

## Supplementary Information for:

### Phylotranscriptomic consolidation of the jawed vertebrate timetree

Iker Irisarri, Denis Baurain, Henner Brinkmann, Frédéric Delsuc, Jean-Yves Sire, Alexander Kupfer, Jörn Petersen, Michael Jarek, Axel Meyer, Miguel Vences, Hervé Philippe

#### Supplementary Materials and Methods

- Transcriptome sequencing and assembly of 23 new transcriptomes
- Assembly of nuclear phylogenomic datasets
- Phylogenetic performance of transcriptomes in relation to sequencing effort
- Assembly of the mitogenomic datasets
- Phylogenetic inference
- Molecular dating
- Correlation of genome size and indels in protein-coding genes
- Correlation of nuclear substitution rates with genome size and species diversity
- Comparison of nuclear and mitochondrial substitution rates

#### Supplementary Tables and Figures

- Supplementary Table 1. Statistics of transcriptome sequencing effort, assembly and completeness
- Supplementary Table 2. Correlation among sequencing effort and transcriptome completeness
- Supplementary Table 3. Effect of alignment length on the recovery of nodes
- Supplementary Table 4. Recovery of 40 selected nodes by single genes from NoDP
- Supplementary Table 5. Topology tests performed on the nuclear test dataset
- Supplementary Table 6. Relationship between species diversity and substitution rates
- Supplementary Table 7. Bayesian correlation analysis performed in coevol
- Supplementary Table 8. Calibrations used in molecular dating analyses
- Supplementary Table 9. Divergence time estimates
- Supplementary Table 10. Taxon sampling of the nuclear datasets
- Supplementary Table 11. Taxon sampling of the mitochondrial datasets
- Supplementary Table 12. Non-vertebrate species used in BLASTP-based decontamination
- Supplementary Table 13. Model cross-validations performed in PhyloBayes

- Supplementary Figure 1. Comparison of datasets by quality and size
- Supplementary Figure 2. Flowchart of the new bioinformatics pipeline
- Supplementary Figure 3. ML phylogeny from NoDP under GTR+ $\Gamma$
- Supplementary Figure 4. ML phylogeny from the NoDP dataset
- Supplementary Figure 5. ASTRAL phylogeny from the NoDP dataset
- Supplementary Figure 6. BI phylogeny from 1DP under GTR+ $\Gamma$
- Supplementary Figure 7. BI phylogeny from 2DP under GTR+ $\Gamma$
- Supplementary Figure 8. ASTRAL phylogeny from the 1DP dataset
- Supplementary Figure 9. ASTRAL phylogeny from the 2DP dataset
- Supplementary Figure 10. BI from mitochondrial dataset (106 species) under CAT-GTR+ $\Gamma$
- Supplementary Figure 11. BI from mitochondrial dataset (106 species) under CAT+ $\Gamma$
- Supplementary Figure 12. ML from mitochondrial dataset (106 species) under GTR+ $\Gamma$
- Supplementary Figure 13. ML from mitochondrial dataset (106 species) under MTREV+ $\Gamma$
- Supplementary Figure 14. BI from mitochondrial dataset (95 species) under CAT-GTR+ $\Gamma$
- Supplementary Figure 15. BI from mitochondrial dataset (95 species) under CAT+ $\Gamma$
- Supplementary Figure 16. ML from mitochondrial dataset (95 species) under GTR+ $\Gamma$
- Supplementary Figure 17. ML from mitochondrial dataset (95 species) under MTREV+ $\Gamma$
- Supplementary Figure 18. Relationship between mitochondrial and nuclear substitution rates
- Supplementary Figure 19. Tree with IDs corresponding to Supplementary Tables 3, 4, 8, 9

#### Supplementary References

## Supplementary Materials and Methods

### Transcriptome sequencing and assembly of 23 new transcriptomes

Transcriptomes for the following ten amphibians (*Andrias davidianus*, *Calotriton asper*, *Discoglossus pictus*, *Geotrypetes seraphini*, *Hymenochirus curticeps*, *Megophrys nasuta*, *Pipa pipa*, *Proteus anguinus*, *Siren lacertina* and *Typhlonectes natans*) were generated at Leibniz Institute-DSMZ (German Collection of Microorganisms and Cell Cultures) as follows. Samples were taken from freshly sacrificed specimens and the variety of organs and tissues was maximized to allow the recovery of genes with tissue-specific expression, except *Andrias*, for which we used a skin biopsy preserved in liquid nitrogen. Samples were immediately homogenized in liquid nitrogen, transferred to TRIzol (Invitrogen, Carlsbad, CA, USA) and RNA was extracted using the guanidinium thiocyanate-phenol-chloroform method. Selection of mRNA used the PolyATtract mRNA Isolation System III (Promega, Mannheim, Germany). Sequencing was performed on Illumina MiSeq (2x250 bp paired-end; v2 chemistry, 500 cycles), each species using half a MiSeq flow cell. Transcriptomes were assembled *de novo* using the Trinity software package<sup>1</sup>. Likely coding regions were extracted using TransDecoder (included in the Trinity pipeline<sup>2</sup>) and functionally annotated by homology search against available sequence databases (using BLAST+ against SwissProt/Uniref90), protein domain identification (using HMMER against PFAM) and protein signal peptide and transmembrane domain prediction (using signalP/tmHMM).

Transcriptomes of eleven additional species (*Acipenser baerii*, *Amia calva*, *Lepisosteus platyrhincus*, *Pleurodeles waltl*, *Polypterus senegalus*, *Protopterus aethiopicus*, *Raja clavata*, *Rhinatrema bivittatum*, *Scyliorhinus canicula*, *Tarentola mauritanica* and *Typhlonectes compressicauda*) were obtained in the Evolution and Development of Skeleton group at the IBPS (Institut de Biologie Paris Seine) using a slightly different protocol. Individuals were sacrificed, their lower jaw dissected and immediately immersed in liquid nitrogen and pulverized. Total RNAs were then extracted and purified using RNeasy fibrous tissue Mini kit (Qiagen, France). Frozen RNAs were sequenced via RNA-seq using Illumina HiSeq2000 2x50 bp paired-end reads at GATC Biotech (Konstanz, Germany). *De novo* transcriptome assemblies from Illumina paired-end reads were performed at ISEM (Institut des Sciences de l'Evolution, Montpellier) using Trinity with default parameters. Expression values were obtained with SOAPdenovo-Trans (<http://soap.genomics.org.cn/SOAPdenovo-Trans.html>) with default settings. Details, assemblies and raw data for these eleven transcriptomes will be published elsewhere.

Transcriptomes for *Lepidosiren paradoxa* and *Neoceratodus forsteri* were obtained at the University of Konstanz (Germany) from a variety of tissues adequately stored in RNAlater. Total RNA was extracted with TRIzol (Invitrogen, Carlsbad, CA, USA), mRNA selected with poly-T and sequenced in a single lane of Illumina HiSeq2000 with 2x100 bp paired-end reads. Transcriptomes were assembled *de novo* using Trinity and Oases<sup>3</sup> under different parameterizations<sup>4</sup>.

Additionally, publicly available RNA-Seq data suitable for this study were downloaded from NCBI SRA (Supplementary Table 10) and transcriptomes were assembled using the Trinity software package with default parameters or the MIRA software package<sup>5</sup> with settings '-job=denovo,est,normal,solexa -SK:mnr=yes:nrr=10'.

### Assembly of nuclear phylogenomic datasets

Our taxon sampling comprised all major lineages of jawed vertebrates, including the 23 newly generated transcriptomes, and improved the sampling for several previously underrepresented lineages, especially within amphibians and lepidosaurians. Particular species were strategically chosen to minimize possible long-branch attraction artefacts, by selecting the

earliest branching and/or slower evolving species of each major clade (based on previous phylogenetic analyses), including the Elephant shark (the slowest-evolving vertebrate<sup>6</sup>), bichir, coelacanth, Australian lungfish, the caecilian *Rhinatrema*, *Andrias* and *Hynobius* salamanders, platypus, emu and ostrich, pleurodiran turtles or tuatara (the only exceptions where earliest-branching taxa could not be included due to sample unavailability were dibamid squamates and *Leiopelma* and *Ascaphus* frogs<sup>7,8</sup>).

Nuclear datasets were assembled using the new proposed pipeline summarized in Supplementary Fig. 2. Twenty one vertebrate proteomes were downloaded from Ensembl v.73 (*Anas platyrhynchos*, *Anolis carolinensis*, *Canis familiaris*, *Danio rerio*, *Dasybus novemcinctus*, *Gallus gallus*, *Homo sapiens*, *Latimeria chalumnae*, *Lepisosteus oculatus*, *Loxodonta africana*, *Macropus eugenii*, *Meleagris gallopavo*, *Monodelphis domestica*, *Mus musculus*, *Ornithorhynchus anatinus*, *Pelodiscus sinensis*, *Petromyzon marinus*, *Sarcophilus harrisii*, *Silurana tropicalis*, *Taeniopygia guttata*, *Takifugu rubripes*) and grouped into putative ortholog groups using USEARCH v.7.0.959<sup>9</sup> with parameters ‘-evaluate 1e-5 -accel 1’ and OrthoMCL v.2.0.3<sup>10</sup> with parameters ‘percentMatchCutoff=50, evaluateExponentCutoff=-5’ when loading UBLAST reports and an inflation parameter ‘-I=1.5’ when performing the Markov clustering itself. This created 24,177 gene clusters, that were filtered by requiring the presence of at least one sequence for each of the following groups: mammals, birds, reptiles + amphibians, spotted gar. This resulted in 11,656 gene clusters, which were aligned with MAFFT v7.127<sup>11</sup> with parameters ‘--quiet --localpair --maxiterate 5000 --reorder’ and positions of unreliable homology were discarded with Gblocks v0.91b<sup>12</sup> with default parameters. Six genes were excluded because Gblocks removed all positions.

In order to identify and address paralogy issues, we implemented a paralog-splitting pipeline based on single-gene trees. Alignments were subjected to phylogenetic inference with RAxML<sup>13</sup> under the LG+ $\Gamma$  model and the resulting trees were analysed using a custom Perl script, ParalogDetector.pl. Genes were classified into one of the following four categories: (i) single-copy genes, (ii) genes containing in-paralogs (i.e., multiple copies in a given species grouped in a monophyletic clade), (iii) genes with teleost-specific paralogs (produced by the teleost-specific whole genome duplication<sup>14,15</sup>), and (iv) genes containing out-paralogs (i.e., multiple copies in a given species not grouped into its respective monophyletic clade). The following clades were considered for paralogy detection: chondrichthyans, actinopterygians, coelacanth, lungfishes, caecilians, frogs, salamanders, turtles, crocodylians, birds, lepidosaurians, and mammals. Genes with out-paralogs were iteratively split into two sub-alignments according to the following procedure, implemented in the C program split-out-paralog.c: (i) the branch that maximizes the taxonomic diversity simultaneously in two subtrees (i.e., the sum of the number of species in each of the two partitions defined by the corresponding bipartition) is identified as putatively separating paralogous groups; if several branches fit this criterion, the longest one is selected; (ii) if and only if the identified branch separating two paralogous groups is sufficiently large (i.e., if the branch length is amongst the 10% longest branches of the tree), the alignment is split into two sub-alignments. The splitting procedure was repeated three times on the resulting alignments, which were discarded whenever the above taxonomic filter was not fulfilled or when the identified branch separating the putative paralogs was too short. Therefore, all alignments containing out-paralogs were either split or discarded. After out-paralog splitting, a total of 9,852 alignments remained.

These core alignments were complemented by mining genomic and transcriptomic data using the software Forty-Two (or “42”; <https://bitbucket.org/dbaurain/42/>), which was designed to enrich pre-existing alignments while controlling for orthology relationships. Briefly, “42” identifies orthologous sequences among homologous candidates in complete genomes or transcriptomes using multiple reference proteomes as controls in strict three-way reciprocal best BLAST hit tests.

To ensure the high quality of the alignments, several decontamination steps were carried out. A first step was based on sequence similarity searches. We used BLASTP<sup>16</sup> against a custom database containing both vertebrate (*Danio rerio*, *Gallus gallus* and *Homo sapiens*) and non-vertebrate (Supplementary Table 12) proteins to remove likely non-vertebrate contaminants. Sequences having a non-vertebrate best hit with an e-value at least 3 log<sub>10</sub> units lower than the e-value of the best vertebrate hit were considered putative contaminants and consequently removed. A further BLASTP search used specific databases for insects, nematodes and flatworms (Platyhelminthes) that were underrepresented in the previous database. Due to the smaller size of these databases, a difference of 1 log<sub>10</sub> unit in e-value was considered sufficient to remove putative contaminant sequences. A third BLAST search was used to specifically identify contamination by human sequences likely to occur during specimen collection or data generation steps. To increase the specificity of BLAST searches, we used the nucleotide rather than the amino acid sequences, and ran BLASTN searches against human CDS sequences (downloaded from Ensembl v.75). All sequences having an identity >98% with human were removed.

A second step consisted in eliminating cross-contaminations among newly generated libraries that were processed together during wet lab work. In order to do so, we searched for sequences that were clearly misplaced in gene trees (e.g., a sequence from a frog clustering within caecilians in gene trees) and extracted pairs of putative cross-contaminants. Specifically we tested cross-contamination between (i) *Tarentola* vs. *Polypterus*, (ii) *Neoceratodus* vs. *Lepidosiren* and (iii) and all pairwise combinations among *Andrias*, *Calotriton*, *Discoglossus*, *Geotrypetes*, *Megophrys*, *Proteus*, *Siren*, and *Typhlonectes*. Because contaminants are expected to occur at a lower concentration than the genuine gene, we estimated mRNA expression values and removed the sequence with the lowest FPKM in each pair. For the *Tarentola-Polypterus* pair, expression values were derived from SOAP assemblies, whereas in all other cases, FPKM values were estimated with RSEM<sup>17</sup> after read mapping with Bowtie v1.1.2<sup>18</sup>. This procedure removed 400 individual sequences.

In a third step, we extended the search for putative contaminants to all other downloaded data, using nuclear ribosomal and mitochondrial proteins as indicators because they are highly expressed and thus more prone to reveal contaminations. Nuclear ribosomal proteins present among our alignments were subjected to BLASTP searches against a reference database of metazoan ribosomal proteins. For mitochondrial sequences in our alignments, the corresponding nucleotide sequences were retrieved and subjected to BLASTN searches against a reference database that included available mitochondrial genomes and partial mitochondrial gene sequences for the species included in the study. This latter analysis was particularly efficient at pinpointing high similarities between unrelated species, given the overall faster evolutionary rate of mitochondrial proteins. Altogether, our analyses identified several instances of contamination, for which we performed additional decontamination steps as above. Finally, sequences from *Emys*, *Micrurus* and *Protopterus* were subjected to BLASTP searches against databases of available proteomes for birds, rat (*Rattus*) and squirrel monkey (*Saimiri*), respectively. In these cases, a slightly lower similarity threshold (95% identity) was used to identify contaminants. A total of 2,743 putative contaminating sequences were removed following these approaches. We assessed the effectiveness of all these decontamination steps by checking that the proportion of putatively contaminants was indeed reduced.

Limitations in assembly software based on De-Bruijn-graphs lead to the reconstruction of numerous highly similar transcripts (often of different length) and chimerical sequences in transcriptome assemblies, mainly due to differences between heterozygous alleles or sequencing errors. Since our dataset is in large part based on transcriptomic data, these problematic sequences were removed from our alignments by sequence similarity searches. For each species represented by at least two sequences in a given alignment, every sequence was

BLASTed against all the other sequences in the alignment, which allowed us to compute an average BLAST bit score for each potentially redundant sequence. A sequence was eliminated upon meeting the following two criteria: average bit score  $\geq 10\%$  lower than the best average bit score of the redundant set and a length overlap of  $\geq 95\%$  between the two sequences. This approach was implemented in the C program `detect-divergent-sequence.c` and resulted in the removal of 119,381 sequences.

In a further step, we used a second taxonomic filter to select alignments with a rich taxon sampling. Specifically, we filtered out all alignments that did not contain data for  $\geq 45$  species and at least 9 out of 10 major clades (chondrichthyans, actinopterygians, coelacanth, lungfishes, amphibians, turtles, crocodylians, birds, lepidosaurians, mammals). After this taxonomic filtering, a total of 7,687 alignments remained. To reduce stochastic error in the estimation of gene trees, we discarded both short sequences and alignments by (after ignoring columns with  $>25\%$  of missing data) requiring that all sequences have  $\geq 50$  amino acids and alignments contain  $\geq 100$  positions.

Even if paralogy had been addressed in the step where alignments contained 21 reference species (see above), new paralogy issues were expected to arise following the taxonomic enrichment of our alignments, either because paralogy only becomes apparent after the addition of further species or due to limitations of the heuristics of the software “42”. Since we were mainly concerned by the relationships among the 10 major vertebrate clades listed above, we considered two types of out-paralogy: (i) out-in-paralogs were defined as sequences of the same species that did not cluster in a monophyletic clade but still grouped within its taxonomic lineage (e.g., within ray-finned fishes), whereas (ii) out-out-paralogs (or deep paralogs) were defined when one of such sequences grouped outside of their corresponding clade. For each alignment, we computed the number of out-in-paralogs and of out-out-paralogs with the C program `detect-problem-arb.c` based on single-gene trees. Phylogenies were inferred using RAxML under the LG+ $\Gamma$  model, after discarding sequences containing  $< 50$  parsimony-informative positions to reduce stochastic error. 498 alignments containing  $\geq 3$  deep paralogs were discarded while the remaining genes were classified into 3 sets: no deep paralogs (4,593 genes), one (1,162) and two (1,434) deep paralogs.

To automatically remove sequencing errors (particularly those generating reading frame shifts) and annotation errors, which are frequent in high-throughput data<sup>19,20</sup>, we used the software HMMCleaner (R. Poujol, unpublished). Briefly, for each sequence of an alignment, a Hidden Markov Model (HMM) profile is computed for the alignment minus the sequence using HMMER (<http://hmmer.org>). Then, every region of that sequence having diverged more than a specific accumulative score from the HMM profile is discarded (we used the empirically estimated threshold of 10, which showed good performance for the sequence divergences in the range of those in our data; Amemiya, et al.<sup>21</sup>). Poorly aligned regions were removed with the software BMGE<sup>22</sup>, which uses entropy-like scores weighted by similarity matrices to identify poorly aligned positions and distinguish them from biologically expected variability. BMGE was run with stringent parameters ‘-h 0.5 -m BLOSSUM95’ and positions with  $>75\%$  of gaps or missing data were removed.

For each gene alignment, we used SCaFoS<sup>23</sup> to merge sequences belonging to the same species and to resolve putative remaining paralogy issues, creating alignments with up to 100 taxa each. In the presence of multiple partial sequences per species, SCaFoS collapses them into a single longer sequence, and if multiple complete sequences per species are present, the slowest-evolving sequence is selected. To maximize completeness of our dataset, we allowed the making of chimerical sequences in two cases, by merging data from closely related species of the same genus (*Rana chensinensis* + *Rana kukunoris* and *Chelonoidis carbonaria* + *Chelonoidis nigra*).

A fourth last decontamination step was performed to pinpoint problematic sequences by using branch lengths estimated from gene trees. This automatic procedure was designed to detect possible problems that might remain after all previous decontamination steps, removal of poorly aligned positions and sequence merging. The rationale for this analysis is that remaining erroneous sequences might become apparent through the generation of extremely long branches after enforcing the tree topology to a putative species phylogeny. First, a reference tree was estimated with RAxML under the LG+ $\Gamma$  model using a subset of genes corresponding to the taxonomically most complete alignments (containing >95 taxa; a total of 191,949 amino acid positions). For each gene, branch lengths were estimated with RAxML under the LG+ $\Gamma$  model after fixing the topology to the reference tree. Terminal branch lengths leading to individual taxa in each single-gene tree were then compared to those in the reference tree. Sequences showing an appreciably ( $\geq 5$  times) longer branch in a single gene tree than in the reference tree were eliminated from the corresponding alignment. A total of 4,891 sequences were removed following this procedure.

Finally, fully decontaminated alignments were concatenated again (as described above) and allowed us to generate three supermatrices: (i) the NoDP dataset for the 4,593 genes without deep paralogs (1,964,439 amino acid positions, 32% missing data), (ii) the 1DP dataset for the 1,162 genes with at most one putative deep paralog (668,132 amino acids; 36% missing data), and (iii) the 2DP dataset for the 1,434 genes with at most two putative deep paralogs (1,158,929 amino acids, 39% missing data). The three gene sets were analysed separately to assess the impact of paralogy on the estimated phylogenetic trees.

#### Phylogenetic performance of transcriptomes in relation to sequencing effort

We generated a total of 23 new transcriptomes, using between 6 and 514 million reads per species. Sequencing effort (measured as total sequenced bp) was correlated with transcriptome assembly measures (mean, median, N50), the proportion of recovered core vertebrate genes (CVG<sup>24</sup>), human orthologs and total number of amino acids in final alignments. The CVG protein set contains 233 reference orthologs and has been specifically designed to assess the completeness of vertebrate transcriptomes. The number of recovered CVG was estimated using the server gVolante (<https://gvolante.riken.jp/>) with the aid of BUSCO<sup>25</sup>. The total numbers of human proteins recovered at full (100%) and nearly-full (70%) length were estimated using the scripts of the Trinity pipeline<sup>2</sup> and BLASTX searches used DIAMOND<sup>26</sup>. As reference, we used the peptides predicted from the human genome (assembly GRCh38) downloaded from ENSEMBL. All 233 CVG genes and 99.3% (7136/7189) of human peptides were present in our dataset, demonstrating that the proportion of CVG and human peptides accurately quantify the completeness of transcriptomes and the number of genes readily usable for phylogenomics, respectively.

#### Assembly of the mitogenomic datasets

Available mitogenomes were downloaded from NCBI with a taxon sampling that essentially mirrors the above nuclear datasets. For species lacking mitogenomes, we replaced them if possible by those from closely related taxa. In addition, the partial mitogenome of *Tupinambis teguixin* was added because of its key phylogenetic position. To this end, mitochondrial proteins were identified through TBLASTN searches<sup>27</sup> against its transcriptome and added to the dataset. A few key taxa not present in the nuclear datasets were also included in the mitochondrial dataset to break long branches and thereby reduce long-branch attraction artefacts expected from fast-evolving and/or distantly related lineages (all taxa and NCBI accession numbers are available in Supplementary Table 11). The mitochondrial dataset was based on a curated alignment of animal mitogenomes (generously provided by Didier Casane, Université Paris Diderot) to which new species were added. All mitochondrial proteins were

concatenated except *nad6*, which was excluded to avoid potential problems generated by compositional heterogeneity (*nad6* is the only protein encoded on the minus strand<sup>7,28</sup>). A first mitochondrial dataset consisted of 106 species, from which only few particularly fast-evolving taxa were removed to generate a second mitochondrial dataset with 95 taxa aimed at reducing strong long-branch attraction artefacts detected in preliminary phylogenetic analyses. The MUST program package was used for sequence manipulation and visualization<sup>29</sup>. Highly variable and/or poorly aligned positions and positions with >50% gaps were eliminated with BMGE<sup>22</sup> using a stringent entropy cut-off ( $h=0.4$ ). The initial dataset had 106 taxa and 2,773 remaining amino acid positions (2,086 variable and 687 constant), while the second dataset consisted of 95 taxa and 2,866 amino acid positions (2,072 variable and 794 constant).

