidae here. Most other studies compared here did not include this taxon, but Feng et al. (2017) placed Micrixalidae as the sister taxon to

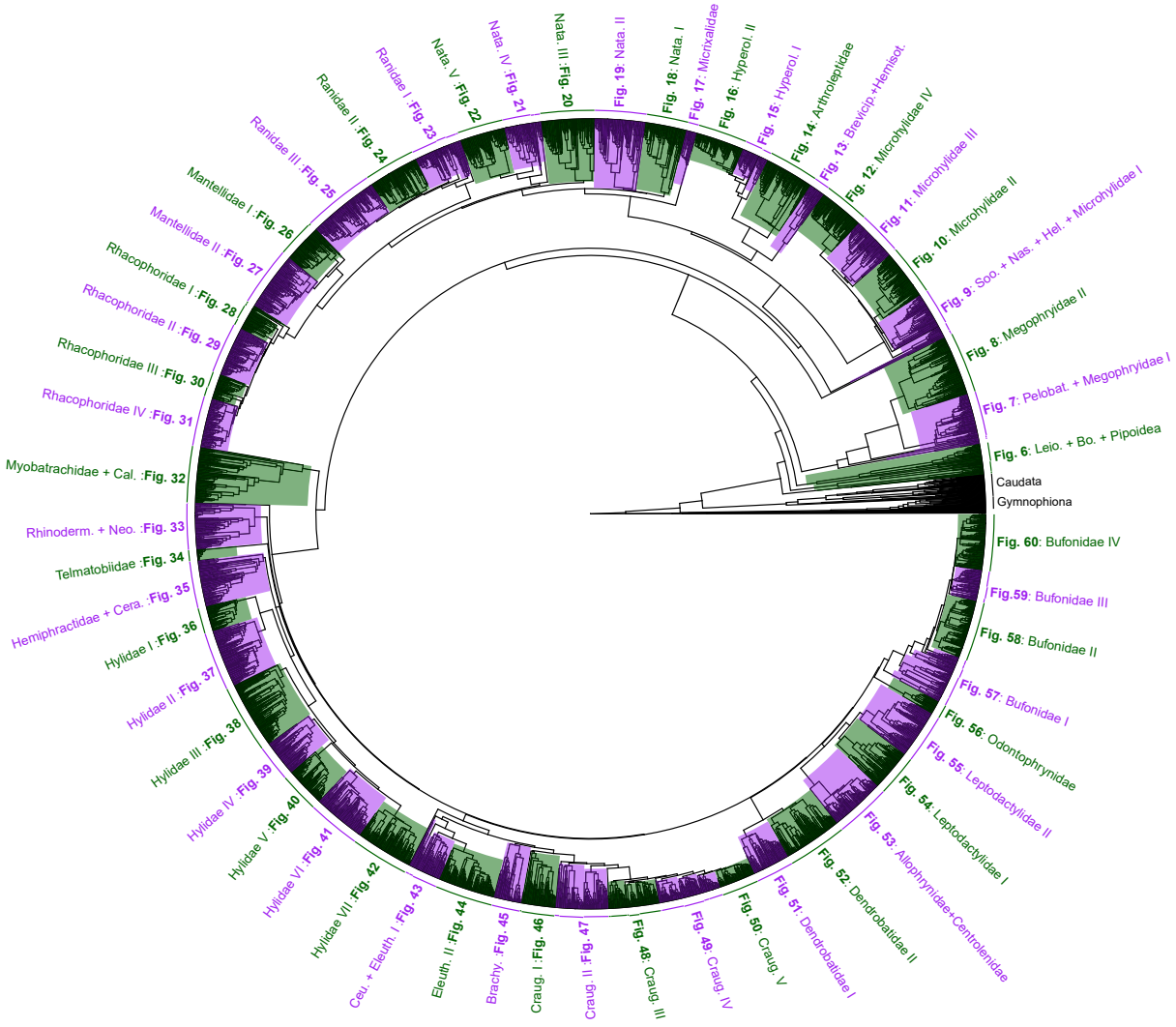

Fig. 5. Legend for detailed phylogeny figures (Figs. 6–60) indicating the placement of each subtree in the global phylogeny. Abbreviations as follows: Le. = Leoipelmatoidea, Bo = Bombinatoridae, Pelobat. = Pelobatoidea, Soo. = Sooglossidae, Nas. = Nasikabatrachidae, Hel. = Heleophrynidae, Brevicip. = Brevicipitidae, Hemiso. = Hemisotidae, Hyper. = Hyperoliidae, Micrixal. = Micrixalidae, Nata. = Natatanura, Cal. = Calyptocephalellidae, Rhinoderm. = Rhinodermatidae, Neo. = Neoaustrarana, Cera. = Ceratophryidae, Ceu. = Ceuthomantidae, Eleuth. = Eleutherodactylidae, Brachy. = Brachycephalidae, Craug. = Craugastoridae.

Ranixalidae. Above Micrixalidae, our tree supports division of Natatanura into a clade of mostly African families ((Pyxicephalidae (Conrauidae, Petropedetidae)), (Odontobatrachidae (Phrynobatrachidae, Ptychadenidae))) and a clade of mostly Asian families ((Nyctibatrachidae, Ceratobatrachidae) ((Dicroglossidae, Ranixalidae) (Ranidae (Rhacophoridae, Mantellidae)))). This basic division between predominantly African and Asian families also appeared in the tree of Feng et al. (2017) and was strongly supported by Hime et al. (2021). Many other components of our tree also appear in these phylogenomic trees. Within the mostly African clade, these include the clade of (Pyxicephalidae (Conrauidae, Petropedetidae)) and the clade (Odontobatrachidae (Phrynobatrachidae, Ptychadenidae)). Within the mostly Asian clade, our tree agrees with that of Hime et al. (2021) in supporting a clade consisting of Nyctibatrachidae + Ceratobatrachidae (bootstrap = 70% here; 100% in that study), which is the sister group to a clade containing (Dicroglossidae (Ranidae (Rhacophoridae, Mantellidae))). However, our tree also included Ranixalidae, which was weakly supported as the sister taxon to Dicroglossidae. Ranixalidae was not sampled by Hime et al. (2021). The tree of Feng et al. (2017) was less congruent.

There was also some disagreement between our tree and that of Jetz and Pyron (2018) within Natatanura. For example, their tree shows the group of mostly African families as being paraphyletic with respect to the group of mostly Asian families (an arrangement that also occurs in the phylogenomic and gigamatrix trees of Portik et al., 2023). However, our tree and theirs are largely congruent regarding relationships among the mostly Asian families, with both trees supporting the relationships: (Nyctibatrachidae + Ceratobatrachidae), ((Dicroglossidae + Ranixalidae) (Ranidae (Rhacophoridae, Mantellidae))).

In summary, our tree was highly congruent with the phylogenomic trees (Feng et al., 2017; Hime et al., 2021) regarding the relationships among the mostly African families of Natatanura (but discordant with that of Jetz and Pyron, 2018). Among the mostly Asian families, it was highly concordant with that of both Hime et al. (2021) and Jetz and Pyron (2018).

3.4. Relationships within families

In the paragraphs that follow, we briefly discuss selected relationships within families. This is not intended as a comprehensive review of our results or previous studies. Instead, we focus particularly on non-monophyletic subfamilies and genera (based specifically on the taxonomy of AmphibiaWeb, 2022), and whether these cases might reflect

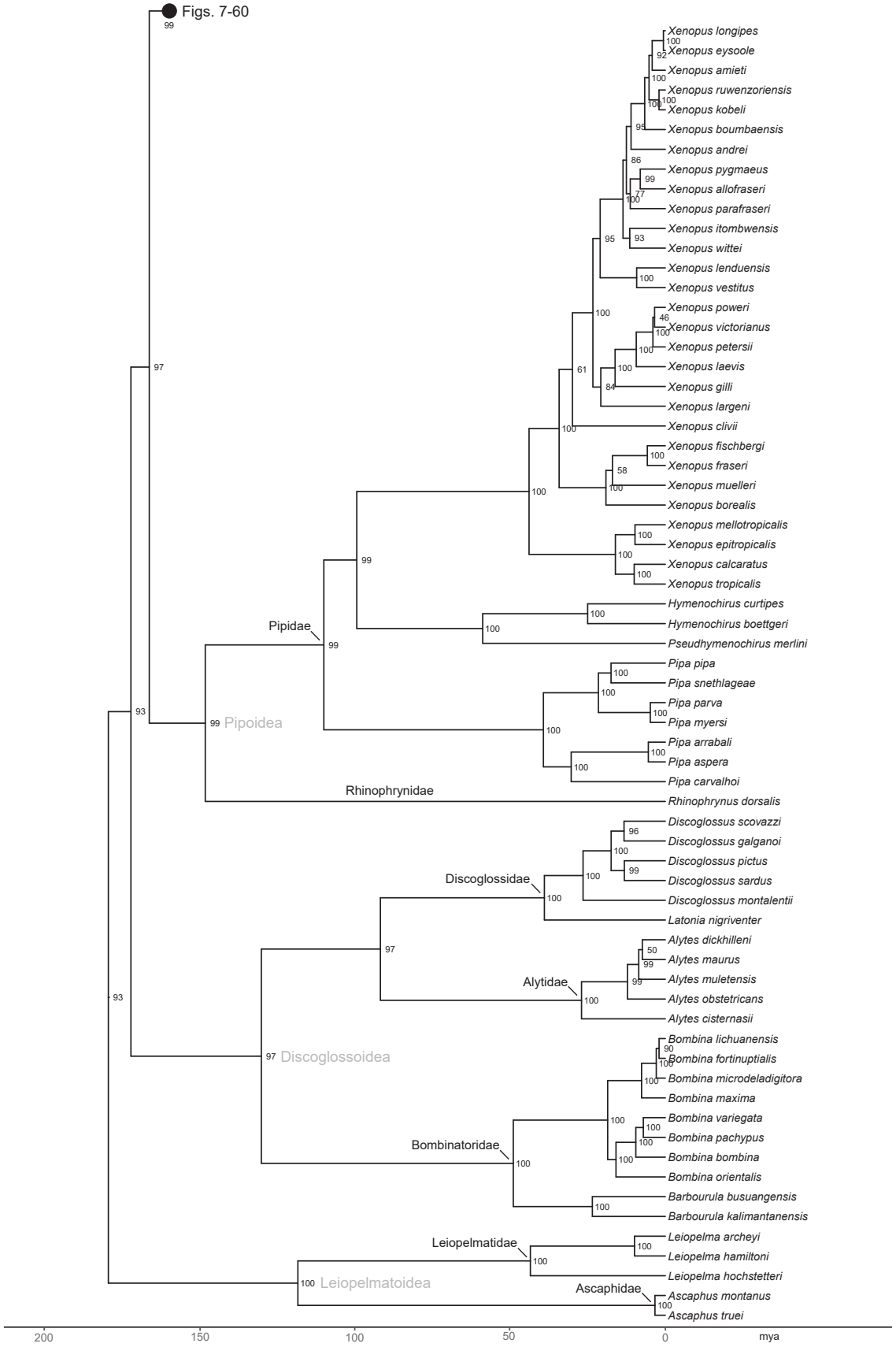

Fig. 6. Archaeobatrachia I subtree including Leiopelmatoidea (Ascaphidae + Leiopelmatidae), Discoglossoidea (Alytidae + Bombinatoridae), and Pipoidea (Rhinophrynidae + Pipidae). Values adjacent to nodes are bootstrap support values from the partitioned maximum-likelihood analysis. Branch lengths are from the divergence-time estimation (in millions of years ago; mya).

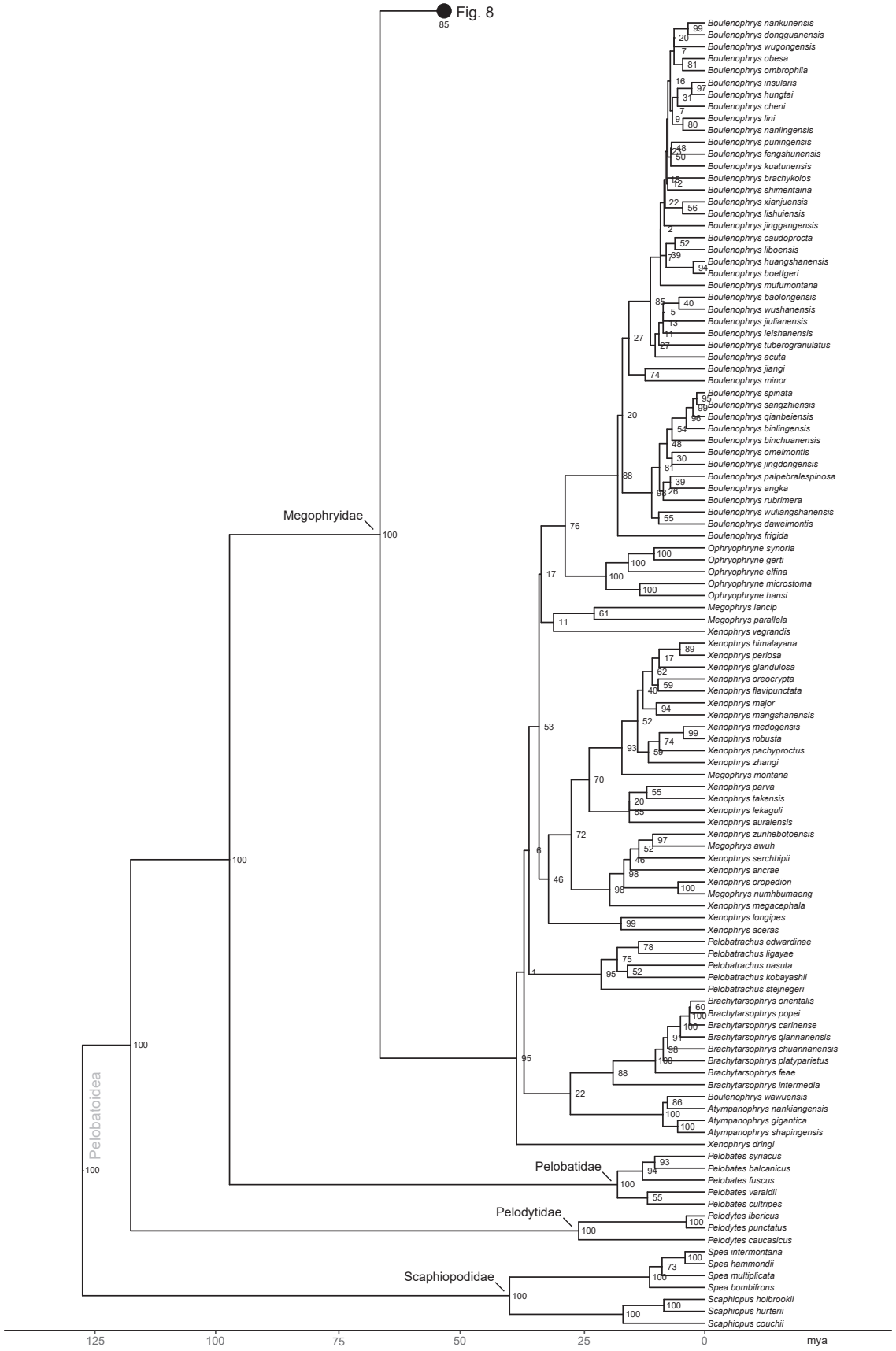

Fig. 7. Archaeobatrachia II subtree including Pelobatoidea (Megophryidae [part] + Pelobatidae + Pelodytidae + Scaphiopodidae). Values adjacent to nodes are bootstrap support values from the partitioned maximum-likelihood analysis. Branch lengths are from the divergence-time estimation (in millions of years ago; mya).

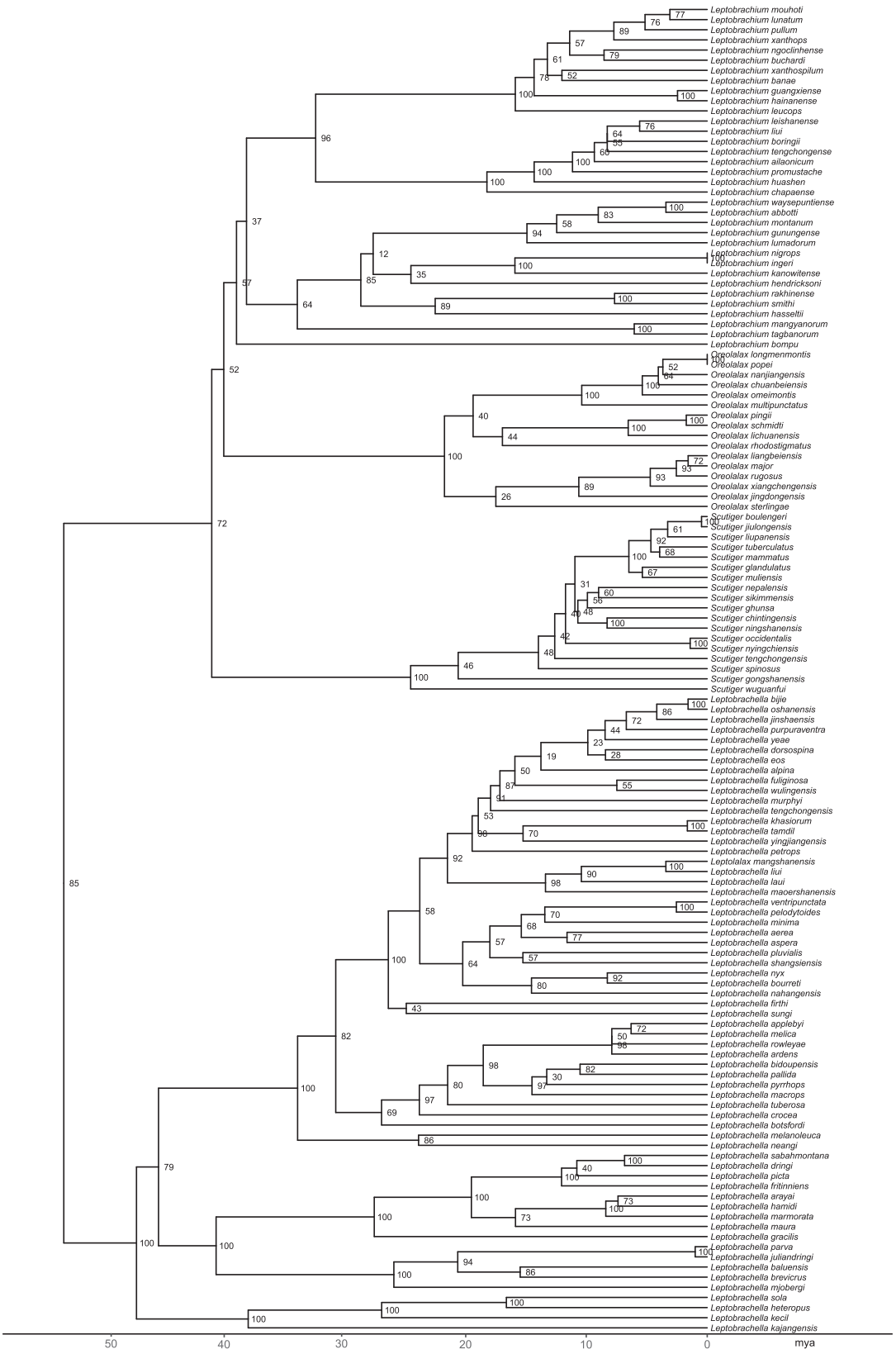

Fig. 8. Archaeobatrachia III subtree including Pelobatoidea (Megophryidae [part]). Values adjacent to nodes are bootstrap support values from the partitioned maximum-likelihood analysis. Branch lengths are from the divergence-time estimation (in millions of years ago; mya).

Fig. 9. Neobatrachia I subtree including Heleophrynidae, Nasikabatrachidae, Sooglossidae and Microhylidae (part: Kalophryninae, Melanobatrachinae, Hoplophryninae, Phrynomerinae, Adelastinae, Otophryninae, and Gastrophryninae). Values adjacent to nodes are bootstrap support values from the partitioned maximum-likelihood analysis. Branch lengths are from the divergence-time estimation (in millions of years ago; mya).

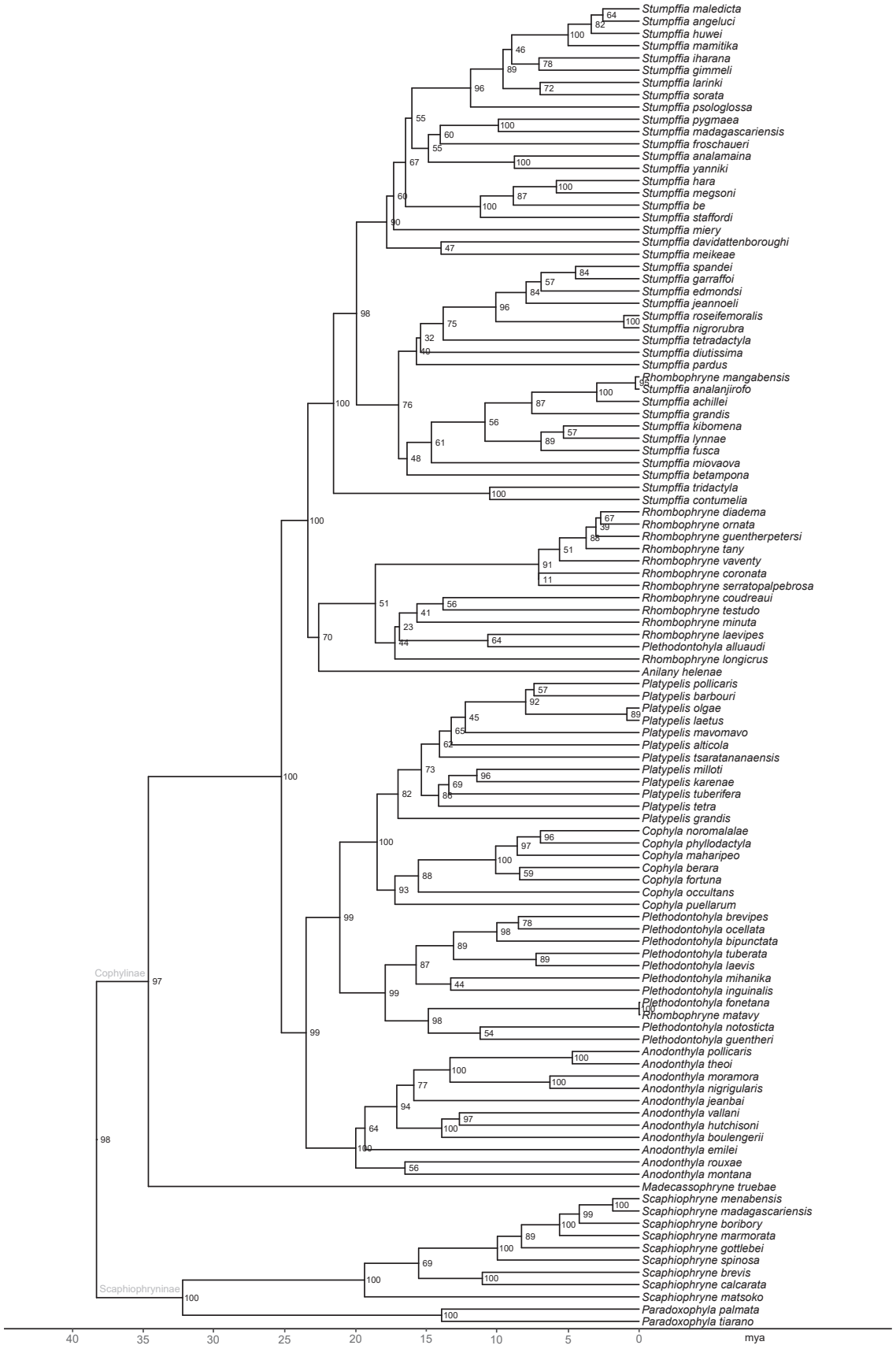

Fig. 10. Neobatrachia II subtree including Microhylidae (part: Scaphiophryninae and Cophylinae). Values adjacent to nodes are bootstrap support values from the partitioned maximum-likelihood analysis. Branch lengths are from the divergence-time estimation (in millions of years ago; mya).

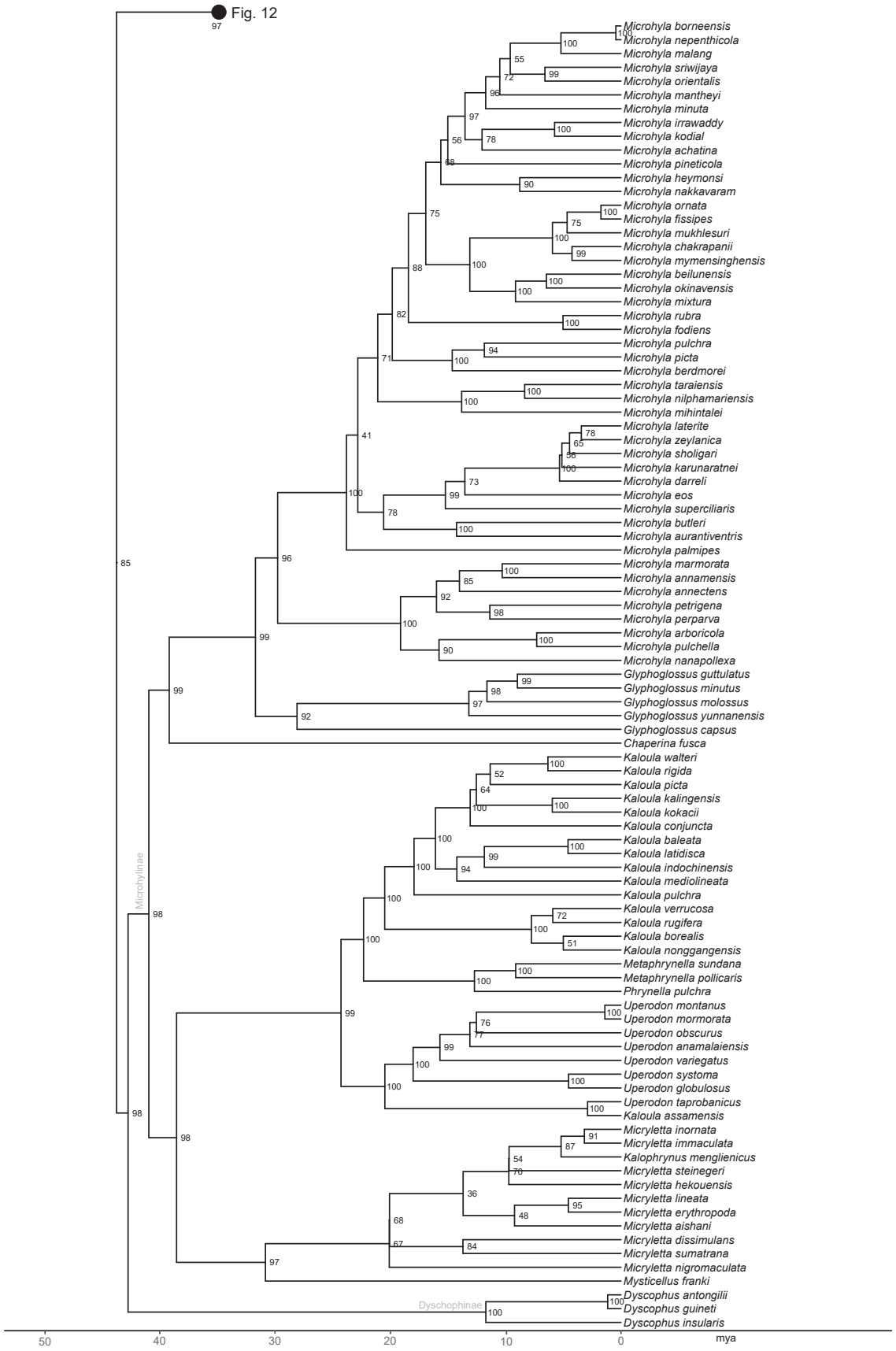

Fig. 11. Neobatrachia III subtree including Microhylidae (part: Dyscophinae and Microhylinae). Values adjacent to nodes are bootstrap support values from the partitioned maximum-likelihood analysis. Branch lengths are from the divergence-time estimation (in millions of years ago; mya).

Fig. 12. Neobatrachia IV subtree including Microhylidae (part: Asterophryinae). Values adjacent to nodes are bootstrap support values from the partitioned maximum-likelihood analysis. Branch lengths are from the divergence-time estimation (in millions of years ago; mya).

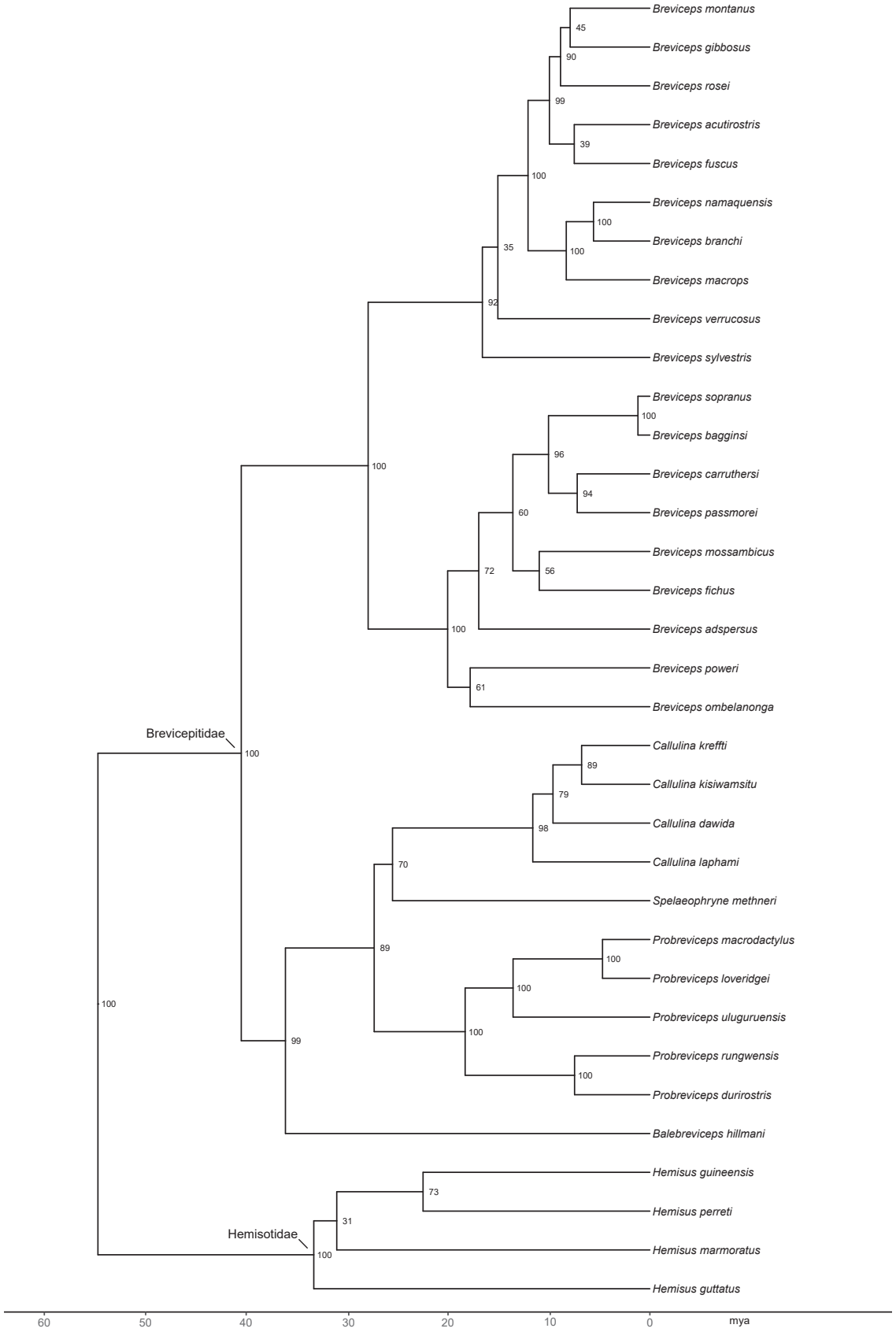

Fig. 13. Neobatrachia V subtree including Brevicipitidae and Hemisotidae. Values adjacent to nodes are bootstrap support values from the partitioned maximum-likelihood analysis. Branch lengths are from the divergence-time estimation (in millions of years ago; mya).

