

Phylotranscriptomic consolidation of the jawed vertebrate timetree

Iker Irisarri^{1,11*}, Denis Baurain², Henner Brinkmann³, Frédéric Delsuc⁴, Jean-Yves Sire⁵, Alexander Kupfer⁶, Jörn Petersen³, Michael Jarek⁷, Axel Meyer¹, Miguel Vences⁸ and Hervé Philippe^{9,10*}

Phylogenomics is extremely powerful but introduces new challenges as no agreement exists on 'standards' for data selection, curation and tree inference. We use jawed vertebrates (Gnathostomata) as a model to address these issues. Despite considerable efforts in resolving their evolutionary history and macroevolution, few studies have included a full phylogenetic diversity of gnathostomes, and some relationships remain controversial. We tested a new bioinformatic pipeline to assemble large and accurate phylogenomic datasets from RNA sequencing and found this phylotranscriptomic approach to be successful and highly cost-effective. Increased sequencing effort up to about 10 Gbp allows more genes to be recovered, but shallower sequencing (1.5 Gbp) is sufficient to obtain thousands of full-length orthologous transcripts. We reconstruct a robust and strongly supported timetree of jawed vertebrates using 7,189 nuclear genes from 100 taxa, including 23 new transcriptomes from previously unsampled key species. Gene jackknifing of genomic data corroborates the robustness of our tree and allows calculating genome-wide divergence times by overcoming gene sampling bias. Mitochondrial genomes prove insufficient to resolve the deepest relationships because of limited signal and among-lineage rate heterogeneity. Our analyses emphasize the importance of large, curated, nuclear datasets to increase the accuracy of phylogenomics and provide a reference framework for the evolutionary history of jawed vertebrates.

Understanding the evolutionary relationships among organisms is a prerequisite for any biological study aiming at explaining key processes such as adaptive radiations or evolutionary convergences. Evolutionary relatedness is generally represented with phylogenetic trees, which need to be robust and accurate if one aims at obtaining credible macroevolutionary inferences. In the past decade, genome-scale datasets (phylogenomics) have revolutionized molecular phylogenetics thanks to their ability to yield precise estimates of phylogeny and more precise divergence times by reducing sampling error, one of the main hurdles in the pre-genomics era. Several methodologies based on high-throughput sequencing have been used to generate raw data for phylogenomics. But as well as the obvious advantages, phylogenetic inference based on genomic data poses numerous challenges. For the assembly of genome-scale datasets, these include the removal of contaminants (from symbionts, pathogens or food items in the original sample; or introduced during processing steps such as human DNA or sample cross-contaminations), misalignments due to erroneous sequence stretches (often produced by sequencing and annotation errors in low-coverage genome assemblies), the effective detection and removal of paralogues, and the presence of large amounts of missing data, often aggravated by the difficulty of identifying orthology in only partially assembled transcripts. Paralogy in particular can have very detrimental effects on phylogenomic analyses¹, but the robustness of tree topology to the inclusion of paralogues is generally not evaluated.

Because phylogenomics relies on hundreds or thousands of genes and taxa, manual data curation has become unfeasible, and automatic solutions need to be devised. Phylogenomic analyses have generally relied on pooling evidence from multiple genes by concatenation or used 'summary' coalescent-based species-tree methods. The size of genomic datasets also makes phylogenomics more sensitive to model misspecification (systematic error), which often translates into long-branch attraction problems². Systematic error may be reduced with complex mixture models, but their application to large-scale phylogenomic matrices can sometimes become computationally intractable. In addition, phylogenomic alignments are known to inflate non-parametric bootstrap support values and Bayesian clade posterior probabilities, a precision not always accompanied by increased accuracy, thus rendering the interpretation of these support metrics difficult.

The above challenges regarding the quality of the data and the robustness of analytical approaches need to be carefully taken into account to produce reliable estimates of both phylogeny and divergence times. Jawed vertebrates (Gnathostomata) represent a good system to benchmark these challenges because of the availability of genomic data for many species but the remarkable absence of several species with key phylogenetic positions, and the relatively good knowledge of their phylogeny except for some nodes that have been controversial. In addition, jawed vertebrates are among

¹Lehrstuhl für Zoologie und Evolutionsbiologie, Department of Biology, University of Konstanz, Universitätsstrasse 10, Konstanz 78464, Germany.