### Phylogenetic inference

The three nuclear gene sets NoDP, 1DP and 2DP were analysed separately. Maximum likelihood (ML) trees were reconstructed in RAxML v.8 under both LG+F+ $\Gamma$  and GTR+ $\Gamma$  amino acid models by analysing the complete concatenated matrices. Bayesian inference (BI) was performed using a gene jackknifing approach<sup>30</sup>, both for computational tractability and to estimate tree topologies and branch lengths that were robust to gene sampling. For each set of genes (NoDP, 1DP, 2DP), 100 replicates of ~50,000 amino acid positions were generated by randomly sampling single-gene alignments without replacement (~180 genes per replicate). We used PhyloBayes MPI v.1.5<sup>31</sup> to run a total of 300 MCMC chains under the CAT+ $\Gamma$  model for 10,000 cycles, saving every 10. For each set (NoDP, 1DP, 2DP), the corresponding 100 MCMC chains were summarized all at once by *bpcomp* after eliminating the first 10% data points as burnin. Convergence was verified using the diagnostic tools implemented in PhyloBayes. Branch support was assessed by the frequency of recovery across the 100 gene jackknife replicates in BI and across 1,000 pseudo-replicates of non-parametric bootstrapping in ML. To assess the robustness of our analyses to gene sampling, we generated jackknife replicates of increasing lengths and analysed them under ML. Specifically, we used the NoDP gene set from which 100 replicates were generated for five length classes (*ca.* 2,500, 5,000, 10,000, 25,000 and 50,000 aligned positions). A total of 500 independent analyses were performed with RAxML under the LG+ $\Gamma$  model, and for each of the four length classes, we calculated jackknife proportions of bipartitions and descriptive statistics on individual branch lengths. The three nuclear datasets were also analysed under a species tree approach consistent with the multispecies coalescent using ASTRAL-II v.4.10.12<sup>32</sup> with ML gene trees as input (estimated under best-fit models in RAxML). Node support of coalescent-based trees was assessed by local quartet-based branch support<sup>33</sup> and multi-locus bootstrapping<sup>34</sup> with 100 replicates.

The two mitochondrial datasets were each analysed by both ML (RAxML v.7.9.5<sup>35</sup>) and BI (PhyloBayes v.3.3<sup>36</sup>). ML analyses were performed with MTREV+ $\Gamma$  and GTR+ $\Gamma$  models, whereas BI analyses used the site-heterogeneous CAT+ $\Gamma$  and CAT-GTR+ $\Gamma$  models. In BI, two independent MCMC chains were run for 10,000 cycles, saving every 10 and convergence was verified using the diagnostic tools implemented in PhyloBayes.

To assess the fit of different models to both the nuclear and mitochondrial datasets, we performed 10-fold model cross-validations. For computational efficiency, model fit of the nuclear dataset was done using training datasets of 50,000 amino acid positions and test datasets of 10,000 amino acid positions. For both the nuclear test and mitochondrial datasets, we used PhyloBayes to perform cross-validation for the following models: LG+ $\Gamma$ , CAT+ $\Gamma$  and CAT-GTR+ $\Gamma$ . PhyloBayes was run after the removal of constant positions ('-dc' option) for 1100 cycles and removing the first 10% as burnin. The performed model cross-validations clearly favoured CAT-GTR+ $\Gamma$  > CAT+ $\Gamma$  > LG+ $\Gamma$  for both the nuclear and mitochondrial datasets (Supplementary Table 13). However, we used the CAT+ $\Gamma$  model to analyse the 300

gene jackknife replicates from the nuclear datasets (NoDP, 1DP, 2DP) as a compromise between good model fit and computational efficiency.

### Molecular dating

As a first step, we compared the performance of different evolutionary models (LG+ $\Gamma$ , CAT+ $\Gamma$ , CAT-GTR+ $\Gamma$ ) in the estimation of divergence times using a 10-fold cross-validation in the same manner as mentioned before, but with the additional assumptions of a birth-death prior on divergence times, a lognormal autocorrelated clock model, a fixed topology (Fig. 2a) and five well-accepted fossil calibrations plus a prior on the root (nodes 100, 102, 104, 105, 106, 107 in Supplementary Table 8). MCMC chains were run for 1,100 cycles and the initial 10% were removed as burnin. Model cross-validation favoured CAT-GTR+ $\Gamma$  as the best-fit model (Supplementary Table 13), and all dating analyses relied on this model. A second cross-validation was performed to assess the fit of different clock models (strict clock, autocorrelated lognormal and uncorrelated), with all other parameters set as in the previous cross-validations. Results were conclusive in rejecting a strict molecular clock and provided equal support for autocorrelated and uncorrelated models (5/10 replicates; Supplementary Table 13). In order to select the clock models that gave more realistic divergence estimates, we compared the obtained divergence times against the literature. To do so, we fetched all available estimates from the Timetree database ([www.timetree.org](http://www.timetree.org)) and for each node of interest we compared the means of divergence times against our estimated dates. Divergence times estimated by autocorrelated, uncorrelated and strict clock models showed average differences of respectively 35.69, 52.09 and 52.39 Myr against those of the literature, all comparisons being highly significant (paired t-test  $p < 0.05$  after square root or log transformation to conform with normality). Because the differences were appreciably smaller for the autocorrelated lognormal model, this model was assumed for further timetree analyses.

In addition to the previous six fossil calibrations, 24 additional well-established calibration points with solid paleontological evidence were selected from the literature and used conservative intervals to account for dating and phylogenetic uncertainty (Supplementary Table 8). On top of that, we assessed possible inconsistencies among calibrations by two cross-validation procedures. A first strategy relied on the method outlined in Near et al.<sup>37</sup>. Independent timetrees were estimated using each of the 29 single calibration points (plus the root calibration) and differences among prior (=mean of the prior interval) and estimated ages were compared as detailed in Near et al.<sup>37</sup>. None of the 29 calibrations were found to be inconsistent by appreciable differences in (i) the sum of squared differences between estimated molecular and fossil ages for all other fossil-calibrated nodes (SS) nor (ii) the variance of the  $s$  parameter (averaged SS after the successive removal of calibrations with highest SS), as assessed by one-tailed F-tests ( $p > 0.05$ ) (see Near et al.<sup>37</sup> for details). A second strategy (leave-one-out) aimed at comparing prior and posterior estimates after removal of one calibration at a time. Otherwise, individual chains were run as above. We compared prior and posterior dates among the 29 fossil-bearing nodes (root calibration was always used) by checking whether (i) prior and posterior confidence intervals overlapped and whether (ii) posterior mean was included within the prior interval. Given the relatively narrow posterior intervals estimated by PhyloBayes, we found 8 and 14 incongruent fossil calibrations, respectively for the above methods.

Based on the results of calibration cross-validations, we used three different calibration schemes in addition to the root calibration: (i) including all 29 calibrations, (ii) 21 and (iii) 15 calibrations. For each of the three calibration schemes, two MCMC chains were run in PhyloBayes until convergence, using the test dataset (14,352 amino acid positions) without constant positions and assuming the CAT-GTR+ $\Gamma$  and autocorrelated lognormal clock models, a birth-death prior and the fixed topology in Fig. 2a. In addition to these analyses, we run



individual MCMC chains on the 100 sub-alignments of ~15,000 amino acids generated from the NoDP dataset by gene jackknifing. These chains were run for >15,000 cycles using the most conservative set of calibrations (15 calibrations plus root) and all other settings as mentioned above. In all cases, the first 10% cycles were excluded as burnin and high effective sample sizes (>100) indicated that timetree analyses reached convergence.

#### Correlation of genome size and indels in protein-coding genes

We tested for the association between genome size with evolutionary rate<sup>38</sup> and species diversity<sup>39</sup>. Genome size data (haploid DNA content c-value) were retrieved from the Animal Genome Size Database ([www.genomesize.com](http://www.genomesize.com)). When several measurements existed for a given species, these were averaged; 18 species lacked information but genome size could be approximated by averaging existing data from other species in the same genus; 17 additional species lacked any genome size information and thus were excluded from this analysis. We calculated the number of gaps in both conserved and variable gene regions after filtering out species with >10% gaps in the regions to correct for misalignments (likely due to bad genome annotations). Conserved and variable gene regions were defined on untrimmed gene alignments using BMGE ('-h 0.5'), which would respectively retain and exclude these regions. The total number of gaps in both conserved and variable regions was calculated for each species. We used Pearson's correlation to study the relationship between genome size and number of gaps both in conserved and variable gene regions after correcting for phylogenetic relatedness using independent contrasts<sup>40</sup>. All calculations were performed in R<sup>41</sup> with the help of the ape package<sup>42</sup>.

As a complementary analysis, we used the Bayesian methods implemented in coevol v.1.4b<sup>43</sup> to study the correlation between parameters of the substitution process, genome size and the number of gaps in conserved and variable gene regions. For this analysis, we used a subset of the 41,360 most complete positions from the NoDP dataset (after removing positions with >5% gaps) and fixed topology and divergence times to those in Fig. 3. Two Kr/Kc amino acid models were used, either '-polvol' or '-charge', which model the rates of radical versus conservative amino acid replacements (analogous to dN/dS), considering as radical replacements those not conserving the polarity and/or volume ('-polvol') or charge ('-charge'). Two MCMC analyses were run for each of the models for 1,000 generations. All chains reached convergence, judged by the high effective sample sizes (>200). Results are available in Supplementary Table 7.

#### Correlation of nuclear substitution rates with genome size and species diversity

Nuclear substitution rates were calculated as root-to-tip paths from the BI tree of Fig. 2a, each divided by the total root age derived from the timetree (Fig. 3) to obtain absolute substitution rates. The correlation between nuclear substitution rates and genome size was performed on a total of 83 taxa (excluding those with no available genome size) using Pearson's correlation after correcting for phylogenetic non-independence. Data on current species diversity was obtained from web resources (Supplementary Table 6) and tabulated for a total of 44 lineages present in our tree, which were defined as lineages that split from their sister groups >150 Mya, an ad-hoc determined cut-off that allowed comparing lineages with obviously different diversity. Substitution rates either averaged for each clade or using the slowest rate per clade, were correlated with the species numbers using Pearson's correlation on phylogenetic independent contrasts.

We used a second approach to study the relationship between substitution rates and species diversity by comparing sister groups. For each of the mentioned 44 lineages, we identified the slowest-evolving species and performed relative-rate tests for all sister groups using the test dataset, as implemented in MEGA v.7<sup>44</sup>. For each sister group, we tabulated whether

evolutionary rates were significantly different ( $p < 0.05$ ) together with their corresponding species diversity, and assessed whether the observed differences were consistent using a Sign test.

### Comparison of nuclear and mitochondrial substitution rates

To visualize the among-lineage differences in the mitochondrial versus nuclear substitution rates, we estimated the nuclear and mitochondrial substitution rates per million year for each individual internal and terminal branch in the tree. This analysis was based on a subset of 78 species that were present in both the nuclear and mitochondrial datasets. We pruned the 22 species not present in the mitochondrial dataset from the BI tree from the NoDP dataset (Fig. 2a) using ape<sup>42</sup>. Species absent in the nuclear tree were removed from the mitochondrial dataset and branch lengths were estimated using the above reference pruned nuclear tree in PhyloBayes 3.3e under the CAT+ $\Gamma$  model. The resulting trees (Supplementary Fig. 18) were plotted in R using phytools<sup>45</sup>.

We further studied the relationship between nuclear and mitochondrial substitution rates by correlating the lengths of all individual internal and external branches in the nuclear and mitochondrial trees. For this analysis, nuclear branch lengths were first optimized using an alignment of approximately the same length as the mitochondrial dataset (2,866 positions) produced by gene jackknifing. Correlation between nuclear and mitochondrial branch lengths were assessed using Pearson's correlation. Using datasets of equal size excludes the effect of estimating branch lengths with different precisions. In addition, we discard the effect of gene sampling on the estimated nuclear branch lengths, as demonstrated by the very strong correlation with branch lengths estimated from 100 independent gene jackknife replicates of *ca.* 2,866 positions drawn from the NoDP dataset (average  $r = 0.94 \pm 0.02$ ). In order to fully understand the extent of mitochondrial to nuclear rate correlation obtained above, we calculated the correlation that can be expected from random datasets characterized by the amino acid frequencies of our nuclear and amino acid datasets. In order to do so, we used Seq-Gen<sup>46</sup> to simulate random gene alignments using the pruned nuclear reference tree and the LG+ $\Gamma$  model, and additionally assuming the amino acid frequencies of either the nuclear and mitochondrial datasets. A total of 100 "random nuclear" and 100 "random mitochondrial" alignments were generated, and estimated branch lengths were correlated using 100 correlation analyses for each pair of random datasets.

## Supplementary Tables and Figures

**Supplementary Table 1.** Statistics of sequencing effort, assembly and completeness of new transcriptomes. Sequencing effort: number of reads (Num.), read length (Length) and total sequenced base pairs (Total bp). Assembly statistics: mean and median of transcript lengths and N50. Number of amino acids in NoDP dataset (AANoDP; 4,593 genes) and for the total of all three nuclear datasets (AATot; 7,189 genes). Assembly completeness: number of recovered Core Vertebrate Genes (CVG). Number of recovered proteins: human genes recovered at full (100%) and nearly full (>70%) length coverage.

Species	Raw clean reads (Millions)			Assembly statistics			Num. amino acids		CVG	Human proteins	
	Num.	Length	Total bp	Mean	Median	N50	AANoDP	AATot	(%)	70%	100%
<i>Megophrys nasuta</i>	6.10	250	1,524.16	503	325	616	1,018,306	1,856,209	196 (84%)	5,118	2,601
<i>Calotriton asper</i>	6.63	250	1,658.62	533	347	685	1,000,947	1,686,050	189 (81%)	5,100	2,778
<i>Typhlonectes natans</i>	8.22	250	2,055.69	472	319	540	678,646	1,101,327	143 (61%)	3,423	1,851
<i>Geotrypetes seraphini</i>	8.10	250	2,244.31	470	318	543	948,260	1,645,775	177 (76%)	4,415	2,255
<i>Proteus anguinus</i>	9.18	250	2,294.97	493	327	581	815,678	1,377,722	167 (72%)	3,968	2,088
<i>Andrias davidianus</i>	11.46	250	2,866.12	365	258	355	1,434,893	2,506,140	137 (59%)	2,274	1,123
<i>Pipa pipa</i>	11.83	250	2,956.80	479	310	590	687,849	1,116,268	158 (68%)	3,562	1,898
<i>Discoglossus pictus</i>	14.14	250	3,534.38	473	308	570	935,026	1,627,075	179 (77%)	4,387	2,175
<i>Siren lacertina</i>	14.22	250	3,554.94	459	295	533	771,029	1,299,717	159 (68%)	3,560	1,800
<i>Hymenochirus curticeps</i>	16.18	250	4,045.05	544	319	758	1,156,199	2,123,126	197 (85%)	6,131	3,241
<i>Neoceratodus forsteri</i>	72.69	100	7,268.54	762	423	1,269	1,304,999	2,285,528	207 (89%)	10,847	6,134
<i>Lepidosiren paradoxa</i>	102.10	100	10,209.47	731	491	1,539	1,673,662	3,141,799	221 (95%)	13,642	7,433
<i>Scyliorhinus canicula</i>	258.90	50	12,944.90	1142	487	2,445	1,600,775	3,040,809	220 (94%)	9,872	6,692
<i>Protopterus annectens</i>	137.39	100	13,739.38	670	434	1,577	1,533,239	2,815,136	215 (92%)	13,186	7,587
<i>Protopterus aethiopicus</i>	328.67	50	16,433.39	1038	480	2,109	1,462,581	2,558,803	198 (85%)	8,238	5,681
<i>Rhinatrema bivittatum</i>	332.50	50	16,625.07	871	437	1,638	1,600,673	2,833,244	215 (92%)	9,343	6,083
<i>Polypterus senegalus</i>	355.43	50	17,771.39	1198	481	2,669	1,644,586	3,128,920	223 (96%)	10,833	7,534
<i>Lepisosteus patyrhincus</i>	359.89	50	17,994.57	1784	835	3,767	1,681,766	3,209,672	233 (100%)	11,860	7,263
<i>Amia calva</i>	368.68	50	18,434.12	1342	638	2,708	1,543,408	2,859,757	221 (95%)	9,435	6,074
<i>Pleurodeles waltl</i>	381.04	50	19,051.82	1447	561	3,298	1,758,559	3,328,982	229 (98%)	11,610	8,467
<i>Typhlonectes compressicauda</i>	383.93	50	19,196.26	843	425	1,558	1,556,391	2,814,845	204 (88%)	8,862	6,020
<i>Tarentola mauritanica</i>	392.79	50	19,639.30	1333	554	2,944	1,787,455	3,378,576	227 (97%)	12,051	7,819
<i>Raja clavata</i>	447.16	50	22,357.93	859	374	1,887	1,535,230	2,907,885	225 (97%)	10,020	6,779
<i>Acipenser baerii</i>	514.43	50	25,721.56	922	525	1,575	1,696,440	3,252,781	223 (96%)	11,260	7,200

**Supplementary Table 2.** Pearson’s correlation statistics (r) among measures of sequencing effort (total number of reads, total sequenced base pairs, read length), statistics of assembled transcripts (mean, median, N50), the total number of amino acids in NoDP (AANoDP) and all three nuclear datasets (AATot), transcriptome completeness (Core Vertebrate Genes, CVG) and human proteins at 70% and 100% levels of coverage (cov). Significant correlations are highlighted in bold. The upper triangle shows correlations across all new transcriptomes, whereas the lower triangle shows correlations only among deeply-sequenced transcriptomes (>10,000 million total sequenced base pairs).

	Total reads	Total bp	Read length	Mean	Median	N50	AANoDP	AATot	CVG	70% cov.	100% cov.
<b>Total reads</b>		<b>0.80</b>	<b>-0.91</b>	<b>0.98</b>	<b>0.71</b>	<b>0.81</b>	<b>0.81</b>	<b>0.82</b>	<b>0.76</b>	<b>0.71</b>	<b>0.82</b>
<b>Total bp</b>	0.35		<b>-0.82</b>	<b>0.78</b>	<b>0.94</b>	<b>0.99</b>	<b>0.75</b>	<b>0.77</b>	<b>0.78</b>	<b>0.72</b>	<b>0.80</b>
<b>Read length</b>	<b>-0.85</b>	-0.54		<b>-0.94</b>	<b>-0.77</b>	<b>-0.86</b>	<b>-0.89</b>	<b>-0.89</b>	<b>-0.85</b>	<b>-0.88</b>	<b>-0.94</b>
<b>Mean</b>	<b>0.95</b>	0.19	<b>-0.64</b>		<b>0.72</b>	<b>0.80</b>	<b>0.85</b>	<b>0.86</b>	<b>0.80</b>	<b>0.79</b>	<b>0.87</b>
<b>Median</b>	0.13	0.86	-0.21	0.07		<b>0.91</b>	<b>0.72</b>	<b>0.73</b>	<b>0.76</b>	<b>0.75</b>	<b>0.77</b>
<b>N50</b>	0.24	<b>0.97</b>	-0.43	0.09	<b>0.79</b>		<b>0.80</b>	<b>0.82</b>	<b>0.81</b>	<b>0.78</b>	<b>0.86</b>
<b>AANoDP</b>	0.18	0.45	-0.08	0.19	0.40	0.47		<b>1.00</b>	<b>0.82</b>	<b>0.86</b>	<b>0.89</b>
<b>AATot</b>	0.21	0.47	-0.07	0.23	0.43	<b>0.50</b>	<b>0.97</b>		<b>0.82</b>	<b>0.86</b>	<b>0.89</b>
<b>CVG</b>	0.19	0.55	-0.07	0.21	0.53	<b>0.60</b>	<b>0.72</b>	<b>0.83</b>		<b>0.91</b>	<b>0.92</b>
<b>70% cov.</b>	-0.47	-0.01	<b>0.70</b>	-0.27	0.22	0.10	<b>0.58</b>	<b>0.62</b>	<b>0.58</b>		<b>0.98</b>
<b>100% cov.</b>	-0.10	0.26	0.29	0.02	0.21	0.39	<b>0.77</b>	<b>0.82</b>	<b>0.71</b>	<b>0.82</b>	

**Supplementary Table 3.** Effect of alignment length on the recovery of the nodes in the reference phylogeny of Fig. 2a, measured as proportions from 100 gene jackknife replicates of increasing lengths: *ca.* 2,500, 5,000, 10,000, 25,000 and 50,000 amino acids (AA). Node IDs correspond to Supplementary Fig. 19.

Node ID	2500 AA	5000 AA	10,000 AA	25,000 AA	50,000 AA	Lineage
101	100	100	100	100	100	Ray-finned fishes (Actinopterygii)
102	100	100	100	100	100	
103	75	99	100	100	100	Sarcopterygians (Sarcopterygii)
104	57	75	87	93	100	Lungfish + tetrapod
105	87	100	100	100	100	Tetrapods
106	96	100	100	100	100	Amniotes
107	94	100	100	100	100	Archosaurs + turtles + Lepidosaurs (Diapsida)
108	87	99	100	100	100	Archosaurians + turtles
109	97	100	100	100	100	Archosaurians (Archosauria)
110	100	100	100	100	100	Crocodyles (Crocodylia)
111	88	96	99	100	100	
112	100	100	100	100	100	Birds (Aves)
113	96	100	100	100	100	
114	95	100	100	100	100	
115	100	100	100	100	100	
116	100	100	100	100	100	
117	99	100	100	100	100	Turtles (Testudines)
118	92	99	100	100	100	
119	96	100	100	100	100	
120	48	46	55	43	51	Position of <i>Caretta</i> within turtles
121	78	95	100	100	100	
122	93	98	99	100	100	
123	91	99	100	100	100	
124	81	96	98	100	100	
125	73	91	96	100	100	Lepidosaurians (Lepidosauria)
126	97	100	100	100	100	Squamates (Squamata)
127	68	94	99	100	100	node within Lepidosauria
128	78	87	98	100	100	
129	38	49	65	95	100	<i>Elgaria</i> + iguanian lizards + snakes
130	29	32	32	45	47	<i>Elgaria</i> + iguanian lizards
131	77	93	97	100	100	
132	99	99	100	100	100	
133	42	41	42	41	56	<i>Anolis</i> + <i>Sceloporus</i>
134	29	31	33	48	58	<i>Iguana</i> + <i>Basiliscus</i>
135	94	98	100	100	100	
136	100	100	100	100	100	Snakes (Serpentes)
137	79	90	92	100	100	
138	100	100	100	100	100	
139	100	100	100	100	100	
140	100	100	100	100	100	
141	100	100	100	100	100	
142	100	100	100	100	100	
143	88	99	100	100	100	
144	99	100	100	100	100	
145	31	46	60	89	98	<i>Podarcis</i> + <i>Tupinambis</i>
146	100	100	100	100	100	
147	100	100	100	100	100	
148	55	78	87	97	98	<i>Lampropholis</i> + <i>Saproscincus</i>
149	100	100	100	100	100	
150	98	100	100	100	100	Mammals (Mammalia)
151	97	100	100	100	100	
152	99	100	100	100	100	
153	62	85	95	100	100	Boreoeutheria
154	100	100	100	100	100	
155	70	83	97	100	100	Human + mouse
156	33	34	36	30	75	Armadillo + Elephant
157	100	100	100	100	100	
158	78	92	100	100	100	
159	89	100	100	100	100	Lissamphibia
160	58	66	85	96	100	Frogs + salamanders
161	89	100	100	100	100	Salamanders (Caudata)
162	41	54	69	71	99	Position of <i>Siren</i> within salamanders
163	77	98	100	100	100	
164	62	83	93	99	100	Position of <i>Ambystoma</i> within salamanders
165	88	100	100	100	100	
166	78	98	100	100	100	
167	52	67	80	97	100	<i>Notophthalmus</i> + <i>Calotriton</i> + <i>Cynops</i>
168	90	94	98	100	100	
169	100	100	100	100	100	
170	100	100	100	100	100	Frogs and toads (Anura)
171	62	69	89	98	100	Position of pipoidean frogs
172	56	57	80	95	95	Pelobatoidean + neobatrachian frogs
173	100	100	100	100	100	
174	100	100	100	100	100	
175	48	53	69	85	91	<i>Atelopus</i> + <i>Espadarana</i>
176	100	100	100	100	100	
177	79	90	94	96	97	
178	99	99	98	100	100	
179	54	65	67	82	98	<i>Hymenochirus</i> + <i>Silurana</i>
180	97	100	100	100	100	
181	100	100	100	100	100	Caecilians (Gymnophiona)
182	100	100	100	100	100	
183	100	100	100	100	100	
184	100	100	100	100	100	
185	100	100	100	100	100	
186	100	100	100	100	100	
187	87	99	100	100	100	Cartilaginous fishes (Chondrichthyes)
188	78	96	100	100	100	
189	98	98	100	100	100	
190	93	97	98	98	100	
191	100	100	100	100	100	
192	68	88	96	100	100	<i>Leucoraja</i> + <i>Neotrygon</i> + <i>Raja</i>
193	82	99	100	100	100	
194	70	75	88	99	100	<i>Acipenser</i> within ray-finned fishes
195	100	100	100	100	100	
196	97	97	100	100	100	
197	100	100	100	100	100	
198	100	100	100	100	100	

**Supplementary Table 4.** Recovery of 40 selected nodes by 4,593 single genes from the NoDP dataset. For each node of interest, the proportion of genes supporting its monophyly is shown. For comparison, the jackknife proportions (GJP) for replicates of size 2,500 and 25,000 amino acids (AA) are shown (extracted from Supplementary Table 3). Node IDs correspond to Supplementary Fig. 19.