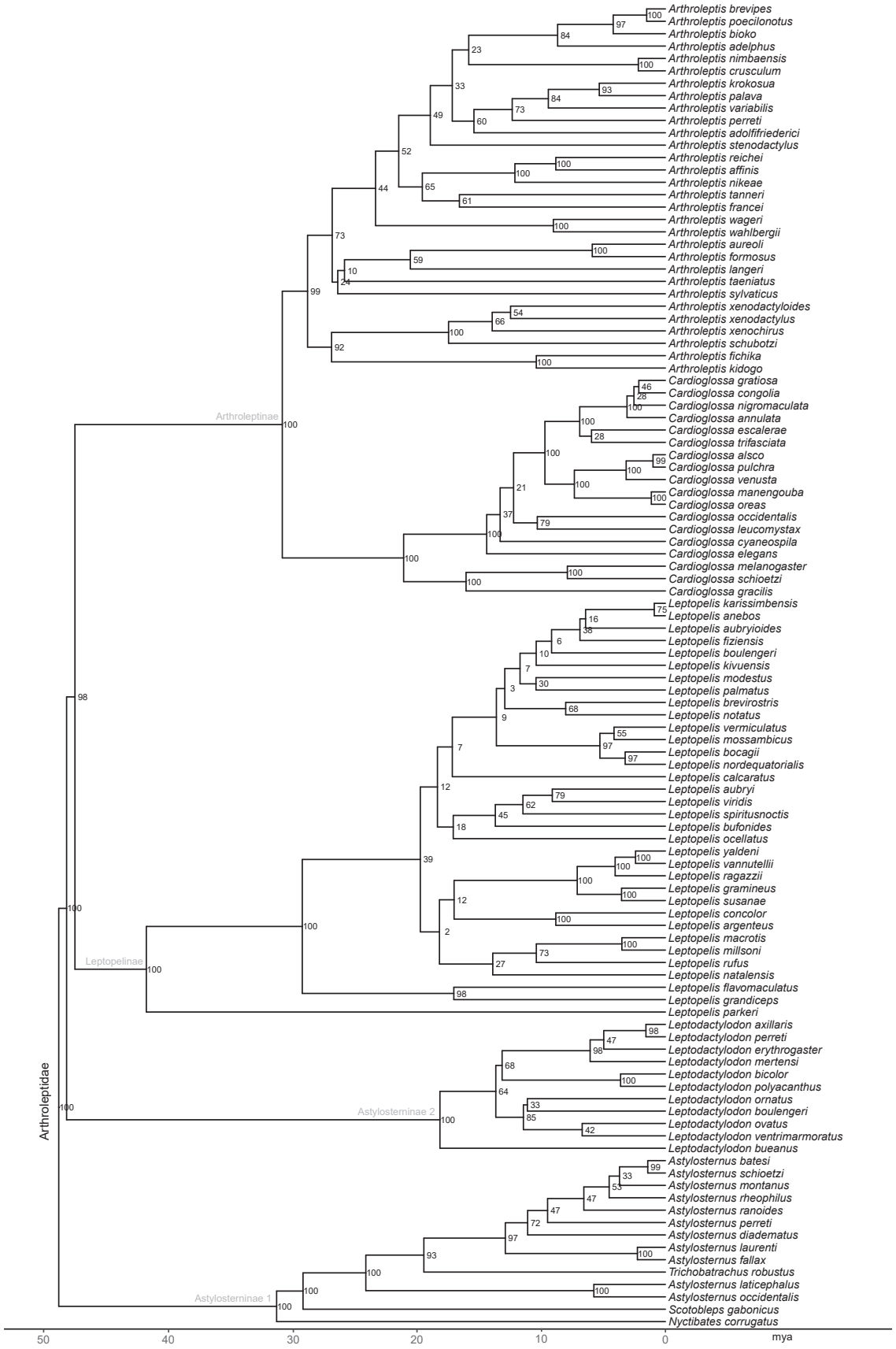

Fig. 14. Neobatrachia VI subtree including Arthroleptidae. Values adjacent to nodes are bootstrap support values from the partitioned maximum-likelihood analysis. Branch lengths are from the divergence-time estimation (in millions of years ago; mya).

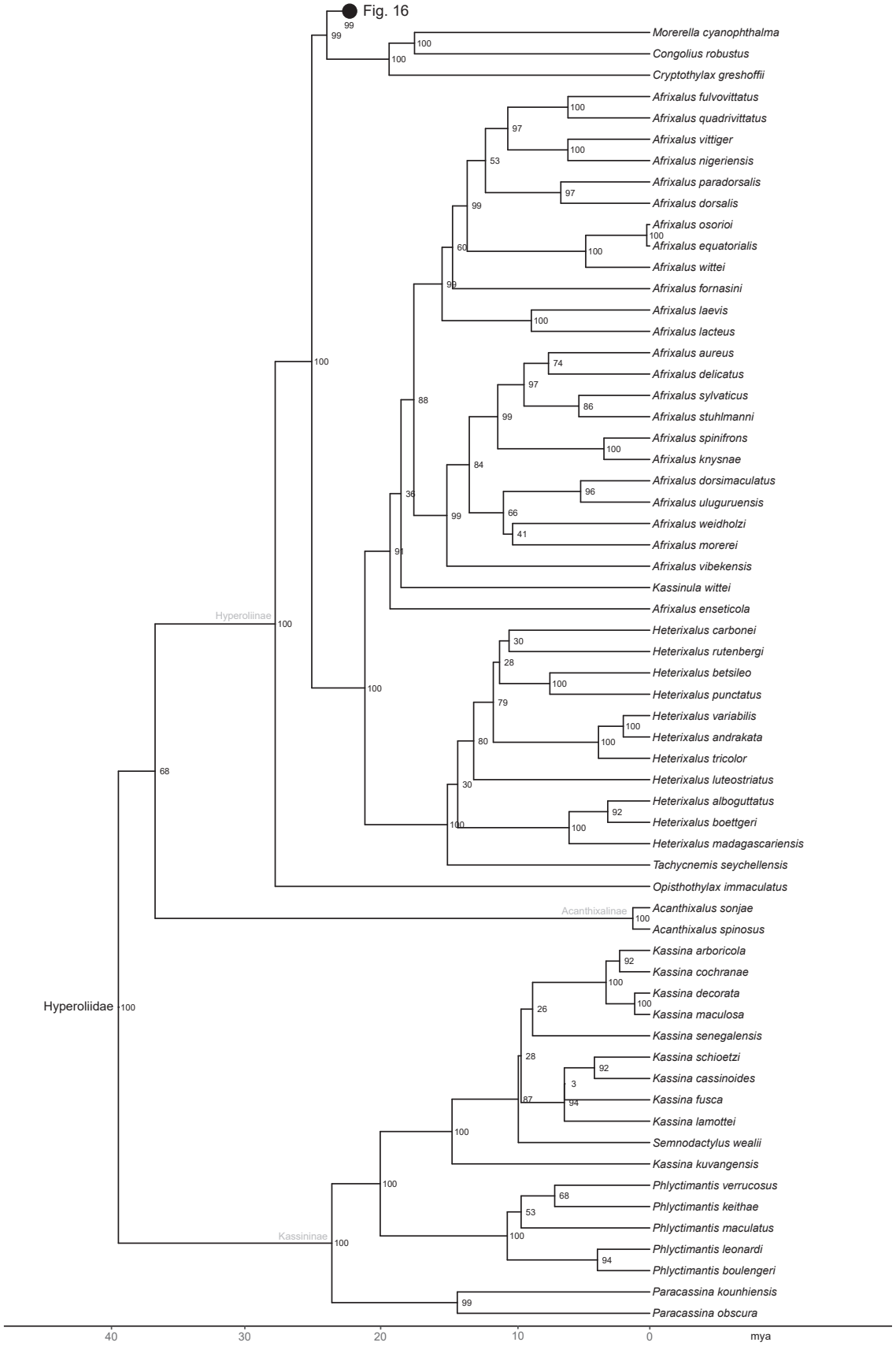

Fig. 15. Neobatrachia VII subtree including Hyperoliidae (part). Values adjacent to nodes are bootstrap support values from the partitioned maximum-likelihood analysis. Branch lengths are from the divergence-time estimation (in millions of years ago; mya).

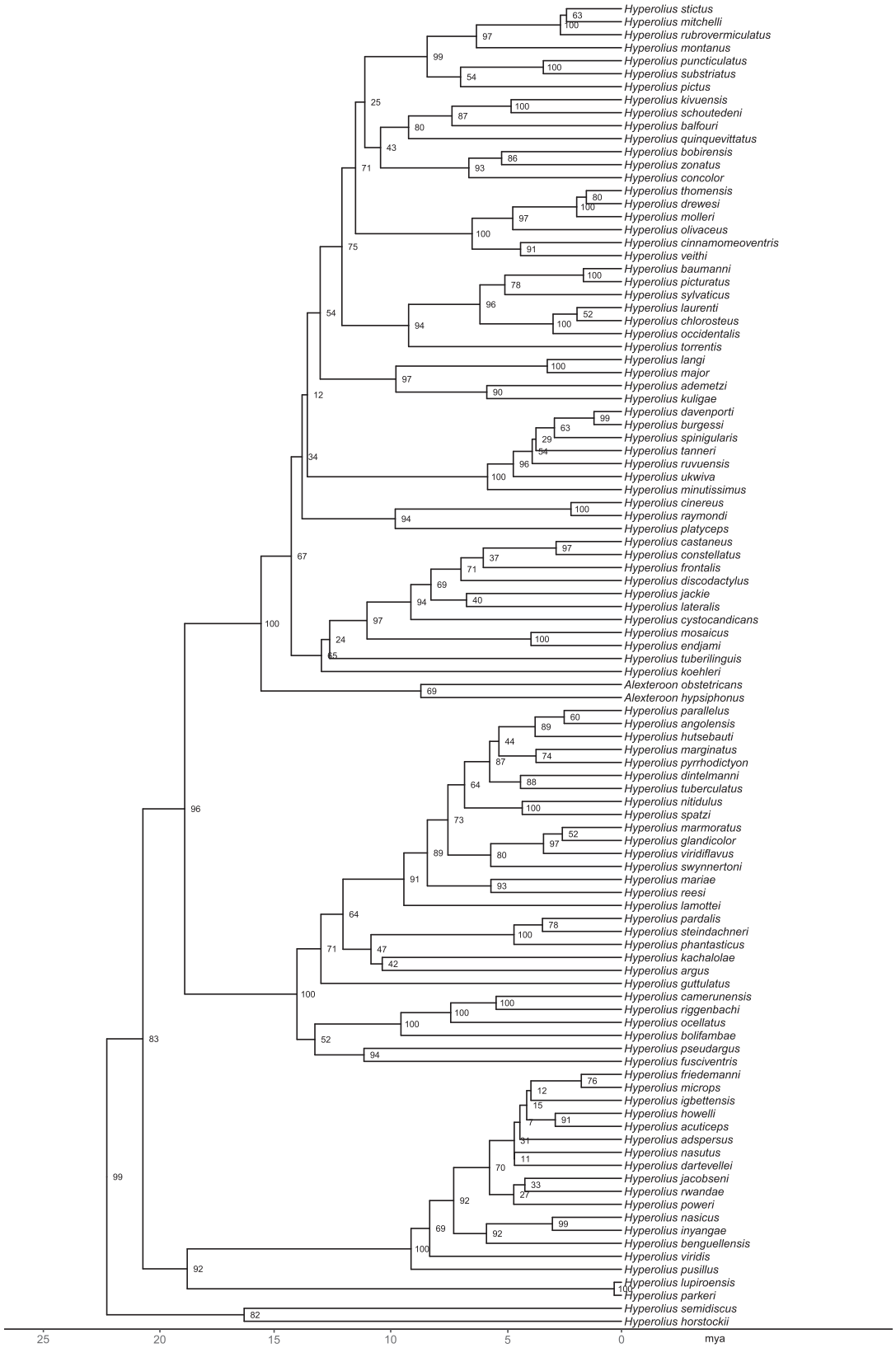

Fig. 16. Neobatrachia VIII subtree including Hyperoliidae (part). Values adjacent to nodes are bootstrap support values from the partitioned maximum-likelihood analysis. Branch lengths are from the divergence-time estimation (in millions of years ago; mya).

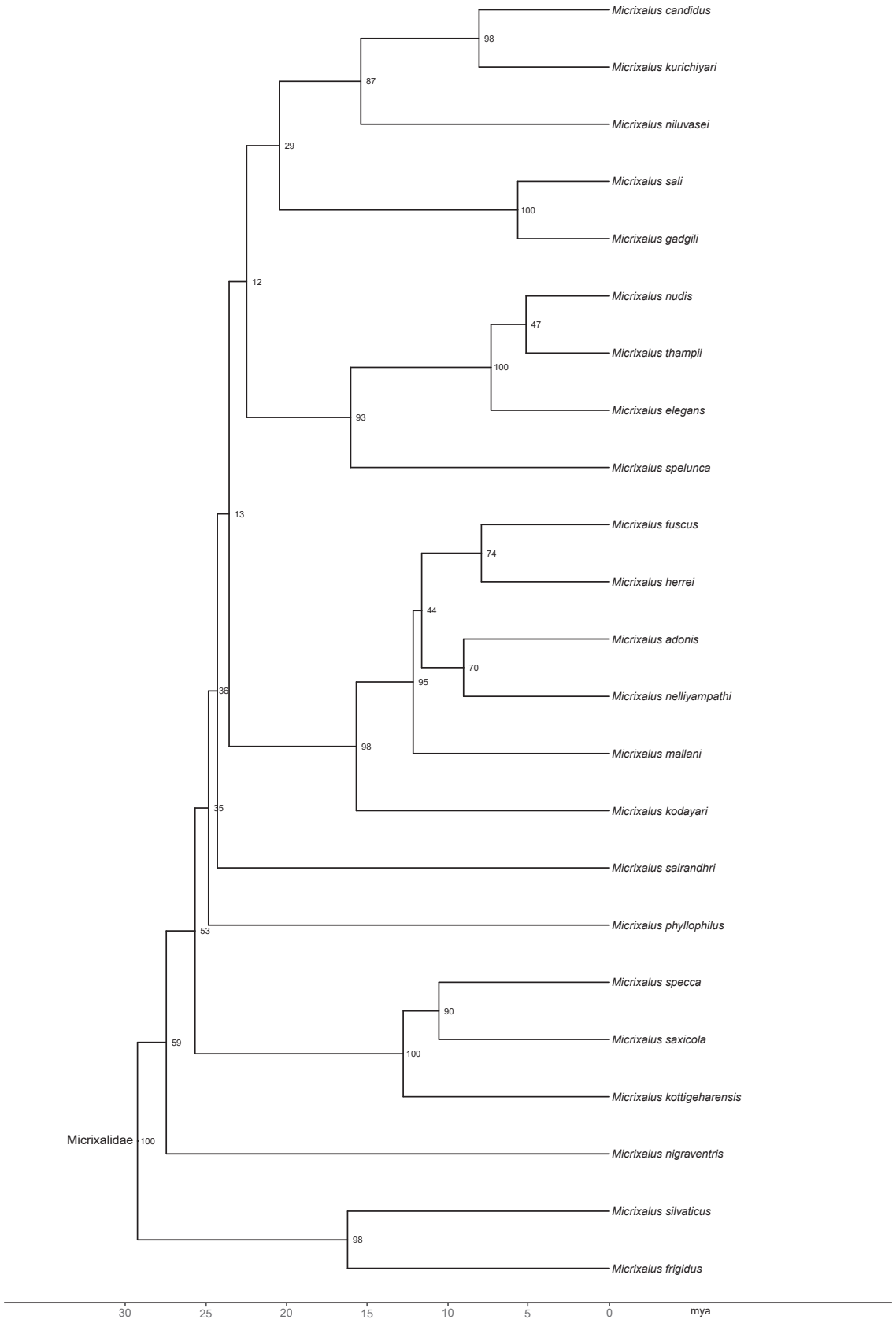

Fig. 17. Neobatrachia IX subtree including Micrixalidae. Values adjacent to nodes are bootstrap support values from the partitioned maximum-likelihood analysis. Branch lengths are from the divergence-time estimation (in millions of years ago; mya).

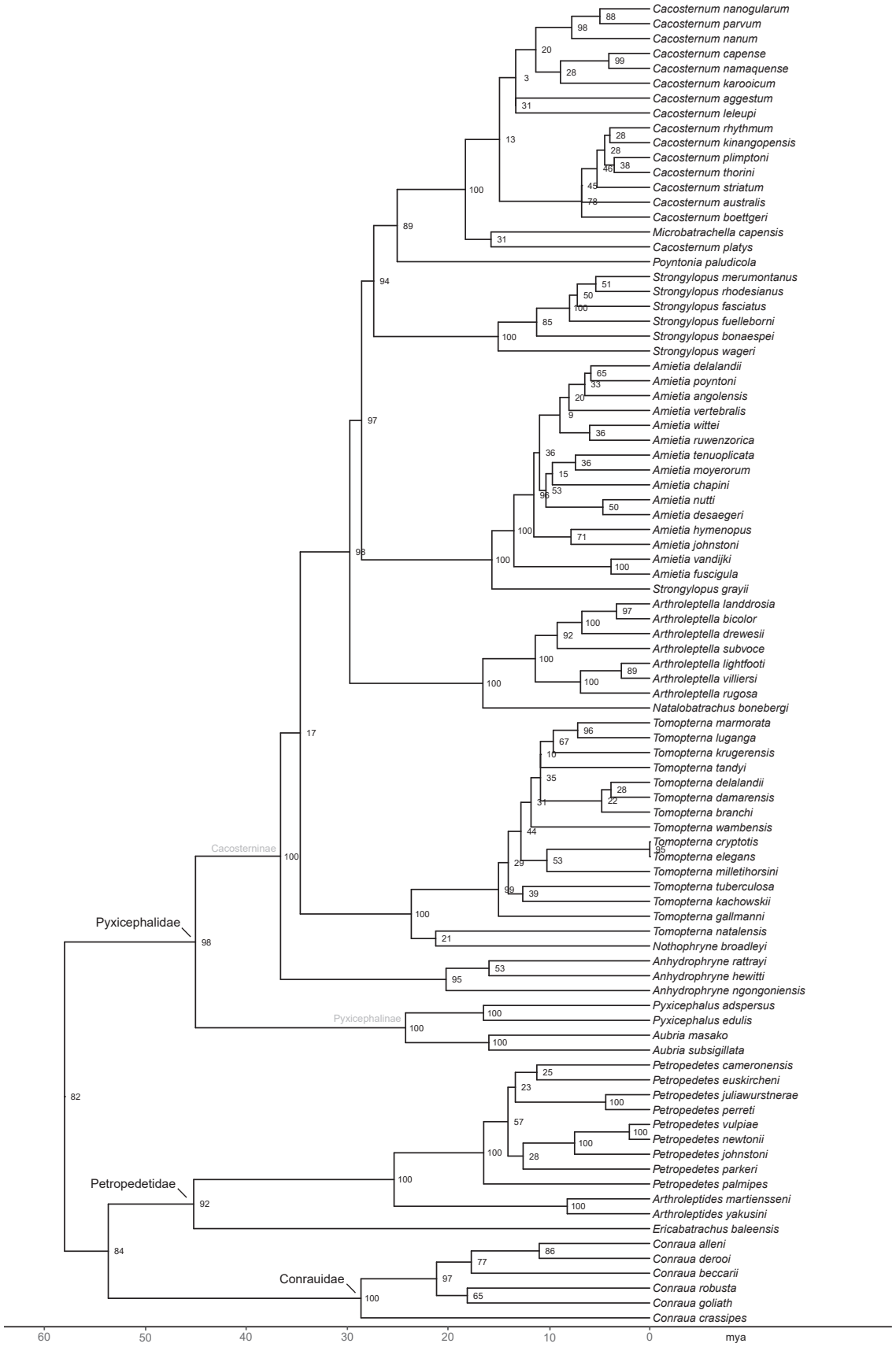

Fig. 18. Neobatrachia X subtree including Conrauidae, Petropedetidae, and Pyxicephalidae. Values adjacent to nodes are bootstrap support values from the partitioned maximum-likelihood analysis. Branch lengths are from the divergence-time estimation (in millions of years ago; mya).

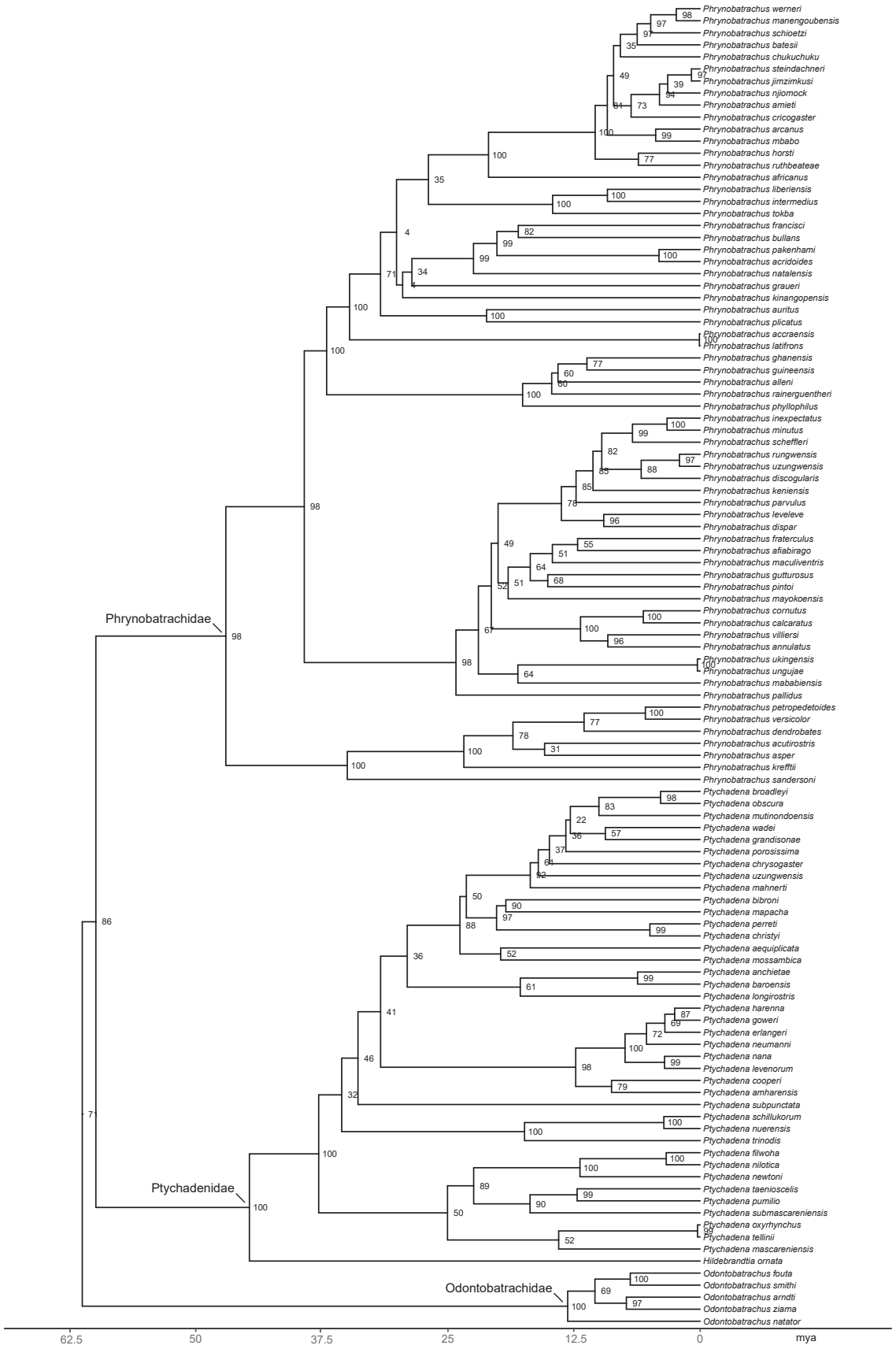

Fig. 19. Neobatrachia X subtree including Odontobatrachidae, Ptychadenidae, and Phrynobatrachidae. Values adjacent to nodes are bootstrap support values from the partitioned maximum-likelihood analysis. Branch lengths are from the divergence-time estimation (in millions of years ago; mya).

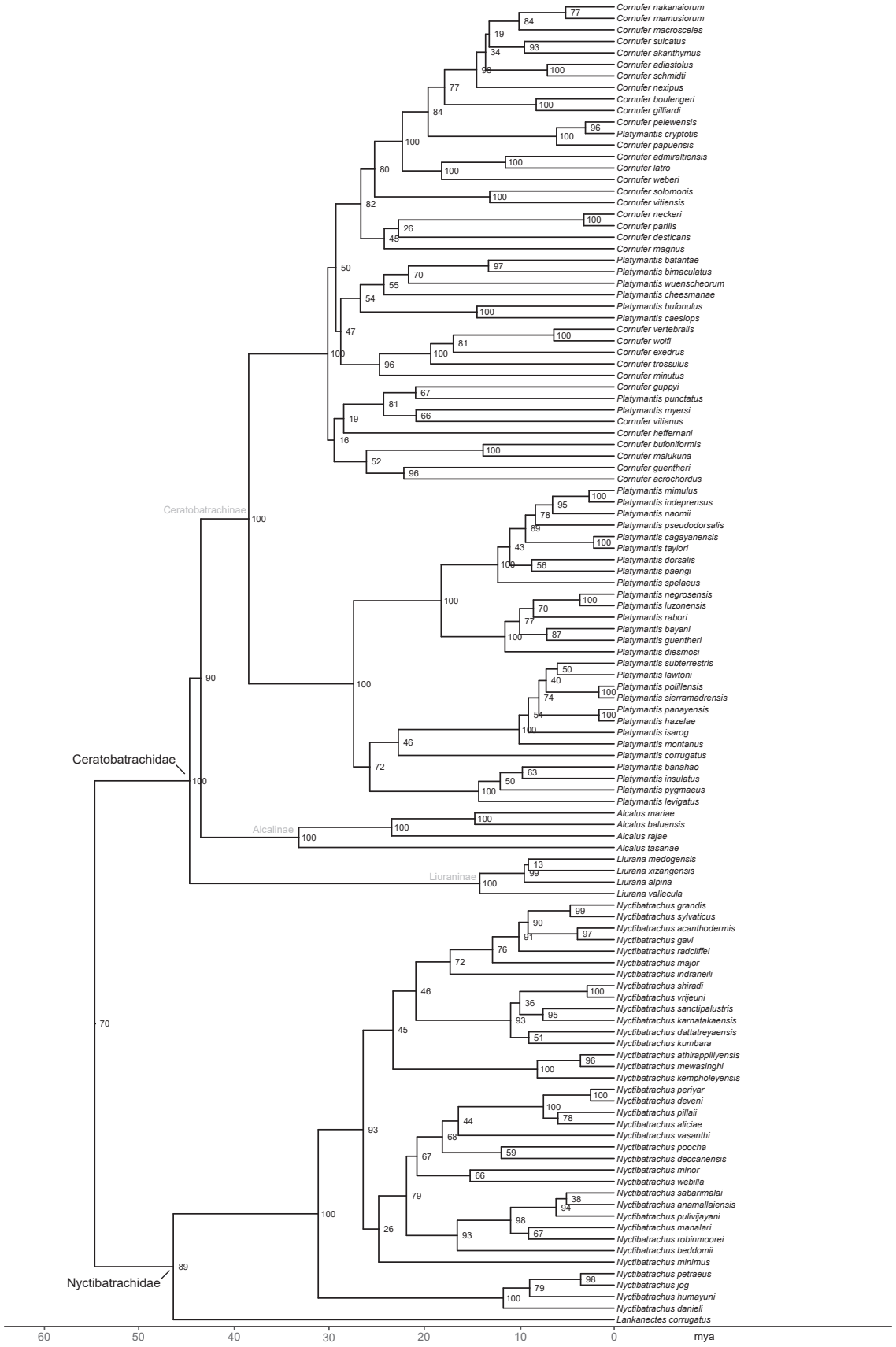

Fig. 20. Neobatrachia XI subtree including Nyctibatrachidae and Ceratobatrachidae. Values adjacent to nodes are bootstrap support values from the partitioned maximum-likelihood analysis. Branch lengths are from the divergence-time estimation (in millions of years ago; mya).