²InBioS-Eukaryotic Phylogenomics, Department of Life Sciences and PhytoSYSTEMS, University of Liège, Liège 4000, Belgium. ³Leibniz-Institut DSMZ-

German Collection of Microorganisms and Cell Cultures, Braunschweig 38124, Germany. ⁴Institut des Sciences de l'Evolution, UMR 5554, CNRS, IRD,

EPHE, Université de Montpellier, Montpellier 34095, France. ⁵Institut de Biologie Paris-Seine, UMR7138, Sorbonne Universities, Paris 75005, France.

⁶Department of Zoology, Stuttgart State Museum of Natural History, Stuttgart 70191, Germany. ⁷Department of Genome Analytics, Helmholtz Centre

for Infection Research, Braunschweig 38124, Germany. ⁸Zoological Institute, Braunschweig University of Technology, Braunschweig 38106, Germany.

⁹Centre for Biodiversity Theory and Modelling, UMR CNRS 5321, Station of Theoretical and Experimental Ecology, Moulis 09200, France.

¹⁰Departement de Biochimie, Université de Montréal, Montréal, QC H3C3J7, Canada. Present address: ¹¹Systematic Biology Program, Department of

Organismal Biology, University of Uppsala, Norbyvägen 18D, Uppsala 75236, Sweden. *e-mail: irisarri.iker@gmail.com; herve.philippe@sete.cnrs.fr

the best-studied organisms and include astonishing examples of convergent evolution (for example flight, echolocation or limb loss) and prominent instances of classic paraphyletic taxa such as ‘fishes’ or ‘reptiles’. Biologists have long been interested in understanding the evolutionary relationships among jawed vertebrates, first using morphological characters and later with sequence data. Molecular phylogenies have greatly contributed towards shaping the jawed vertebrate tree, in many instances corroborating classical morphology-based classifications, but sometimes establishing new hypotheses such as the close relationship of turtles with crocodiles and birds. Studies relying on mitochondrial genomes (mitogenomes) have resolved several controversial issues³ but also recovered some unorthodox relationships⁴. Earlier molecular studies based on multiple genes obtained by classical Sanger-sequencing approaches have generally been limited by the number of genes or taxa, and were generally restricted to particular lineages such as ray-finned fishes⁵, amphibians⁶, squamate reptiles⁷, mammals⁸ or birds⁹. With the rise of genome-scale molecular datasets, it became possible to use ever-larger datasets in an attempt at solving the relationships in the Tree of Life, and many nodes of the jawed vertebrate tree have been confirmed by phylogenomic analyses based on datasets obtained by second-generation sequencing and typically focusing on particular gnathostome clades^{10–17}. Despite this growing consensus, some phylogenomic studies have also challenged important relationships, such as the monophyly and internal relationships of amphibians¹⁸ or the position of turtles¹⁹, demonstrating that crucial aspects of the jawed vertebrate tree still require careful attention. Further evolutionary relationships also remain controversial because of incongruence among molecular phylogenies or with morphological evidence, such as the close relationship of iguanian lizards with snakes^{20–22} or the relationships among tongueless frogs^{23,24}. Convincingly resolving difficult nodes requires more than just a large number of genes, and instead a focus is needed on carefully avoiding and removing contaminations and errors in the data, and avoiding model misspecifications²⁵.

Since their origin in the Ordovician period (~470 million years ago (Ma)), jawed vertebrates have diversified into lineages with markedly different morphologies and life histories, including hyperdiverse radiations such as spiny-rayed fishes, birds, modern frogs (Neobatrachia) and placental mammals. As an appealing hypothesis, the main diversification bursts of these hyperdiverse radiations have been proposed to coincide with the Cretaceous/Palaeogene boundary^{5,15}, but owing to uncertainties in timetree reconstruction and methodological disputes on molecular dating, this hypothesis remains contentious, especially for mammals^{26–28}.