Node ID	Monophyly % genes	GJP 2,500 AA	GJP 25,000 AA	Lineage
101	90.73	100	100	Ray-finned fish (Actinopterygii)
103	42.87	75	100	Sarcopterygians (Sarcopterygii)
104	31.98	57	93	Lungfish + tetrapod
105	82.67	87	100	Tetrapods
106	78.71	96	100	Amniotes
107	66.64	94	100	Archosauroids + turtles + lepidosaurs (Diapsida)
108	43.20	87	100	Archosauroids (Archosauria)
110	92.81	100	100	Crocodylians (Crocodylia)
112	86.85	100	100	Birds (Aves)
117	78.32	99	100	Turtles (Testudines)
120	49.88	48	–	Position of <i>Caretta</i> within turtles
125	67.38	73	100	Lepidosauroids (Lepidosauria)
126	82.03	97	100	Squamates (Squamata)
127	31.76	68	100	node within lepidosauria
129	25.35	38	95	<i>Elgaria</i> + iguanian lizards + snakes
130	16.77	29	–	<i>Elgaria</i> + iguanian lizards
133	27.47	42	–	<i>Anolis</i> + <i>Sceloporus</i>
134	22.76	29	48	<i>Iguana</i> + <i>Basiliscus</i>
136	89.15	100	100	Snakes (Serpentes)
145	15.40	31	89	<i>Podarcis</i> + <i>Tupinambis</i>
148	42.60	55	97	<i>Lampropholis</i> + <i>Saproscincus</i>
150	87.08	98	100	Mammals (Mammalia)
153	29.37	62	100	Boreoeutheria
155	31.72	70	100	Human + mouse
156	21.90	–	–	Armadillo + Elephant
159	57.51	89	100	Lissamphibia
160	33.55	58	96	Frogs + salamanders
161	91.43	89	100	Salamanders (Caudata)
162	39.73	41	71	Position of <i>Siren</i> within salamanders
164	47.96	62	99	Position of <i>Ambystoma</i> within salamanders
167	64.35	78	100	<i>Notophthalmus</i> + <i>Calotriton</i> + <i>Cynops</i>
170	89.59	100	100	Frogs and toads (Anura)
171	28.87	62	98	Position of pipoidean frogs
172	30.85	56	95	Pelobatoidean + neobatrachian frogs
175	42.46	48	85	<i>Atelopus</i> + <i>Espadarana</i>
179	47.22	54	82	<i>Hymenochirus</i> + <i>Silurana</i>
181	93.37	100	100	Caecilians (Gymnophiona)
187	95.98	87	100	Cartilaginous fishes (Chondrichthyes)
192	85.03	98	100	<i>Leucoraja</i> + <i>Neotrygon</i> + <i>Raja</i>
194	45.53	70	99	<i>Acipenser</i> within ray-finned fish

**Supplementary Table 5.** Results of topology tests performed on the nuclear test dataset. Table shows tested topologies and p-values of approximately unbiased (AU) and Shimodaira-Hasegawa (SH) tests. Significant p-values (<0.05) are highlighted in bold italics.

<b>Phylogenetic hypotheses</b>	<b>AU (p)</b>	<b>SH (p)</b>
Unconstrained BI tree from NoDP dataset	0.659	0.982
BI tree from 1DP dataset	0.639	0.981
BI tree from 2DP dataset (= ML tree under GTR+ $\Gamma$ )	0.093	0.394
Paraphyly of Lissamphibia <sup>47</sup>	<b><i>6e-07</i></b>	<b><i>0.002</i></b>
Caecilians + Salamanders are sister group (Procera hypothesis)	<b><i>0.001</i></b>	<b><i>0.031</i></b>
Turtles basal within amniotes	<b><i>9e-40</i></b>	<b><i>2e-04</i></b>
Turtles sister group to diapsids	<b><i>8e-12</i></b>	<b><i>4e-04</i></b>
Turtles sister group to lepidosaurs	<b><i>3e-33</i></b>	<b><i>6e-05</i></b>
Turtles sister group to crocodiles <sup>48</sup>	<b><i>8e-53</i></b>	<b><i>0</i></b>
Coelacanth sister group to tetrapods	<b><i>0.031</i></b>	0.210
Coelacanth + lungfishes basal within sarcopterygians	0.052	0.273
<i>Siren</i> basal in salamanders	<b><i>0.008</i></b>	0.122
<i>Anolis</i> basal to <i>Iguana</i> , <i>Basiliscus</i> and <i>Sceloporus</i>	<b><i>0.001</i></b>	<b><i>0.033</i></b>
<i>Elgaria</i> sister group to snakes	<b><i>0.016</i></b>	0.148
<i>Phrynops</i> basal to all other turtles <sup>49</sup>	<b><i>0.011</i></b>	0.094
( <i>Silurana</i> , ( <i>Pipa</i> + <i>Hymenochirus</i> ))	<b><i>0.002</i></b>	<b><i>0.043</i></b>
<i>Sthernotherus</i> + <i>Caretta</i> basal in turtles	<b><i>0.001</i></b>	<b><i>0.026</i></b>
Elephant (Afrotheria) basal in placental mammals	<b><i>0.001</i></b>	0.086
Armadillo (Xenarthra) basal in mammals	<b><i>0.022</i></b>	0.138

**Supplementary Table 6.** Relationship between species diversity and substitution rates. Representative taxa for 44 clades of ages >150 Myr were used. Rate comparisons refer to comparisons of average branch lengths of sister clades (>, longer branches; <, shorter branches than sister group), and significance values refer to relative rate tests (NS, not significant; \* p<0.05; \*\* p<0.005; \*\*\* p<0.001). Sources for species numbers are abbreviated as follows: CoL: Catalogue, of Life, www.catalogueoflife.org; AW: Amphibiaweb, www.amphibiaweb.org; RD: Reptile Database, www.reptile-database.org; all accessed on 22 November 2015. NA shows cases where the species diversity could not be determined.

Representative species	Clade name	Taxonomical correspondence	Evolutionary rate	Species Number	Rate difference	Diversity differences	Data source for diversity
<i>Acipenser baerii</i>	Sturgeons	Acipenseriformes	0.5993	27	< NS	<	CoL
<i>Ambystoma mexicanum</i>	Ambystomatid salamanders	Ambystomatidae + Dicamptodontidae	0.8502	36	>*	<	AW
<i>Amia calva</i>	Bowfin	Amiiformes	0.6772	1	< NS	<	CoL
<i>Andrias davidianus</i>	Giant salamanders	Cryptobranchidae	0.7855	3	< NS	<	AW
<i>Basiliscus plumifrons</i>	Iguanas	Iguanidae	0.6221	2450	NA	NA	RD
<i>Callorhynchus milii</i>	Ratfishes, rabbitfishes, and elephantfishes	Holocephali	0.3905	50	< NS	<	CoL
<i>Carcharodon carcharias</i>	Mackerel sharks and allies	Lamniformes	0.3960	16	< ***	<	CoL
<i>Crocodylus niloticus</i>	Crocodiles, alligators, and caimans	Crocodylia	0.5243	25	< NS	<	RD
<i>Cyclorana alboguttata</i>	Neobatrachian frogs	Neobatrachia	0.7157	6318	NA	NA	AW
<i>Danio rerio</i>	Teleost fishes	Teleostei	0.9934	31826	> ***	>	CoL
<i>Discoglossus pictus</i>	Fire-bellied toads, midwife toads, and painted frogs	Alytidae + Bombinatoridae	0.9242	21	< ***	<	AW
<i>Elgaria multicarinata</i>	Anguimorph lizards	Anguimorpha	0.8953	225	< NS	<	RD
<i>Emys orbicularis</i>	Hard-shelled cryptodiran turtles and tortoises	Durocryptodira	0.7806	217	NA	NA	RD
<i>Eublepharis macularius</i>	Geckos	Gekkota	0.6560	1650	< NS	<	RD
<i>Ginglymostoma cirratum</i>	Carp sharks and allies	Orectolobiformes	0.6837	44	> **	<	CoL
<i>Homo sapiens</i>	Placental mammals	Placentalia	0.6743	5485	NA	NA	CoL
<i>Hynobius chinensis</i>	Hynobiid salamanders	Hynobiidae	0.7911	64	NA	NA	AW
<i>Latimeria chalumnae</i>	Coelacanths	Coelacanthiformes	0.5012	2	< NS	<	CoL
<i>Lepisosteus platyrhincus</i>	Gars	Lepisosteiformes	0.8647	7	NA	NA	CoL
<i>Megophrys nasuta</i>	Spadefoot toads and allies	Megophryidae + Pelobatidae + Pelodytidae +	1.0450	205	> ***	<	AW
<i>Monodelphis domestica</i>	Marsupial mammals	Marsupialia	0.8727	335	< ***	<	CoL
<i>Neoceratodus forsteri</i>	Australian lungfish	Ceratodontiformes	0.5486	1	< ***	<	CoL
<i>Neotrygon kuhlii</i>	Stingrays and allies	non-rajooid Batoidea	0.4615	293	> NS	<	CoL
<i>Ornithorhynchus anatinus</i>	egg-laying mammals	Monotremata	0.8625	5	< NS	<	CoL
<i>Pelodiscus sinensis</i>	Pignose and softshell turtles	Trionychia (Trionychidae + Carettochelyidae)	0.7687	33	< NS	<	RD
<i>Pelusios castaneus</i>	Pelomedusid and podocnemidid turtles	Podocnemididae + Pelomedusidae	0.7618	35	> NS	<	RD
<i>Phrynops hilarii</i>	Austro-American sideneck turtles	Chelidae	0.7526	56	NA	NA	RD
<i>Podarcis</i> sp	Lacertid lizards and amphisbaenas	Lacertibaenia	0.9038	515	NA	NA	RD
<i>Pogona vitticeps</i>	Agamas and chameleons	Acrodonta	0.9496	661	> ***	<	RD
<i>Polypterus senegalus</i>	Bichirs and ropefish	Polypteriformes	0.6984	14	> NS	<	CoL
<i>Proteus anguinus</i>	Lungless and proteid salamanders, and allies	Proteidae + Rhyacotritonidae + Amphiumidae +	0.8585	461	> NS	>	AW
<i>Protopterus aethiopicus</i>	African and South American lungfishes	Lepidosireniformes	0.8333	5	NA	NA	CoL
<i>Python regius</i>	Snakes	Serpentes	0.8397	3567	> ***	>	RD
<i>Raja clavata</i>	Skates	Rajoidei	0.6005	357	NA	NA	CoL
<i>Rhinatrema bivittatum</i>	Rhinatreimid caecilians	Rhinatremitidae	0.7417	11	< ***	<	AW
<i>Salamandra salamandra</i>	Salamandrid salamanders	Salamandridae	0.7425	112	NA	NA	AW
<i>Scincella lateralis</i>	Skinks and allies	Scinciformata	0.7410	1736	> NS	<	RD
<i>Scyliorhinus canicula</i>	Ground sharks	Carcharhiniformes	0.4224	284	NA	NA	CoL
<i>Siren lacertina</i>	Sirens (salamanders)	Sirenidae	0.8731	4	> ***	<	AW
<i>Sphenodon punctatus</i>	Tuatara	Rhynchocephalia	0.7887	1	< ***	<	RD
<i>Struthio camelus</i>	Birds	Aves	0.7909	10306	NA	NA	CoL
<i>Tuinambis teguixin</i>	Whiptails, tegus, and allies	Teiiformata	0.9389	399	> ***	<	RD
<i>Typhlonectes compressicauda</i>	Non-rhinatreimid caecilians	all caecilian families except	0.9075	194	NA	NA	AW
<i>Xenopus tropicalis</i>	Clawed frogs and allies	Pipidae + Rhinophrynidae	1.0034	34	< NS	<	AW



**Supplementary Table 7.** Results from the Bayesian correlation analysis performed in coevol. Correlation among the following variables was studied: genome size (c-value), the number of gaps in conserved (gapscon) and variable (gapsvar) gene regions, the rate of conservative amino acid replacements (Kc) and the rate between radical and conservative amino acid replacements (Kr/Kc), estimated for two Kr/Kc models (“polvol” and “charge”; see Supplementary Materials and Methods for details). Correlation coefficients and the corresponding Bayesian posterior probabilities are shown in the lower and upper triangles, respectively (highlighted in bold when posterior probabilities >0.95).

<b>Correlation coefficients \ Posterior probabilities</b>					
<b>“polvol” model</b>					
	<b>Kc</b>	<b>Kr/Kc</b>	<b>c-value</b>	<b>gapsvar</b>	<b>gapscon</b>
<b>Kc</b>	–	0.84	0.80	0.22	0.42
<b>Kr/Kc</b>	0.238	–	<b>0.98</b>	0.0055	0.0016
<b>c-value</b>	0.141	<b>0.338</b>	–	0.03	0.23
<b>gapsvar</b>	-0.133	-0.58	-0.244	–	<b>1</b>
<b>gapscon</b>	-0.0305	-0.731	-0.0911	<b>0.361</b>	–
<b>“charge” model</b>					
	<b>Kc</b>	<b>Kr/Kc</b>	<b>c-value</b>	<b>gapsvar</b>	<b>gapscon</b>
<b>Kc</b>	–	0.84	0.73	0.28	0.44
<b>Kr/Kc</b>	0.24	–	0.78	0.02	0.18
<b>c-value</b>	0.121	0.121	–	0.041	0.28
<b>gapsvar</b>	-0.0939	-0.586	-0.23	–	<b>1</b>
<b>gapscon</b>	-0.0231	-0.213	-0.0704	<b>0.353</b>	–

**Supplementary Table 8.** Calibrations used in molecular dating analyses. Dates are in million years and node IDs correspond to Supplementary Fig. 19.

Node ID	Speciation event	Min. age	Max. age	Reference	Calibration scheme
100	Root (ray-finned vs. lobe-finned fishes)	421.75	462.5	<sup>50</sup>	C30, C16
102	Ray-finned vs. lobe-finned fishes	416	439	<sup>51</sup>	C30
104	Lungfishes vs. tetrapods	408	419	<sup>52</sup>	C30, C16
105	Amphibians vs. amniotes	330.4	350.1	<sup>53</sup>	C30, C16
106	Diapsids vs. synapsids	288	338	<sup>54</sup>	C30
107	Lepidosaurians vs. archosaurians + turtles	252	257	<sup>52</sup>	C30
109	Crocodiles vs. birds	243	251	<sup>55</sup>	C30
111	<i>Alligator</i> vs. <i>Caiman</i>	66	75	<sup>55</sup>	C30
113	Chicken vs. zebrafinch	66	86.5	<sup>53</sup>	C30
117	Cryptodiran vs. pleurodiran turtles	210	–	<sup>56</sup>	C30, C16
124	<i>Phrynos</i> vs. <i>Pelusios</i>	25	–	<sup>56</sup>	C30, C16
125	Tuatara ( <i>Sphenodon</i> ) vs. squamates	238	–	<sup>57</sup>	C30, C16
129	Anguimorpha vs. snakes	148	–	<sup>57</sup>	C30, C16
131	Pleurodont vs. acrodont iguanians	165	230	<sup>56</sup>	C30
132	Crown group Iguanidae	125	180	<sup>56</sup>	C30, C16
150	Human vs. platypus (Theria-Monotremata)	162.5	191.4	<sup>53</sup>	C30
151	Human vs. opossum (Placentalia-Marsupialia)	124.6	138.4	<sup>53</sup>	C30
152	Human vs. armadillo (Boreoeutheria-Xenarthra)	95.3	113	<sup>53</sup>	C30
154	Dog vs. cat	42.8	63.8	<sup>58</sup>	C30
155	Human vs. mouse	61.5	100.5	<sup>53</sup>	C30, C16
157	Opossum vs. kangaroo (Ameridelphia-Australidelphia)	61.5	71.2	<sup>53</sup>	C30
160	Frogs vs. salamanders	249	–	<sup>59</sup>	C30, C16
169	Hynobiidae vs. Cryptobranchidae	145.5	–	<sup>60</sup>	C30, C16
170	Discoglossoidea vs. other frogs	161.2	–	<sup>61</sup>	C30, C16
171	Pipoidea vs. other frogs	145.5	–	<sup>62</sup>	C30, C16
178	African vs. American pipids	86	–	<sup>63</sup>	C30, C16
187	Holocephalans vs. elasmobranchs	410	495	<sup>64</sup>	C30
188	Sharks vs. batoids	190	–	<sup>64</sup>	C30, C16
192	Rajidae vs. other batoids	176	–	<sup>64</sup>	C30, C16
195	Crown Neopterygii (Teleostei+ <i>Lepisosteus</i> )	345	392	<sup>51</sup>	C30

**Supplementary Table 9.** Results of divergence time estimates, including genome-wide estimates calculated from 100 independent timetrees based on 100 gene jackknife replicates using calibration scheme C16, and timetrees based on the nuclear test data set using calibration schemes C16 and C30. Divergence times for the nuclear test data set are accompanied by 95% confidence intervals (CI) estimated by PhyloBayes, whereas the genome-wide estimates are accompanied by conservative credibility intervals (CrI) calculated as the absolute minimum and maximum values of 95% CI across the 100 gene jackknife timetrees. Node IDs correspond to Supplementary Fig. 19.