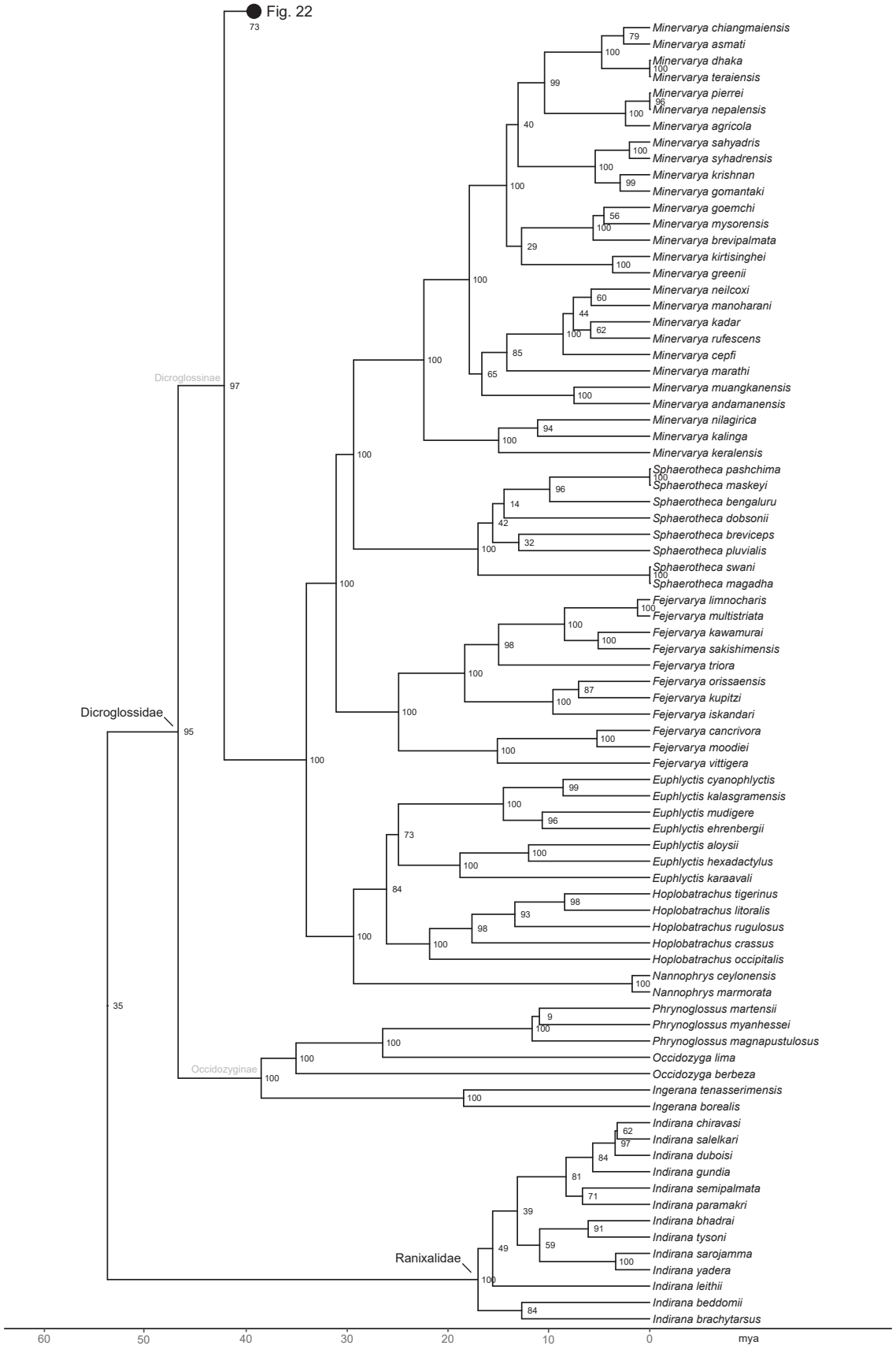

Fig. 21. Neobatrachia XII subtree including Ranixalidae and Dicroglossidae (part). Values adjacent to nodes are bootstrap support values from the partitioned maximum-likelihood analysis. Branch lengths are from the divergence-time estimation (in millions of years ago; mya).

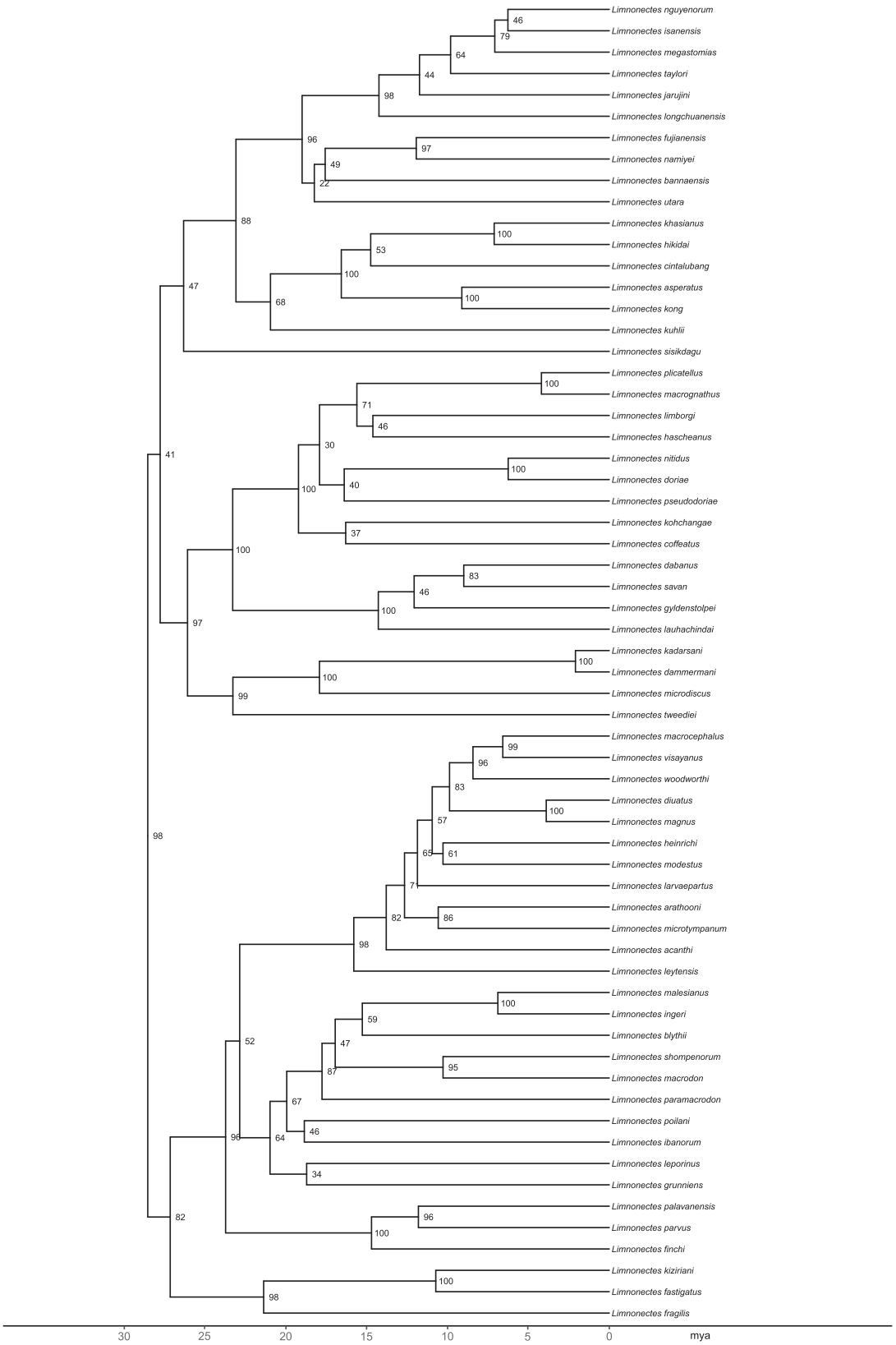

Fig. 22. Neobatrachia XIII subtree including Dicroglossidae (part). Values adjacent to nodes are bootstrap support values from the partitioned maximum-likelihood analysis. Branch lengths are from the divergence-time estimation (in millions of years ago; mya).

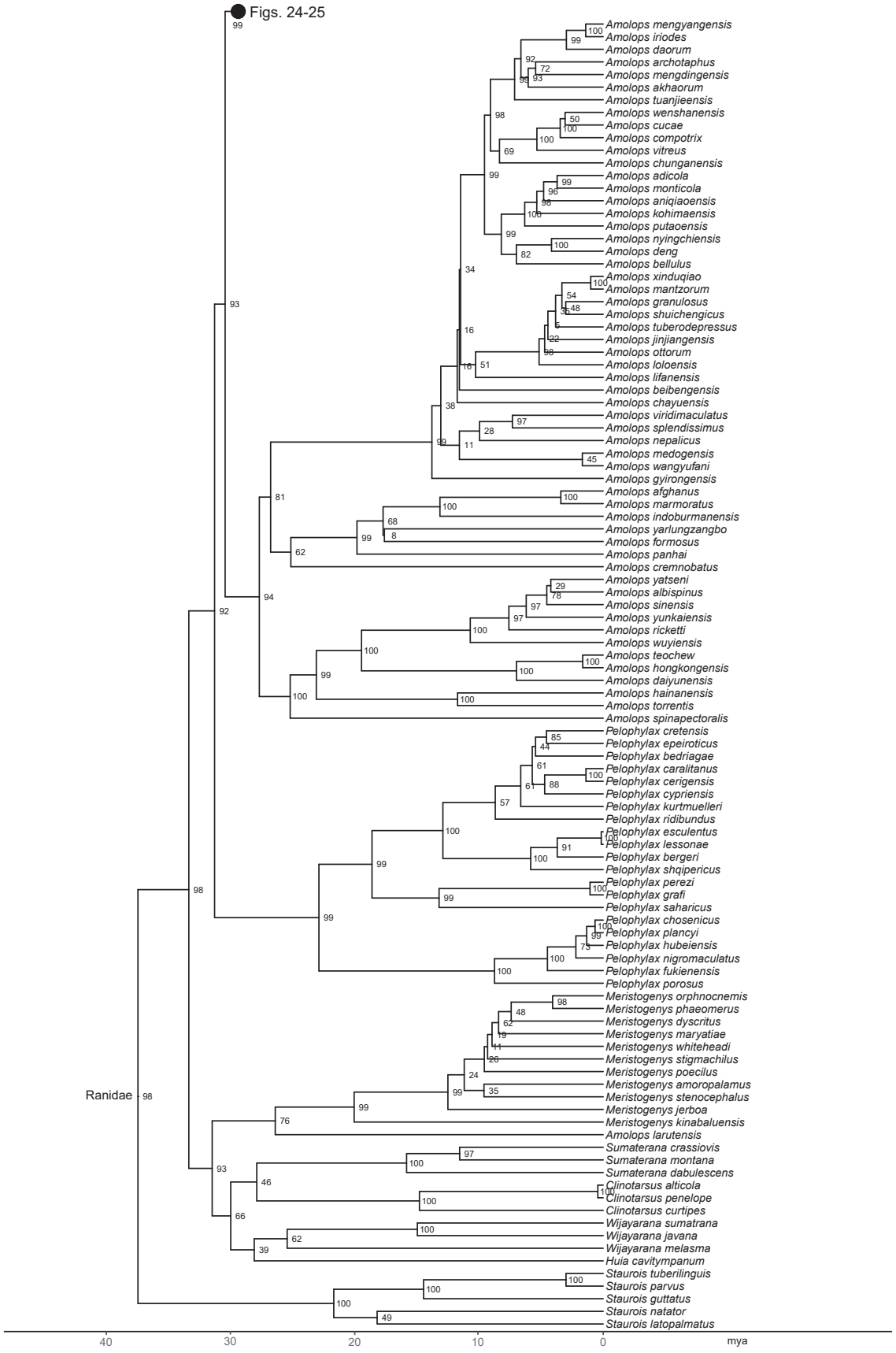

Fig. 23. Neobatrachia XIV subtree including Ranidae (part). Values adjacent to nodes are bootstrap support values from the partitioned maximum-likelihood analysis. Branch lengths are from the divergence-time estimation (in millions of years ago; mya).

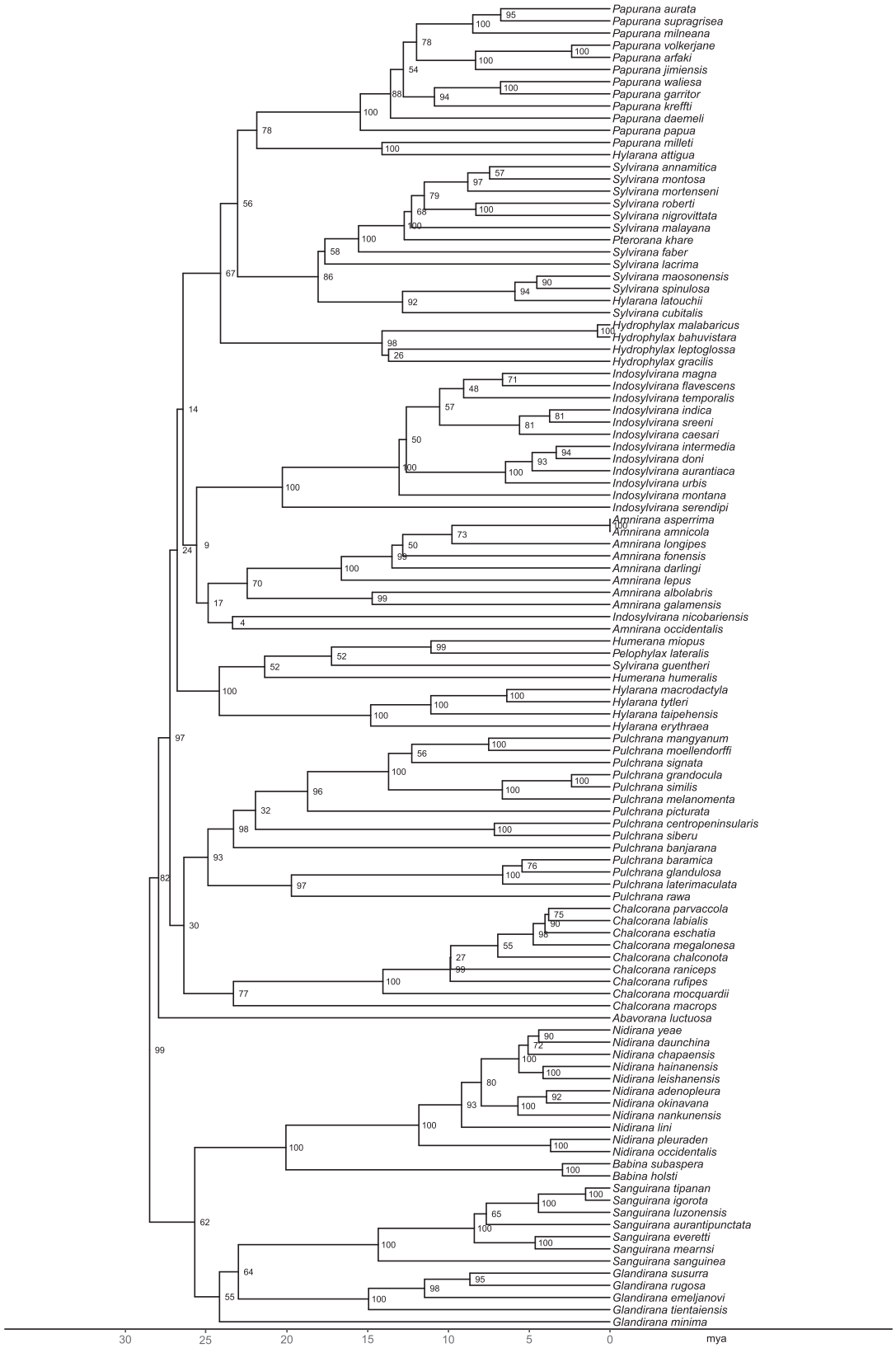

Fig. 24. Neobatrachia XV subtree including Ranidae (part). Values adjacent to nodes are bootstrap support values from the partitioned maximum-likelihood analysis. Branch lengths are from the divergence-time estimation (in millions of years ago; mya).

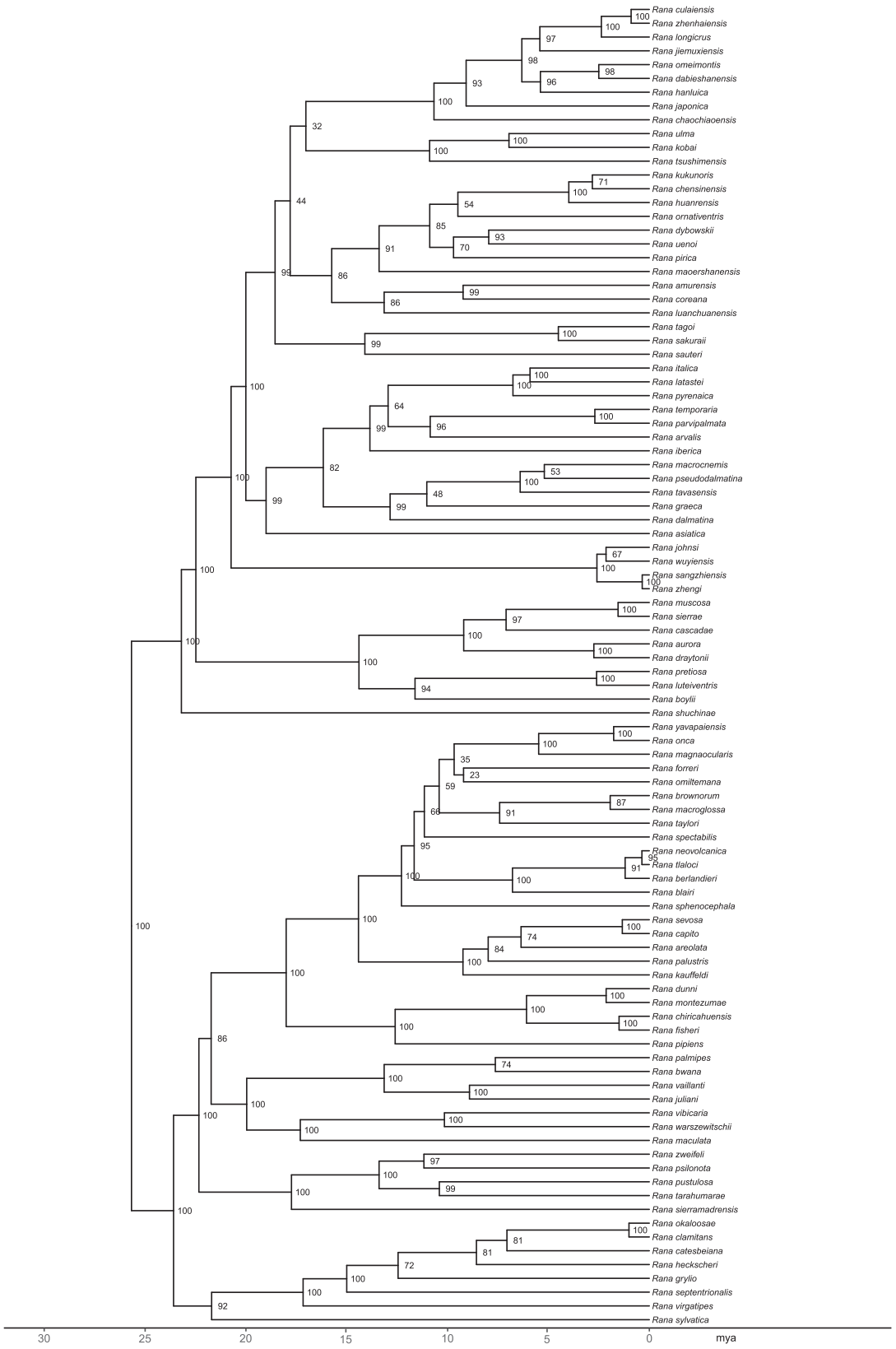

Fig. 25. Neobatrachia XVI subtree including Ranidae (part). Values adjacent to nodes are bootstrap support values from the partitioned maximum-likelihood analysis. Branch lengths are from the divergence-time estimation (in millions of years ago; mya).

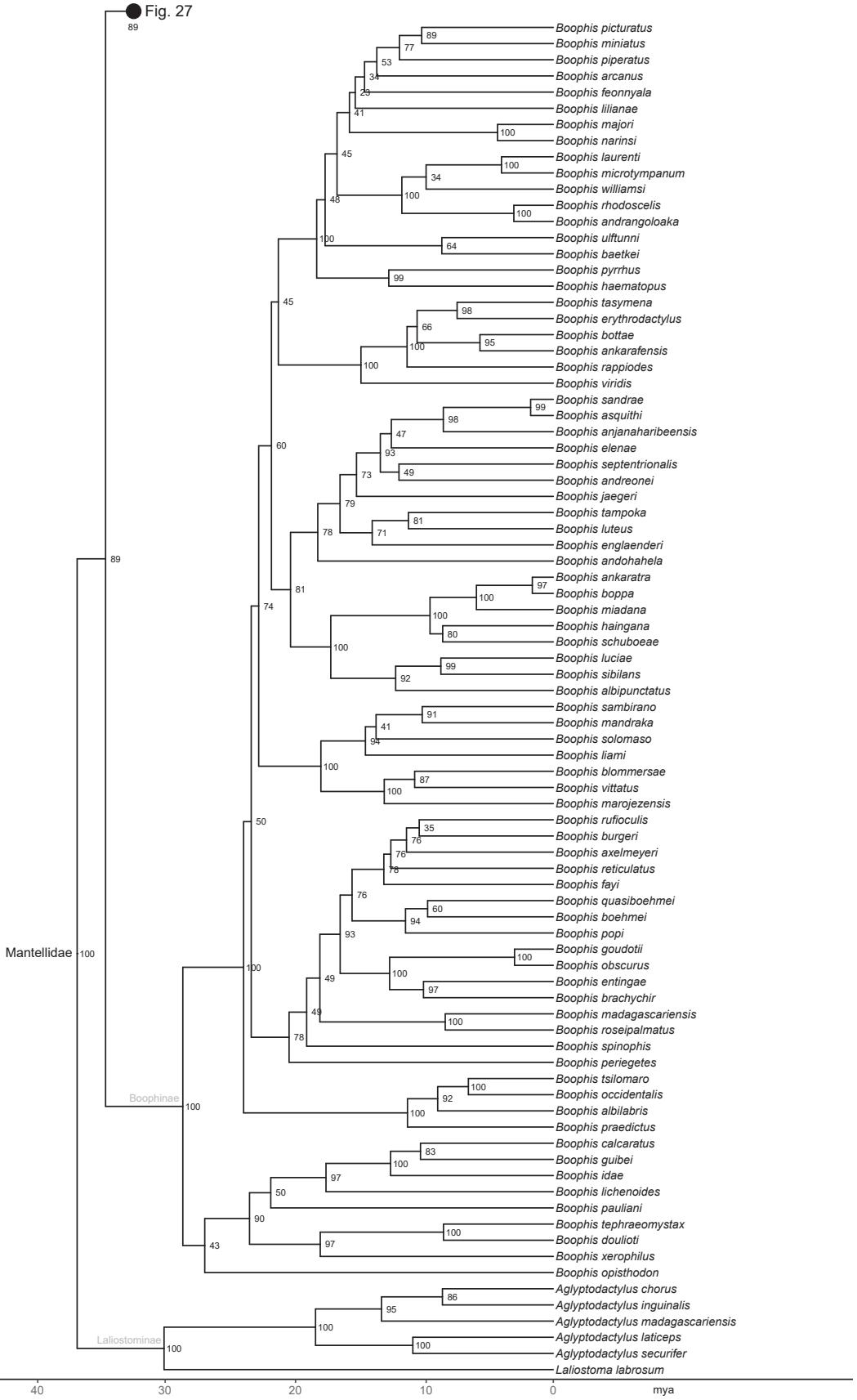

Fig. 26. Neobatrachia XVII subtree including Mantellidae (part). Values adjacent to nodes are bootstrap support values from the partitioned maximum-likelihood analysis. Branch lengths are from the divergence-time estimation (in millions of years ago; mya).

Fig. 27. Neobatrachia XVIII subtree including Mantellidae (part). Values adjacent to nodes are bootstrap support values from the partitioned maximum-likelihood analysis. Branch lengths are from the divergence-time estimation (in millions of years ago; mya).

Fig. 28. Neobatrachia XIX subtree including Rhacophoridae (part). Values adjacent to nodes are bootstrap support values from the partitioned maximum-likelihood analysis. Branch lengths are from the divergence-time estimation (in millions of years ago; mya).

Fig. 29. Neobatrachia XX subtree including Rhacophoridae (part). Values adjacent to nodes are bootstrap support values from the partitioned maximum-likelihood analysis. Branch lengths are from the divergence-time estimation (in millions of years ago; mya).

Fig. 30. Neobatrachia XXI subtree including Rhacophoridae (part). Values adjacent to nodes are bootstrap support values from the partitioned maximum-likelihood analysis. Branch lengths are from the divergence-time estimation (in millions of years ago; mya).

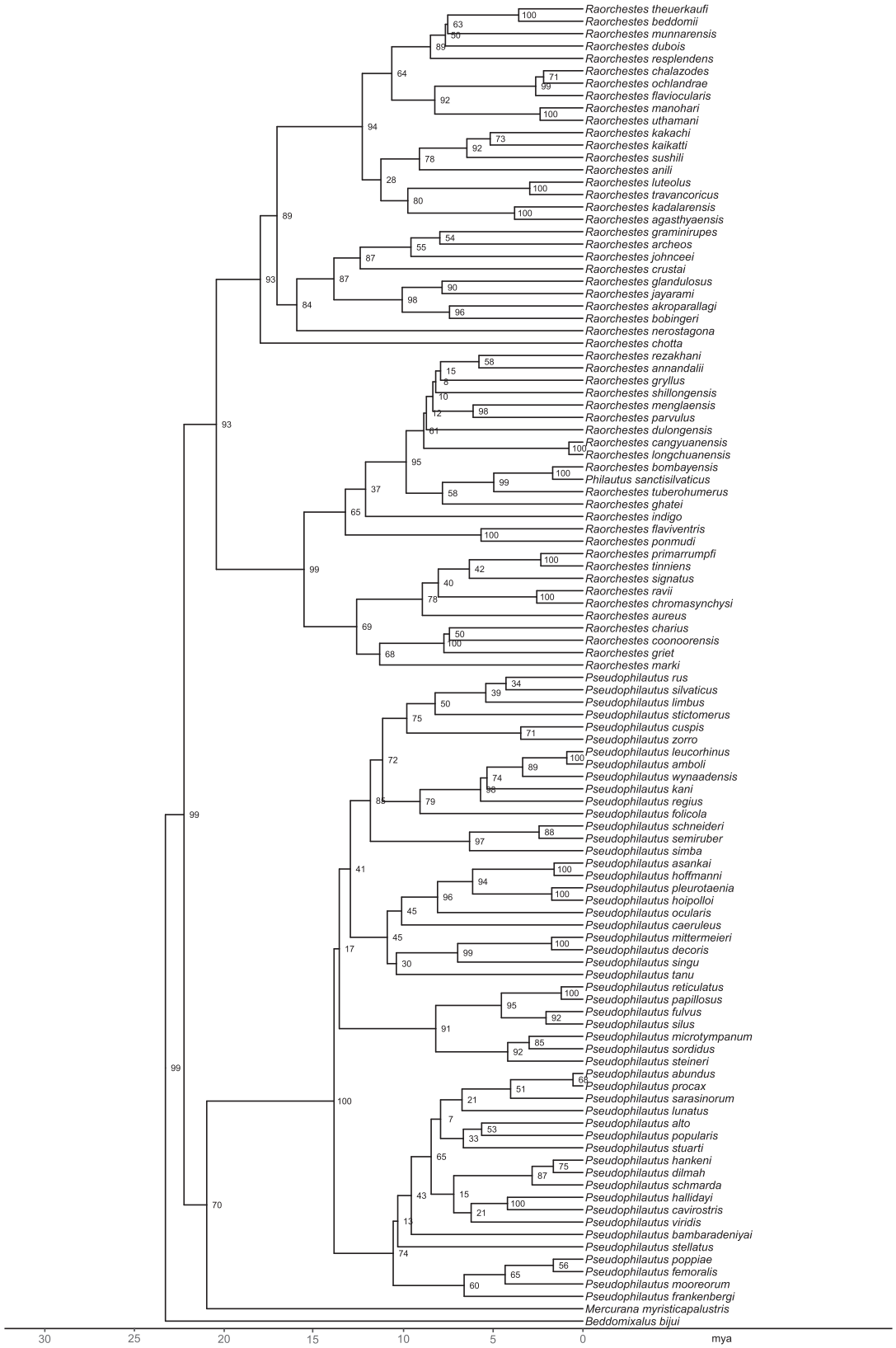

Fig. 31. Neobatrachia XXII subtree including Rhacophoridae (part). Values adjacent to nodes are bootstrap support values from the partitioned maximum-likelihood analysis. Branch lengths are from the divergence-time estimation (in millions of years ago; mya).

Fig. 32. Neobatrachia XXIII subtree including Calyptocephalellidae and Myobatrachidae (Myobatrachinae + Limnodynastinae). Values adjacent to nodes are bootstrap support values from the partitioned maximum-likelihood analysis. Branch lengths are from the divergence-time estimation (in millions of years ago; mya).

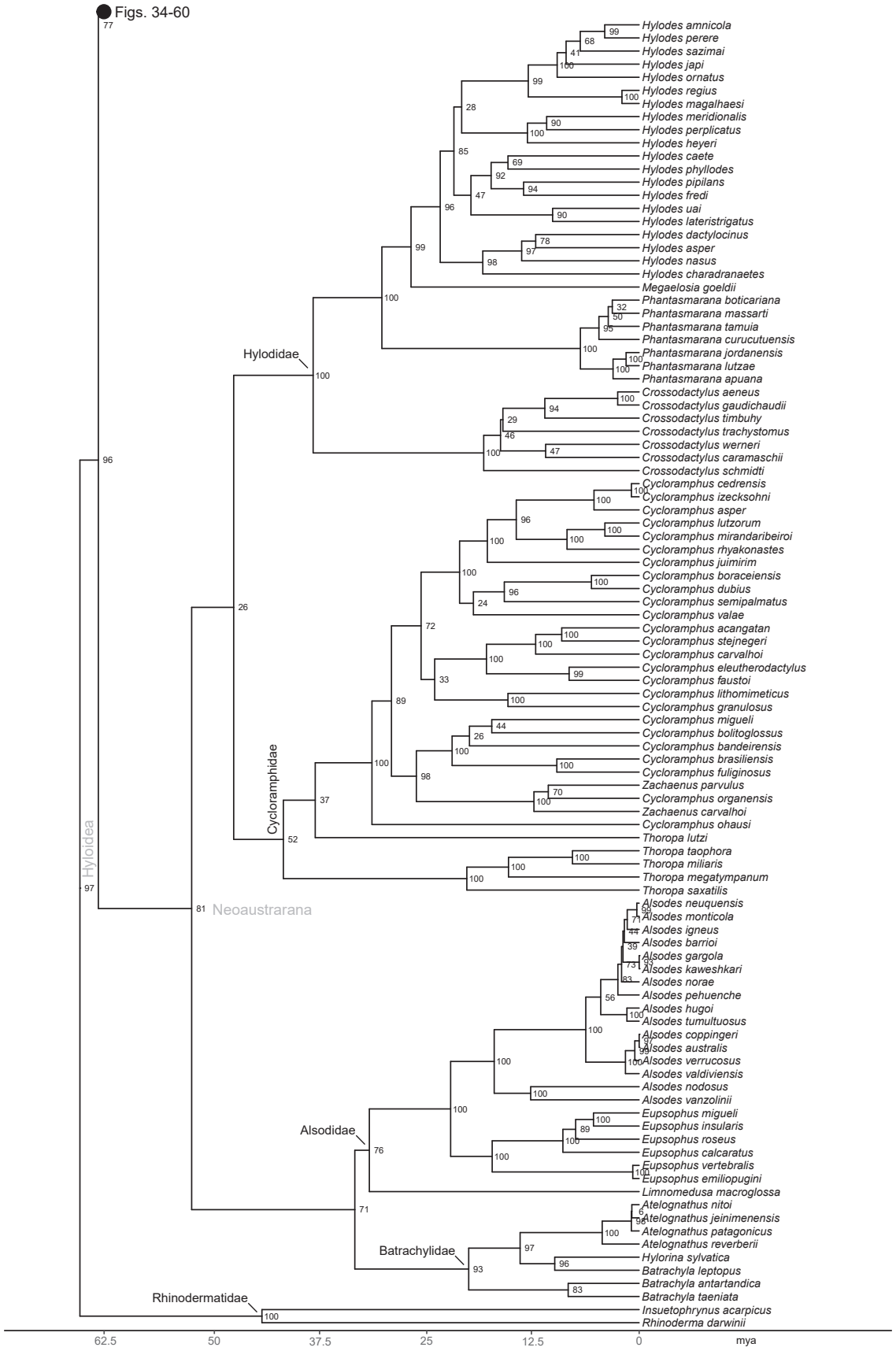

Fig. 33. Neobatrachia XXIV subtree including Rhinodermatidae and Neoaustrarana (Cycloramphidae + Hylodidae + Alsodidae + Batrachylidae). Values adjacent to nodes are bootstrap support values from the partitioned maximum-likelihood analysis. Branch lengths are from the divergence-time estimation (in millions of years ago; mya).