Here, we use a phylotranscriptomic approach to reconstruct the backbone of the jawed vertebrate tree based on a dataset of unprecedented size composed of 7,189 genes for 100 species representing all main gnathostome lineages (a total of 3,791,500 aligned amino acid positions). The dataset includes 23 newly generated transcriptomes from previously underrepresented clades occupying key phylogenetic positions, particularly early-branching ray-finned and lobe-finned fishes, lungfishes, amphibians and squamate reptiles. We devised a new bioinformatic pipeline to assemble the largest and most informative dataset ever analysed for vertebrates (Supplementary Fig. 1) while focusing on the comprehensive removal of contaminants and paralogues. This dataset is subjected to thorough phylogenetic and molecular dating analyses. We present a strongly supported phylogenetic hypothesis, which is fossil-calibrated to yield robust divergence time estimations, thus providing a reference framework for the evolutionary history of jawed vertebrates.

Results and discussion

Phylotranscriptomic pipeline to assemble clean datasets. We developed a new bioinformatic pipeline (Supplementary Fig. 2) to

assemble an informative and ‘clean’ genome-scale dataset of jawed vertebrates using genome and transcriptome sequence data. For this study we collected RNA-Seq data for 23 previously unsampled gnathostome species representing key lineages. Sequencing effort for the new transcriptomes varied considerably among species (total sequenced base pairs ranged from 1.5 to 26 Gbp; Fig. 1 and Supplementary Table 1), and it correlated positively with (i) the average length of reconstructed transcripts ($r=0.78$; $P=8.207\times 10^{-6}$), (ii) transcriptome completeness, measured as the proportion of recovered core vertebrate genes²⁹ ($r=0.78$; $P=6.173\times 10^{-6}$) and (iii) the total number of amino acids in final phylogenomic datasets ($r=0.82$; $P=0.00066$) (Supplementary Table 2). Despite considerable differences in sequencing effort, all transcriptomes were relatively complete (58.8% to 100% of the 233 core vertebrate genes were recovered; Fig. 1) and thousands of genes readily usable for phylogenomics were reconstructed (2,274 to 13,642 high-coverage genes per species, measured as human proteins at $\geq 70\%$ length coverage; Fig. 1). Hence, deeper sequencing increased the completeness of transcriptome assemblies and the number of genes and amino acid positions in final alignments. Nevertheless, this tendency stabilized at approximately 10 Gbp of total data (for example, 50 million 100-bp-long read pairs), after which a higher sequencing effort did not significantly increase the above performance metrics ($r<0.5$ and $P>0.05$ in all correlations for transcriptomes with >10 Gbp of total data; Supplementary Table 2). Interestingly, genes missing in final phylogenomic matrices were essentially not different in species with shallow or deeply sequenced transcriptomes (assessed by gene ontology (GO) enrichment tests with FDR <0.05 against the annotated set of 7,189 genes, run in the platform Blast2GO), which suggests that sequencing effort does not significantly bias the types of genes present in final alignments.

The new bioinformatic pipeline established herein warranted the high quality of alignments by addressing key issues in data integrity²⁵, including several steps to minimize possible contaminations, resolution of paralogy, masking of misalignments and minimizing missing data. During decontamination steps, similarity searches with the Basic Local Alignment Search Tool (BLAST) were used to identify potentially contaminant sequences from non-vertebrates and human sequences (in this latter case requiring high-identity at the nucleotide level). To remove any remaining contamination, we devised a sensitive protocol that identifies extremely long branches estimated on a fixed reference tree to flag possibly erroneous sequences, which were then removed. Per-sequence missing data was minimized by merging conspecific sequences (typically overlapping partially reconstructed transcripts) with the SCAFoS software tool³⁰, and unreliably aligned regions were discarded. A new tool based on profile hidden Markov models (HMM) was used to mask erroneous sequence stretches typically produced by frame shifts in open reading frames or incorrect structural annotation. We implemented an innovative paralogue-splitting pipeline that specifically targets distant paralogues (those particularly problematic for resolving the backbone of the tree) and further assessed the effect in the tree stability of including various levels of deep paralogy in the datasets. To do that, genes were classified into three sets that contained zero (NoDP), one (1DP) and two (2DP) deep paralogues (that is, duplication events pre-dating the origin of major jawed vertebrate lineages), which were then concatenated into three datasets that were separately analysed: NoDP (4,593 genes, 1,964,439 amino acids, 32% missing data), 1DP (1,162 genes, 668,132 amino acids, 36% missing data) and 2DP (1,434 genes, 1,158,929 amino acids, 39% missing data).