Node ID	Genome-wide estimates Calibration: C16			Nuclear test data set Calibration: C16			Nuclear test data set Calibration: C30			Lineage
	Mean	minCrI	maxCrI	Mean	minCI	maxCI	Mean	minCI	maxCI	
100	457.55	438.96	465.40	458.21	446.44	464.65	460.01	451.57	465.31	
101	321.64	227.75	375.77	325.10	281.23	356.79	392.61	382.59	402.91	Ray-finned fish (Actinopterygii)
102	449.07	430.99	461.95	445.53	435.25	455.82	436.97	431.19	440.27	
103	427.46	413.15	443.67	427.67	419.16	437.48	426.35	420.45	431.06	Sarcopterygians (Sarcopterygii)
104	411.98	407.60	418.95	412.09	407.85	418.54	412.02	407.90	418.03	Lungfish + tetrapod
105	346.15	333.07	351.35	345.09	334.21	350.65	340.57	330.91	349.93	Tetrapods
106	316.64	299.00	330.15	312.13	297.40	323.65	288.56	282.58	295.83	Amniotes
107	294.43	272.95	313.03	291.13	274.97	305.51	256.77	255.70	257.44	Diapsids
108	281.71	256.53	304.75	283.67	267.03	298.94	254.38	252.53	255.96	
109	259.38	225.99	288.36	253.08	231.01	273.49	243.21	242.16	244.71	Archosaurs (Archosauria)
110	119.73	41.47	192.54	114.51	65.08	164.93	120.32	90.13	162.47	Crocodyles (Crocodylia)
111	80.47	25.29	160.21	79.09	41.74	122.23	70.61	66.02	75.02	
112	150.34	86.79	211.96	145.62	100.52	187.17	136.57	111.04	173.14	Birds (Aves)
113	116.74	57.90	187.75	107.45	67.19	150.01	83.03	70.24	87.49	Neognathae
114	93.51	42.73	164.71	83.90	49.49	123.43	63.00	47.20	72.81	
115	25.87	8.77	66.09	20.44	9.61	36.96	15.96	8.09	25.30	
116	108.93	50.53	179.38	107.39	65.22	150.93	92.13	65.83	130.11	Palaeognathae
117	227.36	189.06	265.68	223.30	197.95	250.15	224.33	211.24	234.25	Turtles (Testudines)
118	202.41	158.85	242.79	200.32	168.90	230.99	206.17	184.05	221.45	Cryptodira
119	160.38	103.12	207.37	154.12	114.51	187.96	168.59	133.04	187.99	
120	152.86	93.79	200.31	139.01	99.73	173.18	155.40	117.28	176.32	
121	123.40	66.07	173.25	111.47	74.76	146.06	127.13	90.14	150.13	
122	85.15	38.93	148.51	83.01	51.70	116.04	95.02	63.08	124.33	
123	81.02	28.36	133.92	63.66	35.77	96.36	77.52	45.26	107.43	
124	190.64	116.08	247.78	180.99	143.06	215.14	191.94	167.22	210.97	Pleurodira
125	270.76	246.16	298.09	267.21	249.60	283.69	239.06	233.30	244.30	Lepidosaurians (Lepidosauria)
126	203.67	183.05	227.96	202.37	186.10	219.58	199.34	190.37	208.01	Squamates (Squamata)
127	197.68	176.90	222.23	194.31	177.94	211.94	194.75	185.31	203.89	
128	185.83	166.45	210.88	182.22	166.54	199.95	186.55	177.06	196.27	
129	179.30	159.31	203.92	176.74	161.69	194.07	181.98	172.75	191.82	
130	175.16	155.82	198.10	175.29	160.49	192.34	180.61	171.65	190.19	
131	160.14	142.85	183.60	158.63	146.82	172.77	166.36	159.40	174.68	
132	125.47	116.27	138.93	127.19	120.28	137.99	136.57	124.02	150.85	
133	119.87	105.58	134.25	116.50	105.84	128.35	126.84	110.62	142.45	
134	119.84	104.16	136.49	119.53	109.57	131.24	129.94	114.64	145.09	
135	118.10	88.10	147.63	114.69	97.39	130.54	128.27	104.46	143.36	
136	69.52	42.66	106.39	79.33	61.54	99.42	94.20	71.86	118.74	Snakes (Serpentes)
137	63.11	37.84	97.33	71.86	54.27	91.40	87.71	65.73	111.91	
138	37.04	19.03	70.68	47.43	32.32	66.12	63.65	40.49	91.16	
139	22.71	10.60	49.45	31.80	19.92	47.49	46.94	26.10	72.14	
140	6.87	2.20	17.19	6.23	3.03	11.72	11.49	4.11	24.80	
141	27.15	12.16	52.35	31.07	19.79	46.34	45.58	25.12	71.55	
142	14.44	5.37	30.59	17.27	10.27	27.46	26.81	13.02	48.98	
143	23.01	10.95	46.16	25.71	15.95	39.16	38.72	20.67	63.88	
144	13.06	5.60	28.14	14.17	8.24	22.75	22.41	10.63	41.71	
145	171.84	112.44	200.04	172.96	156.09	191.11	178.70	166.81	189.64	<i>Podarcis</i> + <i>Tupinambis</i>
146	141.52	99.87	175.16	137.01	113.25	162.49	155.74	135.45	171.62	
147	44.81	19.07	77.06	42.41	26.58	61.46	57.09	34.23	77.41	
148	37.32	14.79	69.78	31.71	18.87	48.19	44.38	23.96	64.79	
149	162.18	124.66	196.21	144.30	116.81	170.59	164.79	145.89	180.58	
150	254.73	202.83	290.83	242.00	212.52	268.23	164.59	160.61	172.00	Mammals (Mammalia)
151	214.48	168.94	257.13	214.87	187.19	241.41	138.26	136.01	140.17	Theria
152	101.53	73.30	138.52	95.18	77.07	124.44	94.15	90.56	96.45	Placentals
153	90.94	65.71	122.29	85.47	69.82	112.61	88.92	84.82	91.97	Boreoeutheria
154	53.81	32.66	82.10	55.08	40.24	77.75	61.08	52.77	64.90	
155	79.12	60.22	101.56	72.95	60.60	97.40	78.51	71.13	84.44	
156	98.15	68.60	136.47	90.46	72.20	119.51	90.97	87.13	93.67	
157	121.14	65.74	182.28	129.84	96.01	165.45	68.42	62.34	71.54	
158	99.90	52.36	157.98	102.48	70.09	135.49	50.41	38.14	60.15	
159	325.15	306.58	338.38	320.90	306.19	331.57	315.07	300.21	328.40	Amphibians (Lissamphibia)
160	314.85	293.07	332.10	312.75	296.62	325.18	307.38	289.51	322.72	Frogs + salamanders
161	209.05	170.07	254.86	199.82	175.49	226.75	202.00	173.49	237.49	Salamanders (Caudata)
162	196.11	154.62	245.82	187.66	163.37	214.12	191.68	162.58	225.91	
163	174.76	126.67	225.72	170.16	143.87	196.88	177.38	146.40	210.28	
164	161.09	112.51	210.11	158.83	130.61	185.43	167.73	137.34	199.05	
165	88.33	44.90	132.55	103.59	73.82	129.39	117.29	86.21	143.25	
166	63.44	30.27	99.11	77.81	52.00	101.98	91.90	62.39	117.05	
167	52.60	20.91	90.03	62.58	39.82	85.88	77.04	48.84	100.78	
168	33.52	13.14	65.29	40.19	23.86	59.02	53.48	29.57	73.59	
169	163.92	131.78	217.22	153.59	132.59	180.75	161.68	133.89	196.43	
170	186.33	153.03	230.77	193.44	162.42	225.97	200.81	170.20	232.33	Frogs and toads (Anura)
171	174.20	140.83	222.40	182.78	150.84	216.60	192.11	160.56	224.21	
172	165.25	127.79	215.68	175.41	143.38	209.93	186.12	154.33	217.75	
173	123.78	81.80	172.73	137.67	103.74	173.77	154.78	122.64	185.98	
174	72.30	35.05	124.39	85.58	54.60	120.65	105.37	71.10	139.66	
175	64.84	25.31	118.48	74.18	45.85	107.07	93.55	62.29	126.69	
176	40.18	14.71	94.18	48.18	25.26	82.19	70.37	33.40	110.02	
177	30.33	10.61	77.16	33.42	16.17	61.45	53.61	21.69	89.26	
178	130.05	92.33	181.88	140.18	106.95	177.21	155.94	119.47	189.28	
179	119.60	77.77	176.20	127.09	93.67	164.22	143.99	105.86	176.98	
180	158.18	108.62	208.02	145.71	110.58	183.27	160.14	124.95	194.00	
181	213.48	151.09	270.45	211.85	169.20	249.84	212.66	161.89	255.34	Caecilians (Gymnophiona)
182	148.19	82.34	208.59	145.95	103.36	183.65	155.43	105.14	194.83	
183	28.86	4.87	85.33	25.22	10.04	47.60	36.11	12.27	65.41	
184	203.12	129.36	294.63	208.28	157.01	262.06	222.66	165.02	279.27	
185	55.53	25.66	109.72	62.25	37.41	92.11	78.25	47.52	106.88	
186	5.38	1.11	22.62	3.06	1.06	7.60	6.07	1.60	14.62	
187	303.73	238.51	358.70	296.23	250.19	339.78	414.06	402.40	428.01	Chondrichthyes
188	228.07	182.03	275.62	226.72	193.86	258.22	292.57	256.23	331.61	
189	177.93	104.56	238.91	174.93	141.81	208.66	201.69	139.64	269.47	
190	150.79	82.04	204.52	135.66	100.21	172.45	156.27	91.97	222.91	
191	83.88	31.51	154.53	81.48	49.22	120.04	97.58	49.99	167.80	
192	174.14	122.03	225.45	172.14	143.75	201.45	206.52	171.91	262.34	
193	47.92	14.51	114.67	52.46	30.86	77.20	75.92	42.49	110.03	
194	301.37	194.51	362.15	301.58	251.28	336.83	379.79	369.91	390.05	
195	206.58	122.39	284.95	228.13	177.53	272.79	345.36	338.11	352.48	
196	187.84	110.79	269.14	204.59	155.20	252.23	330.04	318.57	339.78	
197	10.16	1.05	46.88	11.35	2.07	34.13	55.44	18.36	91.19	
198	117.15	59.74	191.42	139.42	93.27	187.40	276.64	244.27	297.36	

**Supplementary Table 10.** Taxon sampling of the nuclear data sets, including the taxon ID in multiple sequence alignments and accession numbers for genome assemblies and raw RNA-seq data.

Taxon ID	Species	Data type	Accession
Acipenser_	<i>Acipenser baeri</i>	RNA-seq	SAMN06146200
Alligator_	<i>Alligator mississippiensis</i>	RNA-seq	SRX209114
Ambystoma_	<i>Ambystoma mexicanum</i>	RNA-seq	SRX1149362
Amia_calva	<i>Amia calva</i>	RNA-seq	SAMN06146197
Anas_platy	<i>Anas platyrhynchos</i>	genome	BGI_duck_1.0 (ENSEMBL v.73)
Andrias_da	<i>Andrias davidianus</i>	RNA-seq	SAMN06075442
Anolis_car	<i>Anolis carolinensis</i>	genome	AnoCar2.0 (ENSEMBL v.73)
Atelopus_z	<i>Atelopus zeteki</i>	RNA-seq	SRX339600
Basiliscus	<i>Basiliscus plumifrons</i>	RNA-seq	SRS777494
Boa_constr	<i>Boa constrictor</i>	RNA-seq	SRS777493
Bombina_ma	<i>Bombina maxima</i>	RNA-seq	SRX185157, SRX185159
Caiman_cro	<i>Caiman crocodilus</i>	RNA-seq	Figuet et al. 2015. Genome Biol Evol 7:240-250.
Callorhinc	<i>Callorhynchus milii</i>	genome	PRJNA236996, PRJNA18361
Calotriton	<i>Calotriton asper</i>	RNA-seq	SAMN06075443
Canis_fami	<i>Canis lupus familiaris</i>	genome	CanFam3.1 (ENSEMBL v.73)
Carcharodo	<i>Carcharodon carcharias</i>	RNA-seq	SRX228421
Caretta_ca	<i>Caretta caretta</i>	RNA-seq	Figuet et al. 2015. Genome Biol Evol 7:240-250.
Carlia_rub	<i>Carlia rubrigularis</i>	RNA-seq	SRX213353 to SRX213362
Chamaeleo_	<i>Chamaeleo chamaeleon</i>	RNA-seq	SRP018939
Chelonoidi	<i>Chelonoidis carbonaria</i>	RNA-seq	SRX385216
Chelonoidi	<i>Chelonoidis nigra</i>	RNA-seq	SRX565043-SRX565047; SRX385212-SRX385216
Chiloscyll	<i>Chiloscyllium griseum</i>	RNA-seq	ERX348252, ERX348253
Chinemys_r	<i>Chinemys reevesii</i>	RNA-seq	SRX338126
Crocodylus	<i>Crocodylus niloticus</i>	RNA-seq	SRS777495
Crotalus_a	<i>Crotalus adamanteus</i>	RNA-seq	SRX127425
Cyclorana_	<i>Cyclorana alboguttata</i>	RNA-seq	SRX205680; SRX205682-SRX205684; SRX206002-SRX206005
Cynops_pyr	<i>Cynops pyrrhogaster</i>	RNA-seq	SRX391946
Danio_eri	<i>Danio rerio</i>	genome	Zv9 (ENSEMBL v.73)
Dasybus_no	<i>Dasybus novemcinctus</i>	genome	dasNov2 (ENSEMBL v.73)
Discogloss	<i>Discoglossus pictus</i>	RNA-seq	SAMN06075438
Dromaius_n	<i>Dromaius novaehollandiae</i>	RNA-seq	SRX012419; SRX252410-SRX252413;
Echis_colo	<i>Echis coloratus</i>	RNA-seq	ERX190966-ERX190967; ERX190976-ERX190977; ERX190984; ERX190991
Elgaria_mu	<i>Elgaria multicarinata</i>	RNA-seq	SRX209116
Emys_orbic	<i>Emys orbicularis</i>	RNA-seq	Figuet et al. 2015. Genome Biol Evol 7:240-250.
Espadarana	<i>Espadarana prosoblepon</i>	RNA-seq	SRX648210
Eublephari	<i>Eublepharis macularius</i>	RNA-seq	ERX190969; ERX190971; ERX190980; ERX190981; ERX190987; ERX190990
Felis_catu	<i>Felis catus</i>	genome	ENSEMBL v.73
Gallus_gal	<i>Gallus gallus</i>	genome	Galgal4 (ENSEMBL v.73)
Geotrypete	<i>Geotrypetes seraphini</i>	RNA-seq	SAMN06075446
Ginglymost	<i>Ginglymostoma cirratum</i>	RNA-seq	SRX219865, SRX219866
Hynobius_c	<i>Hynobius chinensis</i>	RNA-seq	SRX386518
Homo_sapie	<i>Homo sapiens</i>	genome	GRCh37 (ENSEMBL v.73)
Hymenochir	<i>Hymenochirus curticeps</i>	RNA-seq	SAMN06075440
Iguana_igu	<i>Iguana iguana</i>	RNA-seq	SRS777492
Lamprophol	<i>Lampropholis coggeri</i>	RNA-seq	SRX213363-SRX213377
Latimeria_	<i>Latimeria chalumnae</i>	genome	LatCha1 (ENSEMBL v.73)
Lepidosire	<i>Lepidosiren paradoxa</i>	RNA-seq	SRR3632079-SRR3632086
Lepisost00	<i>Lepisosteus oculatus</i>	genome	LepOcu1 (ENSEMBL v.73)
Lepisosteus	<i>Lepisosteus platyrhincus</i>	RNA-seq	SAMN06234167
Leucoraja_	<i>Leucoraja erinacea</i>	RNA-seq	SRX1092112-SRX1092134
Loxodont_	<i>Loxodonta africana</i>	genome	loxAfr3 (ENSEMBL v.73)
Macropus_e	<i>Macropus eugenii</i>	genome	Meug_1.0 (ENSEMBL v.73)
Megophrys_	<i>Megophrys nasuta</i>	RNA-seq	SAMN06075447
Meleagris_	<i>Meleagris gallopavo</i>	genome	UMD2 (ENSEMBL v.73)
Micrurus_f	<i>Micrurus fulvius</i>	RNA-seq	SRX209497
Monodelphi	<i>Monodelphis domestica</i>	genome	BORAD05 (ENSEMBL v.73)
Mus_muscul	<i>Mus musculus</i>	genome	GRCm38 (ENSEMBL v.73)
Neoceratod	<i>Neoceratodus forsteri</i>	RNA-seq	SRR3632078
Neotrygon_	<i>Neotrygon kuhlii</i>	RNA-seq	SRX481088
Notophthal	<i>Notophthalmus viridescens</i>	RNA-seq	SRX221724-SRX221727;SRX221730-SRX221731
Opheodrys_	<i>Opheodrys aestivus</i>	RNA-seq	ERX190968; ERX190970; ERX190978; ERX190979; ERX190985; ERX190986
Ophiophagu	<i>Ophiophagus hannah</i>	RNA-seq	SRX365144
Ornithorhy	<i>Ornithorhynchus anatinus</i>	genome	OANA5 (ENSEMBL v.73)
Pantheroph	<i>Pantherophis guttatus</i>	RNA-seq	ERX190963; ERX190972; ERX190973; ERX190982; ERX190988
Pelodiscus	<i>Pelodiscus sinensis</i>	RNA-seq	DRX001538; DRX001541; DRX001551; DRX001578
Pelophylax	<i>Pelophylax lessonae</i>	RNA-seq	SRX472107
Pelophyl00	<i>Pelophylax nigromaculatus</i>	RNA-seq	SRX532382
Pelusios_c	<i>Pelusios castaneus</i>	RNA-seq	SRX209127
Phrynos_h	<i>Phrynosoma hillebrandii</i>	RNA-seq	Figuet et al. 2015. Genome Biol Evol 7:240-250.
Pipa_pipa_	<i>Pipa pipa</i>	RNA-seq	SAMN06075441
Pleurodele	<i>Pleurodeles waltil</i>	RNA-seq	SAMN06146195
Podarcis_s	<i>Podarcis sp.</i>	RNA-seq	Figuet et al. 2015. Genome Biol Evol 7:240-250.
Pogona_vit	<i>Pogona vitticeps</i>	RNA-seq	Tzika et al. 2011. EvoDevo 2:19
Polypterus	<i>Polypterus senegalus</i>	RNA-seq	SAMN06146198
Proteus_an	<i>Proteus anguinus</i>	RNA-seq	SAMN06075444
Protopte00	<i>Protopterus aethiopicus</i>	RNA-seq	SAMN06146196
Protopteru	<i>Protopterus annectens</i>	RNA-seq	SRX152529-SRX152531
Python_reg	<i>Python regius</i>	RNA-seq	ERX190964; ERX190965; ERX190974; ERX190975; ERX190983; ERX190989
Raja_clava	<i>Raja clavata</i>	RNA-seq	SAMN06146201
Rana_chens	<i>Rana chensinensis</i>	RNA-seq	SRX200798
Rana_chens	<i>Rana kukunoris</i>	RNA-seq	SRX200800
Rhinatrema	<i>Rhinatrema bivittatum</i>	RNA-seq	SAMN06146202
Salamandra	<i>Salamandra salamandra</i>	RNA-seq	SRS777496
Saprosinc	<i>Saprosincus basiliscus</i>	RNA-seq	SRS777494
Sarcophilu	<i>Sarcophilus harrisii</i>	genome	DEVIL7.0 (ENSEMBL v.73)
Sceloporus	<i>Sceloporus undulatus</i>	RNA-seq	SRX209117
Scincella_	<i>Scincella lateralis</i>	RNA-seq	SRX793983
Scyliorhin	<i>Scyliorhinus canicula</i>	RNA-seq	SAMN06234165; SAMN06234166
Siren_lace	<i>Siren lacertina</i>	RNA-seq	SAMN06075439
Sistrurus_	<i>Sistrurus miliarius</i>	RNA-seq	SRX1032424
Sphenodon_	<i>Sphenodon punctatus</i>	RNA-seq	SRA051647
Sternother	<i>Sternotherus odoratus</i>	RNA-seq	SRX209126
Struthio_c	<i>Struthio camelus</i>	RNA-seq	SRX790604
Taeniopygi	<i>Taeniopygia guttata</i>	genome	taeGut3.2.4 (ENSEMBL v.73)
Takifugu_r	<i>Takifugu rubripes</i>	genome	FUGU4 (ENSEMBL v.73)
Tarentola_	<i>Tarentola mauritanica</i>	RNA-seq	SAMN06146199
Thamnophis	<i>Thamnophis elegans</i>	RNA-seq	SRA010134
Trachemys_	<i>Trachemys scripta</i>	RNA-seq	SRX565329-SRX565330
Tupinambis	<i>Tupinambis teguixin</i>	RNA-seq	SRS777490
Typhlone00	<i>Typhlonectes compressicauda</i>	RNA-seq	SAMN06146203
Typhlonect	<i>Typhlonectes natans</i>	RNA-seq	SAMN06075445
Xenopus_tr	<i>Silurana tropicalis</i>	genome	JGI_4.2 (ENSEMBL v.73)

**Supplementary Table 11.** Taxon sampling of the mitochondrial datasets, including the taxon ID in multiple sequence alignments and NCBI or RefSeq accession numbers.

Taxon ID	Species	Accession	Taxon ID	Species	Accession
Abronia_gr	<i>Abronia graminea</i>	NC_005958	Leucoraja_	<i>Leucoraja erinacea</i>	NC_016429
Acipenser_	<i>Acipenser baerii</i>	NC_017603	Loxodonta_	<i>Loxodonta africana</i>	NC_000934
Alligator_	<i>Alligator sinensis</i>	NC_004448	Macropus_r	<i>Macropus robustus</i>	NC_001794
Ambystoma_	<i>Ambystoma mexicanum</i>	NC_005797	Mauremys_m	<i>Mauremys mutica</i>	NC_009330
Amia_calva	<i>Amia calva</i>	NC_004742	Meleagris_	<i>Meleagris gallopavo</i>	NC_010195
Amphisbaen	<i>Amphisbaena schmidti</i>	NC_006284	Micrurus_f	<i>Micrurus fulvius</i>	NC_013481
Anas_platy	<i>Anas platyrhynchos</i>	NC_009684	Monodelphi	<i>Monodelphis domestica</i>	NC_006299
Andrias_da	<i>Andrias davidianus</i>	NC_004926	Mus_muscul	<i>Mus musculus musculus</i>	NC_010339
Anolis_car	<i>Anolis carolinensis</i>	NC_010972	Neoceratod	<i>Neoceratodus forsteri</i>	AJ584642
Ascaphus_t	<i>Ascaphus truei</i>	AJ871087	Neotrygon_	<i>Neotrygon kuhlii</i>	NC_021767
Atympanoph	<i>Atympanophrys shapingensis</i>	NC_018785	Nerodia_si	<i>Nerodia sipedon</i>	NC_015793
Basiliscus	<i>Basiliscus vittatus</i>	NC_012829	Notophthal	<i>Notophthalmus viridescens</i>	EU880323
Boa_constr	<i>Boa constrictor</i>	NC_007398	Ophiophagu	<i>Ophiophagus hannah</i>	NC_011394
Bombina_ma	<i>Bombina maxima</i>	NC_011049	Ornithorhy	<i>Ornithorhynchus anatinus</i>	NC_000891
Bufo_japon	<i>Bufo japonicus</i>	NC_009886	Pantheroph	<i>Pantherophis slowinskii</i>	NC_009769
Caiman_cro	<i>Caiman crocodilus</i>	NC_002744	Pelobates_	<i>Pelobates cultripes</i>	NC_008144
Callorhinc	<i>Callorhynchus milii</i>	NC_014285	Pelodiscus	<i>Pelodiscus sinensis</i>	NC_006132
Calotriton	<i>Calotriton asper</i>	EU880307	Pelodytes_	<i>Pelodytes cf. punctatus</i>	NC_020000
Canis_lupu	<i>Canis lupus familiaris</i>	NC_002008	Pelomedusa	<i>Pelomedusa subrufa</i>	NC_001947
Carcharodo	<i>Carcharodon carcharias</i>	NC_022415	Pelophylax	<i>Pelophylax nigromaculatus</i>	NC_002805
Caretta_ca	<i>Caretta caretta</i>	NC_016923	Phrynosop_h	<i>Phrynosoma hilarii</i>	JN999705
Causus_def	<i>Causus defilippi</i>	NC_013479	Pipa_carva	<i>Pipa carvalhoi</i>	NC_015617
Chamaeleo_	<i>Chamaeleo chamaeleon</i>	NC_012427	Plestiodon	<i>Plestiodon elegans</i>	NC_024576
Chiloscyll	<i>Chiloscyllium griseum</i>	NC_017882	Pleurodele	<i>Pleurodeles poireti</i>	EU880329
Chinemys_r	<i>Chinemys reevesi</i>	NC_006082	Podocnemis	<i>Podocnemis unifilis</i>	NC_018865
Chrysemys_	<i>Chrysemys picta bellii</i>	NC_023890	Pogona_vit	<i>Pogona vitticeps</i>	NC_006922
Coleonyx_v	<i>Coleonyx variegatus</i>	NC_008774	Polypterus	<i>Polypterus senegalus senegalus</i>	NC_004418
Crocodylus	<i>Crocodylus niloticus</i>	NC_008142	Proteus_an	<i>Proteus anguinus</i>	NC_023342
Crotalus_h	<i>Crotalus horridus</i>	NC_014400	Protopte00	<i>Protopterus aethiopicus</i>	NC_014764
Cynops_pyr	<i>Cynops pyrrhogaster</i>	EU880313	Protopteru	<i>Protopterus annectens</i>	NC_018822
Danio_eri	<i>Danio rerio</i>	NC_002333	Psammobate	<i>Psammobates pardalis</i>	NC_007694
Dasypus_no	<i>Dasypus novemcinctus</i>	NC_001821	Python_reg	<i>Python regius</i>	NC_007399
Discogloss	<i>Discoglossus galganoi</i>	NC_006690	Rana_cf_c	<i>Rana cf. Chensinensis</i>	NC_023529
Dromaius_n	<i>Dromaius novaehollandiae</i>	NC_002784	Rhinatrema	<i>Rhinatrema bivittatum</i>	NC_006303
Euprepioph	<i>Euprepiophis perlacea</i>	NC_024546	Rhincodon_	<i>Rhincodon typus</i>	NC_023455
Felis_catu	<i>Felis catus</i>	NC_001700	Rhinophryn	<i>Rhinophrynus dorsalis</i>	NC_021849
Gallus_gal	<i>Gallus gallus</i>	NC_001323	Salamandra	<i>Salamandra salamandra</i>	EU880331
Geotrypete	<i>Geotrypetes seraphini</i>	GQ244469	Sarcophilu	<i>Sarcophilus harrisii</i>	NC_018788
Heleophryn	<i>Heleophryne regis</i>	NC_019998	Sceloporus	<i>Sceloporus occidentalis</i>	NC_005960
Hemithecon	<i>Hemitheconyx caudicinctus</i>	NC_018368	Scyliorhin	<i>Scyliorhinus canicula</i>	NC_001950
Homo_sapie	<i>Homo sapiens</i>	AF346972	Silurana_t	<i>Silurana tropicalis</i>	NC_006839
Hyla_japon	<i>Hyla japonica</i>	NC_010232	Siren_inte	<i>Siren intermedia</i>	GQ368661
Hymenochir	<i>Hymenochirus boettgeri</i>	NC_015615	Sooglossus	<i>Sooglossus thomasseti</i>	NC_020001
Hynobius_c	<i>Hynobius chinensis</i>	NC_008088	Sphenodon_	<i>Sphenodon punctatus</i>	NC_004815
Ichthyaetu	<i>Ichthyaetus relictus</i>	NC_023777	Sternother	<i>Sternotherus carinatus</i>	NC_017607
Ichthyophi	<i>Ichthyophis bombayensis</i>	NC_023511	Struthio_c	<i>Struthio camelus</i>	NC_002785
Iguana_igu	<i>Iguana iguana</i>	NC_002793	Taeniopygi	<i>Taeniopygia guttata</i>	NC_007897
Latimeria_	<i>Latimeria chalumnae</i>	AB257297	Takifugu_r	<i>Takifugu rubripes</i>	NC_004299
Leiopelma_	<i>Leiopelma archeyi</i>	NC_014691	Tarentola_	<i>Tarentola mauritanica</i>	NC_012366
Lepidosire	<i>Lepidosiren paradoxa</i>	NC_003342	Trachemys_	<i>Trachemys scripta</i>	NC_011573
Lepisost00	<i>Lepisosteus oculatus</i>	NC_004744	Tupinambis	<i>Tupinambis teguixin</i>	SRS777490
Lepisosteus	<i>Lepisosteus osseus</i>	NC_008104	Typhlonect	<i>Typhlonectes natans</i>	NC_002471
Leptobrach	<i>Leptobranchium boringii</i>	NC_024427	Varanus_ni	<i>Varanus niloticus</i>	NC_008778