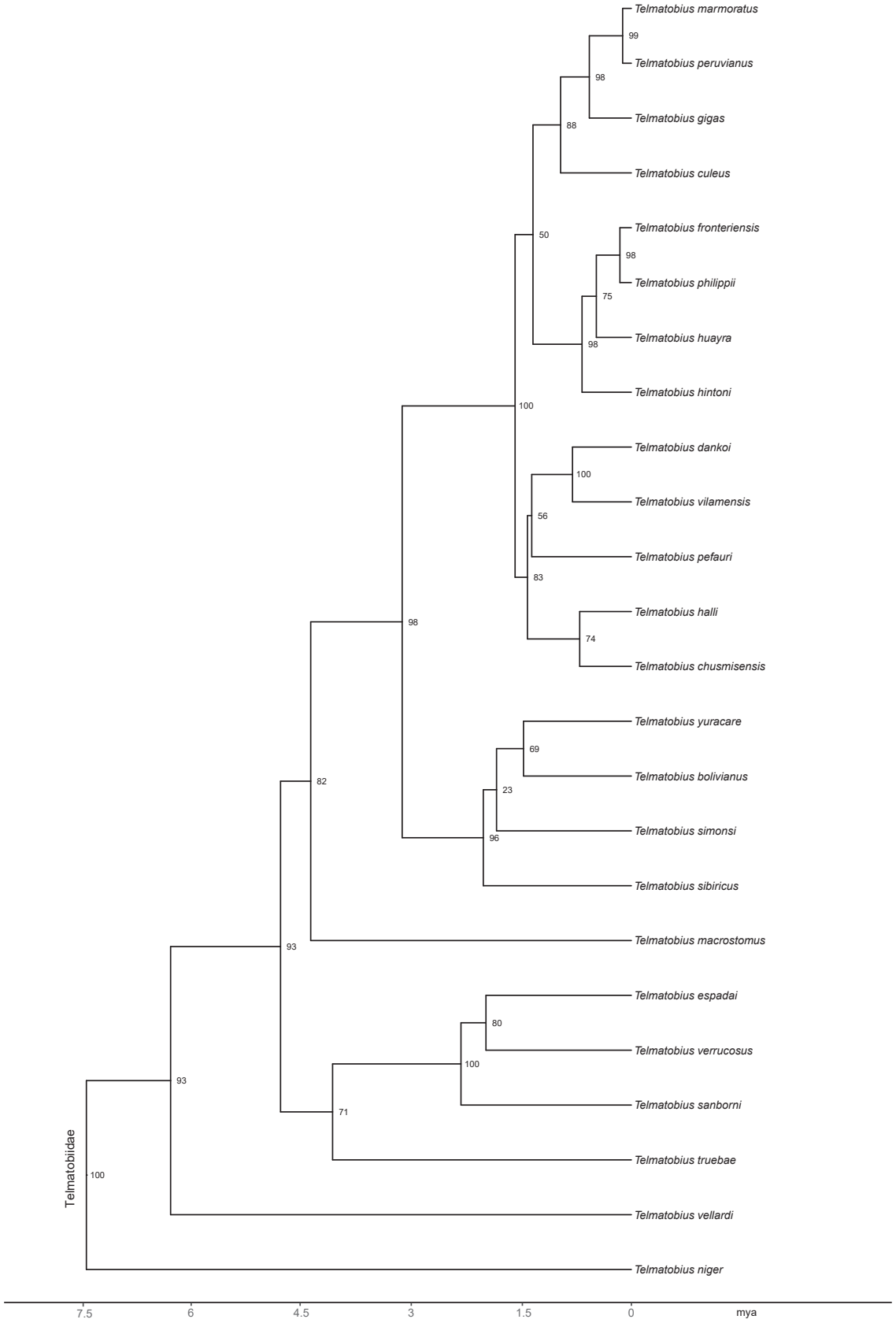

Fig. 34. Neobatrachia XXV subtree including Telmatobiidae. Values adjacent to nodes are bootstrap support values from the partitioned maximum-likelihood analysis. Branch lengths are from the divergence-time estimation (in millions of years ago; mya).

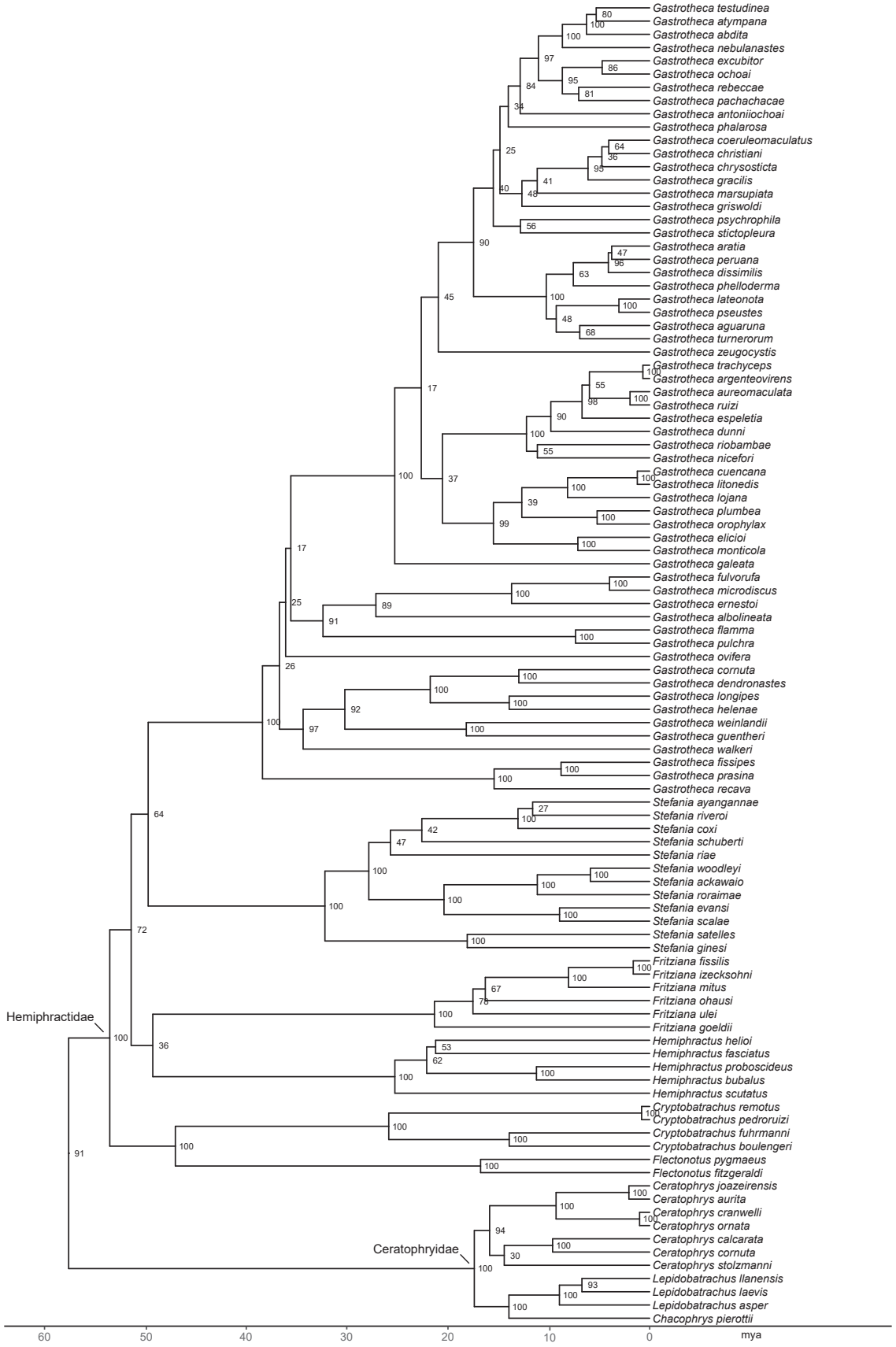

Fig. 35. Neobatrachia XXVI subtree including Ceratophryidae and Hemiphractidae. Values adjacent to nodes are bootstrap support values from the partitioned maximum-likelihood analysis. Branch lengths are from the divergence-time estimation (in millions of years ago; mya).

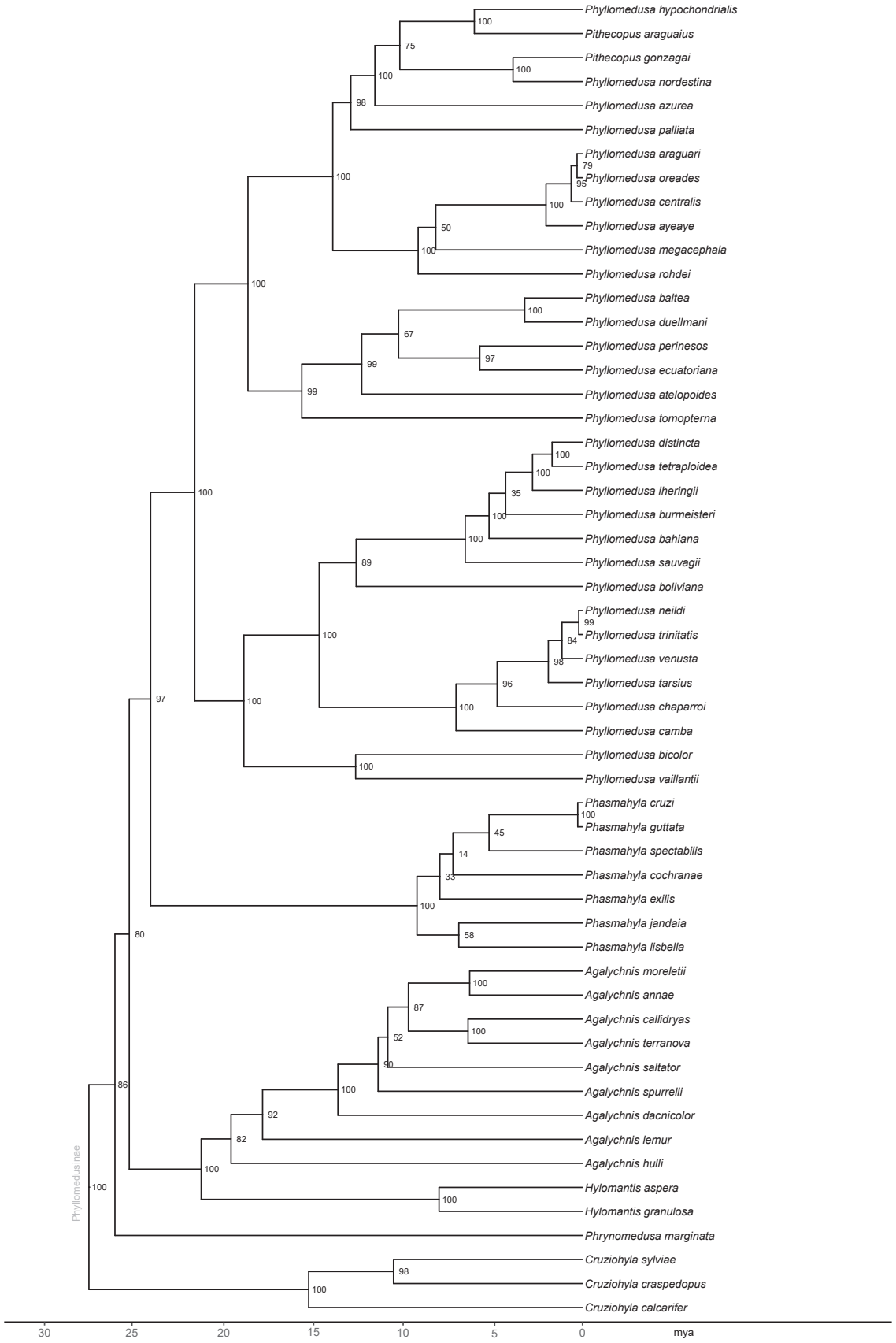

Fig. 36. Neobatrachia XXVII subtree including Hylidae (part: Phyllomedusinae). Values adjacent to nodes are bootstrap support values from the partitioned maximum-likelihood analysis. Branch lengths are from the divergence-time estimation (in millions of years ago; mya).

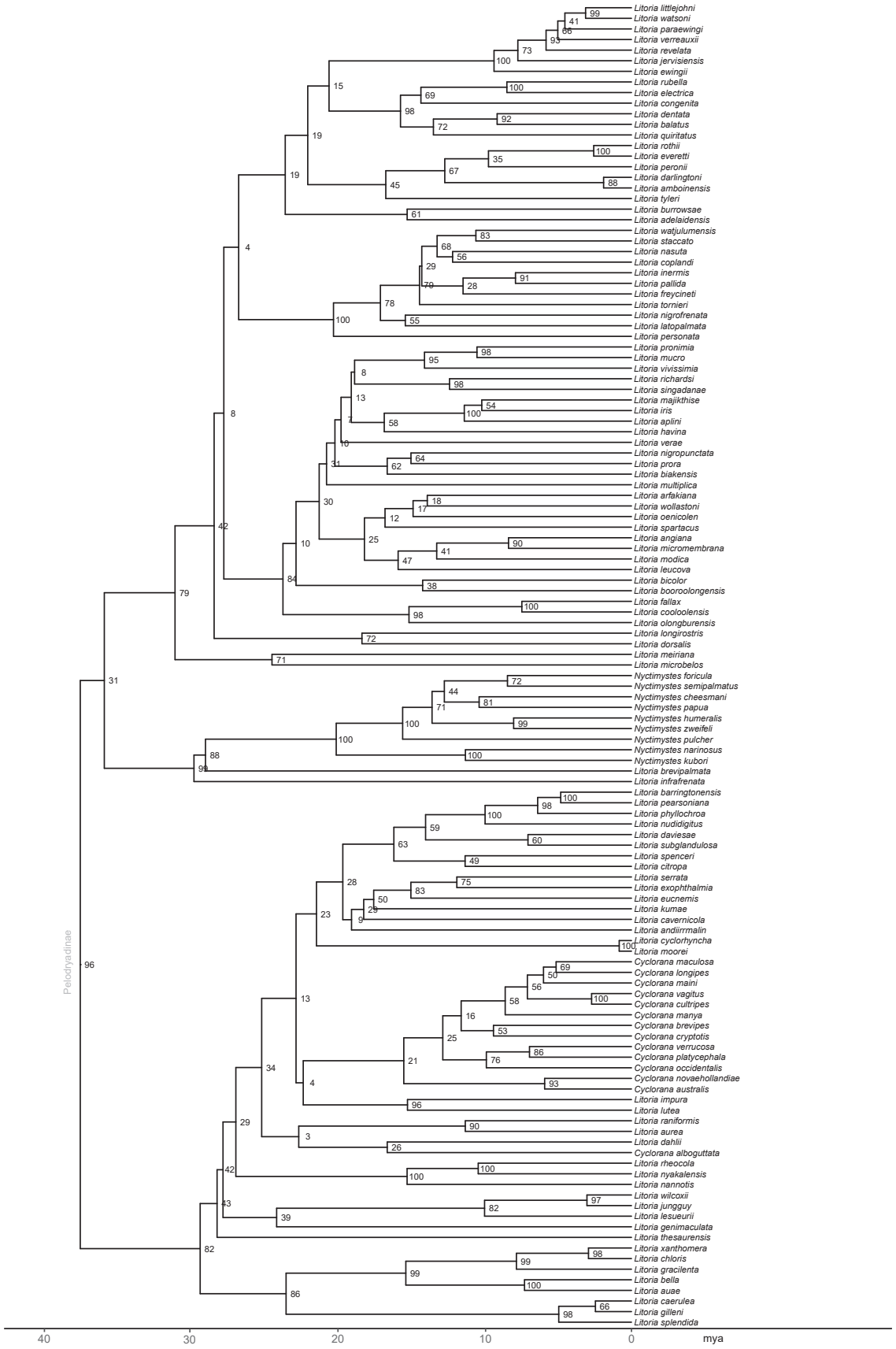

Fig. 37. Neobatrachia XXVIII subtree including Hylidae (part: Pelodryadinae). Values adjacent to nodes are bootstrap support values from the partitioned maximum-likelihood analysis. Branch lengths are from the divergence-time estimation (in millions of years ago; mya).

Fig. 38. Neobatrachia XXIX subtree including Hylidae (part: Hylinae [part]). Values adjacent to nodes are bootstrap support values from the partitioned maximum-likelihood analysis. Branch lengths are from the divergence-time estimation (in millions of years ago; mya).

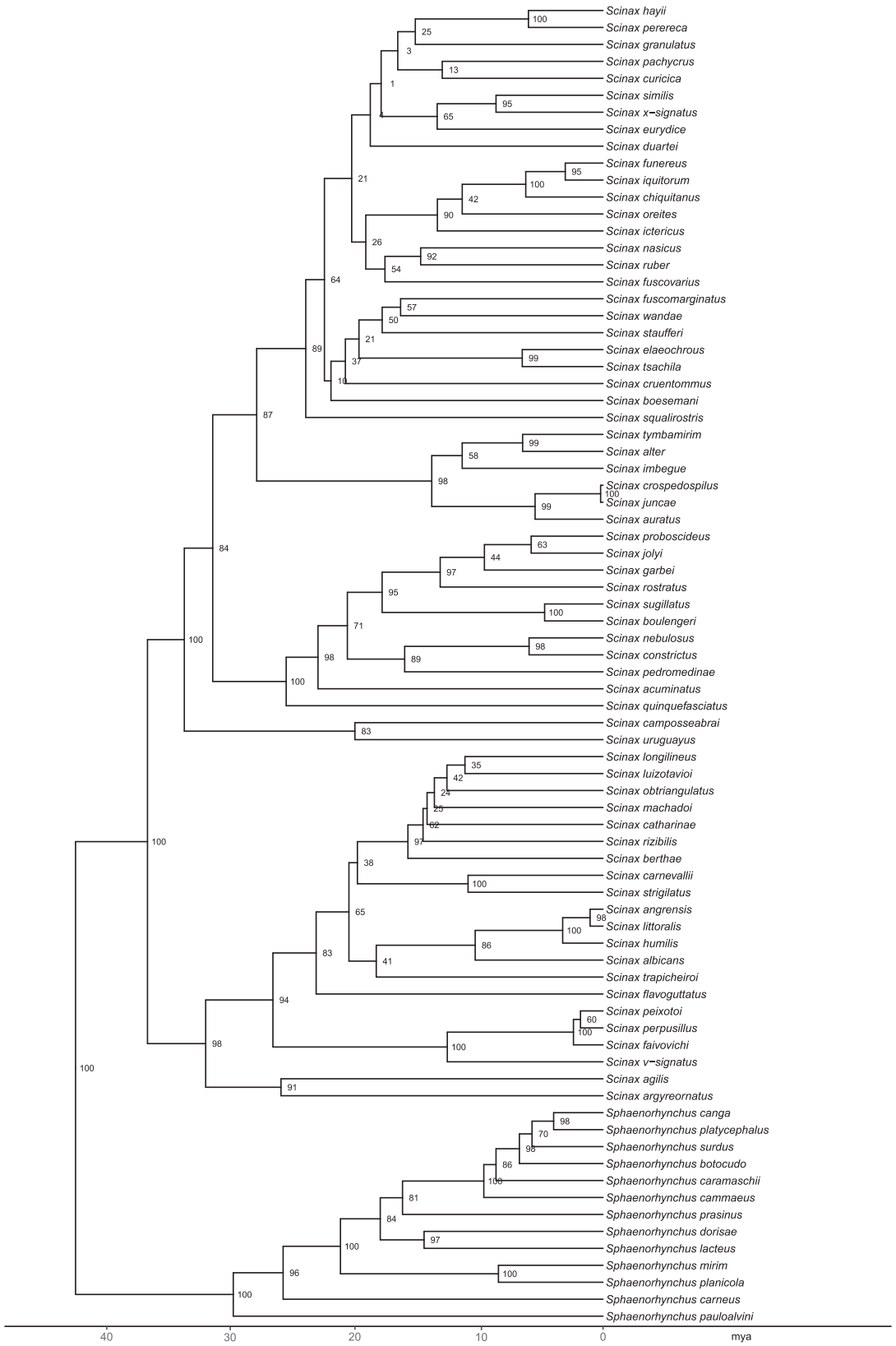

Fig. 39. Neobatrachia XXX subtree including Hylidae (part: Hylinae [part]). Values adjacent to nodes are bootstrap support values from the partitioned maximum-likelihood analysis. Branch lengths are from the divergence-time estimation (in millions of years ago; mya).

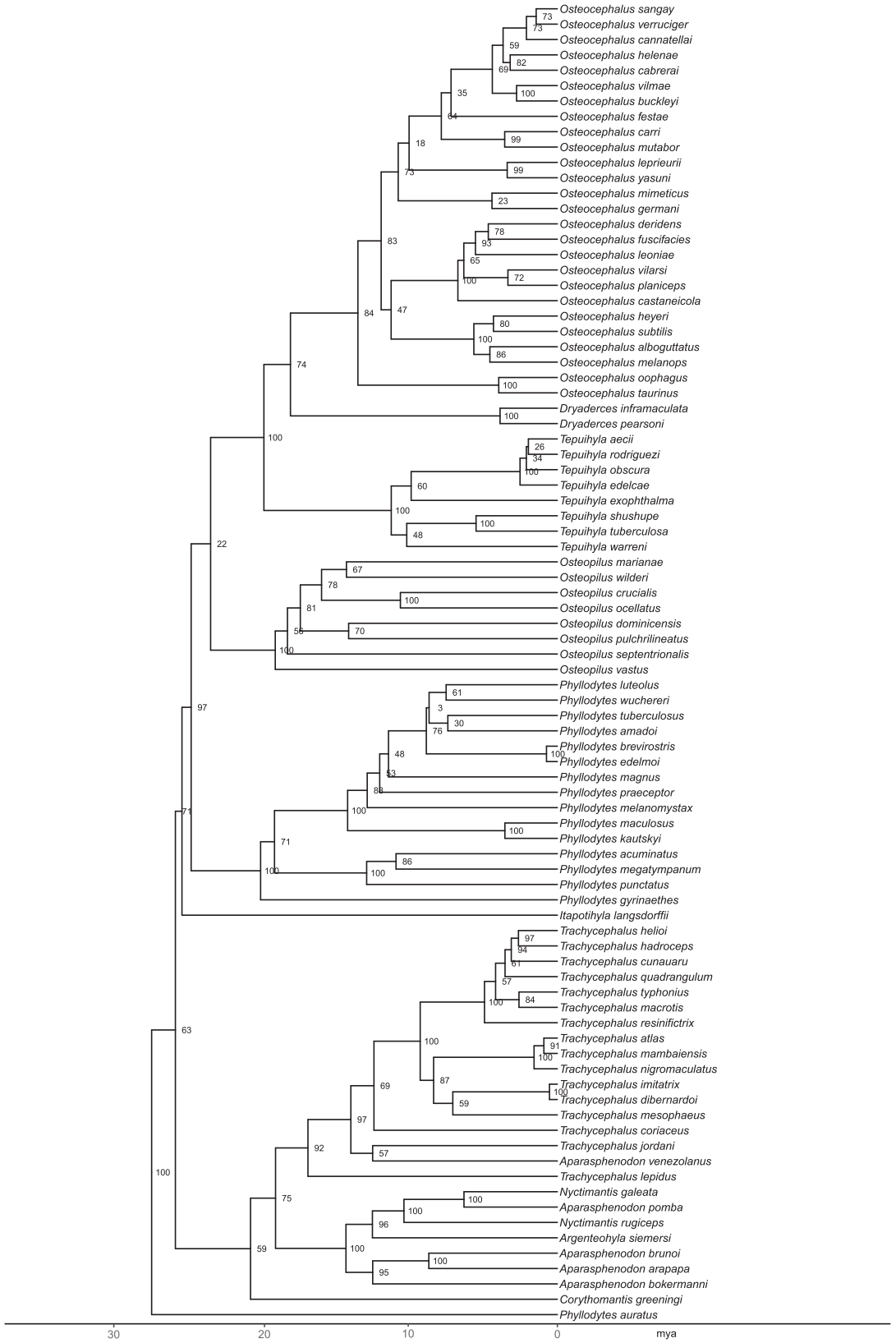

Fig. 40. Neobatrachia XXXI subtree including Hylidae (part: Hylinae [part]). Values adjacent to nodes are bootstrap support values from the partitioned maximum-likelihood analysis. Branch lengths are from the divergence-time estimation (in millions of years ago; mya).

Fig. 41. Neobatrachia XXXII subtree including Hylidae (part: Hylinae [part]). Values adjacent to nodes are bootstrap support values from the partitioned maximum-likelihood analysis. Branch lengths are from the divergence-time estimation (in millions of years ago; mya).

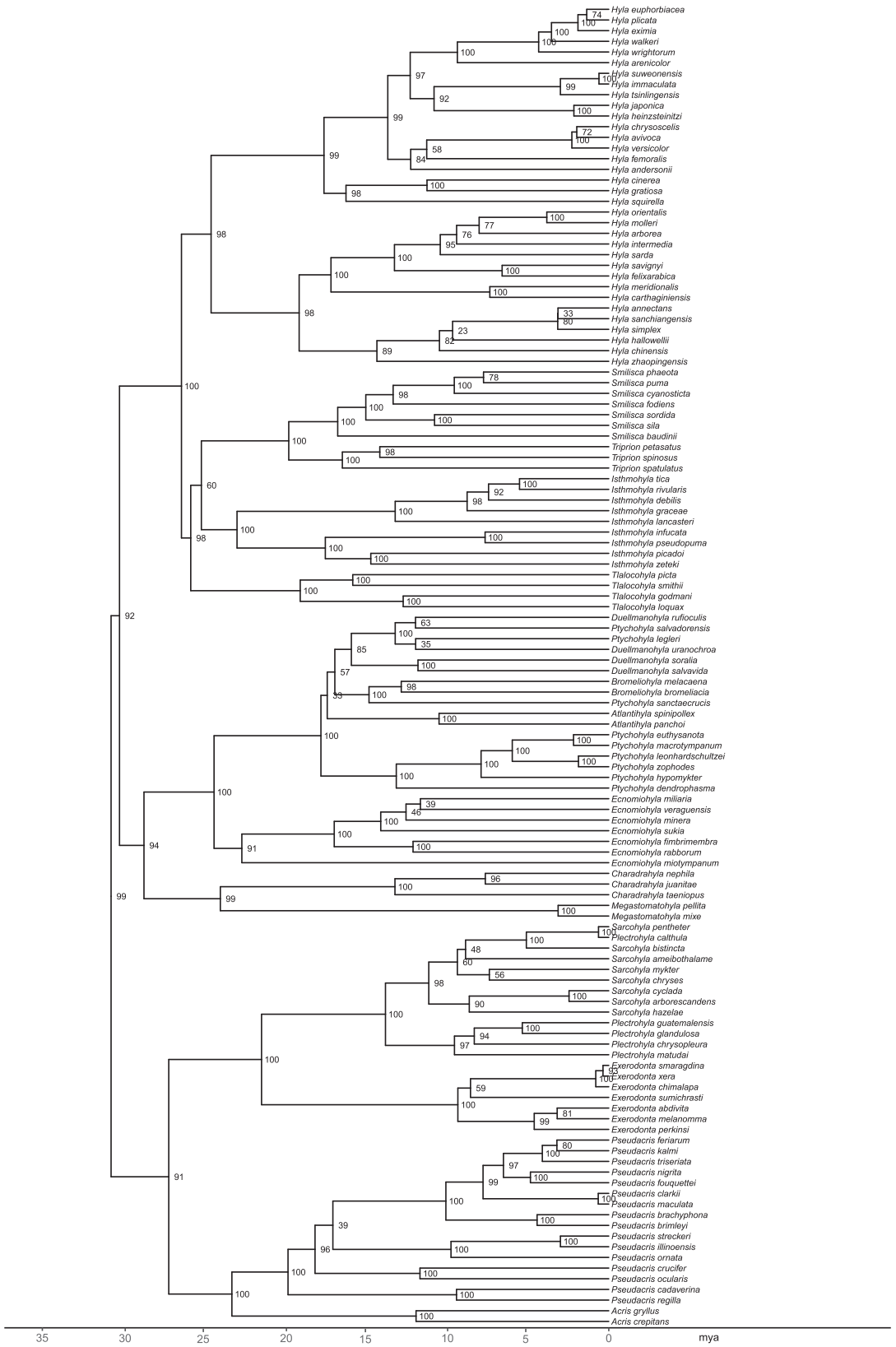

Fig. 42. Neobatrachia XXXIII subtree including Hylidae (part: Hylinae [part]). Values adjacent to nodes are bootstrap support values from the partitioned maximum-likelihood analysis. Branch lengths are from the divergence-time estimation (in millions of years ago; mya).