Backbone phylogeny of jawed vertebrates. The phylogeny was estimated based on concatenated alignments by (i) maximum likelihood (ML) under the site-homogeneous LG+F+ Γ and GTR+ Γ models and 100 bootstrap replicates in RAxML, and (ii) Bayesian

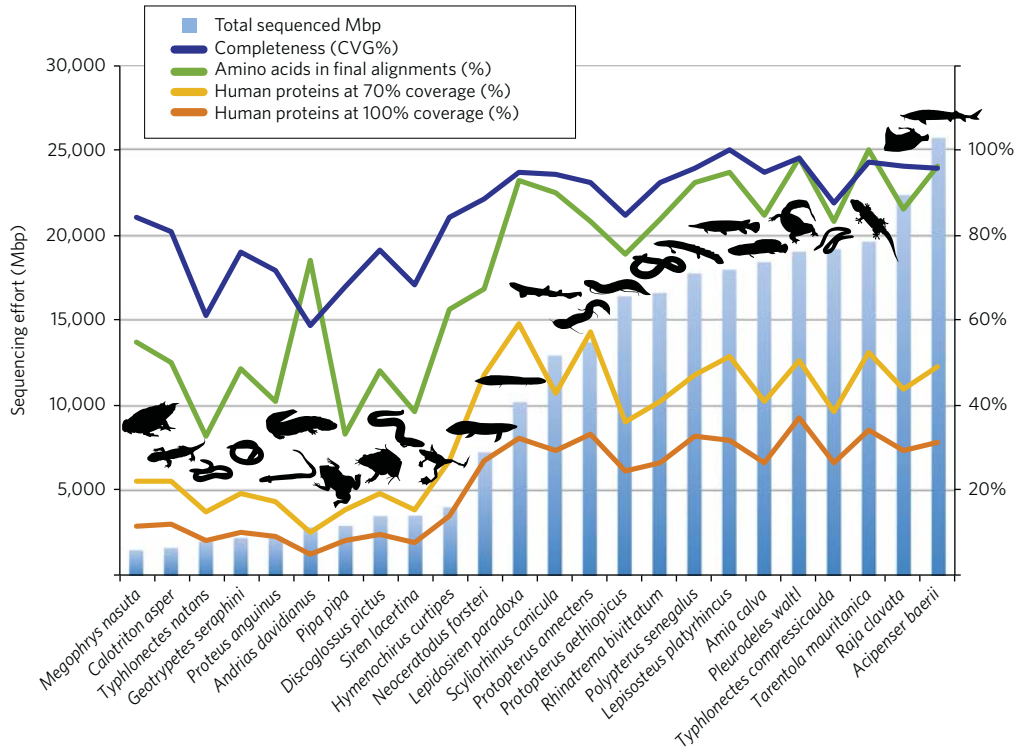


Fig. 1 | Transcriptome sequencing effort and performance in phylogenomic dataset assembly. Histogram represents sequencing effort as total number of sequenced (clean) Mbp (million base pairs). Transcriptome completeness is measured as the proportion of recovered core vertebrate genes (233 CVG; ref. ²⁹). Genes effectively usable for phylogenomics are approximated by the proportion of human proteins reconstructed at full (100%) and nearly full (>70%) lengths (in proportion to a total of 22,964 human genes). The completeness of the relevant species in our final phylogenomic dataset is shown as the proportion of amino acids across all 7,189 genes (3,791,500 aligned amino acids in total).