**Supplementary Table 12.** Non-vertebrate species used in the BLASTP decontamination step.

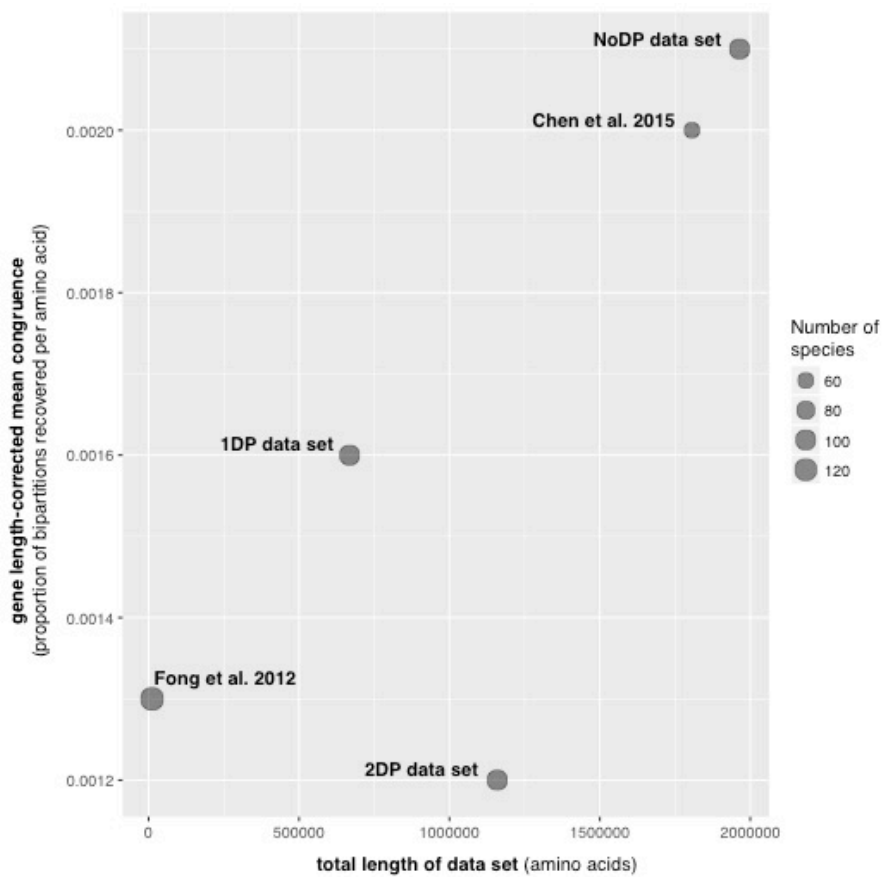
---

<i>Acanthamoeba castellanii</i>	<i>Leishmania major</i>
<i>Acromyrmex echinator</i>	<i>Leptospira borgpetersenii</i> JB197
<i>Amphimedon queenslandica</i>	<i>Loa loa</i>
<i>Anabaena cylindrica</i> PCC7122	<i>Lottia gigantea</i>
<i>Apis mellifera</i>	<i>Magnaporthe oryzae</i>
<i>Aplysia californica</i>	<i>Micromonas</i> RCC299
<i>Arabidopsis thaliana</i>	<i>Monosiga brevicollis</i> ATCC50154
<i>Aspergillus fumigatus</i>	<i>Nannochloropsis gaditana</i> CCMP526
<i>Aureococcus anophagefferens</i>	<i>Nematostella vectensis</i>
<i>Bacillus subtilis</i>	<i>Opisthorchis viverrini</i>
<i>Bacteroides coprocola</i>	<i>Oryza sativa</i>
<i>Batrachomyxium dendrobatidis</i>	<i>Oxytricha trifallax</i>
<i>Bigelowiella natans</i>	<i>Paramecium tetraurelia</i>
<i>Bombyx mori</i>	<i>Pediculus humanus corporis</i>
<i>Branchiostoma floridae</i>	<i>Perkinsus marinus</i>
<i>Caenorhabditis elegans</i>	<i>Physcomitrella patens</i>
<i>Campylobacter concisus</i> 13826	<i>Phytophthora infestans</i>
<i>Candidatus nitrosopumilus</i>	<i>Plancitomyces brasiliensis</i>
<i>Candidatus Protochlamydia amoebophila</i>	<i>Populus trichocarpa</i>
<i>Capitella teleta</i>	<i>Prochlorococcus marinus</i> NATL1A
<i>Capsaspora owczarzaki</i>	<i>Pyrococcus abyssi</i>
<i>Chlamydomonas reinhardtii</i>	<i>Reticulomyxa filosa</i>
<i>Chlorella variabilis</i>	<i>Rhizophagus irregularis</i> DAOM181602
<i>Chlorobium phaeobacteroides</i> DSM266	<i>Rhizopus delemar</i>
<i>Chloroflexus aggregans</i> DSM9485	<i>Rubritalea marina</i>
<i>Chondrus crispus</i>	<i>Rubrobacter radiotolerans</i>
<i>Ciona intestinalis</i>	<i>Saccoglossus kowalevskii</i>
<i>Clostridium akagii</i>	<i>Schistosoma japonicum</i>
<i>Crassostrea gigas</i>	<i>Sinorhizobium meliloti</i>
<i>Cryptococcus neoformans</i>	<i>Streptomyces avermitilis</i> MA4680
<i>Daphnia pulex</i>	<i>Strongylocentrotus purpuratus</i>
<i>Desulfovibrio bastinii</i>	<i>Sulfolobus solfataricus</i>
<i>Dictyostelium discoideum</i>	<i>Symbiodinium</i> spp
<i>Drosophila melanogaster</i>	<i>Tetrahymena thermophila</i>
<i>Emiliana huxleyi</i>	<i>Tetraselmis astigmatica</i>
<i>Escherichia coli</i>	<i>Thalassiosira pseudonana</i>
<i>Fonticula alba</i>	<i>Toxoplasma gondii</i>
<i>Fusarium graminearum</i>	<i>Tribolium castaneum</i>
<i>Fusobacterium ulcerans</i> ATCC49185	<i>Trichinella spiralis</i>
<i>Glycine max</i>	<i>Trichoplax adhaerens</i>
<i>Guillardia theta</i>	<i>Trypanosoma cruzi</i>
<i>Helobdella robusta</i>	<i>Vitrella brassicaformis</i>
<i>Hydra vulgaris</i>	<i>Yarrowia lipolytica</i>
<i>Ixodes scapularis</i>	<i>Zea mays</i>

---

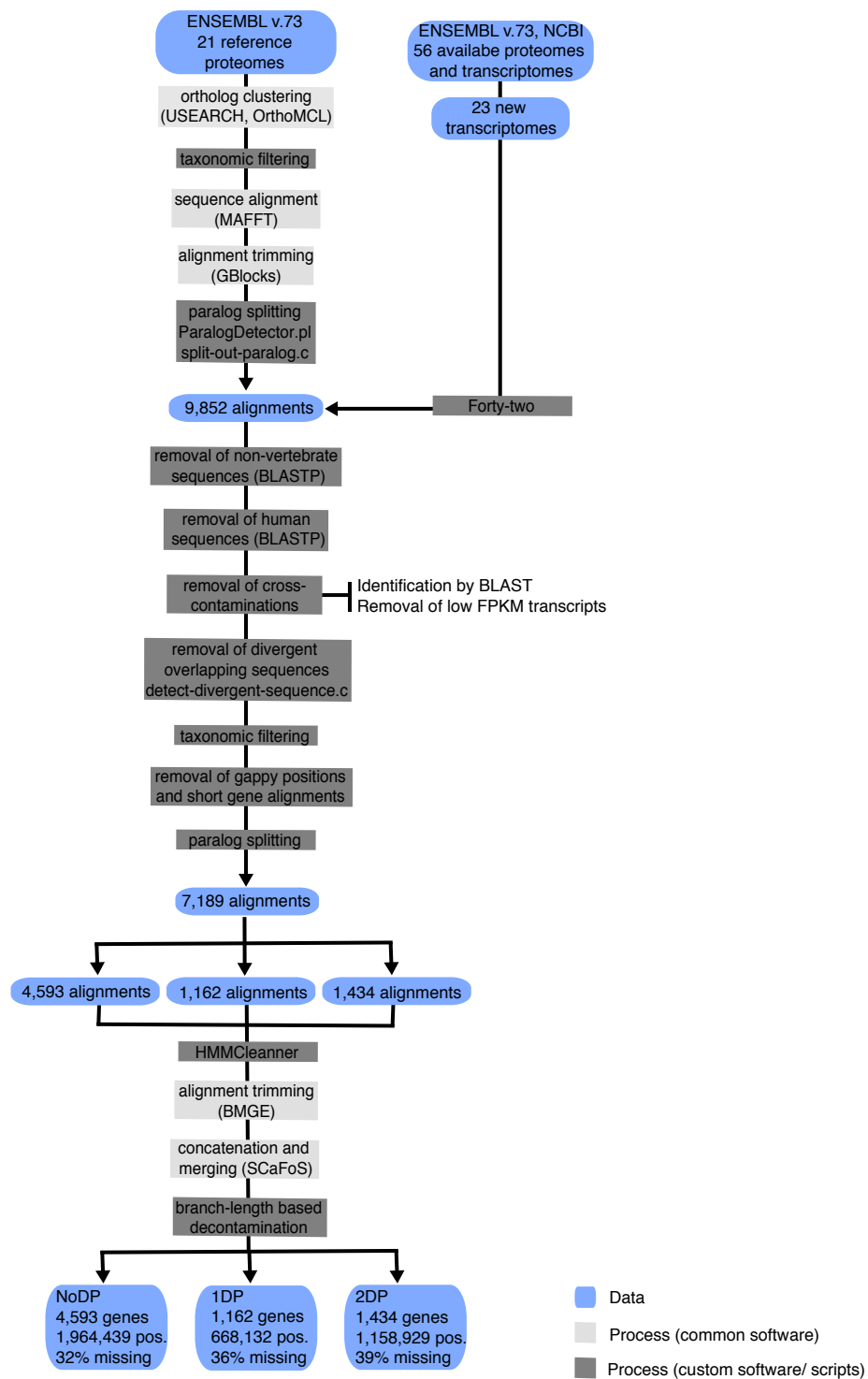
**Supplementary Table 13.** Results from model cross-validations performed in PhyloBayes.

<b>Models compared</b>	<b>Mean score</b>	<b>Standard deviation</b>	<b># times model is best</b>
<b>Amino acid substitution models</b>			
<b>50,000 amino acid-long training datasets and 10,000 amino acid-long test datasets (phylogeny inference)</b>			
CATGTR vs. LG	11636.5	220.54	10
GTR vs. LG	5195	107.232	0
CAT vs. LG	5984.6	277.919	0
<b>Amino acid substitution models</b>			
<b>Nuclear test dataset (timetree inference)</b>			
CATGTR vs. CAT	500.68	112.514	10
CATGTR vs. JTT	1214.68	120.44	10
CAT vs. JTT	714	567.101	10
<b>Molecular clock models</b>			
<b>Nuclear test dataset (timetree inference)</b>			
LN vs. CL	147.63	161.837	10
UGAM vs. CL	148.55	20.445	10
UGAM vs. LN	0.92	152.467	5
<b>Amino acid substitution models</b>			
<b>Mitochondrial dataset 106 species (phylogeny inference)</b>			
CATGTR vs. CAT	318.29	428.483	10
CATGTR vs. MTREV	582.98	45.316	10
CAT vs. MTREV	264.69	47.613	10
GTR vs. MTREV	149.86	31.798	10
MTREV vs. LG	90.02	33.657	10

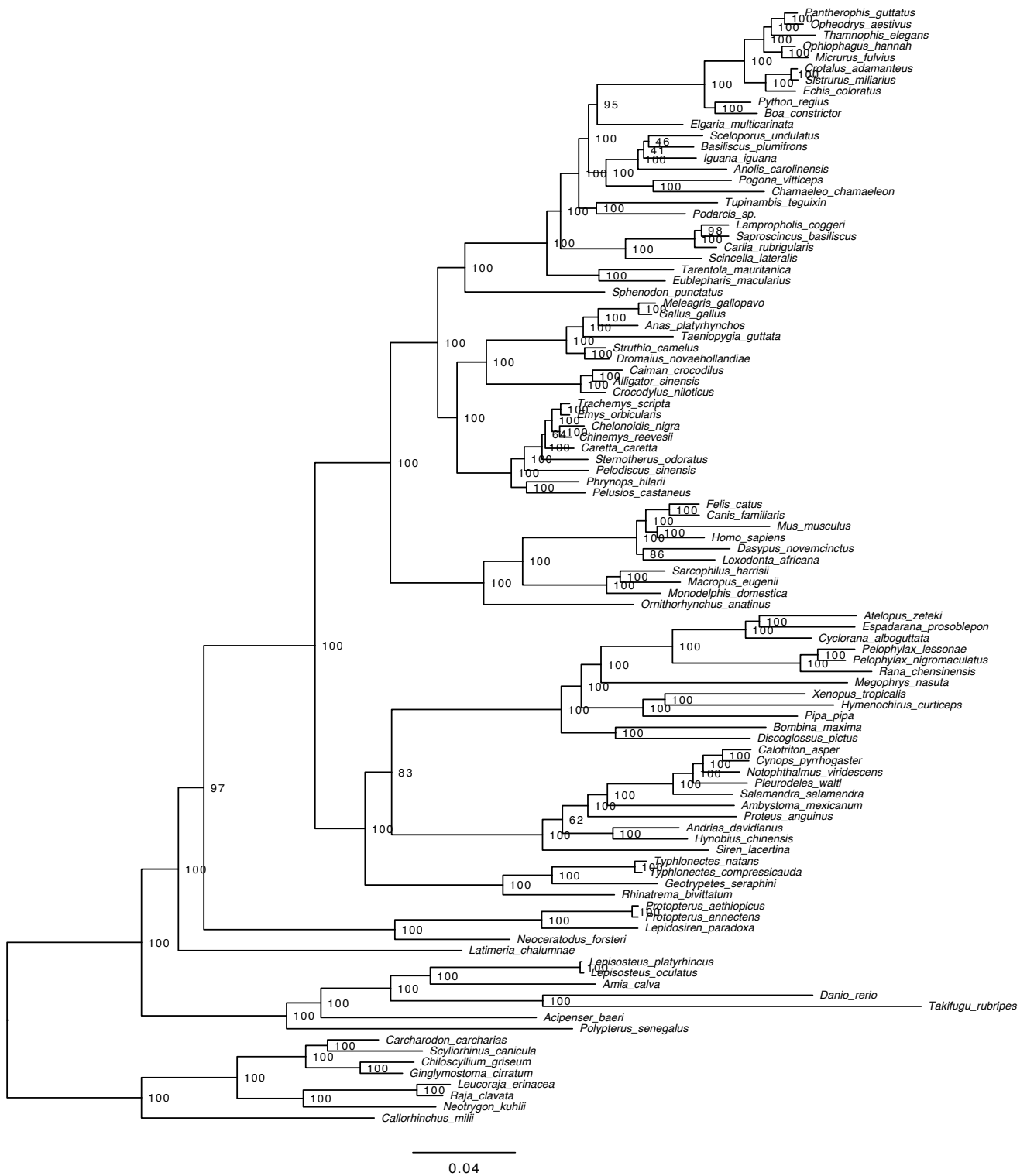


**Supplementary Figure 1.** Comparison of phylogenomic datasets by quality and size. Data quality is measured as mean congruence (proportion of final-tree bipartitions recovered by each gene, and corrected by gene length); dataset size is measured as total amino acid positions and the number of species included (represented by circle size). The graph compares the quality of our three datasets (NoDP, 1DP and 2DP) and the phylogenomic datasets of Fong et al.<sup>47</sup> and the most recent study by Chen et al.<sup>65</sup>.

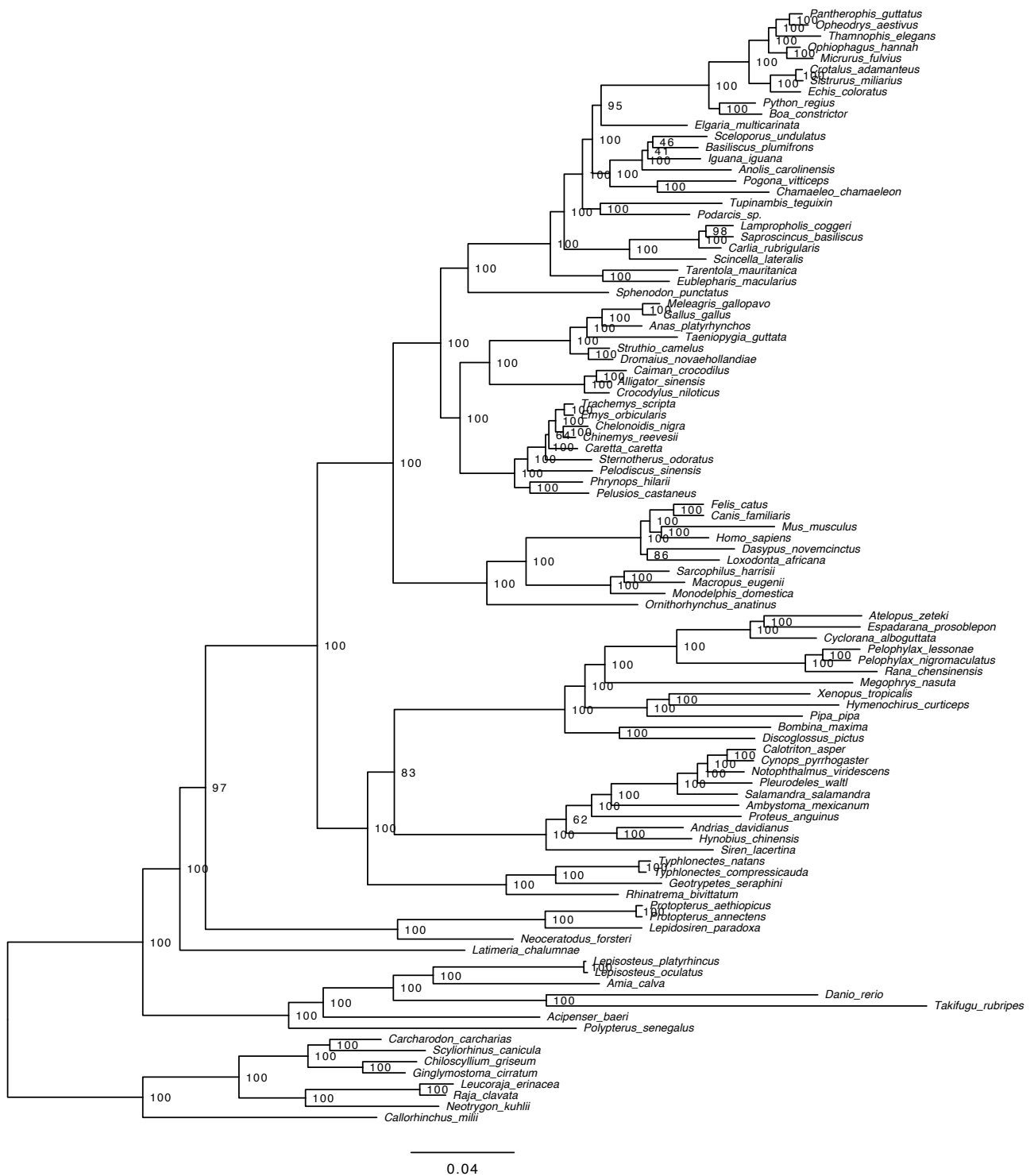




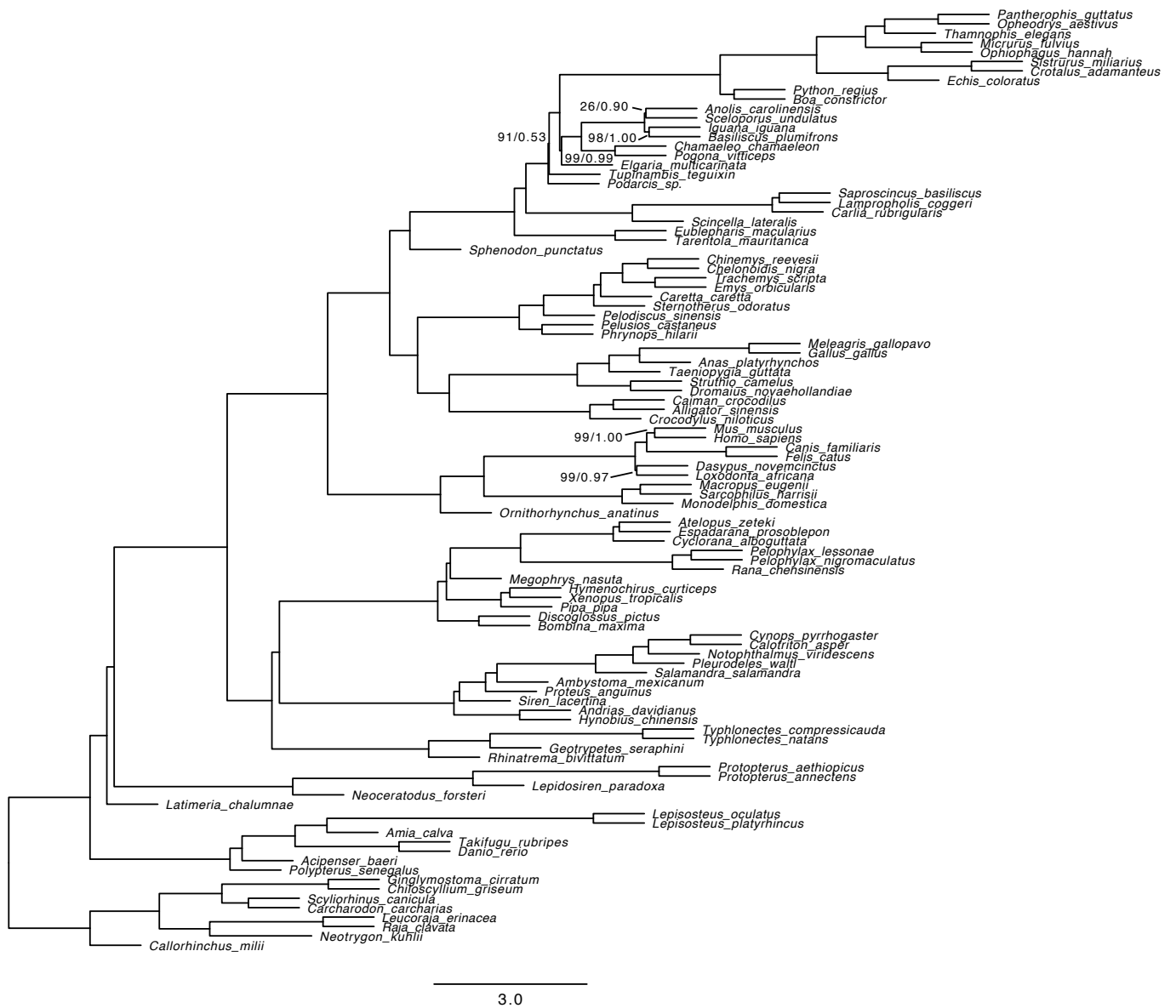
**Supplementary Figure 2.** Flowchart of the bioinformatic pipeline developed for phylogenomic dataset assembly. Abbreviations: FPKM, fragments per Kb per million mapped reads; pos., amino acid positions.