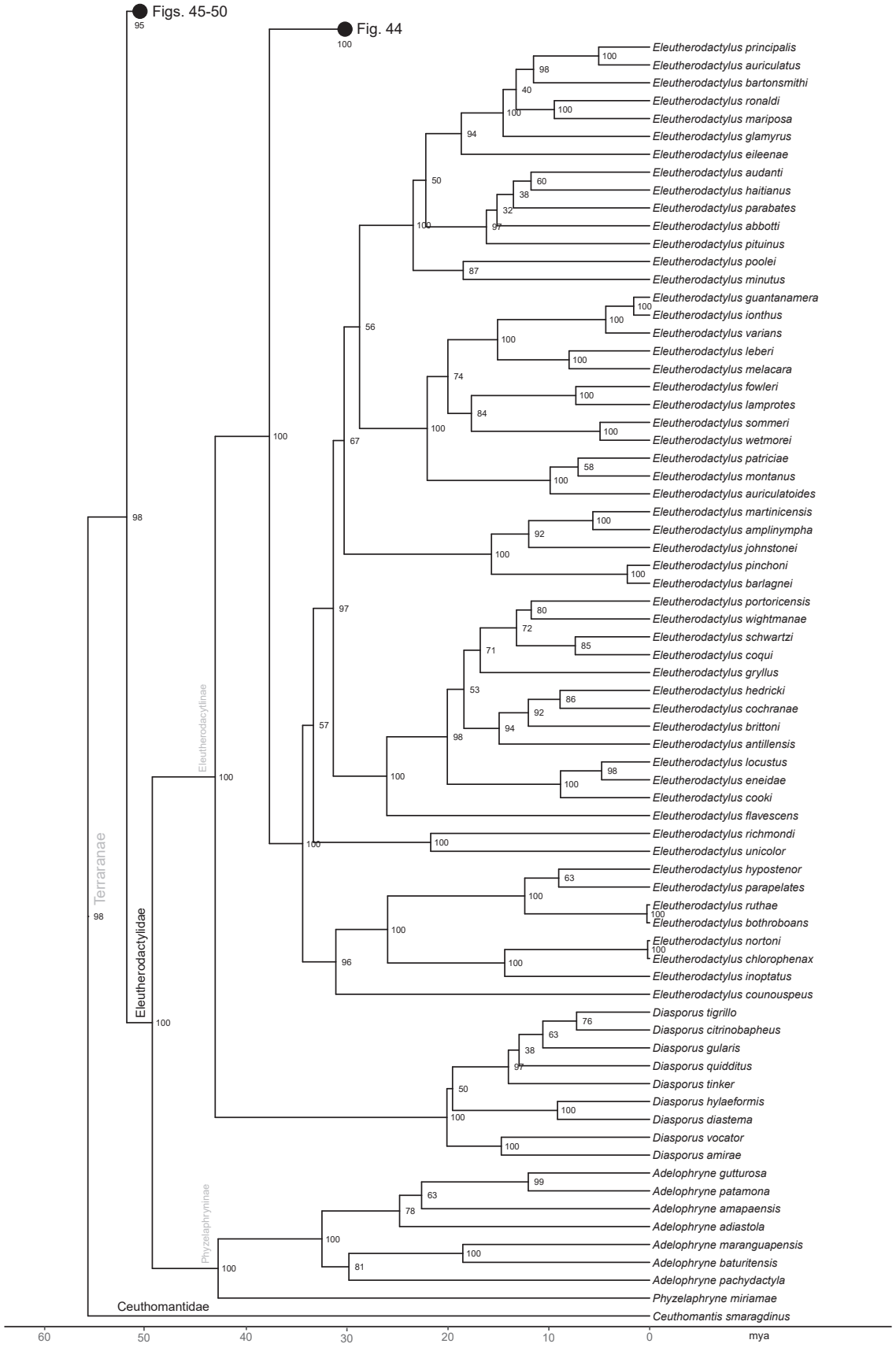

Fig. 43. Neobatrachia XXXVI subtree including Ceuthomantidae and Eleutherodactylidae (part). Values adjacent to nodes are bootstrap support values from the partitioned maximum-likelihood analysis. Branch lengths are from the divergence-time estimation (in millions of years ago; mya).

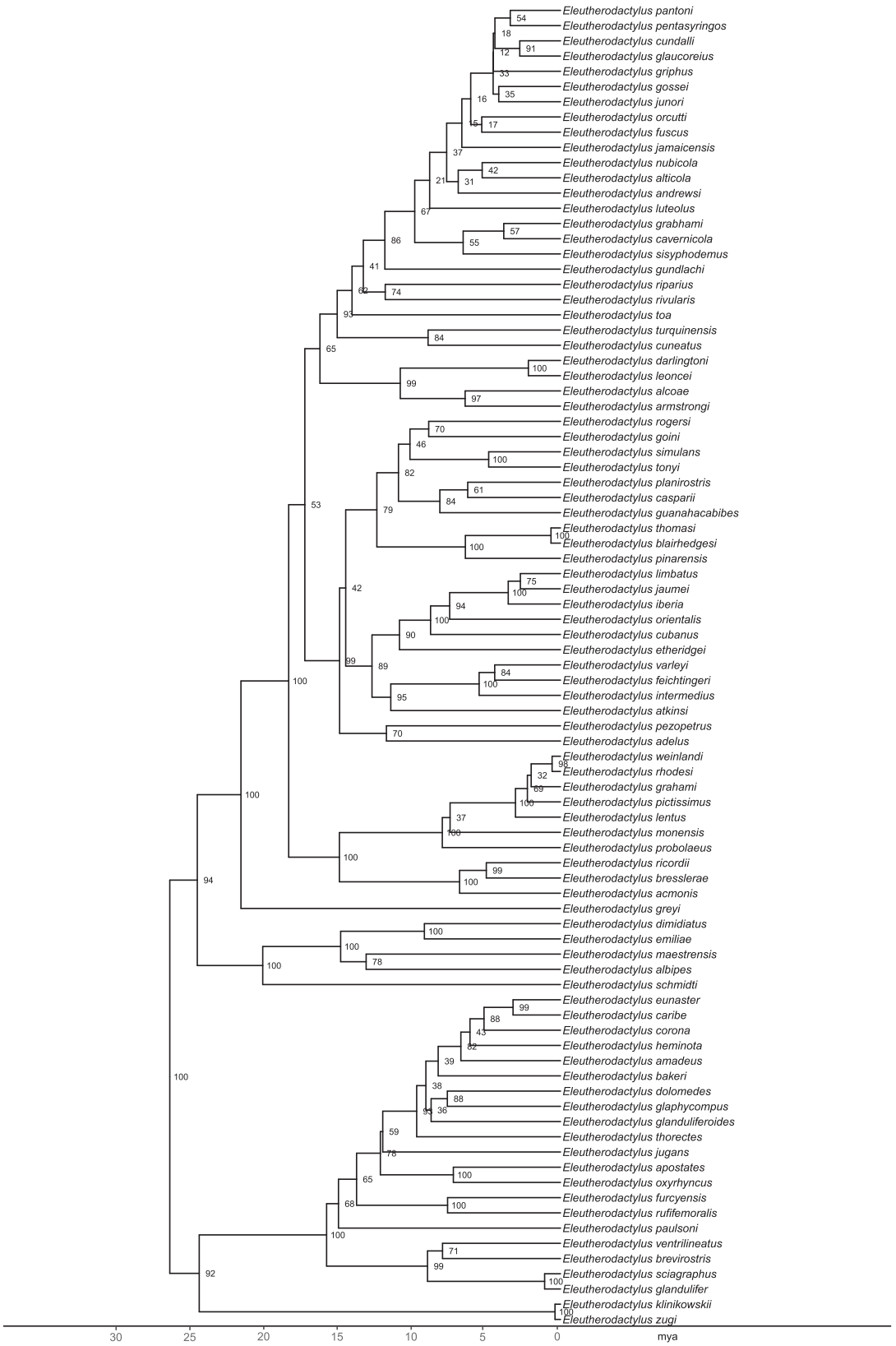

Fig. 44. Neobatrachia XXXVII subtree including Eleutherodactylidae (part). Values adjacent to nodes are bootstrap support values from the partitioned maximum-likelihood analysis. Branch lengths are from the divergence-time estimation (in millions of years ago; mya).

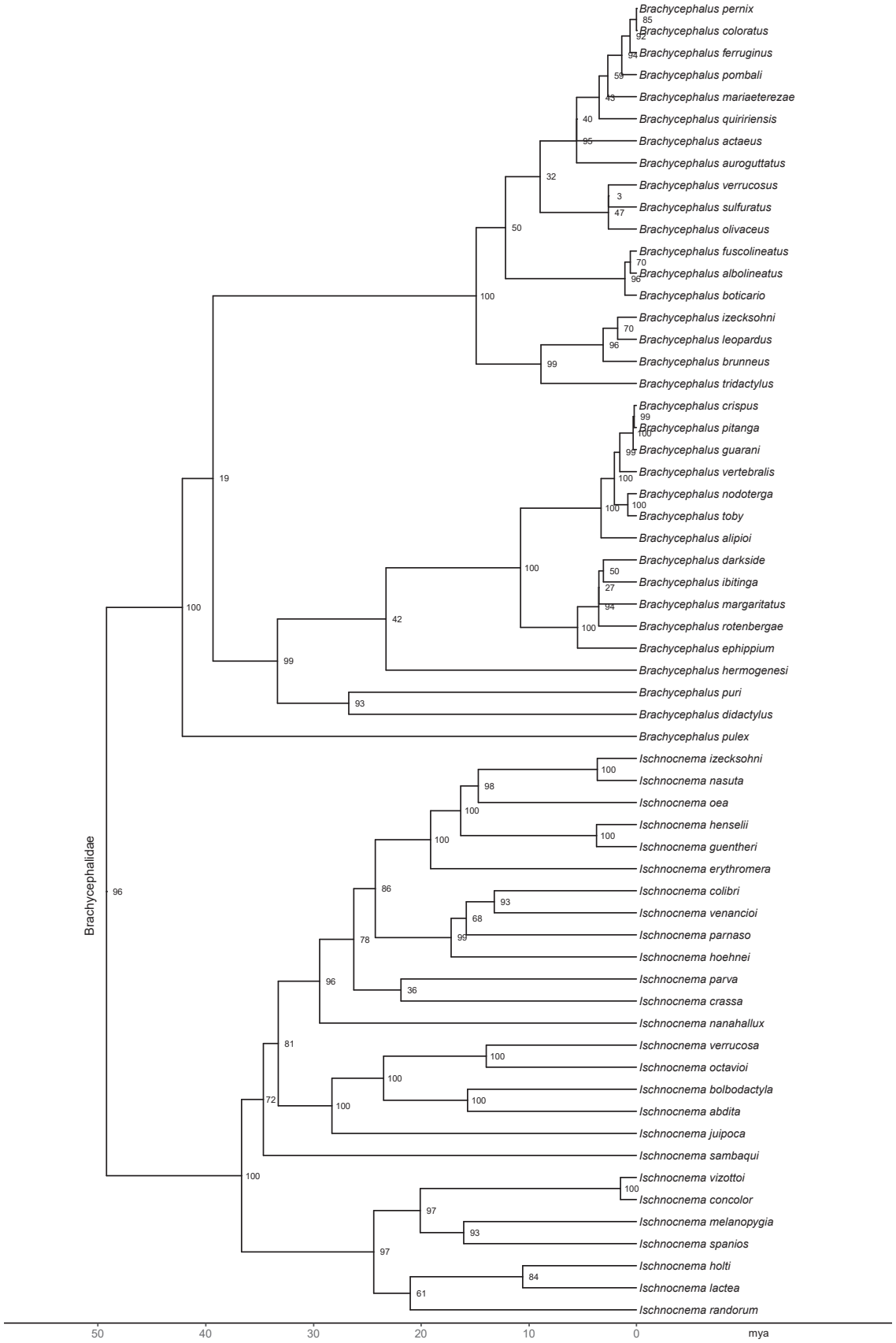

Fig. 45. Neobatrachia XXXVIII subtree including Brachycephalidae. Values adjacent to nodes are bootstrap support values from the partitioned maximum-likelihood analysis. Branch lengths are from the divergence-time estimation (in millions of years ago; mya).

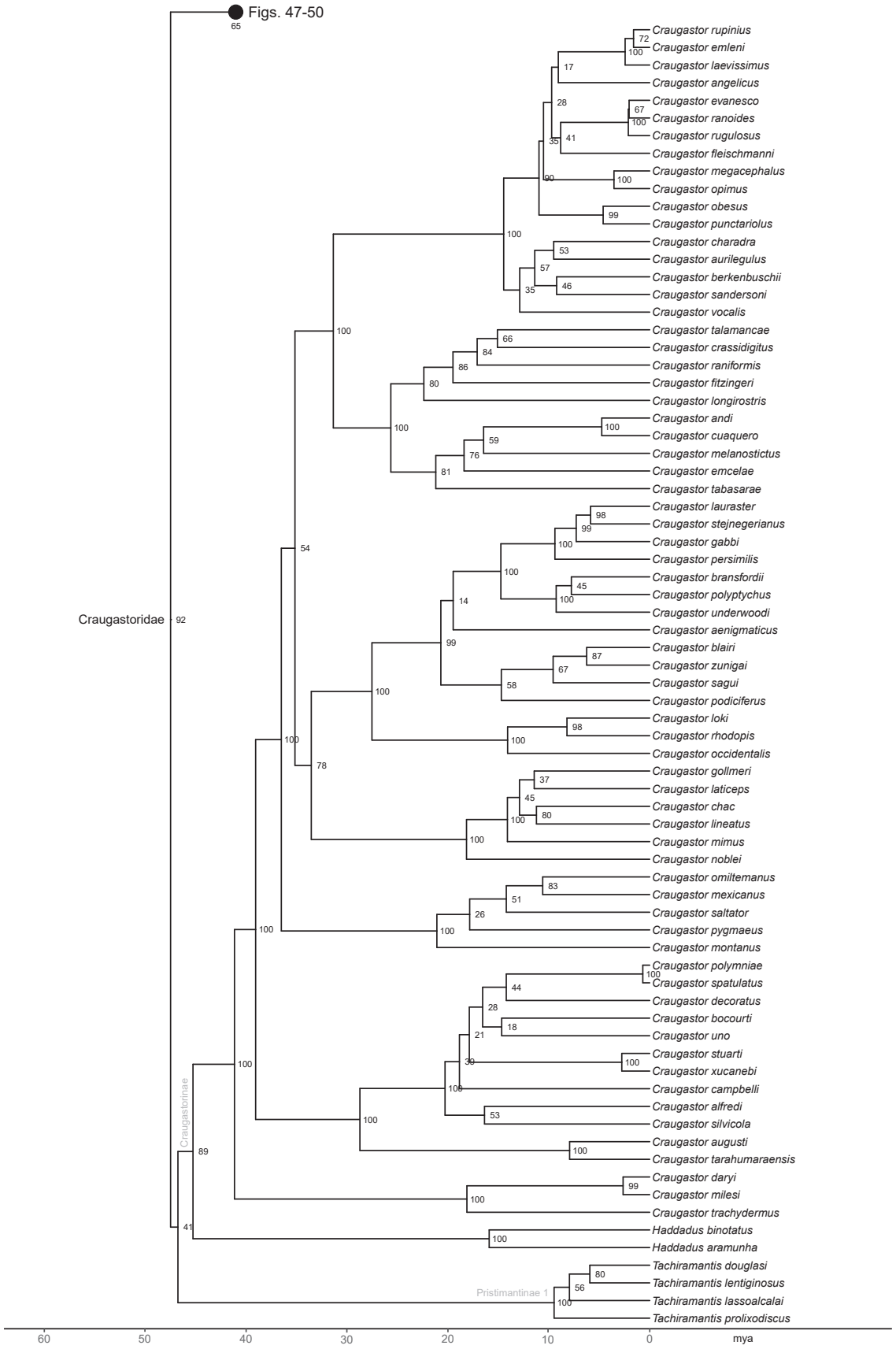

Fig. 46. Neobatrachia XXXIX subtree including Craugastoridae (part). Values adjacent to nodes are bootstrap support values from the partitioned maximum-likelihood analysis. Branch lengths are from the divergence-time estimation (in millions of years ago; mya).

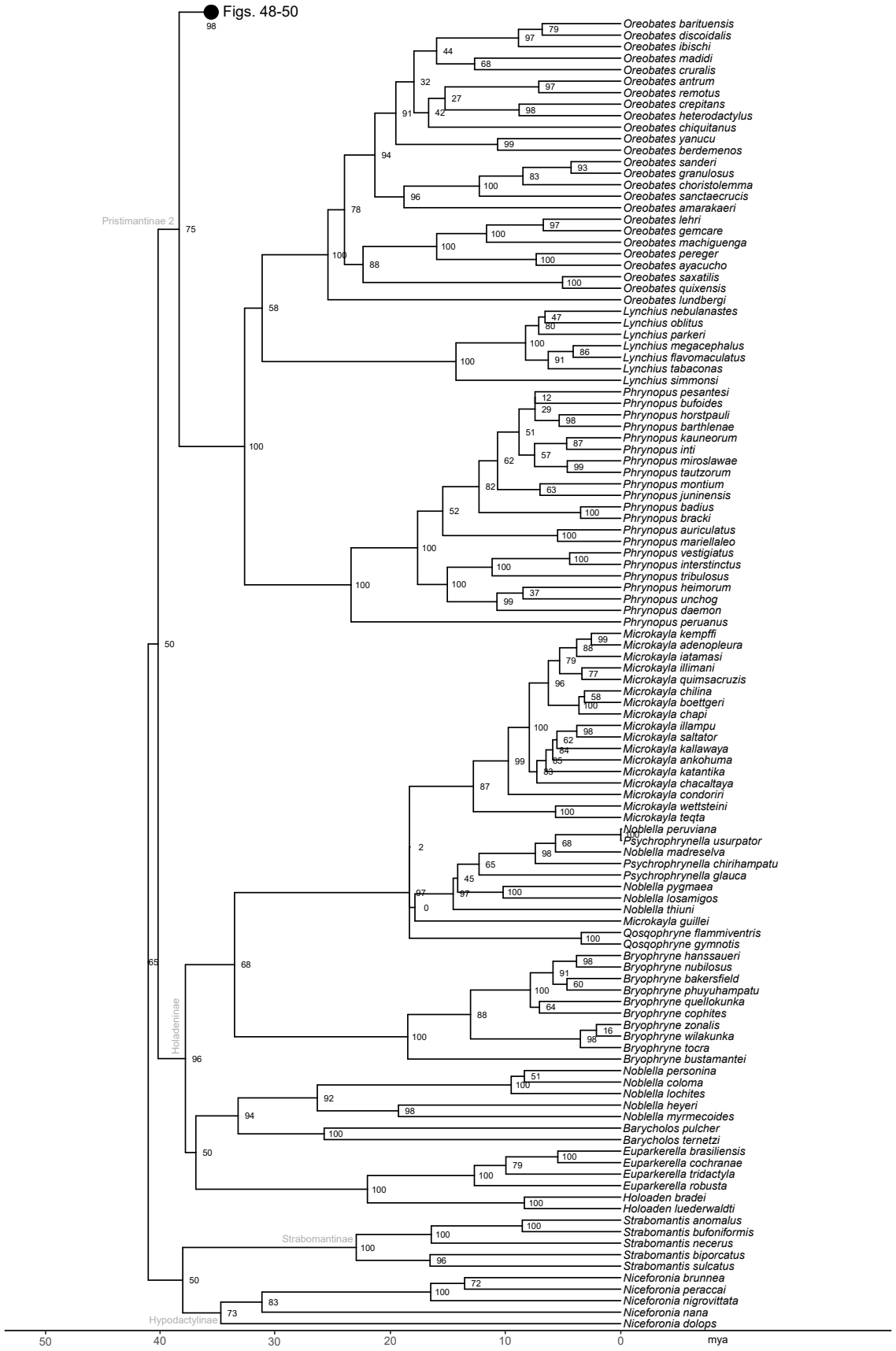

Fig. 47. Neobatrachia XL subtree including Craugastoridae (part). Values adjacent to nodes are bootstrap support values from the partitioned maximum-likelihood analysis. Branch lengths are from the divergence-time estimation (in millions of years ago; mya).

Fig. 48. Neobatrachia XLI subtree including Craugastoridae (part). Values adjacent to nodes are bootstrap support values from the partitioned maximum-likelihood analysis. Branch lengths are from the divergence-time estimation (in millions of years ago; mya).

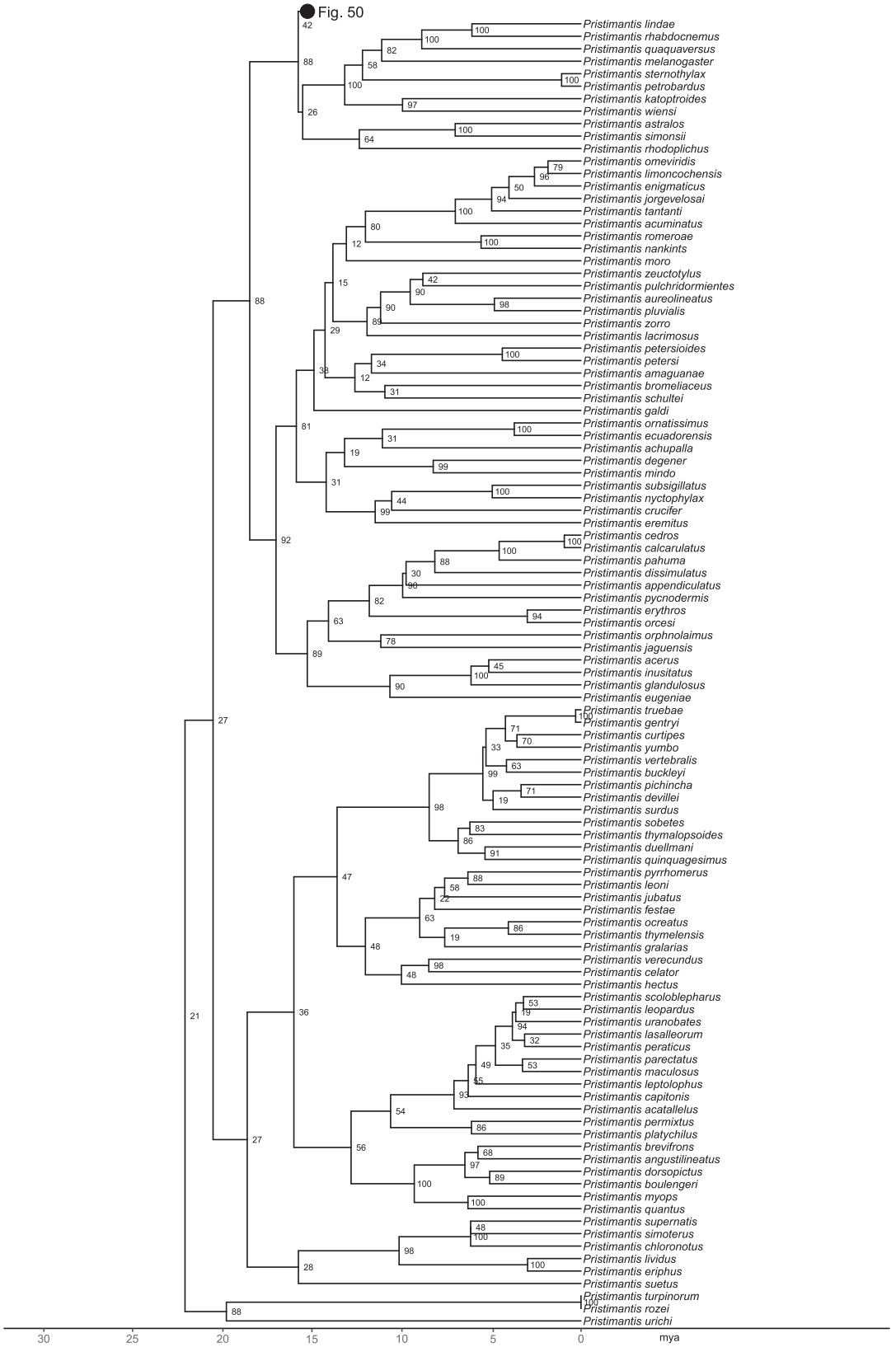

Fig. 49. Neobatrachia XLII subtree including Craugastoridae (part). Values adjacent to nodes are bootstrap support values from the partitioned maximum-likelihood analysis. Branch lengths are from the divergence-time estimation (in millions of years ago; mya).

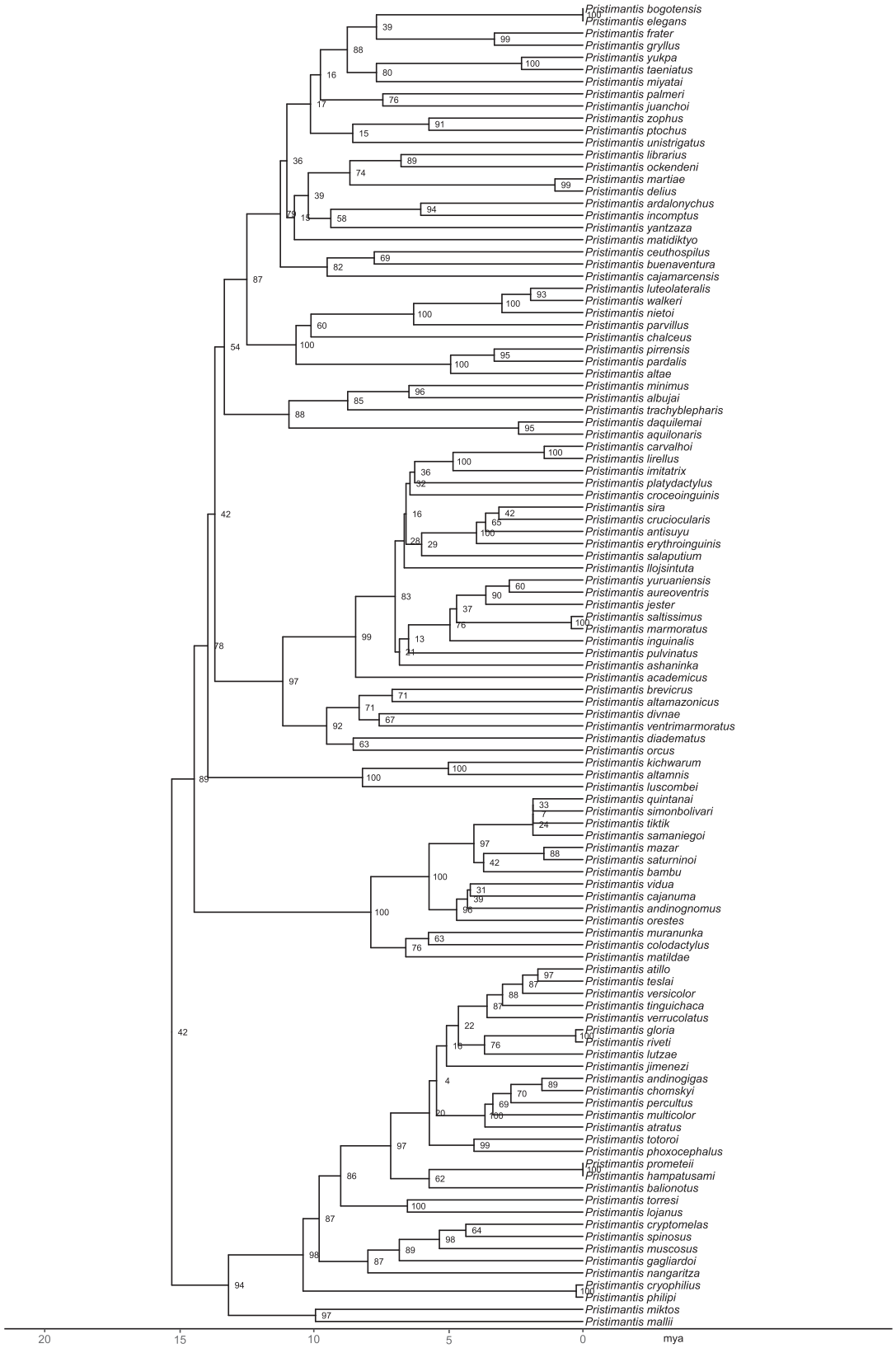

Fig. 50. Neobatrachia XLIII subtree including Craugastoridae (part). Values adjacent to nodes are bootstrap support values from the partitioned maximum-likelihood analysis. Branch lengths are from the divergence-time estimation (in millions of years ago; mya).

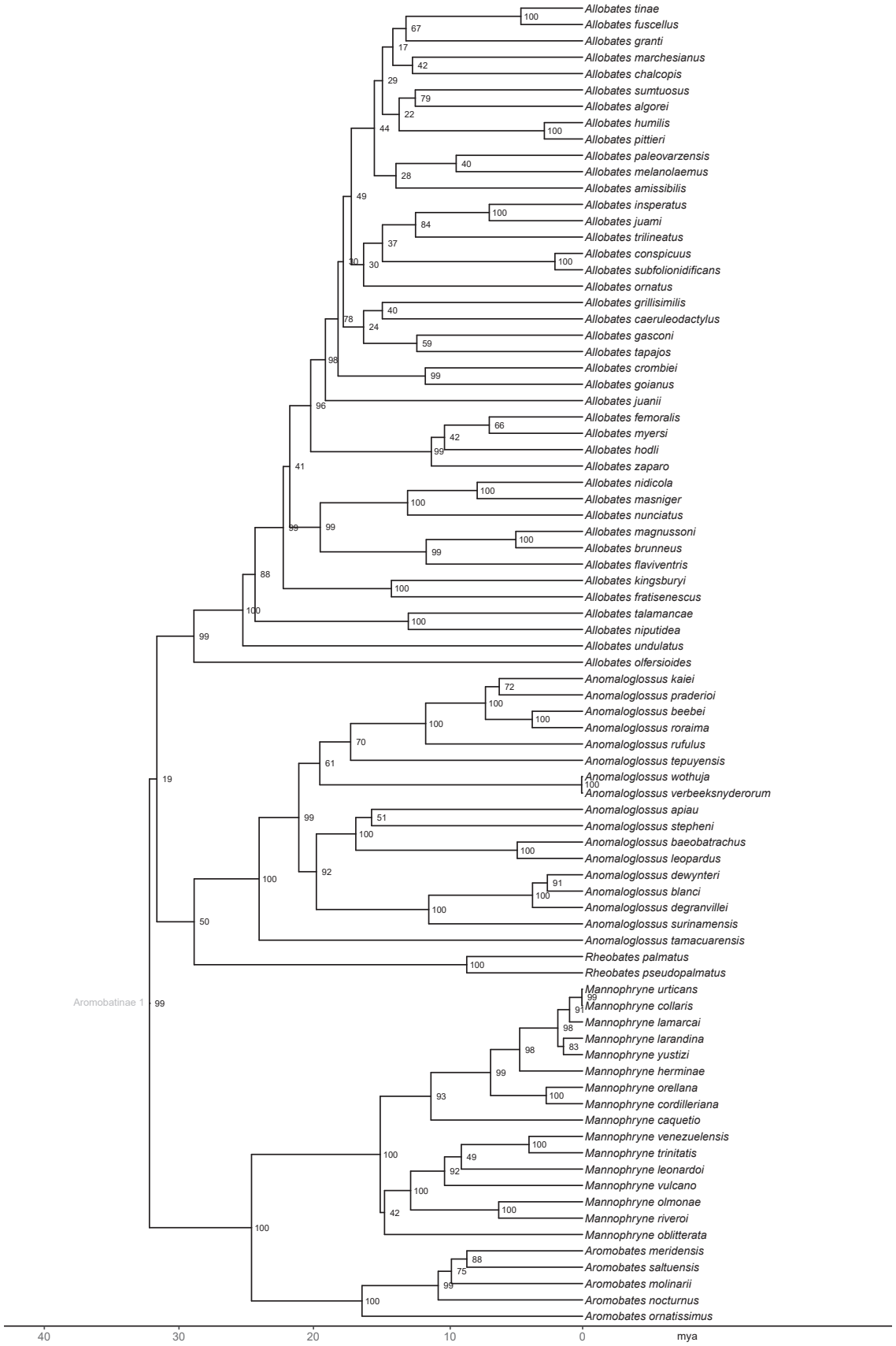

Fig. 51. Neobatrachia XXXIV subtree including Dendrobatidae (part: Aromobatinae). Values adjacent to nodes are bootstrap support values from the partitioned maximum-likelihood analysis. Branch lengths are from the divergence-time estimation (in millions of years ago; mya).

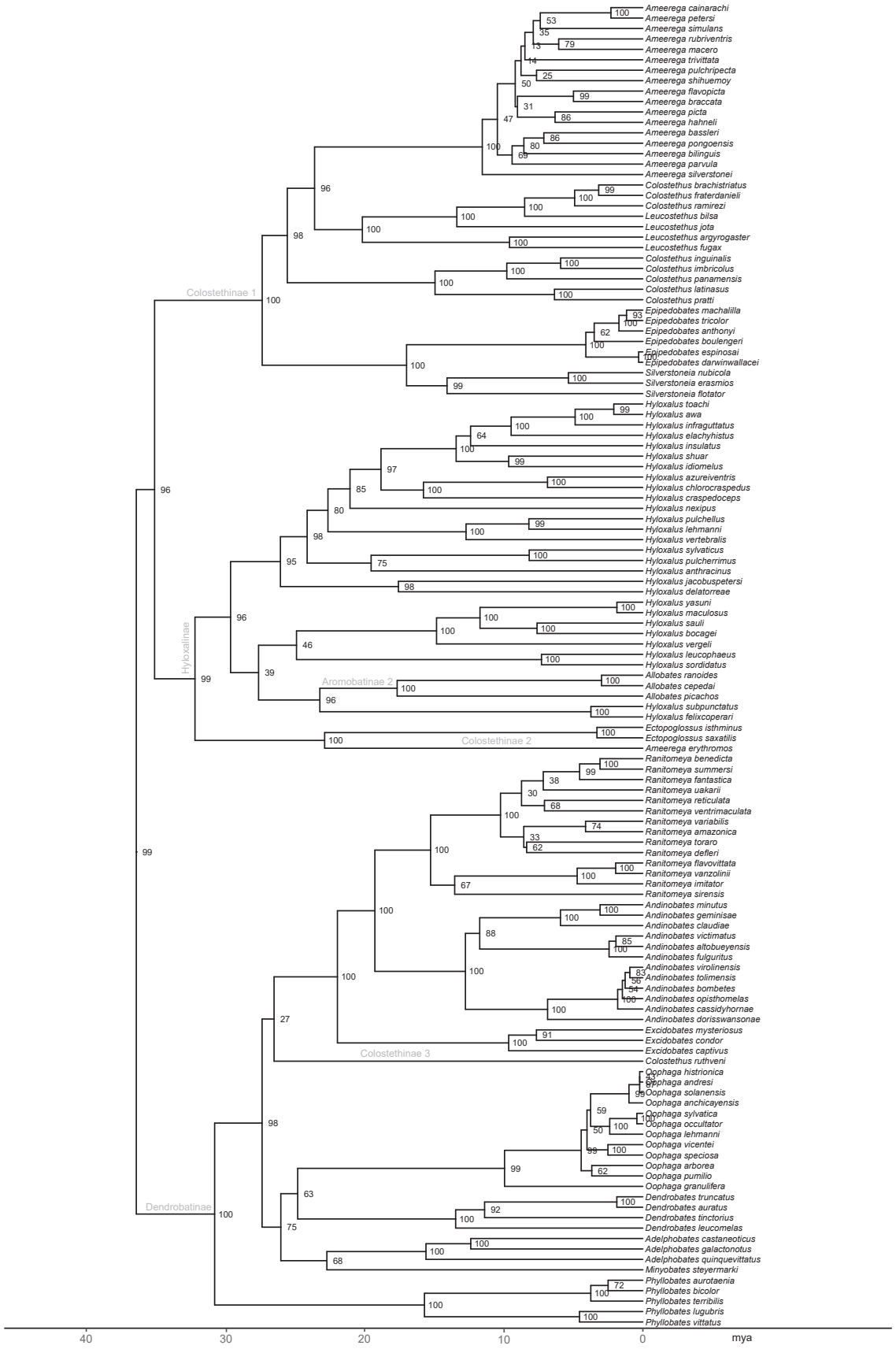

Fig. 52. Neobatrachia XXXV subtree including Dendrobatidae (part: Dendrobatinae). Values adjacent to nodes are bootstrap support values from the partitioned maximum-likelihood analysis. Branch lengths are from the divergence-time estimation (in millions of years ago; mya).

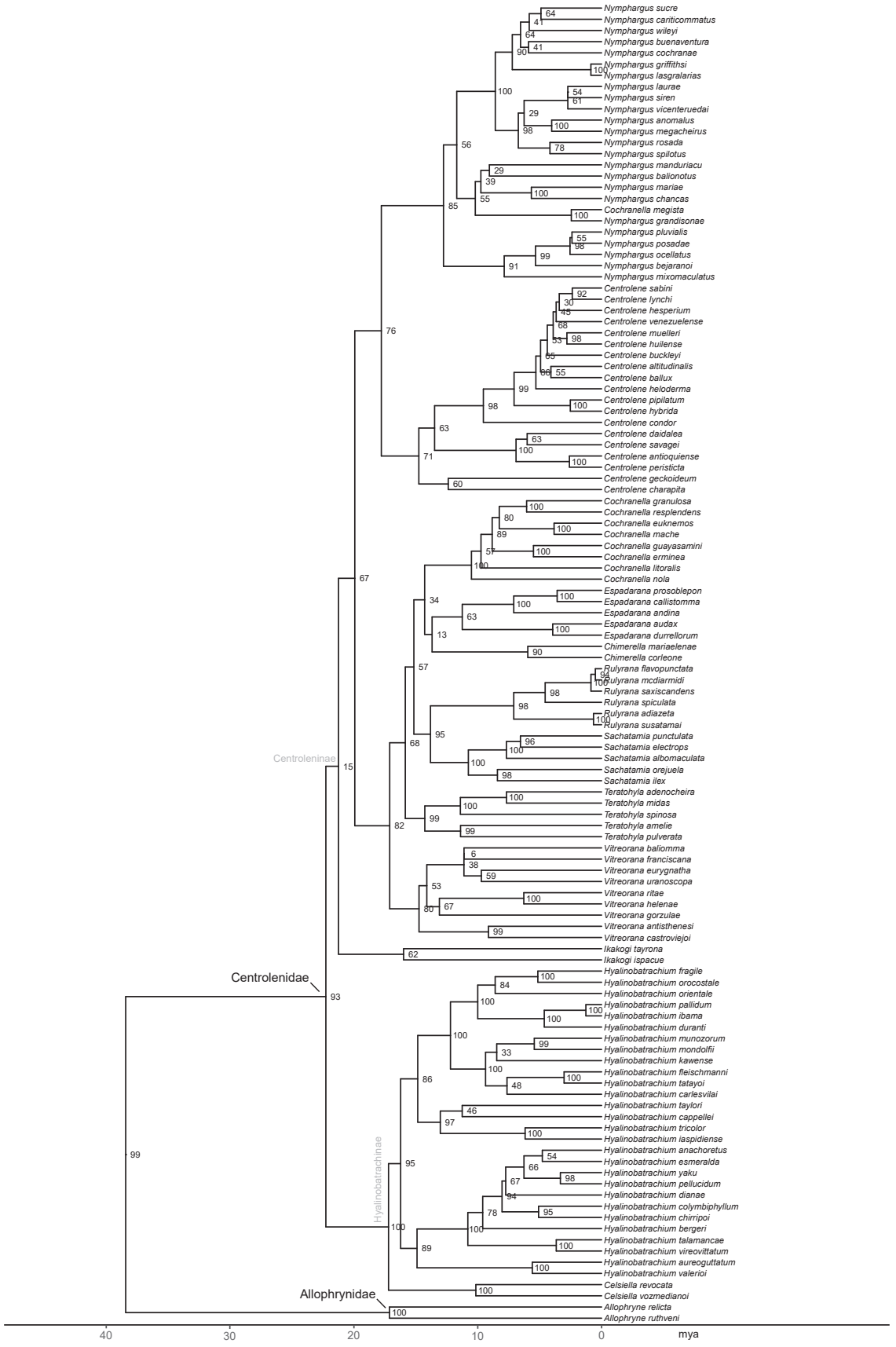

Fig. 53. Neobatrachia XLV subtree including Allophrynidae and Centrolenidae. Values adjacent to nodes are bootstrap support values from the partitioned maximum-likelihood analysis. Branch lengths are from the divergence-time estimation (in millions of years ago; mya).

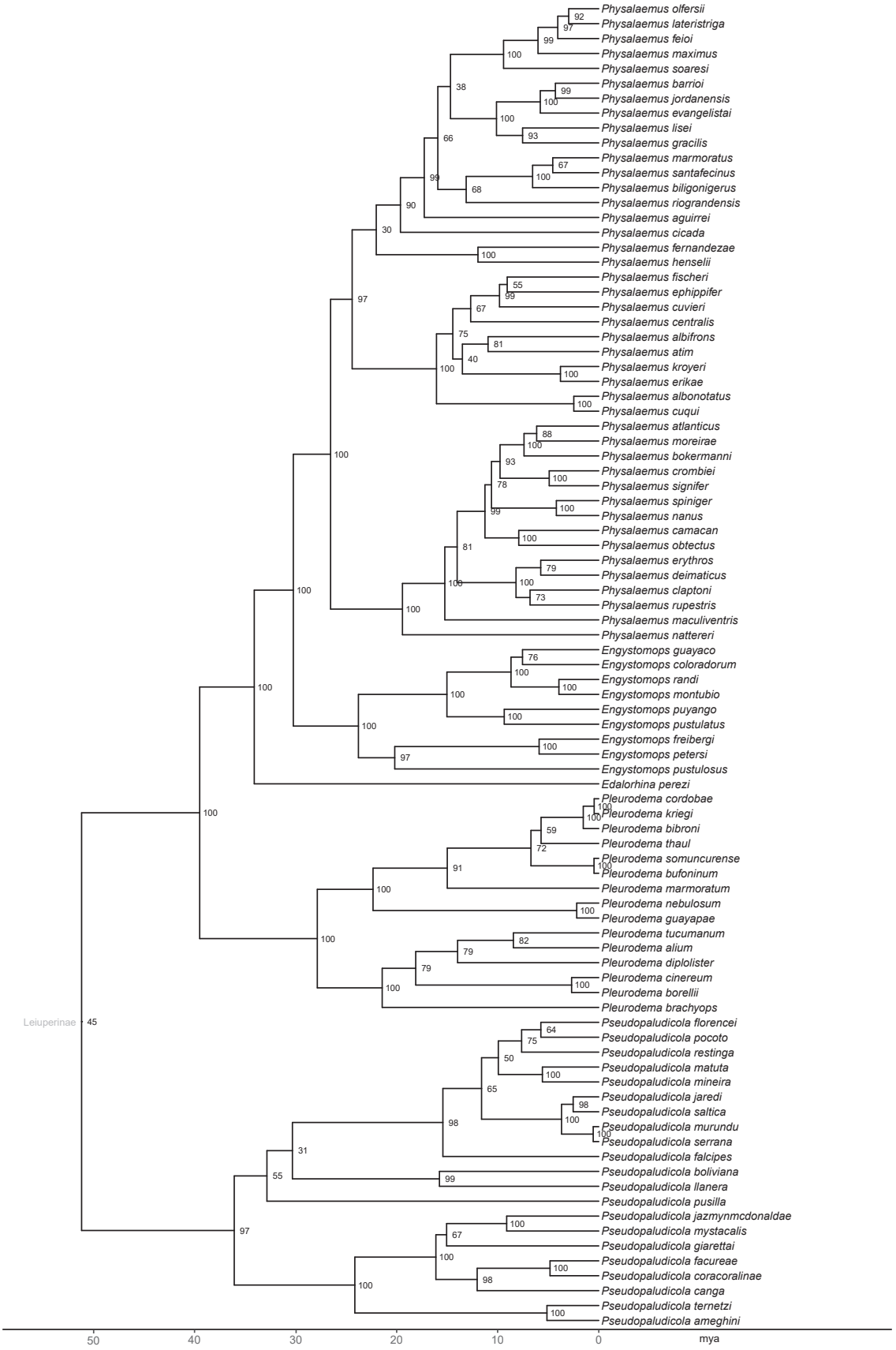

Fig. 54. Neobatrachia XLVI subtree including Leptodactylidae (part: Leiuperinae). Values adjacent to nodes are bootstrap support values from the partitioned maximum-likelihood analysis. Branch lengths are from the divergence-time estimation (in millions of years ago; mya).

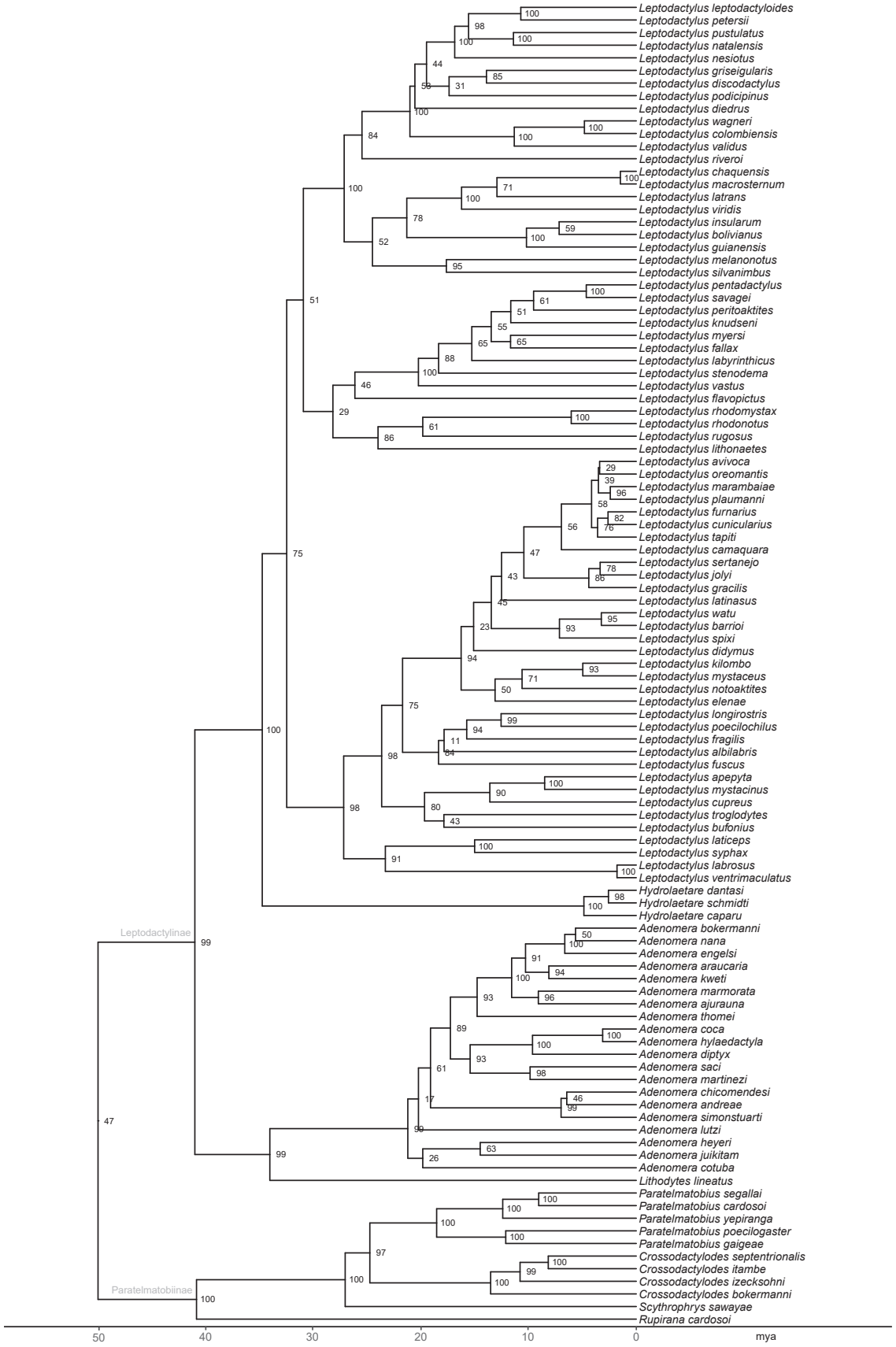

Fig. 55. Neobatrachia XLVII subtree including Leptodactylidae (part: Leptodactylinae and Paratelmatobiinae). Values adjacent to nodes are bootstrap support values from the partitioned maximum-likelihood analysis. Branch lengths are from the divergence-time estimation (in millions of years ago; mya).

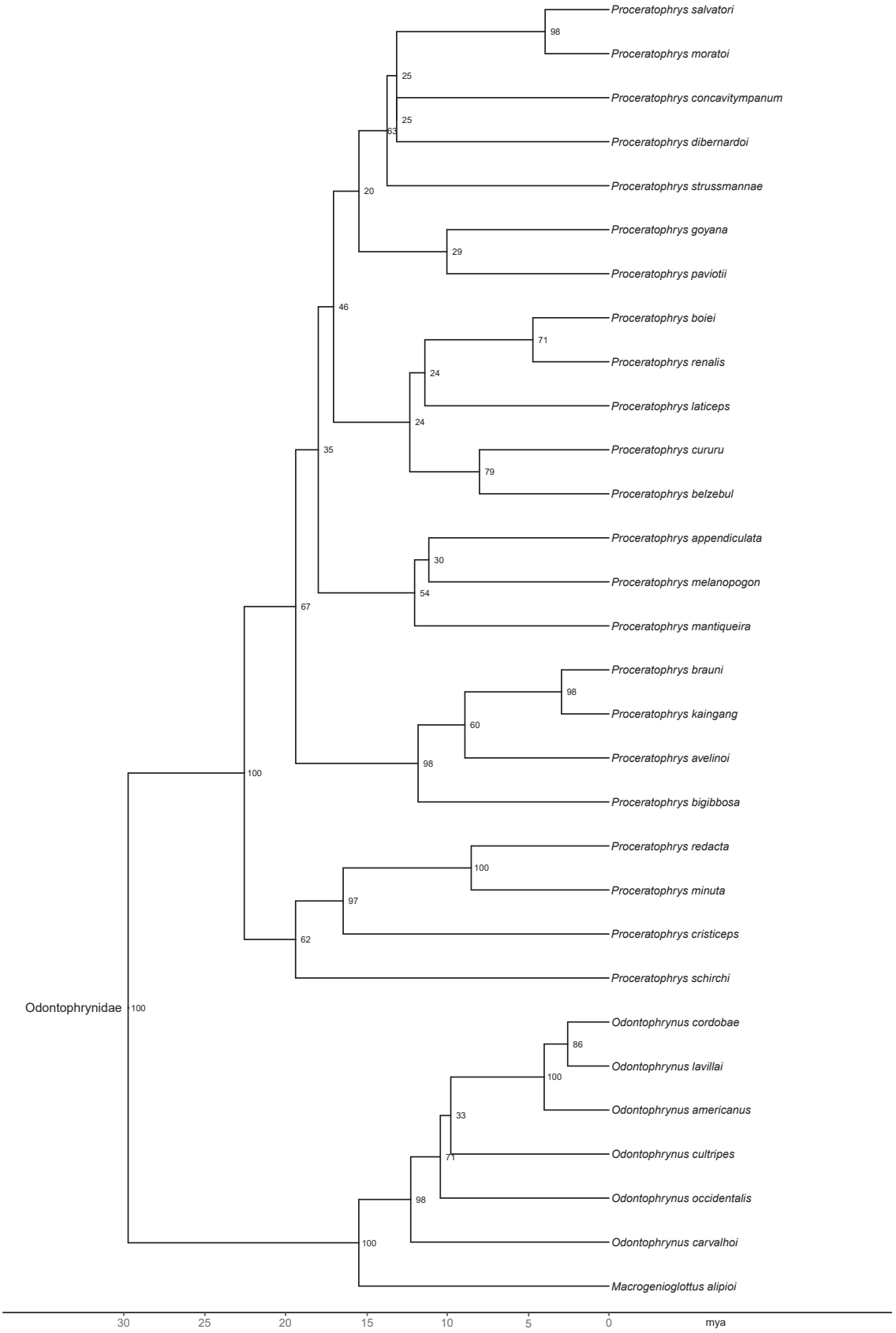

Fig. 56. Neobatrachia XLIV subtree including Odontophrynidae. Values adjacent to nodes are bootstrap support values from the partitioned maximum-likelihood analysis. Branch lengths are from the divergence-time estimation (in millions of years ago; mya).

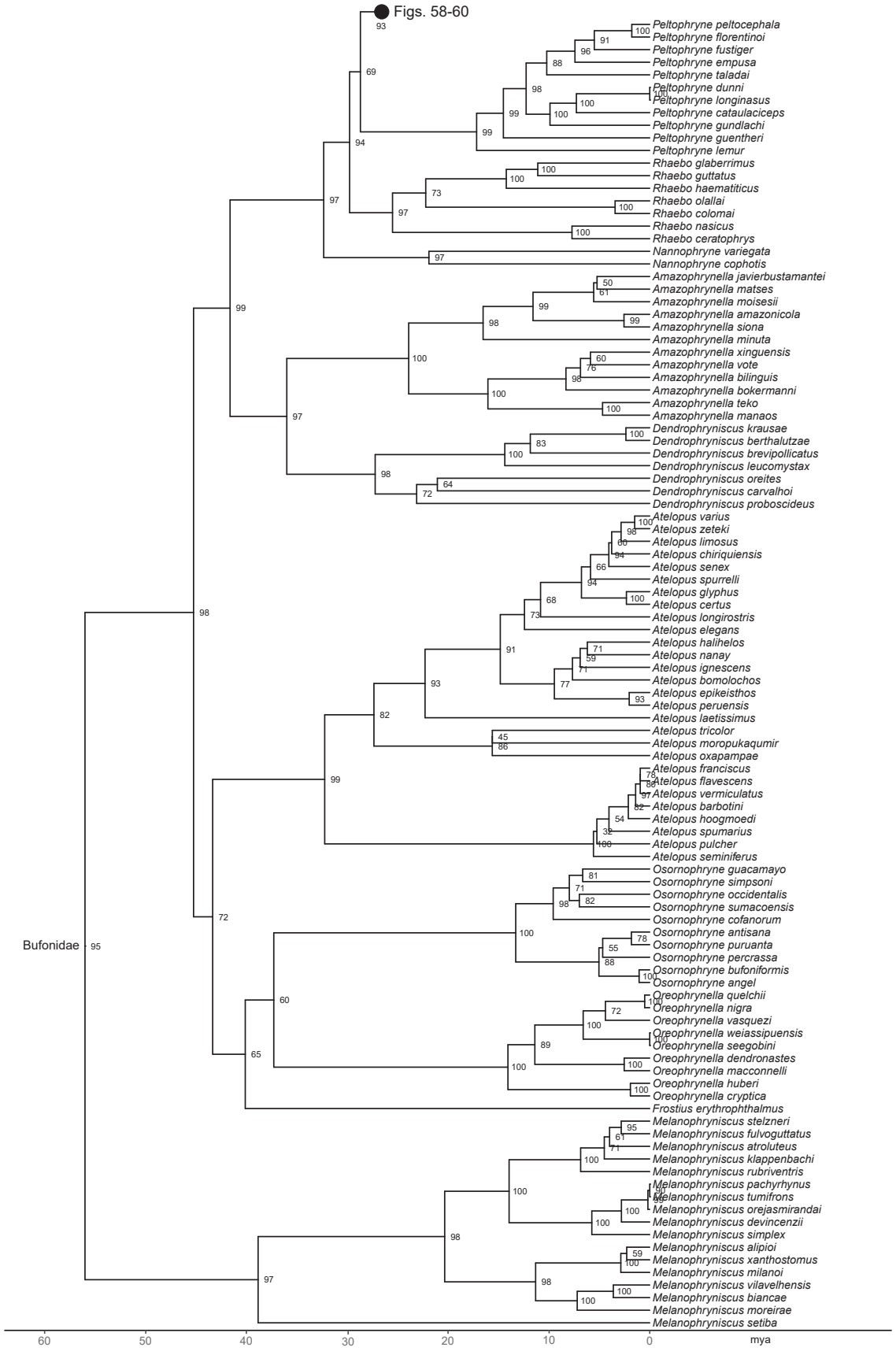

Fig. 57. Neobatrachia XLVIII subtree including Bufonidae (part). Values adjacent to nodes are bootstrap support values from the partitioned maximum-likelihood analysis. Branch lengths are from the divergence-time estimation (in millions of years ago; mya).

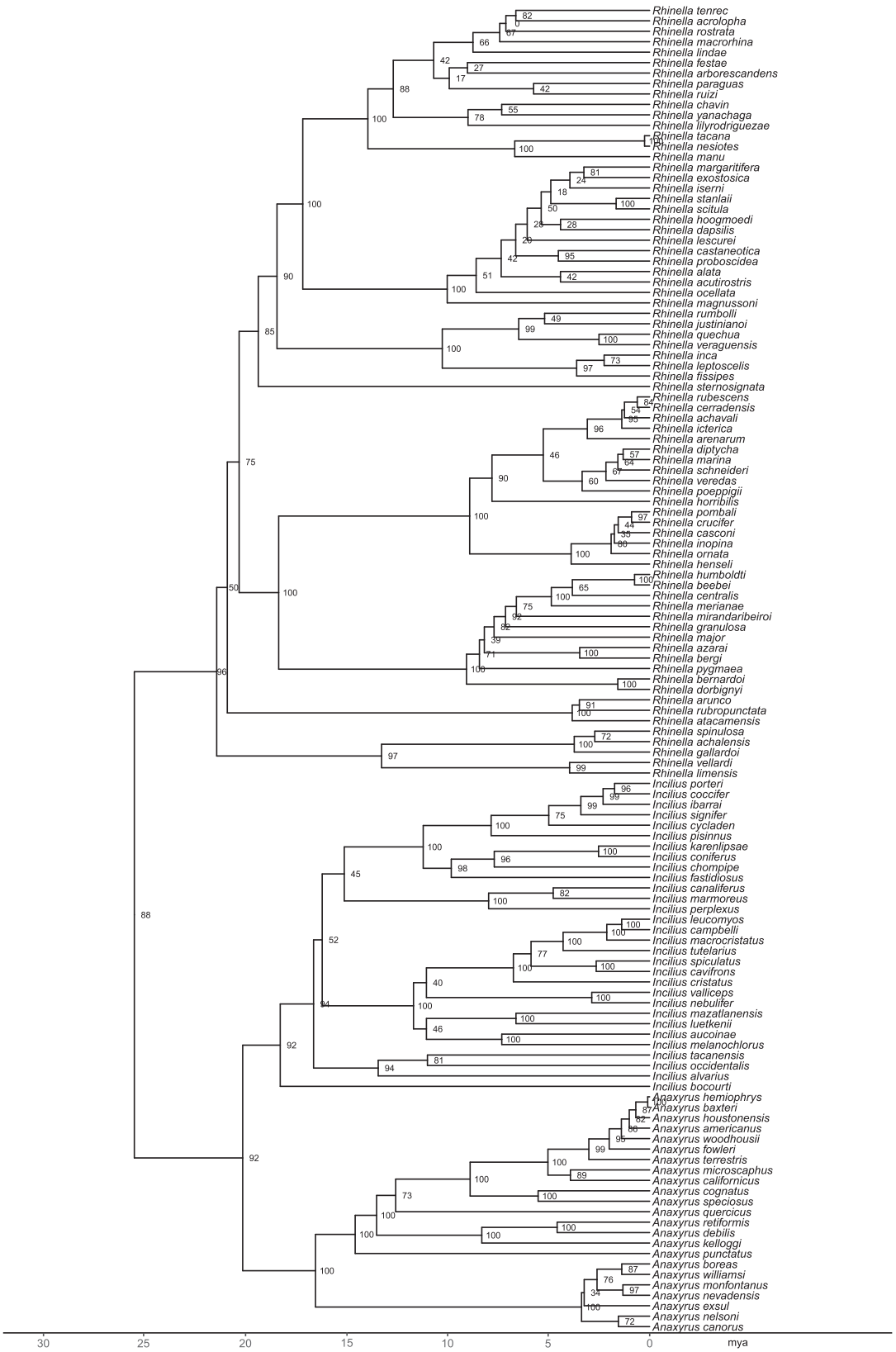

Fig. 58. Neobatrachia XLIX subtree including Bufonidae (part). Values adjacent to nodes are bootstrap support values from the partitioned maximum-likelihood analysis. Branch lengths are from the divergence-time estimation (in millions of years ago; mya).

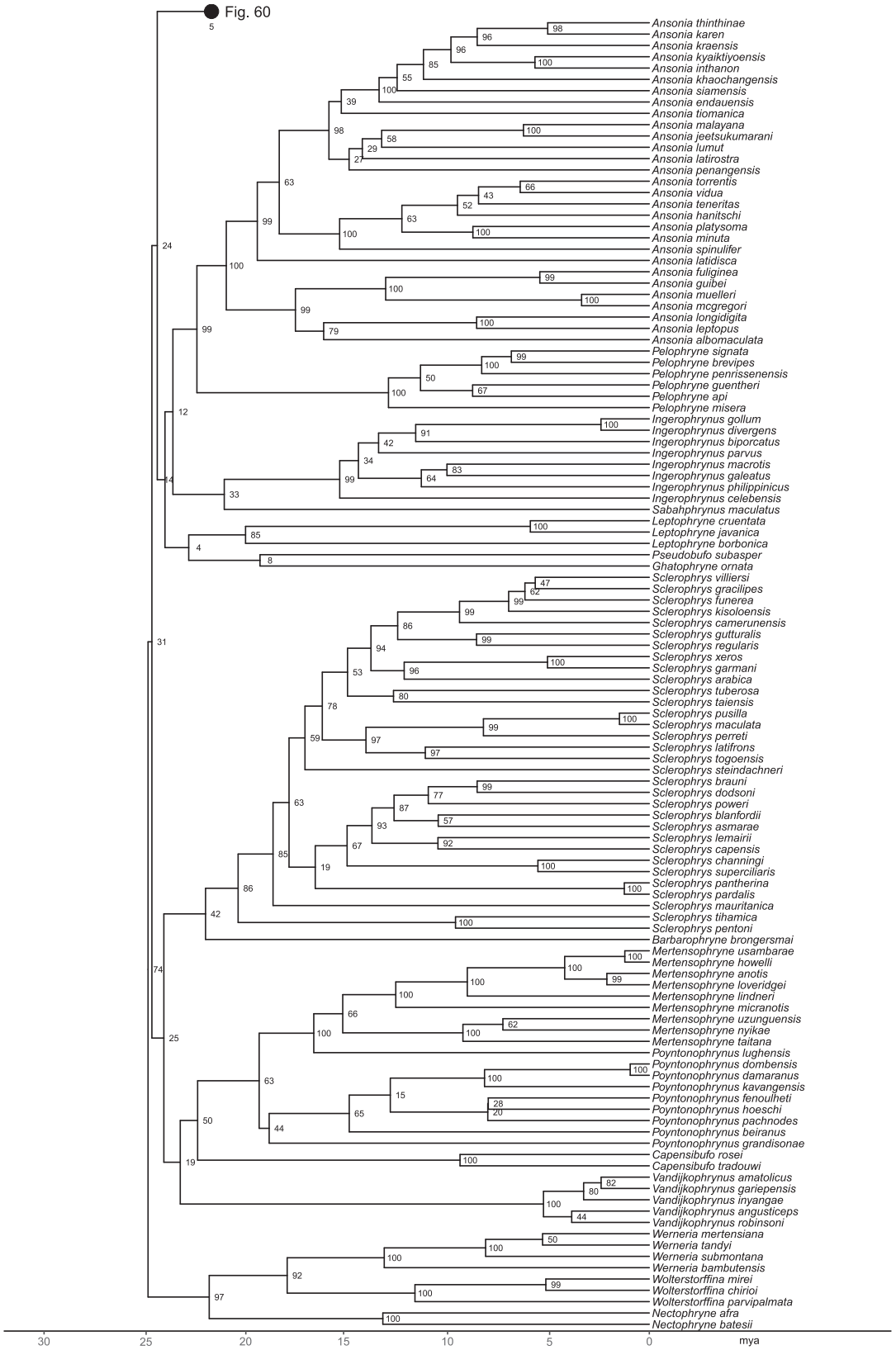

Fig. 59. Neobatrachia L subtree including Bufonidae (part). Values adjacent to nodes are bootstrap support values from the partitioned maximum-likelihood analysis. Branch lengths are from the divergence-time estimation (in millions of years ago; mya).