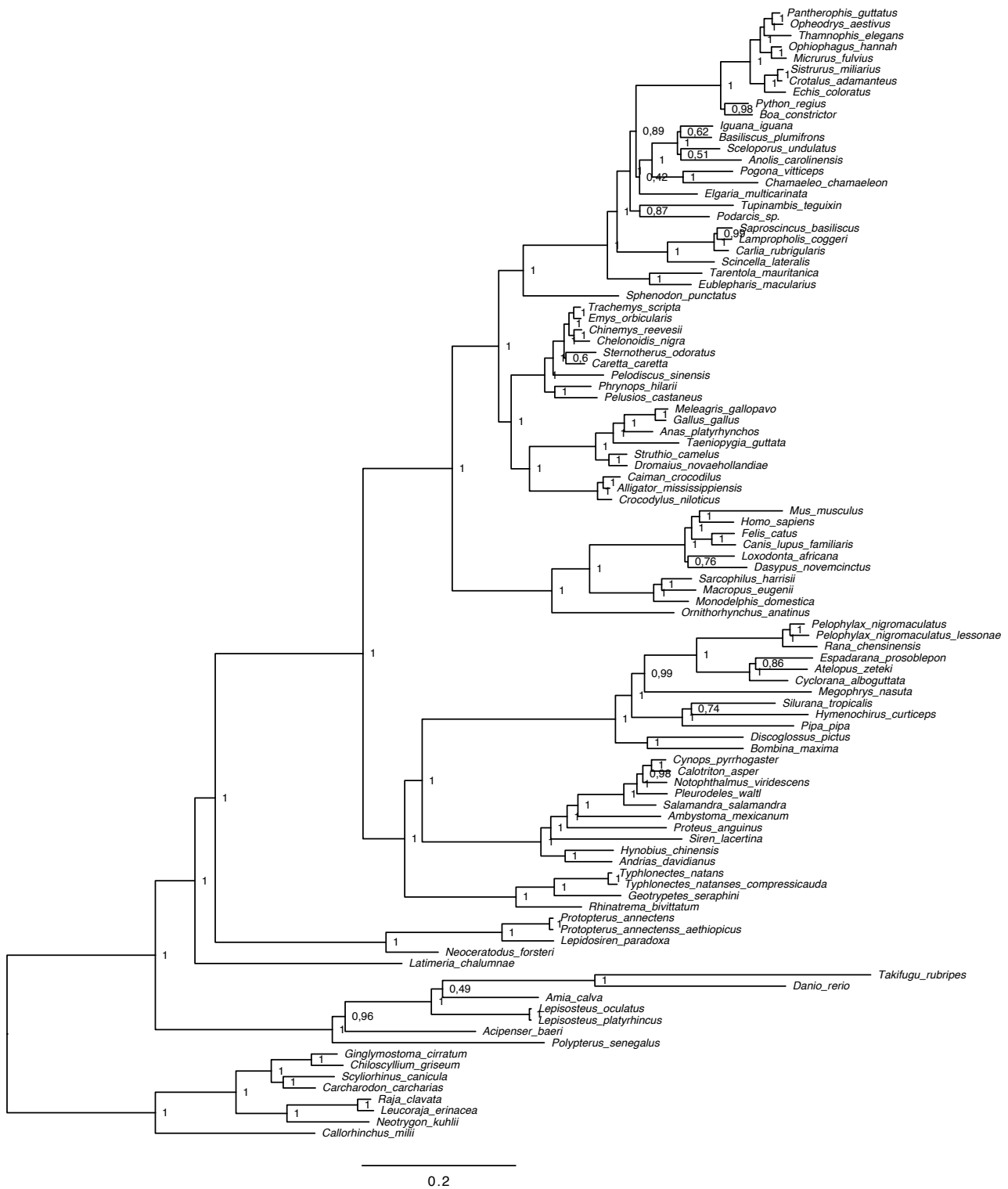
**Supplementary Figure 3.** Maximum likelihood phylogeny estimated from the NoDP dataset under the GTR+ $\Gamma$  model in RAxML. Numbers at nodes are non-parametric bootstrap support and scale bar is in substitutions site<sup>-1</sup>. This topology is identical to that recovered by BI on the 2DP dataset under the CAT+ $\Gamma$  model, and differs from that in Fig. 2a (BI analysis of the NoDP dataset under CAT+ $\Gamma$ ) on the position of *Elgaria*, which is recovered as sister group of snakes, and *Sternotherus*, which is recovered as sister group of *Caretta*.



**Supplementary Figure 4.** Maximum likelihood phylogeny estimated from the NoDP dataset under the LG+ $\Gamma$ +F model in RAxML. Numbers at nodes are non-parametric bootstrap support and scale bar is substitutions site<sup>-1</sup>. The ML tree estimated under LG+ $\Gamma$ +F differs from that in Fig. 2a (BI analysis of the NoDP dataset under CAT+ $\Gamma$ ) on the position of *Elgaria*, which is recovered as sister group of snakes, the position of *Anolis*, which is basal to *Iguana*, *Basiliscus* and *Sceloporus*, and the position of *Siren*, which is recovered as the earliest branching salamander.



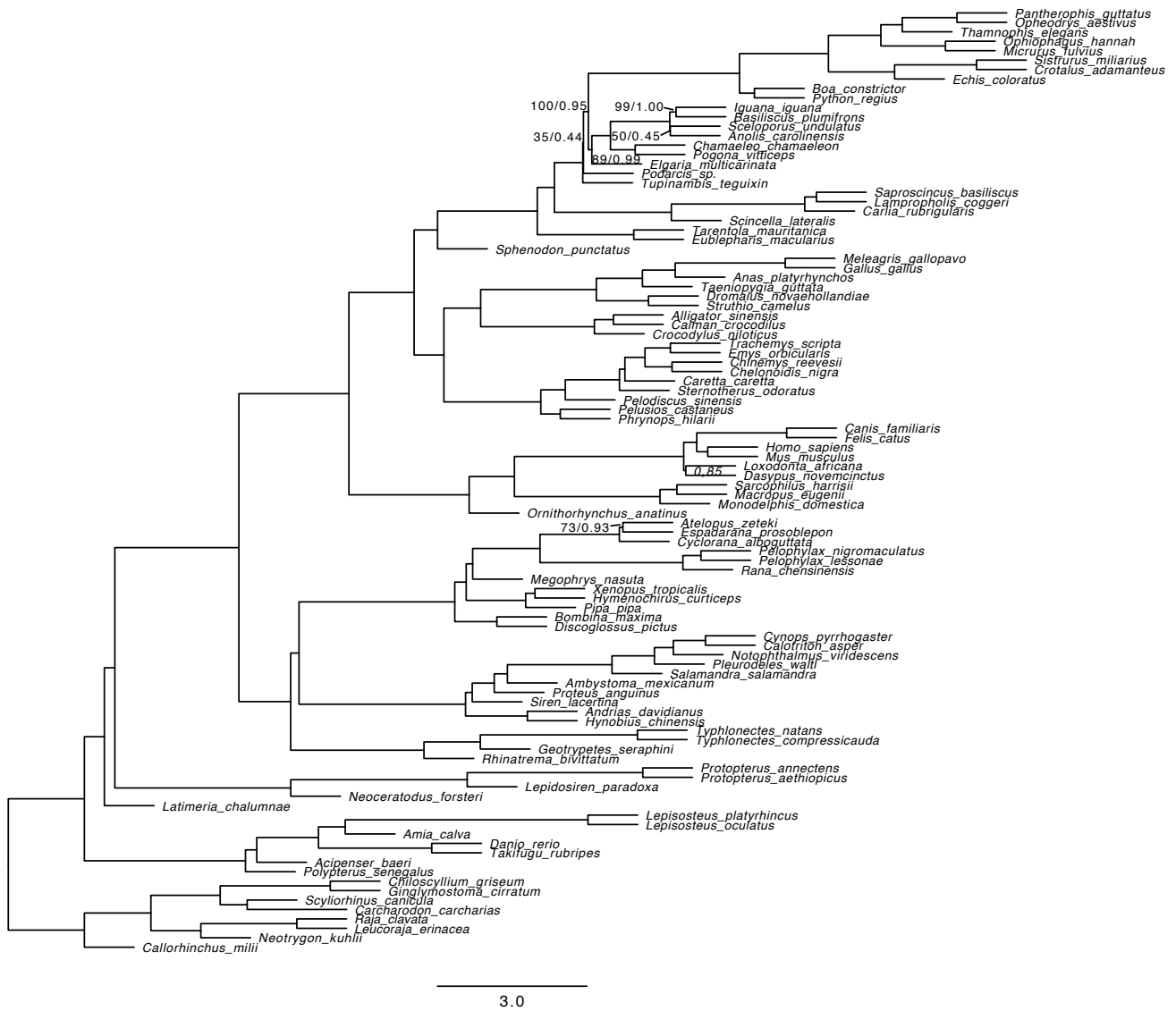
**Supplementary Figure 5.** Coalescent-based tree estimated from the NoDP dataset using ASTRAL-II. All nodes received maximum support from multi-locus bootstrapping and quartet-based local node support, except for nodes where the actual numbers are shown. Scale bar is in coalescent units. This species tree differs from the BI tree obtained from a concatenated matrix of the same dataset (Fig. 2a) in the relative positions of *Lepisosteus* (and incongruence that is strongly supported), and the relative positions of *Tupinambis* with *Podarcis* and *Caretta* with *Pelodiscus*; both disagreements that are lowly supported.



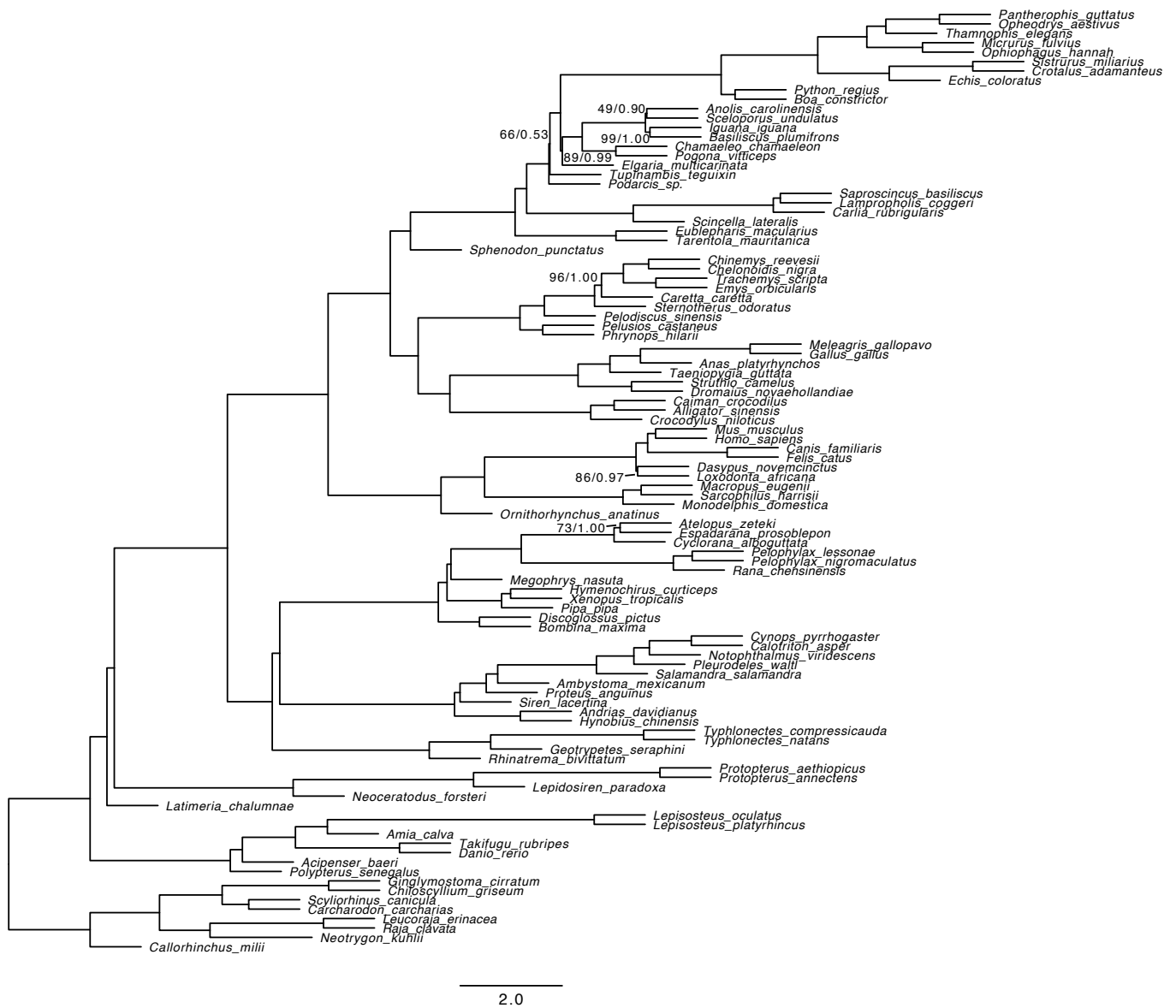
**Supplementary Figure 6.** Bayesian majority-rule consensus tree reconstructed from the 1DP dataset, estimated from 100 gene-jackknife replicates (~50,000 amino acid positions each) under the CAT+ $\Gamma$  model in PhyloBayes. Numbers at nodes are posterior probabilities and scale bar is in substitutions site<sup>-1</sup>. The BI tree estimated from the 1DP dataset differs from that in Fig. 2a (NoDP dataset) on the position of *Amia*, which is recovered as sister group of *Takifugu* + *Danio*, and *Sternotherus*, which is recovered as sister group of *Caretta*.



**Supplementary Figure 7.** Bayesian majority-rule consensus tree reconstructed from the 2DP dataset, estimated from 100 gene-jackknife replicates (~50,000 amino acid positions each) under the CAT+ $\Gamma$  model in PhyloBayes. Numbers at nodes are posterior probabilities and scale bar is in substitutions site<sup>-1</sup>. This topology is identical to that recovered by ML on the NoDP dataset under the GTR+ $\Gamma$  model, and differs from that in Fig. 2a (BI analysis of the NoDP dataset under CAT+ $\Gamma$ ) on the position of *Elgaria*, which is recovered as sister group of snakes, and *Sternotherus*, which is recovered as sister group of *Caretta*.

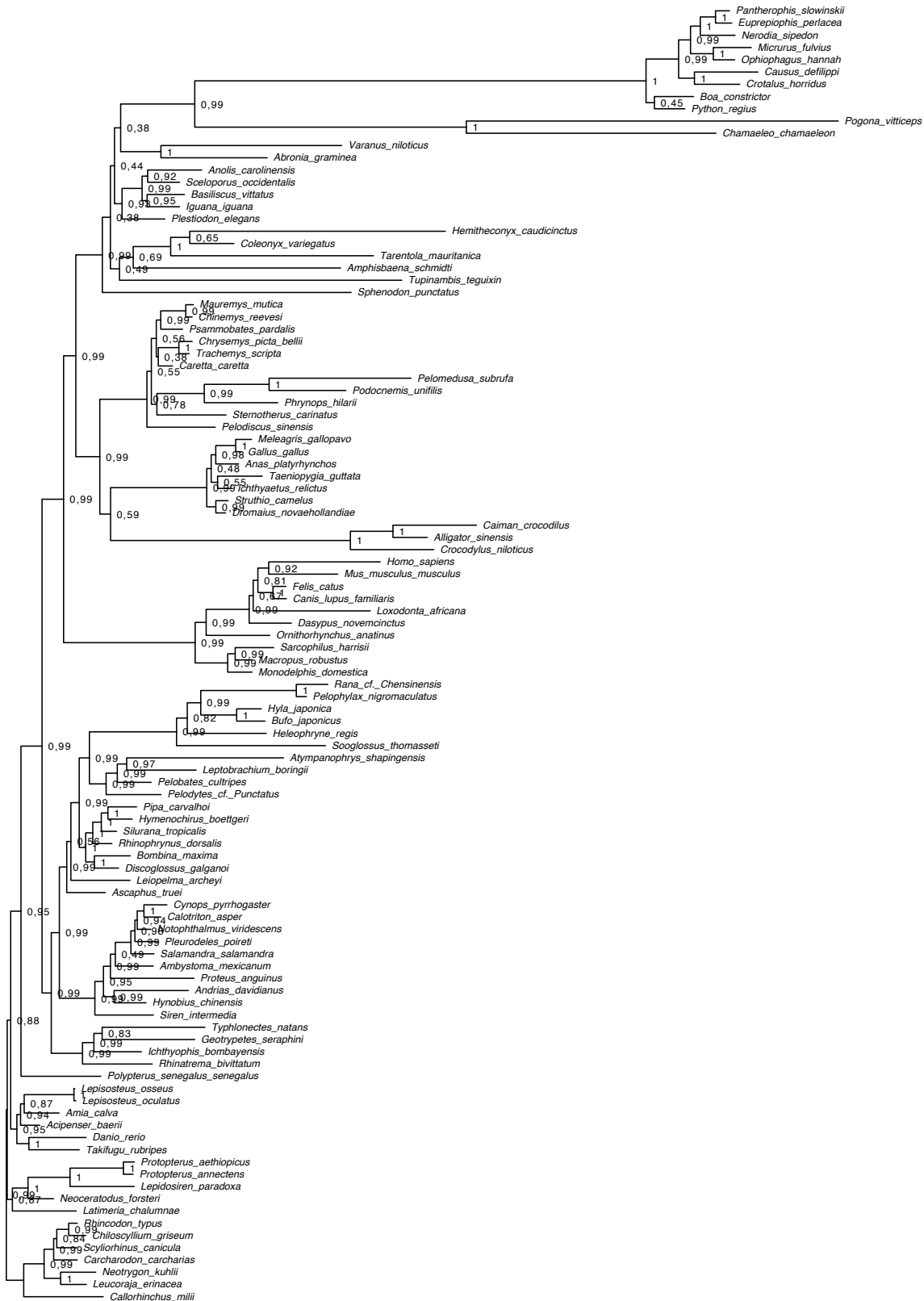


**Supplementary Figure 8.** Coalescent-based tree estimated from the 1DP dataset using ASTRAL-II. All nodes received maximum support from multi-locus bootstrapping and quartet-based local node support, except for nodes where the actual numbers are shown. Scale bar is in coalescent units. This species tree differs from the BI tree obtained from a concatenated matrix of the same dataset (Supplementary Fig. 6) in the relative positions of *Lepisosteus* (and incongruence that is strongly supported), and the relative positions of *Tupinambis* with *Podarcis* and *Caretta* with *Pelodiscus*; both disagreements that are lowly supported.



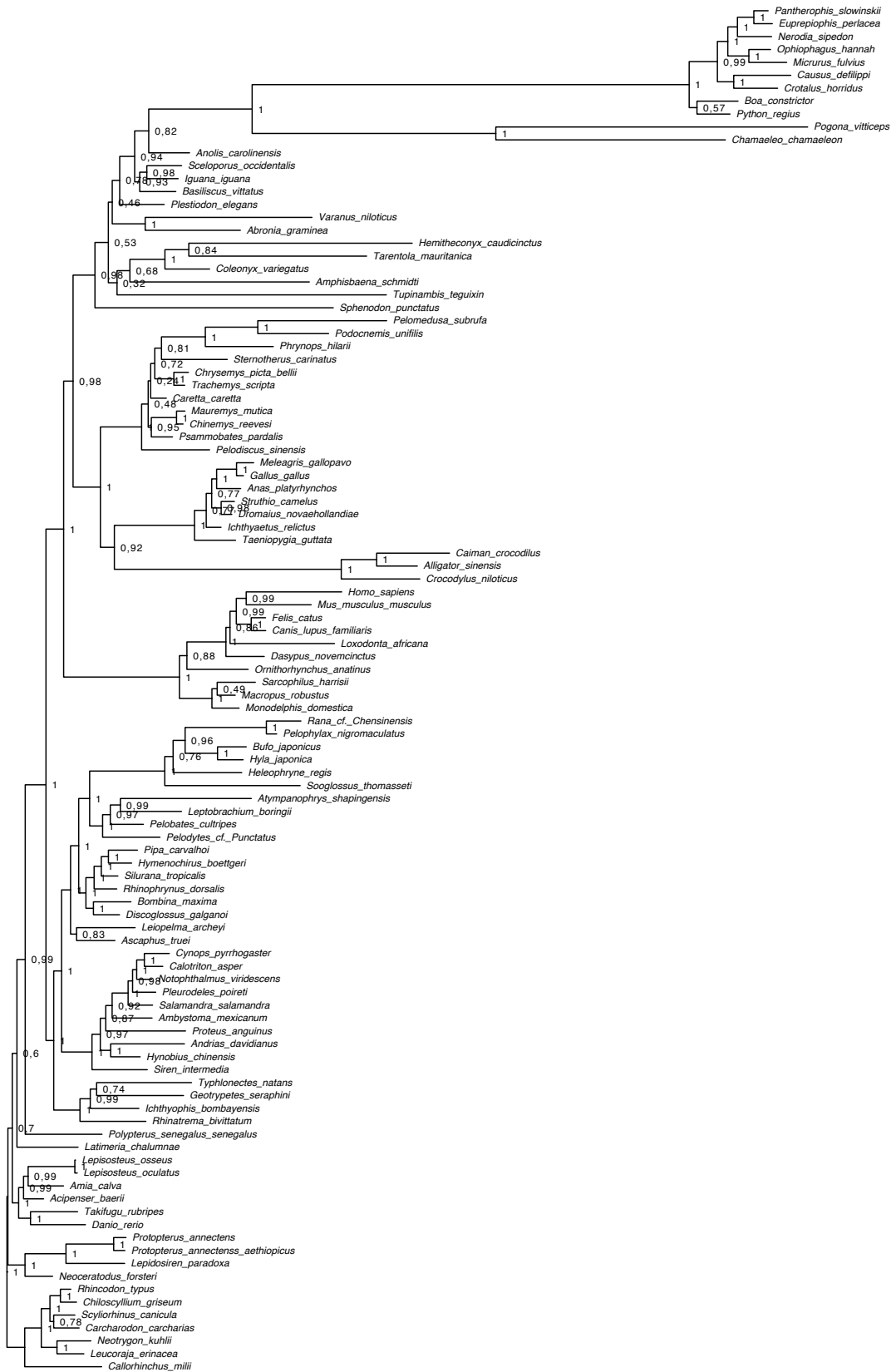
**Supplementary Figure 9.** Coalescent-based tree estimated from the 2DP dataset using ASTRAL-II. All nodes received maximum support from multi-locus bootstrapping and quartet-based local node support, except for nodes where the actual numbers are shown. Scale bar is in coalescent units. This species tree differs from the BI tree obtained from a concatenated matrix of the same dataset (Supplementary Fig. 7) in the relative positions of *Lepisosteus* (and incongruence that is strongly supported), and the relative positions of *Tupinambis* with *Podarcis* and *Caretta* with *Pelodiscus*; both disagreements that are lowly supported.





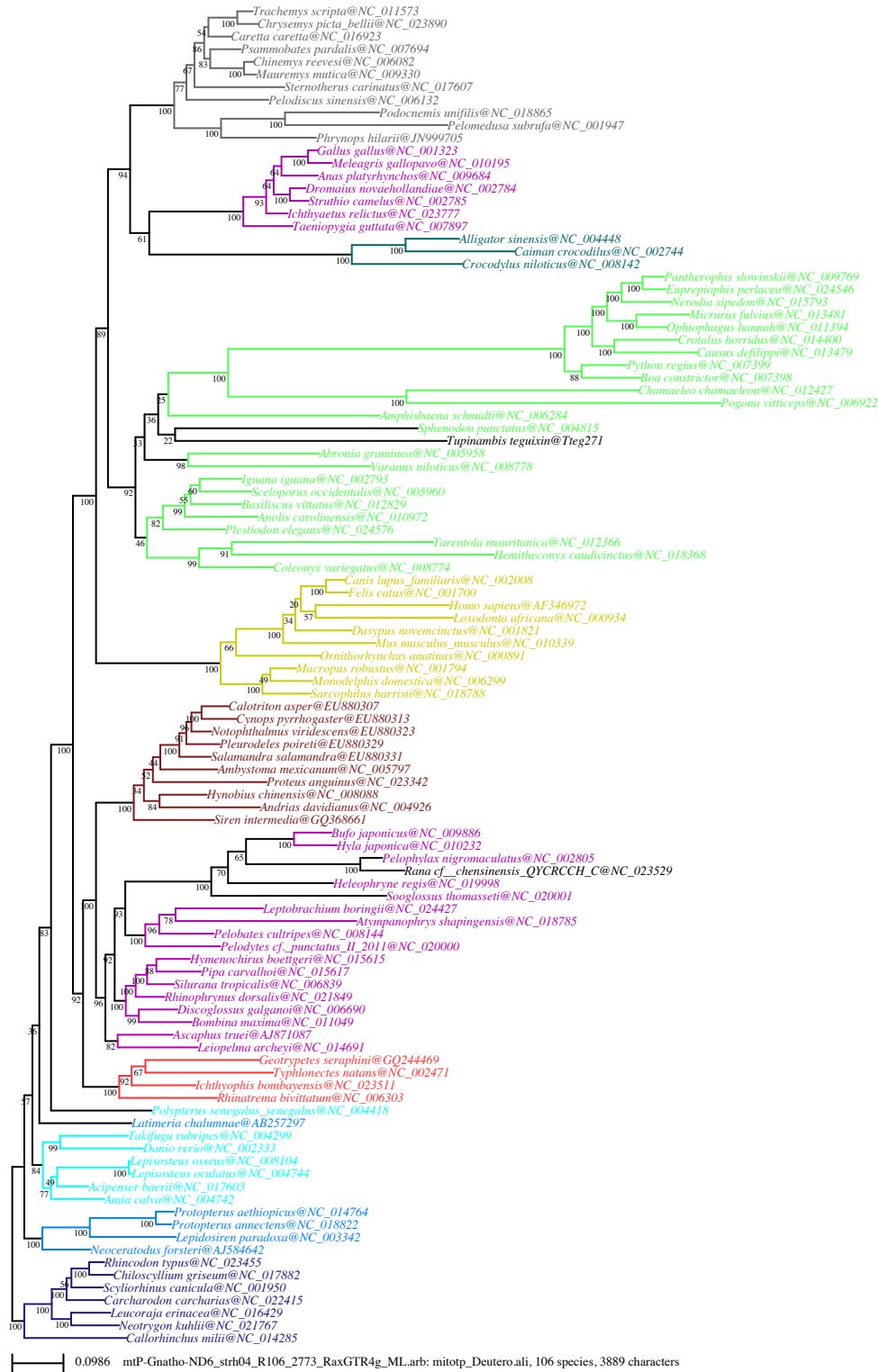
0.4

**Supplementary Figure 10.** Bayesian majority-rule consensus tree reconstructed from the mitochondrial dataset with 106 species (2,773 amino acid positions) under the CAT-GTR+Γ model in PhyloBayes. Numbers at nodes are posterior probabilities and scale bar is in substitutions site<sup>-1</sup>.

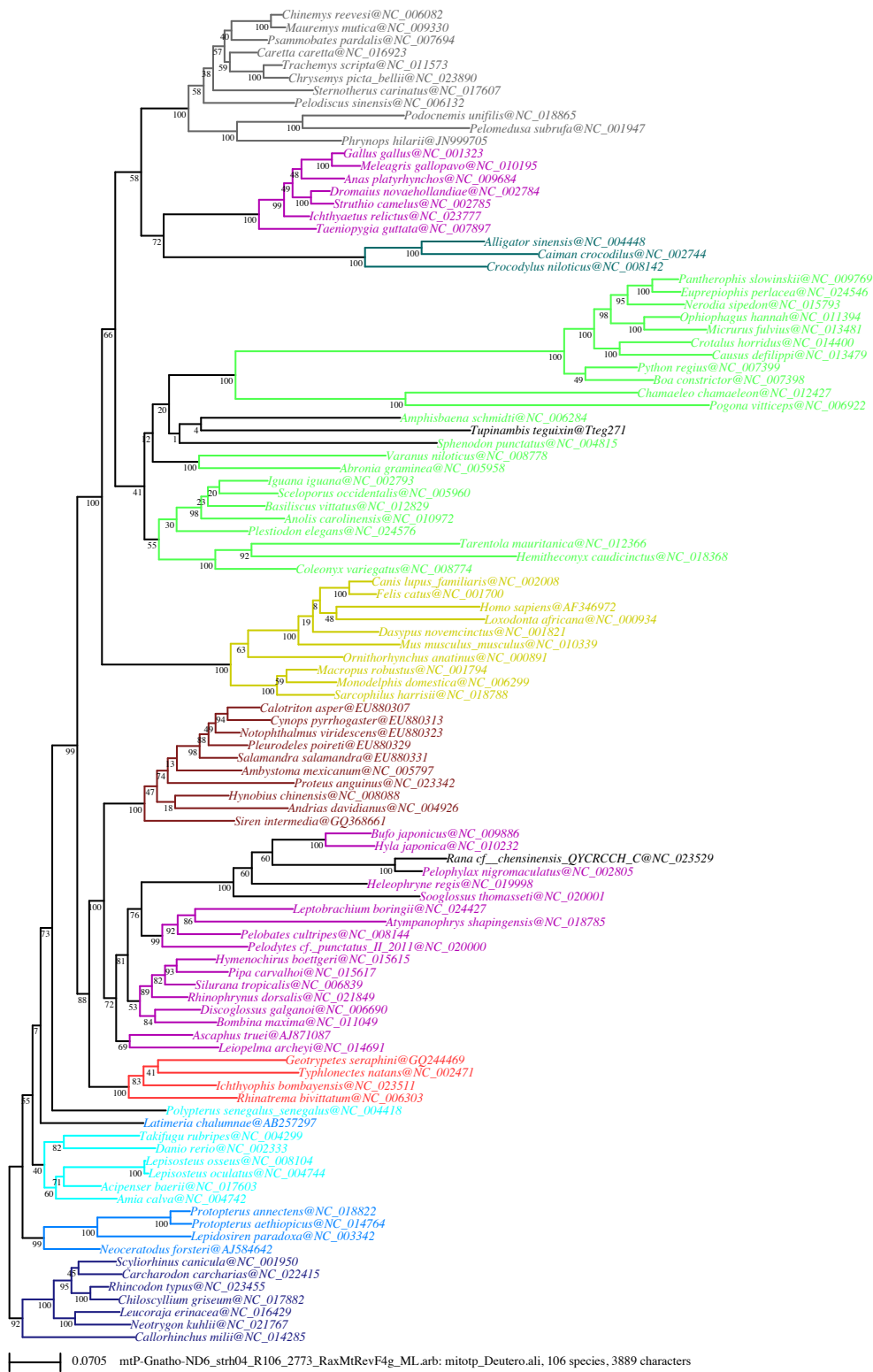


0.3

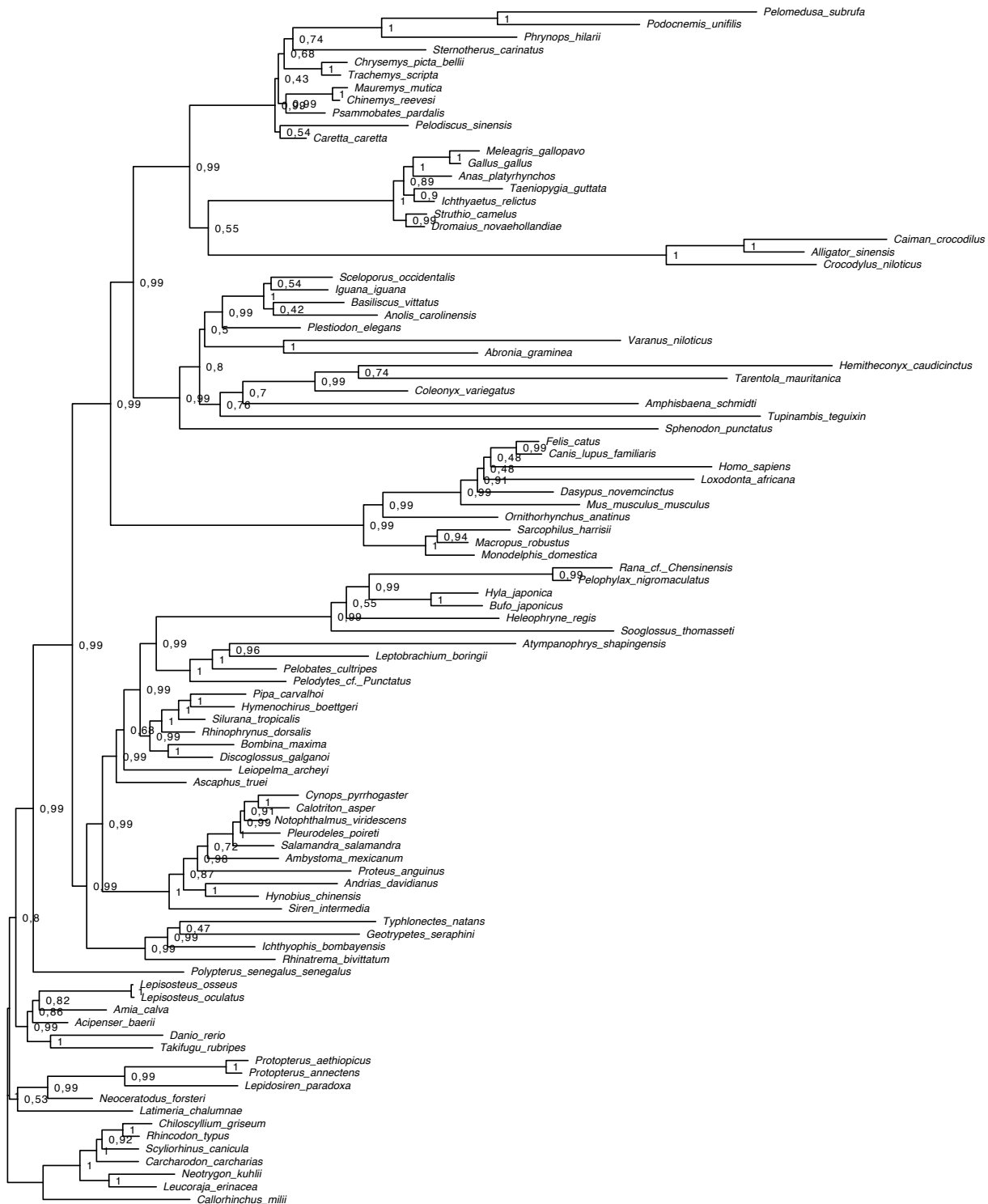
**Supplementary Figure 11.** Bayesian majority-rule consensus tree reconstructed from the mitochondrial dataset with 106 species (2,773 amino acid positions) under the CAT+Γ model in PhyloBayes. Numbers at nodes are posterior probabilities and scale bar is in substitutions site<sup>-1</sup>.



**Supplementary Figure 12.** Maximum likelihood phylogeny estimated from the mitochondrial dataset with 106 species (2,773 amino acid positions) under the GTR+ $\Gamma$  model in RAxML. Numbers at nodes are non-parametric bootstrap support and scale bar is substitutions site<sup>-1</sup>.

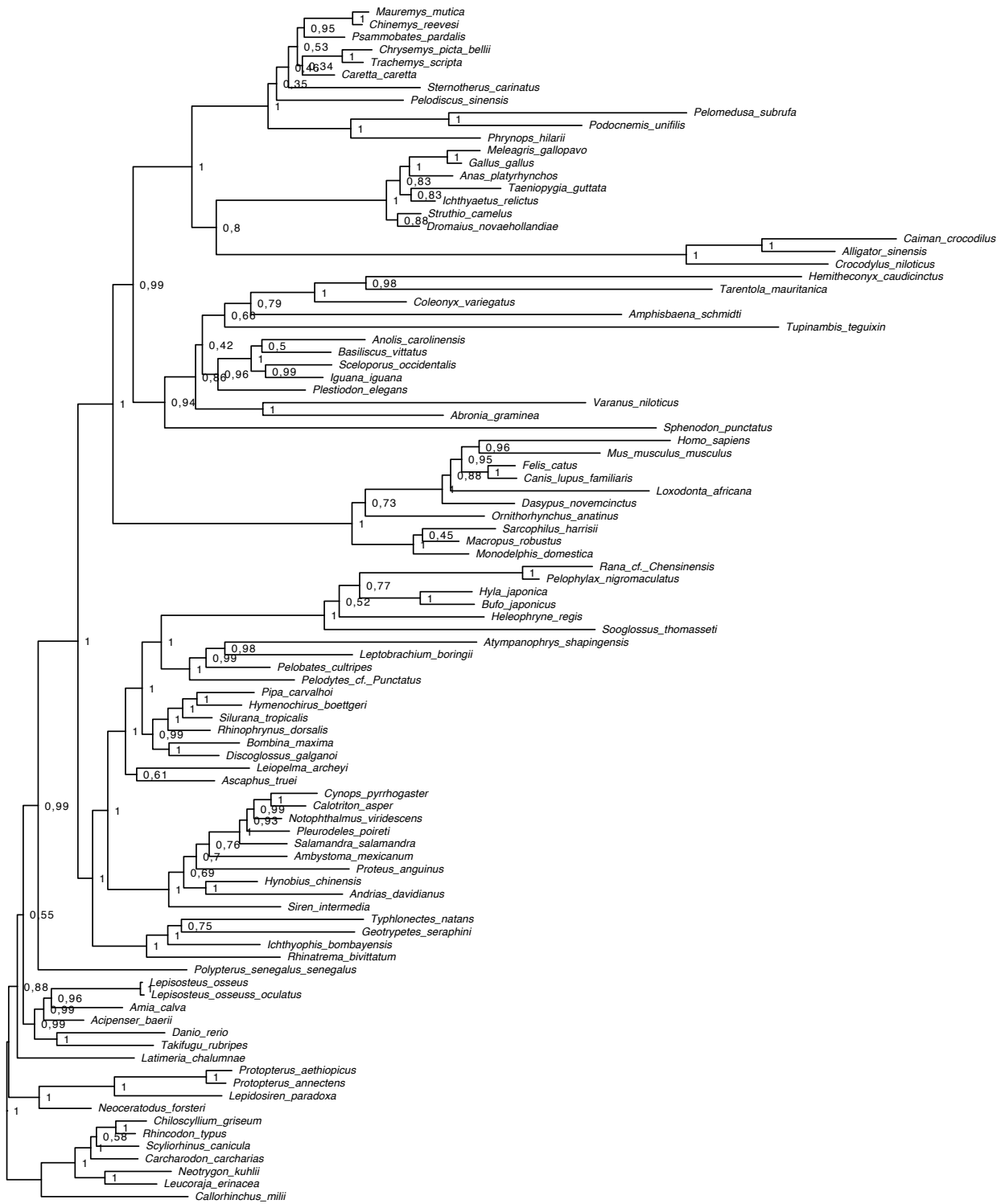


**Supplementary Figure 13.** Maximum likelihood phylogeny estimated from the mitochondrial dataset with 106 species (2,773 amino acid positions) under the MTREV+ $\Gamma$  model in RAxML. Numbers at nodes are non-parametric bootstrap support and scale bar is substitutions site<sup>-1</sup>.



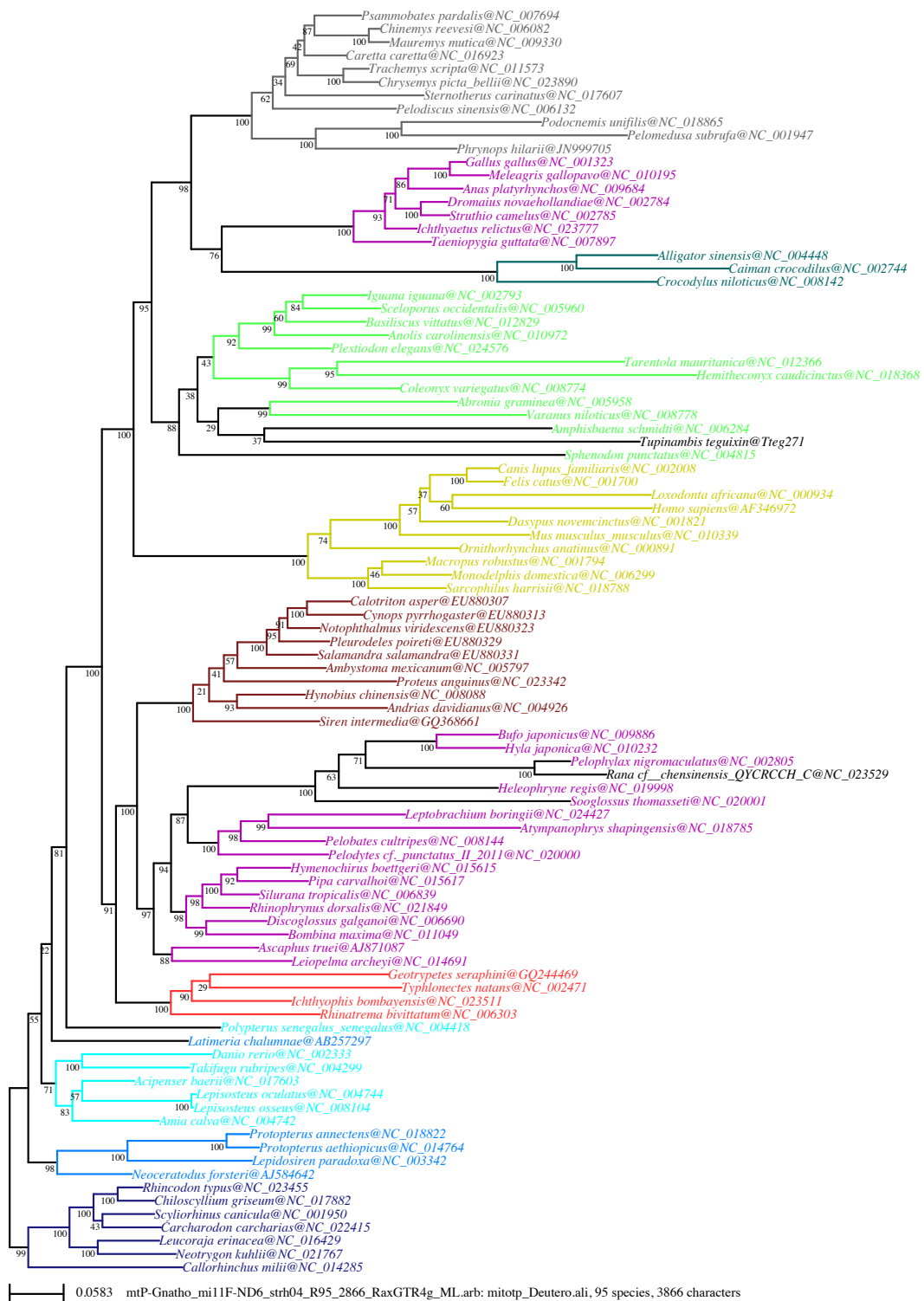
0.3

**Supplementary Figure 14.** Bayesian majority-rule consensus tree reconstructed from the mitochondrial dataset with 95 species (2,866 amino acid positions) under the CAT-GTR+ $\Gamma$  model in PhyloBayes. Numbers at nodes are posterior probabilities and scale bar is in substitutions site<sup>-1</sup>.