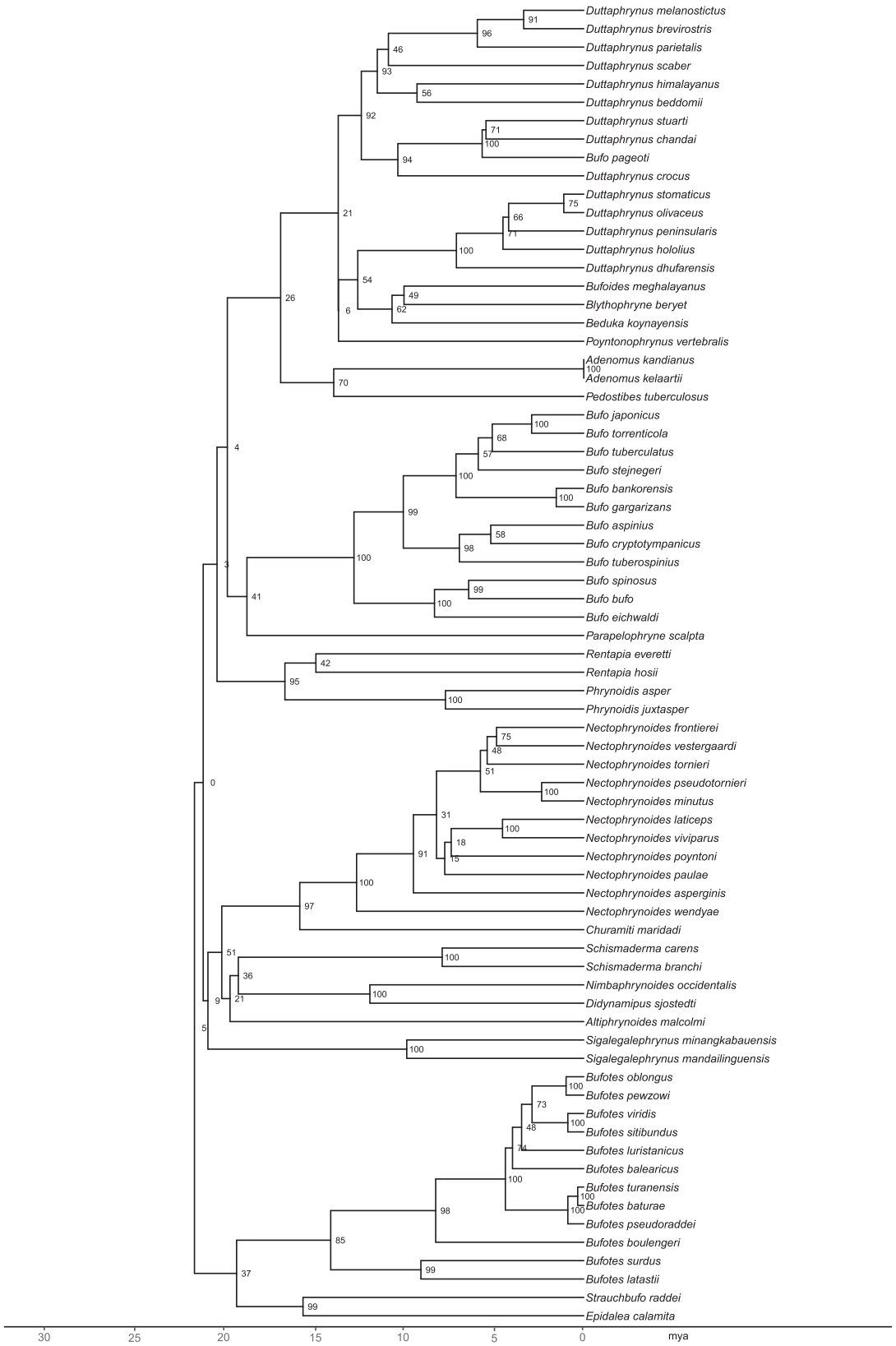

Fig. 60. Neobatrachia LI subtree including Bufonidae (part). Values adjacent to nodes are bootstrap support values from the partitioned maximum-likelihood analysis. Branch lengths are from the divergence-time estimation (in millions of years ago; mya).

problems in taxonomy or in our analyses.

Archaeobatrachians. Most relationships found within archaeobatrachian (non-neobatrachian) families were consistent with earlier studies, and most genera were monophyletic (Figs. 6–8). Nevertheless, we found extensive non-monophyly of genera within Megophryidae (Figs. 7–8), especially *Megophrys* and *Xenophrys* (Fig. 7). Much of this may represent taxonomic artifacts rather than problematic phylogeny estimation. For example, the placement of *Boulenophrys wawuensis* within *Atympanophrynus* has been rectified in some taxonomies by considering this species as part of *Atympanophrynus* (e.g., Chen et al., 2017). This change makes both *Atympanophrynus* and *Boulenophrys* monophyletic. Other megophryid genera were supported as monophyletic (Figs. 7–8), including *Brachytarsophrys, Leptobrachella, Leptobrachium, Ophryophryne, Oreolalax, Pelobatrachus,* and *Scutiger*.

Ranoidea. In Microhylidae (Figs. 9–12), the monophyly of all subfamilies was supported. We also found support for the three major clades of subfamilies that were recovered by previous phylogenomic analyses (Feng et al., 2017; Tu et al., 2018; Streicher et al., 2020; Hime et al., 2021; Portik et al., 2023). These consisted of the clade of (Adelastinae (Gastrophryninane, Otophryninae)), the clade (Cophylinae, Scaphiophryninae), and the clade (Asterophryinae (Dyscophinae, Microhylinae)). We weakly supported an endemic African clade of subfamilies (Hoplophryninae + Phrynomerinae; Fig. 9), as in Feng et al. (2017). As found by Tu et al. (2018) and Streicher et al. (2020), *Chaperina fusca* was nested within Microhylinae (Fig. 11), showing that it is not a distinct, monotypic subfamily (see Peloso et al., 2016). Within Cophylinae (Fig. 10), *Rhombophryne* and *Plethodontohyla* were non-monophyletic, with *R. mangabensis* nested within *Stumpffia*, *R. matavy* nested within *Plethodonothyla*, and *P. alluaudi* nested within *Rhomobophryne*. These results were explained (at least in part) by mislabeled samples from Peloso et al. (2016) that Scherz et al. (2016) previously highlighted, but which our protocol missed (e.g., *R. mangabensis*). However, some of the non-monophyly found here (e.g., *P. alluaudi*) may be explained by taxonomic problems associated with *Plethodontohyla*, *Rhomophryne*, and *Stumpffia* (Peloso et al., 2017).

Among asterophryines (Fig. 12), four genera were non-monophyletic: *Austrochaperina*, *Cophixalus*, *Copiula*, and *Oreophryne*. Hill et al. (2022) also found these genera to be non-monophyletic. Furthermore, our tree showed *Oninia* as nested within *Asterophrys*. These results show that the genus-level taxonomy of asterophryines needs revision.

In microhylines, *Kaloula* was non-monophyletic, with *K. assamensis* nested within *Uperodon* (Fig. 11). The species description of *K. assamensis* did not use molecular data (Das et al., 2004). However, this species was considered to be a close relative of "*Kaloula taprobanica*" (=*Uperodon taprobanicus*), to which *K. assamensis* is the sister taxon in our tree. Therefore, we transfer *K. assamensis* to the genus *Uperodon* creating the new combination *Uperodon assamensis* (Das et al., 2004).

Within Afrobatrachia, the clade ((Hemisotide, Brevicipitidae) (Arthroleptidae, Hyperoliidae)), we supported the monophyly of all the genera in Brevicipitidae (Fig. 13). Within Arthroleptidae (Fig. 14), we found that the traditionally recognized subfamily Astylosterninae is paraphyletic, with the genus *Leptodactylodon* more closely related to Leptopelinae + Arthroleptinae than to other genera of Astylosterninae (*Astylosternus, Nyctibates, Scotobleps, Trichobatrachus*). The placement of *Leptodactylodon* was strongly supported here. Within Astylosterninae, we found *Trichobatrachus* nested inside *Astylosternus*. Dubois et al. (2021) considered *Trichobatrachus* to be part of *Astylosternus*. We found strong support for the monophyly of Leptopelinae + Arthroleptinae, *Leptopelis*, *Arthroleptis*, and *Cardioglossa*. Note that these subfamilies of Arthroleptidae were not included in our initial taxonomy (AmphibiaWeb, 2022).

In hyperoliids (Fig. 15), we strongly supported a clade consisting of *Paracassina*, *Phlyctimantis* (or *Hylambates*), *Kassina*, and *Semnodactylus* as the sister group to other hyperoliids (Kassininae of Dubois et al., 2021). *Semnodactylus* was nested inside *Kassina*. We placed *Acanthixalus* (Acanthixalinae of Nečas et al., 2022) with weak support as the sister taxon to the remaining hyperoliids (Hyperoliinae; Nečas et al., 2022). We supported the monophyly of most of the remaining hyperoliid genera. However, we found *Kassinula* nested inside *Afrixalus* (Fig. 15). Further, *Alexteroon* was nested inside *Hyperolius*, with strong support (Fig. 16). Relationships among species within Micrixalidae were generally weakly supported (Fig. 17).

Within Pyxicephalidae (Fig. 18), we strongly supported the monophyly of the two subfamilies, Pyxicephalinae (*Aubria, Pyxicephalus*) and Cacosterninae (all other genera). Within Cacosterninae (Fig. 18), we placed *Anyhydrophryne* as the sister taxon to all other genera. We found some cacosternine genera to be non-monophyletic, including *Tomopterna* (with *Nothophryne* placed inside it), *Strongylopus* (with *S. grayi* closer to *Amietia* than other *Strongylopus*), and *Cacosternum* (with *Microbatrachella* placed inside it as the sister taxon to *C. platys*, as found previously, e.g., Pyron and Wiens, 2011). We found strong support for many species-level relationships within Phrynobatrachidae and for some within Ptychadenidae (Fig. 19).

Within Ceratobatrachidae (Fig. 20), the tree placed Liuraninae as the sister taxon of Alcalinae + Ceratobatrachinae. We strongly supported monophyly of Alcalinae (genus *Alcalus*) and Ceratobatrachinae (*Cornufer, Platymantis*). Within Ceratobatrachinae, we strongly supported two clades, one that contained most *Platymantis* species and one that contained most *Cornufer* species (following Brown et al., 2015). However, some species placed in *Platymantis* by the taxonomy used here were placed within *Cornufer* in our phylogeny.

Within Dicroglossidae (Figs. 21–22), we strongly supported the monophyly of the two subfamilies, Occidozyginae (*Ingerana, Occidozyga, Phrynoglossus*) and Dicroglossinae (*Euphlyctis, Fejervarya, Hoplobatrachus, Limnonectes, Minervarya, Nannophrys, Sphaerotheca*). We also supported the monophyly of most genera, but *Occidozyga* was paraphyletic with respect to *Phrynoglossus* (Fig. 21).

Within Ranidae (Figs. 23–25), we supported *Staurois* as the sister taxon to all other genera (Fig. 23), as in many previous analyses. However, as also found in many previous analyses, many genera were non-monophyletic, including *Amnirana, Amolops* (given placement of *A. larutensis* as the sister taxon to *Meristogenys*), *Glandirana, Humerana, Hylarana, Indosylvirana, Papuarana,* and *Sylvirana*. Non-monophyly of *Pelophylax* could be resolved by considering *Pelophylax lateralis* (Fig. 24) part of *Humerana* (as in Frost, 2022). We also supported the monophyly of many ranid genera, including *Babina, Chalcorana, Clinotarsus, Hydrophylax, Meristogenys, Nidirana, Odorrana, Pulchrana, Rana, Sanguirana, Staurois, Sumaterana,* and *Wijayrana*.

Within Mantellidae (Figs. 26–27), we supported the monophyly of the three subfamilies, with Laliostominae (*Aglyptodactylus, Laliostoma*) as the sister group to the clade of Boophinae (*Boophis*) and Mantellinae (*Blommersia, Boehmantis, Gephyromantis, Guibemantis, Mantella, Mantidactylus, Spinomantis, Tsingymantis, Wakea*). All the non-monotypic genera were supported as monophyletic, generally with strong support.

In Rhacophoridae (Figs. 28–31), we strongly supported the monophyly of the two subfamilies: Buergerinae (*Buergeria*) and Rhacophorinae (all other genera). We supported the monophyly of most genera, with some exceptions. *Gracixalus* was largely monophyletic (Fig. 28), but one species of *Gracixalus* (*G. medogensis*) was in *Feihyla* (Fig. 29). *Kurixalus ananjevae* was placed in *Gracixalus* (Fig. 28) and not with other *Kurixalus* (Fig. 30). Similarly, most species of *Rhacophorus* formed a clade (Fig. 29), but *R. calcadensis* was with *Chiromantis*, some *Rhacophorus* were within *Leptomantis*, and *R. vampyrus* was within *Zhangixalus*. In summary, we supported the monophyly of *Buergeria, Chiromantis, Ghatixalus, Liuixalus, Nasutixalus, Nyctixalus, Philautus, Polypedates, Pseudophilautus, Raorchestes, Taruga,* and *Theloderma*, but not *Feihyla, Gracixalus, Kurixalus, Leptomantis, Rhacophorus,* or *Zhangixalus*.

Hyloidea (and relatives). Within Myobatrachidae (Fig. 32), we found that both subfamilies (Limnodynastinae, Myobatrachinae) were non-monophyletic. This pattern occurred because: (1) the limnodynastine genus *Mixophyes* was placed (but with weak support) as the sister group

to all other myobatrachids, rather than with other members of Limnodynastinae. (2) The myobatrachine genus *Rheobatrachus* was placed (with weak support) as the sister group to Limnodynastinae, instead of with other members of Myobatrachinae. (3) The limnodynastine genus *Pseudophryne* was within Myobatrachinae (in a clade with *Arenophryne, Metacrinia,* and *Myobatrachus*) with strong support. Within the myobatrachid subfamily Limnodynastinae, we supported the monophyly of some genera (*Limnodynastes, Philoria, Platyplectrum*), but we also found weak support for non-monophyly of *Notaden* (with *N. nichollsi* distantly related to the other two sampled *Notaden* species) and of *Neobatrachus* (paraphyletic with respect to *Heleioporus*). In Myobatrachinae, we found that *Geocrinia victoriana* is within *Crinia* and is only distantly related to the other two sampled species of *Geocrinia*. We supported monophyly of *Assa, Pseudophryne,* and *Uperoleia*.

In Cycloramphidae (Fig. 33), we found that all three genera are non-monophyletic, with *Zachaeneus* and *Cycloramphus* placed inside of *Thoropa*, *Zachaeneus* inside of *Cycloramphus*, and *C. organensis* inside of *Zachaeneus*. By contrast, all genera were supported as monophyletic in Hylodidae and Alsodidae (Fig. 33). In Batrachylidae (Fig. 33), the genus *Batrachyla* was paraphyletic with respect to *Hylorina* and *Atelognathus*. Although our sampling of Telmatobiidae was limited, our results showed that this family radiated very rapidly relative to other frog families (Fig. 34), with almost all splits occurring within the last 5 million years.

In Hemiphractidae (Fig. 35), we strongly supported the monophyly of all genera, with the clade *Cryptobatrachus* + *Flectonotus* placed as the sister group to all other hemiphractids. Within the latter clade, there was a weakly supported clade linking *Hemiphractus* and *Fritziana*. *Stefania* was the sister group to *Gastrotheca*. The latter genus contains the majority of hemiphractid species.

In Hylidae (Figs. 36–42), we supported the monophyly of the three traditionally recognized subfamilies, with the clade Pelodryadinae + Phyllomedusinae placed as the sister group to the Hylinae (as found in many previous studies). Within Phyllomedusinae (Fig. 36), most genera were supported as monophyletic, but the genus *Pithecopus* was both polyphyletic and nested inside of *Phyllomedusa*. Within Pelodryadinae (Fig. 37), the genus *Litoria* was paraphyletic with respect to *Cyclorana* and *Nyctimystes*. We found that *Cyclorana alboguttata* did not group with other *Cyclorana*, but the relevant relationships were weakly supported.

Within Hylinae (Figs. 38–42), we supported the Cophomantini as the sister group to the other traditionally recognized tribes (tribe-level taxonomy following Faivovich et al., 2005; Frost, 2022), and the relationships: (Cophomantini (Sphaenorhynchini (Lophiohylini (Dendropsophini, Hylini)))). There is moderate support for these relationships (bootstrap = 83% for all three among-tribe clades). These relationships are consistent with the recent gigamatrix analyses of Portik et al. (2023), based on 4,091 markers, and the concatenated phylogenomic analyses of Hutter and Duellman (2023), based on ~9,000 nuclear markers. Both of these previous studies show very strong support for these clades.

We supported the monophyly of Cophomantini, the tribe including *Aplastodiscus, Boana, Bokermannohyla, Hyloscirtus, Myersiohyla*, and *Nesorohyla* (Fig. 38). We also supported the monophyly of these genera (except that *Bokermannohyla claresignata* is placed within *Boana*, rendering both genera non-monophyletic). We placed *Scinax* and *Sphaenorhynchus* as sister taxa (tribe Sphaenorhynchini; Fig. 39), with strong support for the monophyly of each genus.

We found some problems within the tribe Lophiohylini (Fig. 40), but these seem to be taxonomic issues and not necessarily errors in the phylogeny. We strongly supported non-monophyly of *Phyllodytes*, with *Phyllodytes auratus* as the sister taxon to other lophiohyliines (but this species is also regarded as a distinct genus, *Phytoriades*; Frost, 2022). We found that the traditionally recognized genus *Aparasphenodon* is polyphyletic and *Nyctimantis* is paraphyletic (Fig. 40). One species of *Aparasphenodon* (*A. venezolanus*) is within *Trachycephalus* (Fig. 40). However, many of these taxonomic problems (based on the taxonomy of AmphibiaWeb, 2022) were recognized (and resolved) by Blotto et al. (2021), who placed *A. venezolanus* in *Trachycephalus*, and placed all other species of *Aparasphenodon* into *Nyctimantis* (along with *Argenteohyla*). Nevertheless, there was relatively strong support for the monophyly of many genera, including *Dryaderces, Osteocephalus, Osteopilus,* and *Tepuihyla*.

Within the tribe Dendropsophini (Fig. 41) we found the relationships: ((*Dendropsophus, Xenohyla*), (*Scarthyla* (*Pseudis, Lysapsus*))). We found weak support for placing *Lysapsus* inside of *Pseudis* and *Xenohyla* inside of *Dendropsophus* (Fig. 41).

Finally, we supported the monophyly of the tribe Hylini (Fig. 42), and most included genera (including *Acris, Atlantihyla, Bromeliohyla, Charadrahyla, Ecnomiohyla, Exerodonta, Hyla, Isthmohyla, Megastomatohyla, Plectrohyla, Pseudacris, Sarcohyla, Smilisca, Tlalocohyla,* and *Triprion*). However, the genera *Atlantihyla, Bromeliohyla,* and *Duellmanohyla* were nested inside of *Ptychohyla*, and both *Ptychohyla* and *Duellmanohyla* were non-monophyletic (Fig. 42). One species of *Plectrohyla* (*P. calthula*) appeared to be nested inside of *Sarcohyla*, but this putative species is a synonym of *S. labeculata* (Campbell et al., 2018; Frost, 2022).

Within Eleutherodactylidae (Figs. 43–44), we strongly supported the monophyly of the subfamilies Eleutherodactylinae and Phyzelaphryninae, and the genera *Adelophryne*, *Diasporus*, and *Eleutherodactylus*. Within Brachycephalidae (Fig. 45), we supported monophyly of *Brachycephalus* and *Ischnocnema*.

In Craugastoridae (Figs. 46–50; here including Strabomantidae), we supported the monophyly of Craugastorinae and the genera *Craugastor* and *Haddadus* (Fig. 46). We supported (Fig. 47) the monophyly of *Strabomantis* (Strabomantinae) and *Niceforonia* (Hypodactylinae). We also supported the monophyly of many sampled genera in Holoadeninae, including *Barycholos*, *Bryophryne*, *Euparkerella*, *Holoaden*, and *Oosqophryne* (Fig. 47). However, we found that *Noblella* and *Psychophrynella* may be non-monophyletic (Fig. 47). One problem is that some sequences for *N. peruviana* may be from a population subsequently described as *Psychophrynella usurpator* (Frost, 2022; M. Heinicke, pers. comm.). Nevertheless, some of the other species of *Noblella* were closely related to *Barycholos* (including *N. coloma, N. heyeri, N. lochites, N. myrmecoides*, and *N. personina*), whereas another set of species (including *N. losamigos, N. madraselva, N. pygmaea,* and *N. thiuni*) were grouped with species of *Psychophrynella* and *Mikrokayla*. Furthermore, we found that the three sampled species of *Psychophrynella* were polyphyletic within this latter clade of *Noblella*. We also found that one species of *Mikrokayla* (*M. guillei*) is placed more closely related to these species of *Noblella* and *Psychophrynella* than to other species of *Mikrokayla*, but with very weak support (Fig. 47).

We did not support monophyly of the craugastorid subfamily Pristimantinae. We found that the pristimantine genus *Tachiramantis* was weakly supported as the sister group to the clade *Craugastor* + *Haddadus* in Craugastorinae (Fig. 46). Other sampled members of Pristimantinae did form a weakly supported clade (Figs. 47–50). Within this clade, one strongly supported subclade consisted of the genera *Lynchius*, *Oreobates*, and *Phrynopus*, each of which was strongly supported as monophyletic (Fig. 47). The other subclade consisted of the monophyletic genera *Pristimantis* and *Yunganastes* (Figs. 48–50).

We found several taxonomic problems in Dendrobatidae (Figs. 51–52). First, all four subfamilies were non-monophyletic. However, this seemed to be due to the taxonomy of a limited number of species. First, the subfamily Aromobatinae (Fig. 51) was strongly supported as non-monophyletic because some species of *Allobates* were placed within Hyloxalinae (Fig. 52). Second, Colostethinae was strongly supported as non-monophyletic because of the placement of *Colostethus ruthveni* within Dendrobatinae and *Ameerega erythromos* in Hyloxalinae (Fig. 52). This also renders Dendrobatinae non-monophyletic. Third, Hyloxalinae was strongly supported as non-monophyletic because it contained three species of *Allobates* (*A. cepadai, A. picachos, A. ranoides*), whereas other members of this genus were placed in Aromobatinae, and because it contained *Ameerega erythromos*, since the other members of this genus are placed in Colostethinae (Fig. 52). In all of these cases,

these species did not appear to be incorrectly placed because of missing data: all of these species were relatively strongly supported in their placements. Instead, these seem to reflect taxonomic issues noted by previous authors, as we describe below. Again, we followed the AmphibiaWeb (2022) taxonomy throughout.

In Aromobatinae (Fig. 51), the genus *Allobates* was monophyletic except that three species of *Allobates* (*A. cepedai, A. picachos, A. ranoides*) were placed with *Hyloxalus* (in Hyloxalinae), but some have already been transferred to *Hyloxalus* (*A. cepedai, A. picachos*; Frost, 2022). The genera *Anomaloglossus, Aromobates, Mannophryne,* and *Rheobates* were each monophyletic (Fig. 51). In Colostethinae, the genera *Amereega, Epipedobates*, and *Silverstoneia* and were each monophyletic (Fig. 52). One species of *Ameerega*, *A. erythromos*, was placed in Hyloxalinae, but this species is now recognized as belonging to the hyloxaline genus *Paruwrobates* (Frost, 2022). However, *Colostethus* and *Leucostethus* were non-monophyletic with respect to each other and to *Ameerega* (Fig. 52). Many of these relationships were very well supported. In Hyloxalinae (Fig. 52), three species of *Allobates* (Aromobatinae) were nested inside of *Hyloxalus* (but should be considered *Hyloxalus*) and one species of *Ameerega* (Colostethinae) was the sister group to *Ectopoglossus* (but this should be recognized as *Paruwrobates*). Finally, in Dendrobatinae (Fig. 52), we supported the monophyly of all the genera (*Adelphobates, Andinobates, Dendrobates, Excidobates, Oophaga, Phyllobates, Ranitomeya*). However, one species of *Colostethus* (*C. ruthveni*) was also placed in this clade.

Within Centrolenidae (Fig. 53), we supported the monophyly of the two subfamilies (Centroleninae, Hyalinobatrachinae). Within Hyalinobatrachinae, we supported the monophyly of each of the two genera (*Celsellia, Hyalinobatrachium*). In Centroleninae, we supported the monophyly of all the sampled genera, including *Centrolene, Chimerella, Cochranella, Espadarana, Ikagoi, Nymphargus, Rulyrana, Sachatamia, Teratohyla*, and *Vitreorana*. However, many intergeneric relationships were not strongly supported. Note that both *Cochranella* and *Nymphargus* appear to be non-monophyletic in our tree because of the placement of *C. megista* in *Nymphargus*, but this taxonomic problem has been resolved by placing this species in *Nymphargus* (Traseger et al., 2021).

In Leptodactylidae (Figs. 54–55), we supported the monophyly of the three subfamilies (Leiuperinae, Leptodactylinae, and Paratelmatobiinae). In Leiuperinae (Fig. 54), we strongly supported the monophyly of all genera and the relationships among them. We placed *Pseudopaludicola* as the sister taxon to the other genera, *Pleurodema* as the sister taxon to the remaining genera, and *Edalorhina* as the sister taxon to *Engystomops* + *Physalaemus*. We found weak support for grouping Leptodactylinae and Paratelmatobiinae as sister taxa (Fig. 55), but each subfamily was strongly supported. In Paratelmatobiinae (Fig. 55), the monotypic genus *Rupirana* was the sister group to the remaining species (with the monotypic genus *Scythrophrys* as the sister group to those excluding *Rupirana*), and the genera *Paratelmatobius* and *Crossodactylodes* were strongly supported as sister taxa and as monophyletic. In Leptodactylinae (Fig. 55), we found strong support for the clade consisting of *Adenomera* and the monotypic *Lithodytes*, and a clade consisting of *Hydrolaetare* + *Leptodactylus*. The support for monophyly of *Adenomera* and *Hydrolaetare* was relatively strong but was moderate for *Leptodactylus* (bootstrap = 75%). We found strong support for the monophyly of Odontophrynidae and the genera within it (Fig. 56).

Finally, we strongly supported monophyly of Bufonidae, and many of the major intergeneric relationships within it (Figs. 57–60). The sister group to all other bufonids was *Melanophryniscus* (Fig. 57). The sister group to the remaining bufonids was a clade consisting of *Atelopus*, *Frostius*, *Oreophrynella*, and *Osornophryne* (Fig. 57). Relationships among these four genera were only weakly supported, but there was strong support for the monophyly of the genera *Atelopus, Oreophrynella*, and *Osornophryne* (note: only one species of *Frostius* was included here). The sister group to the remaining bufonids was a strongly supported clade containing *Amazophrynella* and *Dendrophryniscus* (Fig. 57).

The species traditionally assigned to the genus *Bufo* (and their relatives) formed a strongly supported clade (Fig. 57). Within this clade, the relationships were (*Nannophryne* (*Rhaebo* (*Peltophryne* (all remaining bufonids)))). The genera *Anaxyrus*, *Incilius*, and *Rhinella* together formed a clade (Fig. 58), with each genus strongly supported as monophyletic. Many other intergeneric relationships were only weakly supported, but most genera were supported as monophyletic (Figs. 59–60). There was a well-supported clade including *Nectophryne*, *Werneria*, and *Wolterstorffina* (Fig. 59). The genera *Capensibufo*, *Mertensophryne*, and *Poyntonophrynus* formed a weakly supported clade (Fig. 59), but with *Poyntonophrynus* paraphyletic with respect to *Mertensophryne*, because of the placement of *P. lughensis* with *Mertensophryne*. Furthermore, *Poyntonophrynus vertebralis* was placed in a clade within *Duttaphrynus* (Fig. 60). The species *Bufo pageoti* was also nested inside *Duttaphrynus*, as were the genera *Beduka*, *Blythophryne*, and *Bufoides* (Fig. 60). We do not discuss all relationships within Bufonidae in detail, especially since many relationships among genera were only weakly supported. Nevertheless, many genera were strongly supported in our tree (Figs. 59–60), including *Ansonia, Ingerophrynus, Nectophrynoides, Pelophryne,* and *Schismaderma*.

3.5. Comparing estimated ages of major clades

We compared the ages of major clades estimated here (Fig. 4) to those of several other recent large-scales estimates for anurans (Table 3), including the phylogenomic studies of Feng et al. (2017) and Hime et al. (2021) and the supermatrix study by Jetz and Pyron (2018). In general, we found that our estimates were generally younger than those of these other studies for the same clades. The differences for crown Anura and Leiopelmatoidea were particularly large. However, our estimates were typically within 10–20 million years of most other estimates for the other five clades. Furthermore, the confidence intervals from the 100 time-calibrated bootstrap trees mostly overlapped with these estimates from other studies for these five clades (Table 3). Moreover, the estimates from these other studies were not identical to each other. Although the estimates of Feng et al. (2017) and Hime et al. (2021) were very similar to each other, the estimates from Jetz and Pyron (2018) were sometimes much younger (e.g., Leiopelmatoidea) or much older (e.g., for Discoglossoidea, Pelobatoidea, Neobatrachia, Hyloidea and Ranoidea; Table 3).

4. Discussion

In this study we provide a new large-scale, time-calibrated

Table 3
Summary of estimated clade ages from recent time-calibrated phylogenies of anurans. Ages are approximate for most clades, since many studies did not provide labeled, time-calibrated trees. Ages for this study are presented as the value for the best tree, followed by the range of the 95% confidence interval from the 100 time-calibrated bootstrap trees in parentheses.