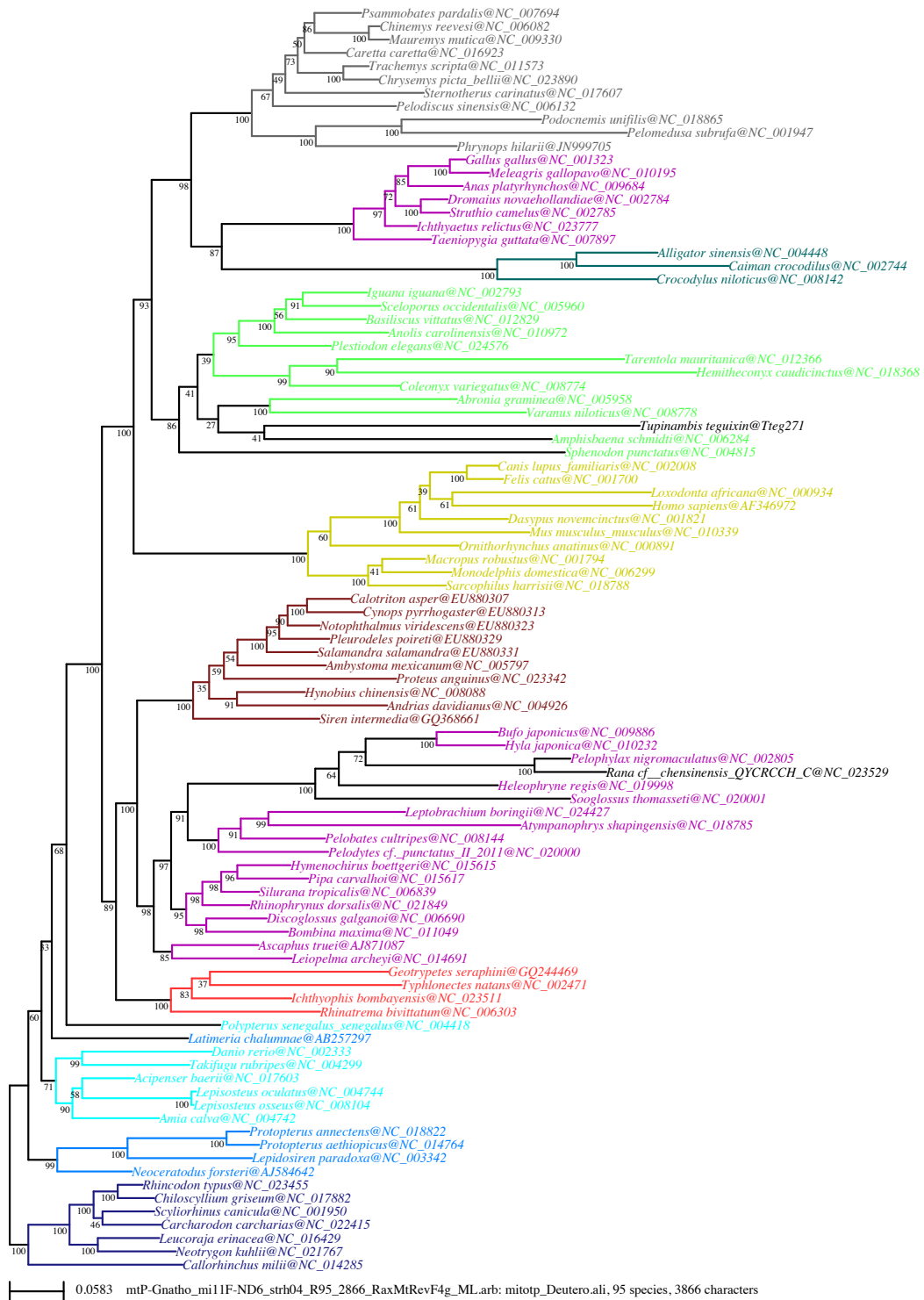


0.2

**Supplementary Figure 15.** Bayesian majority-rule consensus tree reconstructed from the mitochondrial dataset with 95 species (2,866 amino acid positions) under the CAT+ $\Gamma$  model in PhyloBayes. Numbers at nodes are posterior probabilities and scale bar is in substitutions site<sup>-1</sup>.

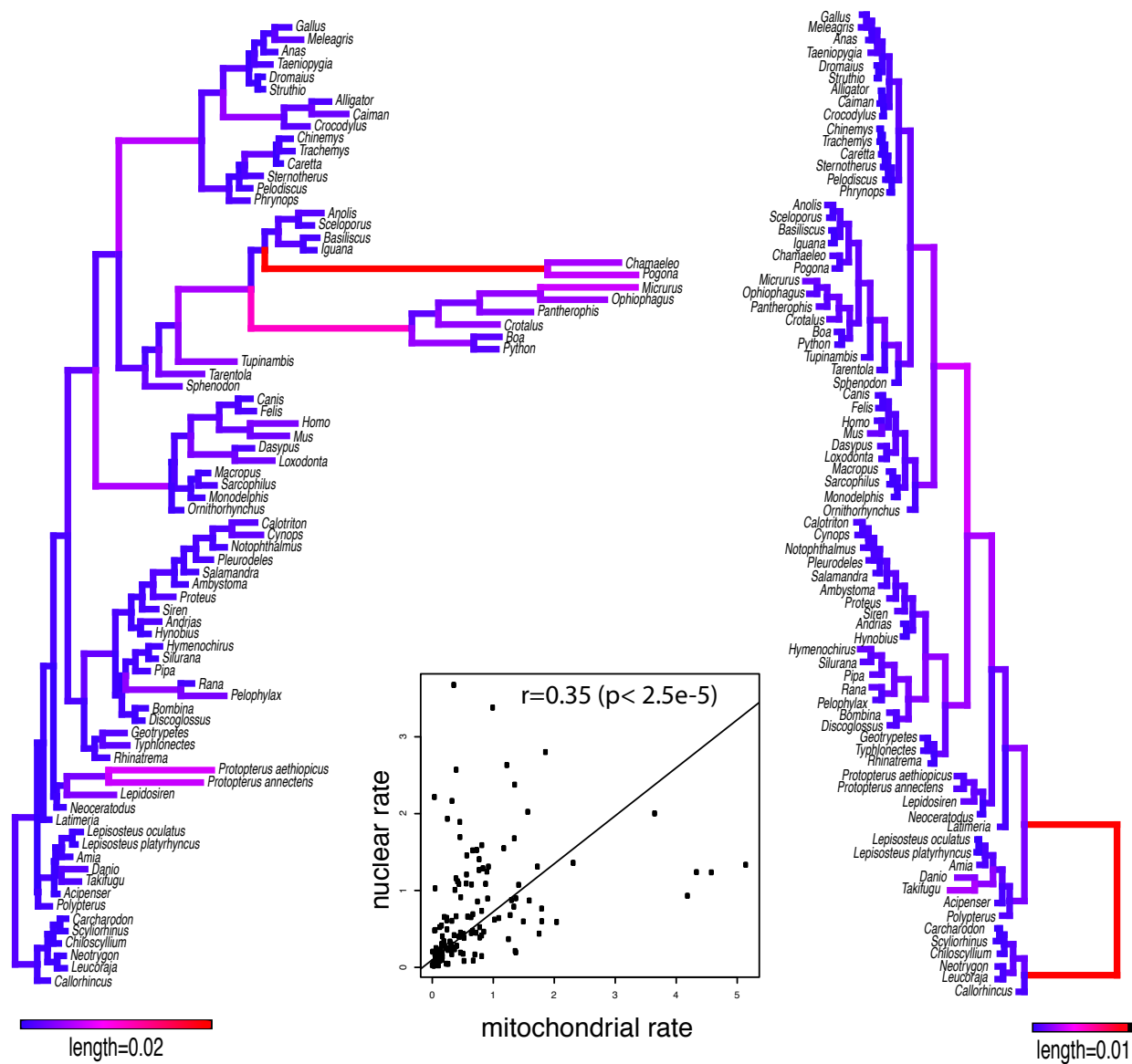


**Supplementary Figure 16.** Maximum likelihood phylogeny estimated from the mitochondrial dataset with 95 species (2,866 amino acid positions) under the GTR+ $\Gamma$  model in RAxML. Numbers at nodes are non-parametric bootstrap support and scale bar is substitutions site<sup>-1</sup>.

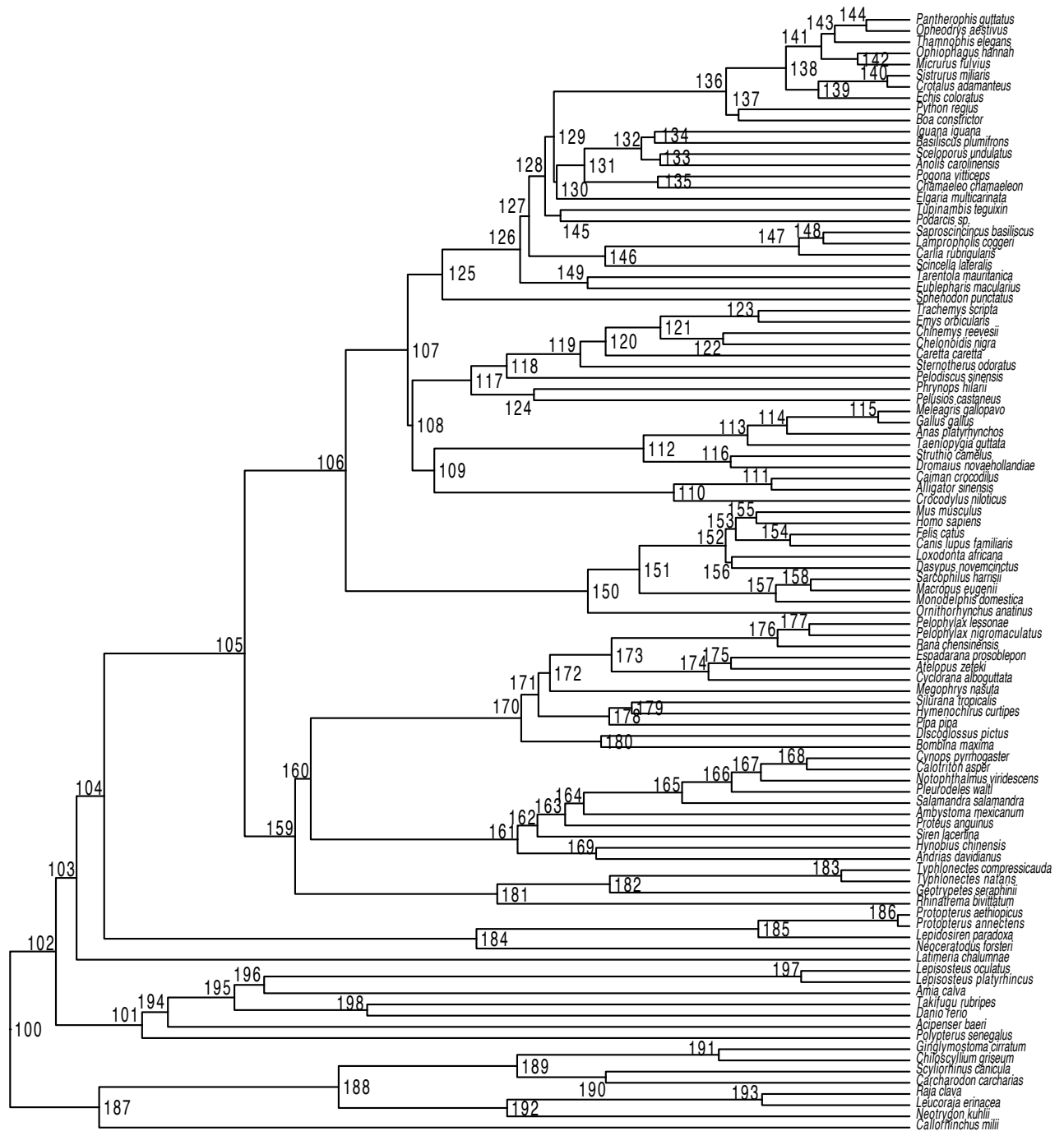


**Supplementary Figure 17.** Maximum likelihood phylogeny estimated from the mitochondrial dataset with 95 species (2,866 amino acid positions) under the MTREV+ $\Gamma$  model in RAXML. Numbers at nodes are non-parametric bootstrap support and scale bar is substitutions site<sup>-1</sup>.





**Supplementary Figure 18.** Relationship between mitochondrial (left tree) and nuclear (right tree) substitution rates. Trees show the 78 species that are present in both the mitochondrial and nuclear datasets, and branch lengths represent the number of substitution site<sup>-1</sup> million year<sup>-1</sup>. Scatter plot shows the correlation between mitochondrial and nuclear rates as measured from individual branches in the trees (N=137) and shows the Pearson's correlation.



**Supplementary Figure 19.** Tree showing the node IDs corresponding to Supplementary Tables 3, 4, 8 and 9.

## Supplementary References

- 1 Grabherr, M. G. *et al.* Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* **29**, 644-652 (2011).
- 2 Haas, B. J. *et al.* *De novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protoc.* **8**, 1494-1512 (2013).
- 3 Schulz, M. H., Zerbino, D. R., Vingron, M. & Birney, E. Oases: Robust *de novo* RNA-seq assembly across the dynamic range of expression levels. *Bioinformatics* **28**, 1086-1092 (2012).
- 4 Irisarri, I. & Meyer, A. The identification of the closest living relative(s) of tetrapods: Phylogenomic lessons for resolving short ancient internodes. *Syst. Biol.* **65**, 1057-1075 (2016).
- 5 Chevreur, B., Wetter, T. & Suhai, S. in *Computer Science and Biology: Proceedings of the German Conference on Bioinformatics (GCB)* p.45-56.
- 6 Venkatesh, B. *et al.* Elephant shark genome provides unique insights into gnathostome evolution. *Nature* **505**, 174-179 (2014).
- 7 Irisarri, I., San Mauro, D., Green, D. M. & Zardoya, R. The complete mitochondrial genome of the relict frog *Leiopelma archeyi*: Insights into the root of the frog tree of life. *Mitochondr. DNA* **21**, 173-182 (2010).
- 8 Pyron, R. A., Burbrink, F. T. & Wiens, J. J. A phylogeny and revised classification of Squamata, including 4161 species of lizards and snakes. *BMC Evol. Biol.* **13**, 93 (2013).
- 9 Edgar, R. C. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**, 2460-2461 (2010).
- 10 Li, L., Stoeckert, C. J. & Roos, D. S. OrthoMCL: Identification of ortholog groups for eukaryotic genomes. *Genome Res.* **13**, 2178-2189 (2003).
- 11 Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772-780 (2013).
- 12 Castresana, J. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol. Biol. Evol.* **17**, 540-552 (2000).
- 13 Stamatakis, A. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312-1313 (2014).
- 14 Meyer, A. & Schartl, M. Gene and genome duplications in vertebrates: The one-to-four (-to-eight in fish) rule and the evolution of novel gene functions. *Curr. Opin. Cell Biol.* **11**, 699-704 (1999).
- 15 Taylor, J. S., Braasch, I., Frickey, T., Meyer, A. & Van de Peer, Y. Genome duplication, a trait shared by 22,000 species of ray-finned fish. *Genome Res.* **13**, 382-390 (2003).
- 16 Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403-410 (1990).
- 17 Li, B. & Dewey, C. RSEM: Accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12**, 323 (2011).
- 18 Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009).
- 19 Philippe, H. *et al.* Resolving difficult phylogenetic questions: Why more sequences are not enough. *PLoS Biol.* **9**, e1000602 (2011).
- 20 Prosdocimi, E. M. *et al.* Errors in ribosomal sequence datasets generated using PCR-coupled 'panbacterial' pyrosequencing, and the establishment of an improved approach. *Mol. Cell Probes* **27**, 65-67 (2013).
- 21 Amemiya, C. T. *et al.* The African coelacanth genome provides insights into tetrapod evolution. *Nature* **496**, 311-316 (2013).

- 22 Criscuolo, A. & Gribaldo, S. BMGE (Block Mapping and Gathering with Entropy): A new software for selection of phylogenetic informative regions from multiple sequence alignments. *BMC Evol. Biol.* **10**, 210 (2010).
- 23 Roure, B., Rodriguez-Ezpeleta, N. & Philippe, H. SCaFoS: A tool for selection, concatenation and fusion of sequences for phylogenomics. *BMC Evol. Biol.* **7**, S2 (2007).
- 24 Hara, Y. *et al.* Optimizing and benchmarking de novo transcriptome sequencing: from library preparation to assembly evaluation. *BMC Genomics* **16**, 977 (2015).
- 25 Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V. & Zdobnov, E. M. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**, 3210-3212 (2015).
- 26 Buchfink, B., Xie, C. & Huson, D. H. Fast and sensitive protein alignment using DIAMOND. *Nat. Meth.* **12**, 59-60 (2015).
- 27 Altschul, S. *et al.* Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucl. Acids Res.* **25**, 3389 - 3402 (1997).
- 28 Bernt, M. *et al.* A comprehensive analysis of bilaterian mitochondrial genomes and phylogeny. *Mol. Phylogenet. Evol.* **69**, 352-364 (2013).
- 29 Philippe, H. MUST, a computer package of Management Utilities for Sequences and Trees. *Nuc. Acids Res.* **21**, 5264-5272 (1993).
- 30 Delsuc, F., Tsagkogeorga, G., Lartillot, N. & Philippe, H. Additional molecular support for the new chordate phylogeny. *Genesis* **46**, 592-604 (2008).
- 31 Lartillot, N., Rodrigue, N., Stubbs, D. & Richer, J. PhyloBayes MPI: Phylogenetic reconstruction with infinite mixtures of profiles in a parallel environment. *Syst. Biol.* **62**, 611-615 (2013).
- 32 Mirarab, S. & Warnow, T. ASTRAL-II: coalescent-based species tree estimation with many hundreds of taxa and thousands of genes. *Bioinformatics* **31**, i44-i52 (2015).
- 33 Sayyari, E. & Mirarab, S. Fast coalescent-based computation of local branch support from quartet frequencies. *Mol. Biol. Evol.* **33**, 1654-1668 (2016).
- 34 Seo, T.-K. Calculating bootstrap probabilities of phylogeny using multilocus sequence data. *Mol. Biol. Evol.* **25**, 960-971 (2008).
- 35 Stamatakis, A. RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22**, 2688-2690 (2006).
- 36 Lartillot, N., Lepage, T. & Blanquart, S. PhyloBayes 3: A Bayesian software package for phylogenetic reconstruction and molecular dating. *Bioinformatics* **25**, 2286-2288 (2009).
- 37 Near, T. J., Meylan, Peter A. & Shaffer, H. B. Assessing concordance of fossil calibration points in molecular clock studies: An example using turtles. *Am. Nat.* **165**, 137-146 (2005).
- 38 Gregory, T. R. A bird's-eye view of the c-value enigma: Genome size, cell size, and metabolic rate in the class Aves. *Evolution* **56**, 121-130 (2002).
- 39 Kraaijeveld, K. Genome size and species diversification. *Evol. Biol.* **37**, 227-233 (2010).
- 40 Felsenstein, J. Phylogenies and the comparative method. *Am. Nat.* **125**, 1-15 (1985).
- 41 R: A language and environment for statistical computing (R Foundation for Statistical Computing, Vienna, Austria, 2017).
- 42 Paradis, E., Claude, J. & Strimmer, K. APE: Analyses of phylogenetics and evolution in R language. *Bioinformatics* **20**, 289-290 (2004).
- 43 Lartillot, N. & Poujol, R. A phylogenetic model for investigating correlated evolution of substitution rates and continuous phenotypic characters. *Mol. Biol. Evol.* **28**, 729-744 (2011).
- 44 Kumar, S., Stecher, G. & Tamura, K. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* **33**, 1870-1874 (2016).
- 45 Revell, L. J. phytools: An R package for phylogenetic comparative biology (and other things). *Meth. Ecol. Evol.* **3**, 217-223 (2012).

- 46 Rambaut, A. & Grass, N. C. Seq-Gen: An application for the Monte Carlo simulation of DNA sequence evolution along phylogenetic trees. *Comput. Appl. Biosci.* **13**, 235-238 (1997).
- 47 Fong, J. J., Brown, J. M., Fujita, M. K. & Boussau, B. A phylogenomic approach to vertebrate phylogeny supports a turtle-archosaur affinity and a possible paraphyletic Lissamphibia. *PLoS ONE* **7**, e48990 (2012).
- 48 Hedges, S. B. & Poling, L. L. A molecular phylogeny of reptiles. *Science* **283**, 998-1001 (1999).
- 49 Shaffer, H. B., Meylan, P. & McKnight, M. L. Tests of turtle phylogeny: Molecular, morphological, and paleontological approaches. *Syst. Biol.* **46**, 235-268 (1997).
- 50 Benton, M. G., Donoghue, Philip C. J. & Asher, R. J. in *The Timetree of Life* (eds S. Blar Hedges & Sudhir Kumar) p. 35-86 (Oxford University Press, 2009).
- 51 Hurley, I. A. *et al.* A new time-scale for ray-finned fish evolution. *Proc. R. Soc. B-Biol. Sci.* **274**, 489-498 (2007).
- 52 Reisz, R. R. & Müller, J. Molecular timescales and the fossil record: A paleontological perspective. *Trends Genet.* **20**, 237-241 (2004).
- 53 Benton, M. J. & Donoghue, P. C. J. Paleontological evidence to date the Tree of Life. *Mol. Biol. Evol.* **24**, 26-53 (2007).
- 54 Graur, D. & Martin, W. Reading the entrails of chickens: Molecular timescales of evolution and the illusion of precision. *Trends Genet.* **20**, 80-86 (2004).
- 55 Müller, J. & Reisz, R. R. Four well-constrained calibration points from the vertebrate fossil record for molecular clock estimates. *Bioessays* **27**, 1069-1075 (2005).
- 56 Noonan, Brice P. & Chippindale, Paul T. Vicariant origin of Malagasy reptiles supports late Cretaceous antarctic land bridge. *Am. Nat.* **168**, 730-741 (2006).
- 57 Jones, M. *et al.* Integration of molecules and new fossils supports a Triassic origin for Lepidosauria (lizards, snakes, and tuatara). *BMC Evol. Biol.* **13**, 208 (2013).
- 58 Wesley, G. D. & Flynn, J. J. A revision of *Tapocyon* (Carnivoramorpha), including analysis of the first cranial specimens and identification of a new species. *J. Paleontol.* **77**, 769-783 (2003).
- 59 Rage, J. & Roček, Z. Redescription of *Triadobatrachus massinoti* (Piveteau, 1936) an anuran amphibian from the early Triassic. *Palaeontogr. Abt. A* **206**, 1-16 (1989).
- 60 Gao, K.-Q. & Shubin, N. H. Earliest known crown-group salamanders. *Nature* **422**, 424-428 (2003).
- 61 Evans, S. E., Milner, A. R. & Mussett, F. A discoglossid frog from the Middle Jurassic of England. *Palaeontology* **33**, 299-311 (1990).
- 62 Henrici, A. C. A new pipoid anuran from the late Jurassic Morrison formation at Dinosaur National Monument, Utah. *J. Vertebr. Paleontol.* **18**, 321-332 (1998).
- 63 Pitman III, W., Cande, S., LaBrecque, J. & Pindell, J. in *Biological relationships between Africa and South America* (ed P. Goldblatt) 405-439 (Yale University Press, 1993).
- 64 Heinicke, M. P., Naylor, G. J. P. & Hedges, S. B. in *The Timetree of Life* (eds S. Blar Hedges & Sudhir Kumar) p. 320-327 (Oxford University Press, 2009).
- 65 Chen, M. Y., Liang, D. & Zhang, P. Selecting question-specific genes to reduce incongruence in phylogenomics: A case study of jawed vertebrate backbone phylogeny. *Syst. Biol.* **64**, 1104-1120 (2015).