Clade	This study (95% CI)	Feng et al. (2017)	Hime et al. (2021)	Jetz and Pyron (2018)
Crown Anura	179 (174–201)	210	210	222
Crown Leiopelmatoidea	118 (113–137)	195	200	160
Crown Discoglossoidea	130 (125–147)	140	140	168
Crown Pipoidea	148 (148–166)	160	160	168
Crown Pelobatoidea	127 (118–154)	125	120	151
Crown Neobatrachia	121 (114–151)	140	140	178
Crown Hyloidea	66 (63–83)	70	70	157
Crown Ranoidea	87 (82–114)	100	105	143

phylogeny of anurans. The previous largest analysis of this type (Jetz and Pyron, 2018) included 3,067 frog species with up to 15 genes each (concatenated alignment 15,091 bp in length). Here, we included 5,242 frog species with up to 307 markers each, representing a 71% increase in the number of species included. Furthermore, we included 90,196 total sequences here (relative to 14,962 in Jetz and Pyron, 2018), an increase of 603%. We found that the higher-level relationships in this new tree were generally more similar to those of recent phylogenomic studies than to this previous supermatrix study, especially within Hyloidea. Moreover, the estimated ages of higher-level clades were often more similar to those of the recent phylogenomic studies (Table 3). We also provide a set of 100 time-calibrated topologies for use in comparative studies (Supplementary File S4). Our results not only provide an improved species-level estimate for frog phylogeny, but further demonstrate that supermatrix construction can be improved by including large numbers of markers, even if those markers have data for only a small subset of taxa (e.g., Cho et al., 2011; Wiens et al., 2005; Zheng and Wiens, 2016; Talavera et al., 2022; Portik et al., 2023).

Our supermatrix was dominated by missing data cells (~95% overall). The impact of missing data cells on phylogenetic analyses has been the subject of considerable debate (e.g., Philippe et al., 2004; Cho et al., 2011; Lemmon et al., 2009; Sanderson et al., 2011; Wiens and Morrill, 2011; Jiang et al., 2014; Hosner et al., 2016; Streicher et al., 2016; Xi et al., 2016; Talavera et al., 2022). However, we found few obvious artifacts in our results associated with these extensive missing data. For example, all species were placed in the expected families. Furthermore, the divergence dates for major clades were broadly similar to those estimated from more complete, phylogenomic datasets (Table 3), suggesting that there were not strong impacts of missing data on these estimates (see also Zheng and Wiens, 2015). Previous studies also suggest that missing data may have limited impact on branch-length estimation (e.g., Wiens and Morrill, 2011; Jiang et al., 2014; Portik et al., 2023). Nevertheless, bootstrap support values in many parts of the tree were far from optimal, especially relationships in the ancient, rapid radiations among the families of Hyloidea and Natatanura. Importantly, many of these same relationships were also weakly supported in the supermatrix tree of Pyron and Wiens (2011), and so cannot be explained by the overall increase in missing data cells associated with adding the phylogenomic markers. The support for the phylogenetic placement of many species is still constrained by those species having data for only one or a few genes. Thus, the uncertainty in their placement seems to be associated with the limited data that they have, not the amount of missing data per se (e.g., Wiens, 2003). Moreover, despite the weak support for many higher-level relationships, these relationships still generally resemble those from phylogenomic studies, and not previous supermatrix studies. Therefore, the addition of these hundreds of phylogenomic markers was clearly worthwhile, even if some relationships remain weakly supported (for similar results from smaller datasets see Zheng and Wiens, 2016; Talavera et al., 2022).

If adding some phylogenomic markers is beneficial, why not add more? In anurans, some datasets include thousands of nuclear markers (e.g., Streicher et al., 2018; Barrientos et al., 2021; Hutter et al., 2022). We initially attempted to do this, yielding a matrix with thousands of markers and thousands of taxa. Unfortunately, after many months of effort, we found that this combined matrix was too large to be analyzed with currently available computational resources (Portik et al., 2023). Nevertheless, we suspect that it will become possible to analyze such matrices in the next few years. Furthermore, a recent study showed that it is possible to add thousands of phylogenomic markers to a dataset in which many taxa have only one or a few markers (Portik et al., 2023). This combination yielded strongly supported higher-level relationships that were very similar to those from the phylogenomic markers alone. However, for computational tractability, that analysis included only one terminal taxon per genus, such that the matrix included hundreds of species, but not thousands.

Another alternative approach to analyzing these data might be to constrain the higher-level relationships based on the results of phylogenomic analyses (e.g., Álvarez-Carretero et al., 2021). However, this would have some disadvantages. First, although phylogenomic analyses agree on many aspects of higher-level frog phylogeny, they do not agree on everything (see Section 3.3 of the Results). Therefore, it is not always clear which relationships should be estimated as opposed to assumed. Second, our combined analysis of the data from Feng et al. (2017) and Hime et al. (2021) has the potential to offer an improved estimate relative to either one separately. Third, although phylogenomic analyses may generally offer the best estimate because of the large number of markers, extensive taxon sampling can also be important for correctly resolving higher-level relationships, and even highly incomplete taxa can be helpful in subdividing long branches (e.g., Wiens, 2005; Wiens and Tiu, 2012). Our study design allows relationships to be determined by both extensive taxon sampling and/or extensive sampling of genes.

We note that other studies have generated taxonomically comprehensive trees for some groups based on randomly adding unsampled species to branches within genera (e.g., Jetz and Pyron, 2018). We preferred not to do so here, and ensured that all species included had at least some molecular data. We found that numerous genera were not monophyletic. Therefore, adding species by assuming that the current taxonomy reflects the phylogeny is somewhat problematic. Nevertheless, it should still be straightforward to add unsampled species to our tree, if one is willing to make this assumption.

Our study also highlights an important issue for supermatrix construction: the accuracy of the sequence data in GenBank. Issues may arise through incorrect uploading of data, changes in taxonomy since the sequences were uploaded, misidentified specimens, contamination, and other lab errors (Nilsson et al., 2006; Leray et al., 2019; Mulcahy et al., 2022). For example, we identified and removed several human *12S* and *16S* mitochondrial sequences labeled as frogs in our analysis. Misidentified specimens are a challenging problem, particularly when the problem occurs at a shallow taxonomic scale (e.g., the sequence is assigned to the wrong species within the same genus or family). Similarity filtering can be used to establish gene identities, yet parallel approaches for identifying inaccurate taxon labeling within the focal group are currently lacking. Finally, there are known cases of mislabeled sequences in NCBI, which have been robustly identified by molecular and taxonomic experts in the research community (Nilsson et al., 2006; Mulcahy et al., 2022). However, given that only the original submitter can update existing GenBank records, these sequences remain available and problematic (Mulcahy et al., 2022). To facilitate removing such problematic sequences, we developed a specific feature for SuperCRUNCH for this project (the –accessions_exclude feature in *Filter_Seqs_and_Species.py*), which can be used to specify accession numbers to exclude from the final supermatrix. This feature will exclude a set of previously identified, problematic sequences during the sequence-selection step and automatically substitute them with other sequences instead. Importantly, without this feature, there is no automated way to avoid problematic sequences during supermatrix construction. We performed initial screens per marker by inspecting individual gene trees produced with a given set of sequences, and subsequently identified problematic sequences that we excluded from our final analysis. However, the initial identification of problematic sequences and manual curation of a list of these sequences still presents difficulties. Overall, inaccurate sequence records remain a challenging problem that would be a useful focus for future studies.

Our study provides several key resources. The molecular datasets that we assembled here can be used in a variety of ways, and because SuperCRUNCH is highly modular it offers access to these molecular datasets at multiple entry points. For example, new analyses can be started before or after sequence selection, pre or post-alignment, and using the full marker set or a subset of markers. Potential uses include the gathering of outgroup sequences, obtaining sequences for a marker for a particular clade of interest, adding unpublished sequences to the phylogeny, and creating new combinations of phylogenetic or

phylogeographic datasets. Importantly, the sequence data from each step of the SuperCRUNCH analysis are freely available on the Open Science Framework (OSF) project page (e.g., prior to similarity filtering, sequence selection, alignment, relabeling), making it easy to obtain the correct resources for a particular project.

Our study also presents a new, modified method for aligning more problematic sequences (TaxonomyAlign). This method is a modification of MAFFT that initially aligns sequences based on taxonomy (e.g., genera, families) before including all sequences in an overall alignment. We applied this new approach to the *12S* and *16S* sequences in our study and found that TaxonomyAlign yields similar sequence lengths to MAFFT alone, but with smaller numbers of informative sites. Importantly, TaxonomyAlign yielded a substantial improvement in the number of families recovered as monophyletic (27% and 15% improvements for *12S* and *16S*), with smaller improvements in the number of genera and subfamilies recovered. Further, most anuran families are supported as monophyletic by phylogenomic analyses (e.g., Feng et al., 2017; Streicher et al., 2018; Hime et al., 2021; Portik et al., 2023). Thus, the greater number of informative sites from MAFFT alone may be associated with more homoplastic sites and more problematic phylogenetic results. However, the improvements associated with TaxonomyAlign might depend on an overall match between the taxonomy used and the phylogeny. Overall, these results seem promising, but further study of the performance of TaxonomyAlign would be valuable.

Finally, we acknowledge that this study will not be the last large-scale estimate of species-level frog phylogeny. These estimates should continue to improve as more species are added that previously lacked molecular data and more genes are added for species that are already represented. These new molecular data can be added to our existing supermatrix to produce an updated phylogeny, or to a new supermatrix with potentially different combinations of sequences. Nevertheless, our estimate should be a substantial improvement over the previous largest estimate for anurans (Jetz and Pyron, 2018), in terms of including many more species and resolving higher-level relationships based on phylogenomic data. Moreover, our results further suggest the possibilities of combining supermatrix and phylogenomic approaches to build large-scale phylogenies.

Funding

This work was supported by U.S. National Science Foundation grant DEB 1655690 to JJW.

CRediT authorship contribution statement

Daniel M. Portik: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Writing – original draft, Writing – review & editing. **Jeffrey W. Streicher:** Conceptualization, Formal analysis, Investigation, Methodology, Writing – original draft, Writing – review & editing. **John J. Wiens:** Conceptualization, Funding acquisition, Investigation, Methodology, Writing – original draft, Writing – review & editing.

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.

Data availability

The data are available via a link at: https://osf.io/3cbsh/

Acknowledgments

We thank A. Larson, M. Heinicke, and an anonymous reviewer for many helpful comments on the manuscript. For funding we acknowledge National Science Foundation grant DEB 1655690 to JJW. This study was also supported by the UK Crop Diversity Bioinformatics HPC Resource (https://www.cropdiversity.ac.uk/). Above all, we thank the many researchers who generated the sequence data that made our study possible.

Appendix A. Supplementary material

Supplementary data for this article includes Supplementary Files S1–S6 and Supplementary Tables S1-S5 (uploaded with this submission). https://osf.io/3cbsh/ https://osf.io/3cbsh/?view_only. All analyses, datasets, and trees produced in this study are freely available on the Open Science Framework project page: https://osf.io/3cbsh/ . This resource includes all starting sequences, inputs and outputs from each SuperCRUNCH step, final sequence alignments, RAxML trees and time-calibrated trees, and instructions to replicate our results.

Appendix B. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ympev.2023.107907.

References

Álvarez-Carretero, S., Tamuri, A.U., Battini, M., Nascimento, F.F., Carlisle, E., Asher, R. J., Yang, Z., Donoghue, P.C.J., dos Reis, M., 2021. A species-level timeline of mammalian evolution integrating phylogenomic data. Nature 602, 263–267.

AmphibiaWeb. 2022. <https://amphibiaweb.org> University of California, Berkeley, CA, USA. Accessed 2 July 2022.

Barrientos, L.S., Streicher, J.W., Miller, E.C., Pie, M.R., Wiens, J.J., Crawford, A.J., 2021. Phylogeny of terraranan frogs based on 2,665 loci and impacts of missing data on phylogenomic analyses. Systematics and Biodiversity 19, 818–833.

Blotto, B.L., Lyra, M.L., Cardoso, M.C.S., Rodrigues, M.T., Dias, I.R., Marciano Jr., E., Vechio, F.D., Orrico, V.G.D., Brandão, R.A., de Assis, C.L., Lantyer-Silva, A.S.F., Rutherford, M.G., Gagliardi-Urrutia, L.A.G., Solé, M., Baldo, D., Nunes, I., Cajade, R., Torres, A., Grant, T., Jungfer, K.-H., da Silva, H.R., Haddad, C.F.B., Faivovich, J., 2021. The phylogeny of the casque-headed treefrogs (Hylidae: Hylinae: Lophiohylini). Cladistics 37, 36–72.

Bossuyt, F., Brown, R.M., Hillis, D.M., Cannatella, D.C., Milinkovitch, M.C., 2006. Phylogeny and biogeography of a cosmopolitan frog radiation: Late Cretaceous diversification resulted in continent-scale endemism in the family Ranidae. Systematic Biology 55, 579–594.

Brown, R.M., Siler, C.D., Richards, S.J., Diesmos, A.C., Cannatella, D.C., 2015. Multilocus phylogeny and a new classification for Southeast Asian and Melanesian forest frogs (family Ceratobatrachidae). Zoological Journal of the Linnean Society 174, 130–168.

Campbell, J.A., Brodie Jr., E.D., Caviedes-Solis, I.W., Nieto-Montes de Oca, A., Luja-Molina, V.H., Flores-Villela, O., García-Vázquez, U.O., Sarker, G.C., Wostl, E., 2018. Systematics of the frogs allocated to *Sarcohyla bistincta* sensu lato (Cope, 1877), with description of a new species from Western Mexico. Zootaxa 4422, 366–384.

Chen, J.-M., Zhou, W.-W., Poyarkov Jr., N.A., Stuart, B.L., Brown, R.M., Lathrop, A., Wang, Y., Yuan, Z.-Y., Jiang, K., Hou, M., Chen, H.-M., Suwannapoom, C., Nguyen, S. N., Duong, T.V., Papenfuss, T.J., Murphy, R.W., Zhang, Y.-P., Che, J., 2017. A novel multilocus phylogenetic estimation reveals unrecognized diversity in Asian horned toads, genus *Megophrys* sensu lato (Anura: Megophryidae). Molecular Phylogenetics and Evolution 106, 28–43.

Cho, S., Zwick, A., Regier, J.C., Mitter, C., Cummings, M.P., Yao, J., Du, Z., Zhao, H., Kawahara, A.Y., Weller, S., Davis, D.R., Baixeras, J., Brown, J.W., Parr, C., 2011. Can deliberately incomplete gene sample augmentation improve a phylogeny estimate for the advanced moths and butterflies (Hexapoda: Lepidoptera)? Systematic Biology 60, 782–796.

Das, I., Sengupta, S., Ahmed, M.F., Dutta, S.K., 2004. A new species of *Kaloula* (Anura: Microhylidae) from north-eastern India. Hamadryad 29, 101–109.

De Lisle, S.P., Rowe, L., 2015. Independent evolution of the sexes promotes amphibian diversification. Proceedings of the Royal Society of London B, Biological Sciences 282, 20142213.

Drummond, A.J., Rambaut, A., 2007. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evolutionary Biology 7, 214.

Faivovich, J., Haddad, C.F.B., de A. Garcia, P.C., Frost, D.R., Campbell, J.A., Wheeler, W. C.,, 2005. Systematic review of the frog family Hylidae, with special reference to Hylinae: a phylogenetic analysis and taxonomic revision. Bulletin of the American Museum of Natural History 294, 1–240.

Feng, Y.-J., Blackburn, D.C., Liang, D., Hillis, D.M., Wake, D.B., Cannatella, D.C., 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 of the United States of America 114, E5864–E5870.

Frost, D.R., Grant, T., Faivovich, J., Bain, R., Haas, A., Haddad, C.F.B., de Sá, R., Channing, A., Wilkinson, M., Donnellan, S.C., Raxworthy, C., Campbell, J.A., Blotto, B.L., Moler, P., Drewes, R.C., Nussbaum, R.A., Lynch, J.D., Green, D.M., Wheeler, W., 2006. The amphibian tree of life. Bulletin of the American Museum of Natural History 297, 1–370.

Frost, D.R., 2022. Amphibian Species of the World 6.1, An Online Reference. https://amphibiansoftheworld.amnh.org/. Accessed June 2022.

Furness, A.I., Capellini, I., 2019. The evolution of parental care diversity in amphibians. Nature Communications 10, 4709.

Furness, A.I., Venditti, C., Capellini, I., 2022. Terrestrial reproduction and parental care drive rapid evolution in the trade-off between offspring size and number across amphibians. PLoS Biology 20, e3001495.

Hedges, S.B., Duellman, W.E., Heinicke, M.P., 2008. New World direct-developing frogs (Anura: Terrana): molecular phylogeny, classification, biogeography, and conservation. Zootaxa 1737, 1–182.

Heinicke, M.P., Lemmon, A.R., Moriarty, E., Mcgrath, K., Hedges, S.B., 2018. Phylogenomic support for evolutionary relationships of New World direct-developing frogs (Anura: Terraranae). Molecular Phylogenetics and Evolution 118, 145–155.

Hill, E.C., Fraser, C.J., Gao, D.F., Jarman, M.J., Henry, E.R., Iova, B., Allison, A., Butler, M.A., 2022. Resolving the deep phylogeny: implications for early adaptive radiation, cryptic, and present-day ecological diversity of Papuan microhylid frogs. Molecular Phylogenetics and Evolution 177, 107618.

Hime, P.M., Lemmon, A.R., Lemmon, E.C.M., Prendini, E., Brown, J.M., Thomson, R.C., Kratovil, J.D., Noonan, B.P., Pyron, R.A., Peloso, P.L.V., Kortyna, M.L., Keogh, J.S., Donnellan, S.C., Mueller, R.L., Raxworthy, C.J., Kunte, K., Ron, S.R., Das, S., Gaitonde, N., Green, D.M., Labisko, J., Che, J., Weisrock, D.W., 2021. Phylogenomics reveals ancient gene tree discordance in the amphibian tree of life. Systematic Biology 70, 49–66.

Hosner, P.A., Faircloth, B.C., Glenn, T.C., Braun, E.L., Kimball, R.T., 2016. Phylogenomic inference of landfowl (Aves: Galliformes): effects of missing data in concatenated and coalescent inference and evidence for a bias in gene tree reconciliation approaches. Molecular Biology and Evolution 33, 1110–1125.

Hutter, C.R., Guayasamin, J.M., Wiens, J.J., 2013. Explaining Andean megadiversity: the evolutionary and ecological causes of glassfrog elevational richness patterns. Ecology Letters 16, 1135–1144.

Hutter, C.R., Lambert, S.M., Wiens, J.J., 2017. Rapid diversification and time explain amphibian species richness at different scales in the Tropical Andes, Earth's most biodiverse hotspot. American Naturalist 190, 828–843.

Hutter, C.R., Cobb, K.A., Portik, D.M., Travers, S.L., Wood Jr., P.L., Brown, R.M., 2022. FrogCap: a modular sequence capture probe-set for phylogenomics and population genetics for all frogs, assessed across multiple phylogenetic scales. Molecular Ecology Resources 22, 1100–1119.

Hutter, C.R., Duellman, W.E., 2023. Filtration of gene trees from 9,000 exons, introns, and UCEs disentangles conflicting phylogenomic relationships in tree frogs (Hylidae). Genome Biology and Evolution 15, evad070. https://doi.org/10.1093/gbe/evad070.

Jetz, W., Pyron, R.A., 2018. The interplay of past diversification and evolutionary isolation with present imperilment across the amphibian tree of life. Nature Ecology and Evolution 2, 850–858.

Jetz, W., Thomas, G.H., Joy, J.B., Hartmann, K., Mooers, A.O., 2012. The global diversity of birds in space and time. Nature 491, 444–448.

Jiang, W., Chen, S.-Y., Wang, H., Wiens, J.J., 2014. Should genes with missing data be excluded from phylogenetic analyses? Molecular Phylogenetics and Evolution 80, 308–318.

Katoh, K., Misawa, K., Kuma, K., Miyata, T., 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Research 30, 3059–3066.

Katoh, K., Standley, D.M., 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Molecular Biology and Evolution 30, 772–780.

Lanfear, R., Frandsen, P.B., Wright, A.M., Senfeld, T., Calcott, B., 2016. PartitionFinder2: new methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. Molecular Biology and Evolution 34, 772–773.

Lemmon, A.R., Brown, J.M., Stanger-Hall, C., Lemmon, E.M., 2009. The effect of ambiguous data on phylogenetic estimates obtained by maximum-likelihood and Bayesian inference. Systematic Biology 58, 130–145.

Lemmon, A.R., Emme, S.A., Lemmon, E.M., 2012. Anchored hybrid enrichment for massively high-throughput phylogenomics. Systematic Biology 61, 727–744.

Leray, M., Knowlton, N., Ho, S.-L., Machida, R.J., 2019. GenBank is a reliable resource for 21st century biodiversity research. Proceedings of the National Academy of Sciences of the United States of America 116, 22651–22656.

Li, W., Godzik, A., 2006. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics 22, 1658–1659.

Liedtke, H.C., Wiens, J.J., Gomez-Mestre, I., 2022. The evolution of reproductive modes and life cycles in amphibians. Nature Communications 13, 7039.

Miller, M.A., Pfeiffer, W., Schwartz, T., 2010. "Creating the CIPRES Science Gateway for inference of large phylogenetic trees" *in* Proceedings of the Gateway Computing Environments Workshop (GCE), 14 Nov. 2010, New Orleans, LA pp 1–8.

Moen, D.S., Wiens, J.J., 2017. Microhabitat and climatic-niche change explain patterns of diversification among frog families. American Naturalist 190, 29–44.

Mulcahy, D.G., Ibáñez, R., Jaramillo, C.A., Crawford, A.J., Ray, J.M., Gotte, S.W., Jacobs, J.F., Wynn, A.H., Gonzalez-Porter, G.P., McDiarmid, R.W., Crombie, R.I., Zug, G.R., de Queiroz, K., 2022. DNA barcoding of the National Museum of Natural History reptile tissue holdings raises concerns about the use of natural history collections and the responsibilities of scientists in the molecular age. PLoS One 17, e0264930.

Nečas, T., Kielgast, J., Nagy, Z.T., Chifundera, Z.K., Gvoždík, V., 2022. Systematic position of the clicking frog (*Kassinula* Laurent, 1940), the problem of chimeric sequences and the revised classification of the family Hyperoliidae. Molecular Phylogenetics and Evolution 174, 1–11.

Nguyen, L.T., Schmidt, H.A., von Haeseler, A., Minh, B.Q., 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Molecular Biology and Evolution 32, 268–274.

Nilsson, R.H., Ryberg, M., Kristiansson, E., Abarenkov, K., Larsson, K.-H., Körljalg, U., 2006. Taxonomic reliability of DNA sequences in public sequence databases: a fungal perspective. PLoS One 1, e59.

Peloso, P.L.V., Raxworthy, C.J., Wheeler, W.C., Frost, D.R., 2017. Nomenclatural stability does not justify recognition of paraphyletic taxa: a response to Scherz et al. (2016). Mol. Phylogenet. Evol. 111, 56–64.

Peloso, P.L.V., Frost, D.R., Richards, S.J., Rodrigues, M.T., Donnellan, S., Matsui, M., Raxworthy, C.J., Biju, S.D., Moriarty Lemmon, E., Lemmon, A.R., Wheeler, W.C., 2016. The impact of anchored phylogenomics and taxon sampling on phylogenetic inference in narrow-mouthed frogs (Anura, Microhylidae). Cladistics 32, 113–140.

Philippe, H., Snell, E.A., Bapteste, E., Lopez, P., Holland, P.W.H., Casane, D., 2004. Phylogenomics of eukaryotes: impact of missing data on large alignments. Molecular Biology and Evolution 21, 1740–1752.

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). Molecular Ecology Resources 16, 1069–1083.

Portik, D.M., Wiens, J.J., 2020. SuperCRUNCH: a bioinformatics toolkit for creating and manipulating supermatrices and other large phylogenetic datasets. Methods in Ecology and Evolution 11, 763–772.

Portik, D.M., Bell, R.C., Blackburn, D.C., Bauer, A.M., Barratt, C.D., Branch, W.R., Burger, M., Channing, A., Colston, T.J., Conradie, W., Dehling, J.M., Drewes, R.C., Ernst, R., Greenbaum, E., Gvoždík, V., Harvey, J., Hillers, A., Hirschfeld, M., Jongsma, G.F.M., Kielgast, J., Kouete, M.T., Lawson, L., Leaché, A.D., Loader, S.P., Lötters, S., van der Meijden, A., Menegon, M., Müller, S., Nagy, Z.T., Ofori-Boateng, C., Ohler, A., Papenfuss, T.J., Rößler, D., Sinsch, U., Rödel, M.-O., Veith, M., Vindum, J., Zassi-Boulou, A.-G., McGuire, J.A., 2019. Sexual dichromatism drives diversification within a major radiation of African amphibians. Systematic Biology 68, 859–875.

Portik, D.M., Papenfuss, T.J., 2015. Historical biogeography resolves the origins of endemic Arabian toad lineages (family Bufonidae): evidence for ancient vicariance and dispersal events with the Horn of Africa and South Asia. BMC Evolutionary Biology 15, 152.

Portik, D.M., Streicher, J.W., Blackburn, D.C., Moen, D.S., Hutter, C.R., Wiens, J.J., 2023. Redefining possible: combining phylogenomic and supersparse data in frogs. Molecular Biology and Evolution 40, msad109. https://doi.org/10.1093/molbev/msad109.

Portik, D.M., Wiens, J.J., 2021. Do alignment and trimming methods matter for phylogenomic (UCE) analyses? Systematic Biology 70, 440–462.

Pyron, R.A., 2014. Biogeographic analysis reveals ancient continental vicariance and recent oceanic dispersal in amphibians. Systematic Biology 63, 779–797.

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. Molecular Phylogenetics and Evolution 61, 543–583.

Pyron, R.A., Wiens, J.J., 2013. Large-scale phylogenetic analyses reveal the causes of high tropical amphibian diversity. Proceedings of the Royal Society of London B, Biological Sciences 280, 20131622.

Ranwez, V., Douzery, E.J.P., Cambon, C., Chantret, N., Delsuc, F., 2018. MACSE v2: toolkit for the alignment of coding sequences accounting for frameshifts and stop codons. Molecular Biology and Evolution 35, 2582–2584.

Roelants, K., Gower, D.J., Wilkinson, M., Loader, S.P., Biju, S.D., Guillaume, K., Moriau, L., Bossuyt, F., 2007. Global patterns of diversification in the history of modern amphibians. Proceedings of the National Academy of Sciences of the United States of America 104, 887–892.

Sanderson, M.J., 2001. Estimating absolute rates of molecular evolution and divergence times: a penalized likelihood approach. Molecular Biology and Evolution 19, 101–109.

Sanderson, M.J., McMahon, M.M., Steel, M., 2011. Terraces in phylogenetic tree space. Science 333, 448–450.

Scherz, M.D., Vences, M., Rakotoarison, A., Andreone, F., Köhler, J., Glaw, F., Crottini, A., 2016. Reconciling molecular phylogeny. Morphological divergence and classification of Madagascan narrow-mouthed frogs (Amphibia: Microhylidae). Molecular Phylogenetics and Evolution 100, 372–381.

Schliep, K.P., 2011. phangorn: Phylogenetic analysis in R. Bioinformatics 27, 592–593.

Shen, X.X., Liang, D., Feng, Y.J., Chen, M.Y., Zhang, P., 2013. A versatile and highly efficient toolkit including 102 nuclear markers for vertebrate phylogenomics, tested by resolving the higher level relationships of the Caudata. Molecular Biology and Evolution 30, 2235–2248.

Smith, S.A., O'Meara, B.C., 2012. treePL: divergence time estimation using penalized likelihood for large phylogenies. Bioinformatics 28, 2689–2690.

Stamatakis, A., 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30, 1312–1313.

Streicher, J.W., Schulte, J.A., Wiens, J.J., 2016. How should genes and taxa be sampled for phylogenomic analyses with missing data? An empirical study in iguanian lizards. Systematic Biology 65, 128–145.

Streicher, J.W., Miller, E.C., Guerrero, P.C., Correa, C., Ortiz, J.C., Crawford, A.J., Pie, M. R., Wiens, J.J., 2018. Evaluating methods for phylogenomic analyses, and a new

phylogeny for a major frog clade (Hyloidea) based on 2,214 loci. Molecular Phylogenetics and Evolution 119, 128–143.
Streicher, J.W., Loader, S.P., Varela-Jaramillo, A., Montoya, A., de Sá, R.O., 2020. Analysis of ultraconserved elements supports African origins of narrow-mouthed frogs. Molecular Phylogenetics and Evolution 146, 106771.
Talavera, G., Lukhtanov, V., Pierce, N.E., Vila, R., 2022. DNA barcodes combined with multilocus data of representative taxa can generate reliable higher-level phylogenies. Systematic Biology 71, 382–395.
Tonini, J.F.R., Beard, K.H., Ferreira, R.B., Jetz, W., Pyron, R.A., 2016. Fully-sampled phylogenies of squamates reveal evolutionary patterns in threat status. Biological Conservation 204, 23–31.
Traseger, S.J., Maynard, R.J., Culebras, J., Kohn, S., Quezada, A., Guayasamin, J.M., 2021. Phylogenetic position of "*Cochranella*" *megista* (Anura: Centrolenidae) and first records for Ecuador. Phyllomedusa 20, 27–35.
Tu, N., Yang, M.H., Liang, D., Zhang, P., 2018. A large-scale phylogeny of Microhylidae inferred from a combined dataset of 121 genes and 427 taxa. Molecular Phylogenetics and Evolution 126, 85–91.
Upham, N.S., Esselstyn, J.A., Jetz, W., 2019. Inferring the mammal tree: species-level sets of phylogenies for questions in ecology, evolution, and conservation. PLoS Biology 17, e3000494.
Van Bocxlaer, I., Biju, S.D., Loader, S.P., Bossuyt, F., 2009. Toad radiation reveals into-India dispersal as a source of endemism in the Western Ghats-Sri Lanka biodiversity hotspot. BMC Evolutionary Biology 9, 131.
Wiens, J.J., 2003. Missing data, incomplete taxa, and phylogenetic accuracy. Systematic Biology 52, 528–538.
Wiens, J.J., 2005. Can incomplete taxa rescue phylogenetic analyses from long-branch attraction? Systematic Biology 54, 731–742.
Wiens, J.J., Fetzner, J.W., Parkinson, C.L., Reeder, T.W., 2005. Hylid frog phylogeny and sampling strategies for speciose clades. Systematic Biology 54, 778–807.
Wiens, J.J., Morrill, M.C., 2011. Missing data in phylogenetic analysis: reconciling results from simulations and empirical data. Systematic Biology 60, 719–731.
Wiens, J.J., Tiu, J., 2012. Highly incomplete taxa can rescue phylogenetic analyses from the negative impacts of limited taxon sampling. PLoS One 7, 42925.
Xi, Z., Liu, L., Davis, C.C., 2016. The impact of missing data on species tree estimation. Molecular Biology and Evolution 33, 838–860.
Zheng, Y., Wiens, J.J., 2015. Do missing data influence the accuracy of divergence-time estimation with BEAST? Molecular Phylogenetics and Evolution 85, 41–49.
Zheng, Y., Wiens, J.J., 2016. Combining phylogenomic and supermatrix approaches, and a time-calibrated phylogeny for squamate reptiles (lizards and snakes) based on 52 genes and 4162 species. Molecular Phylogenetics and Evolution 94, 537–547.
Zhou, X., Shen, X.-X., Hittinger, C.T., Rokas, A., 2018. Evaluating fast maximum likelihood-based phylogenetic programs using empirical phylogenomic data sets. Molecular Biology and Evolution 35, 486–503.