inference (BI) under the more realistic site-heterogeneous CAT + Γ model in PhyloBayes. For computational tractability of large datasets under complex and computationally expensive models and to further assess the effect of gene sampling, BI analyses were performed on 100 gene jackknife replicates (~50,000 amino acids and ~180 genes per replicate), which were summarized in a final majority-rule consensus tree. Gene jackknifing measures the repeatability of the phylogenetic relationships across genes, which are randomly sampled without replacement from the total set of genes³¹. We used gene jackknife proportions (GJP) as a stringent test for the robustness of the obtained relationships because they were estimated under the more realistic CAT model and based on virtually independent gene sets, each containing ~2.5% of the total alignment, as compared to the ~66% of the total alignment used in non-parametric bootstrapping. In addition, we carried out coalescent-based species tree analyses with ASTRAL-II with 100 replicates of multi-locus bootstrapping on the three nuclear datasets separately. All phylogenetic analyses of the paralogue-free dataset (NoDP), including BI (Fig. 2a) and ML on the concatenated super-matrix and species tree analyses (Supplementary Figs. 3–5), reconstructed fully resolved and almost identical trees that were highly supported: 88% and 95% of the nodes in Fig. 2 received, respectively, full (100%) or high (>75%) GJP. All major uncontroversial vertebrate clades were recovered with full support: cartilaginous fishes (Chondrichthyes) were the sister group of bony fishes (Osteichthyes), including ray-finned (Actinopterygii) and lobe-finned (Sarcopterygii) fishes; within sarcopterygians, tetrapods (Tetrapoda) were monophyletic and encompassed amphibians (Lissamphibia), mammals (Mammalia), turtles (Testudines), birds (Aves), crocodiles (Crocodylia), lepidosaurian reptiles (Lepidosauria) and snakes (Serpentes). Even when using relatively small alignments of ~5,000 amino acids (Fig. 2b, Supplementary Table 3), all the above nodes were recovered with strong support. In

fact, these uncontroversial nodes were also recovered by a large proportion of single-gene trees (58–96% of the genes; Supplementary Table 4) although with varying levels of support.

In contrast, some of the relationships that have remained hotly discussed during the past decades were not unambiguously recovered by single genes nor by relatively small-sized gene jackknife replicates (Fig. 2b and Supplementary Tables 3, 4). Thanks to the use of a larger dataset, however, our analyses effectively resolved these controversial relationships with maximum support (Fig. 2a). (i) Lungfishes (Dipnoi) were the sister group of tetrapods, in agreement with the latest phylogenomic results^{12,32}, and topology tests rejected the alternative hypothesis in which coelacanth and tetrapods are sister taxa³³ (Supplementary Table 5). (ii) Amphibians (Lissamphibia) were monophyletic, and salamanders (Caudata) were the sister group of frogs (Anura) to the exclusion of caecilians (Gymnophiona) (Batrachia hypothesis^{34,35}). Both the paraphyly of amphibians and the alternative sister group of caecilians and salamanders (Procera hypothesis¹⁸) were rejected by topological tests. (iii) Turtles were the sister group of crocodiles and birds (Archosauria), in agreement with most previous phylogenomic studies^{10,11} and the latest morphological evidence³⁶. Topology tests rejected the traditional view of turtles as primarily anapsids (early-branching within ‘reptiles’) as well as turtles as the sister group to either lepidosaurians or crocodiles¹⁸. (iv) The earliest offshoot within salamanders was *Andrias* (Cryptobranchidae) plus *Hynobius* (Hynobiidae)³⁴, and the alternative position of *Siren* (Sirenidae) as the earliest-branching salamander clade³⁷ was statistically rejected. (v) Lastly, our BI tree supports a close relationship between snakes and iguanian and anguimorph lizards (*Elgaria*) (‘Toxicofera’).

Only 4 out of 98 nodes in our phylogeny received relatively low support (<75% GJP; Fig. 2), and we consider these nodes in need of further confirmation. Besides relationships within

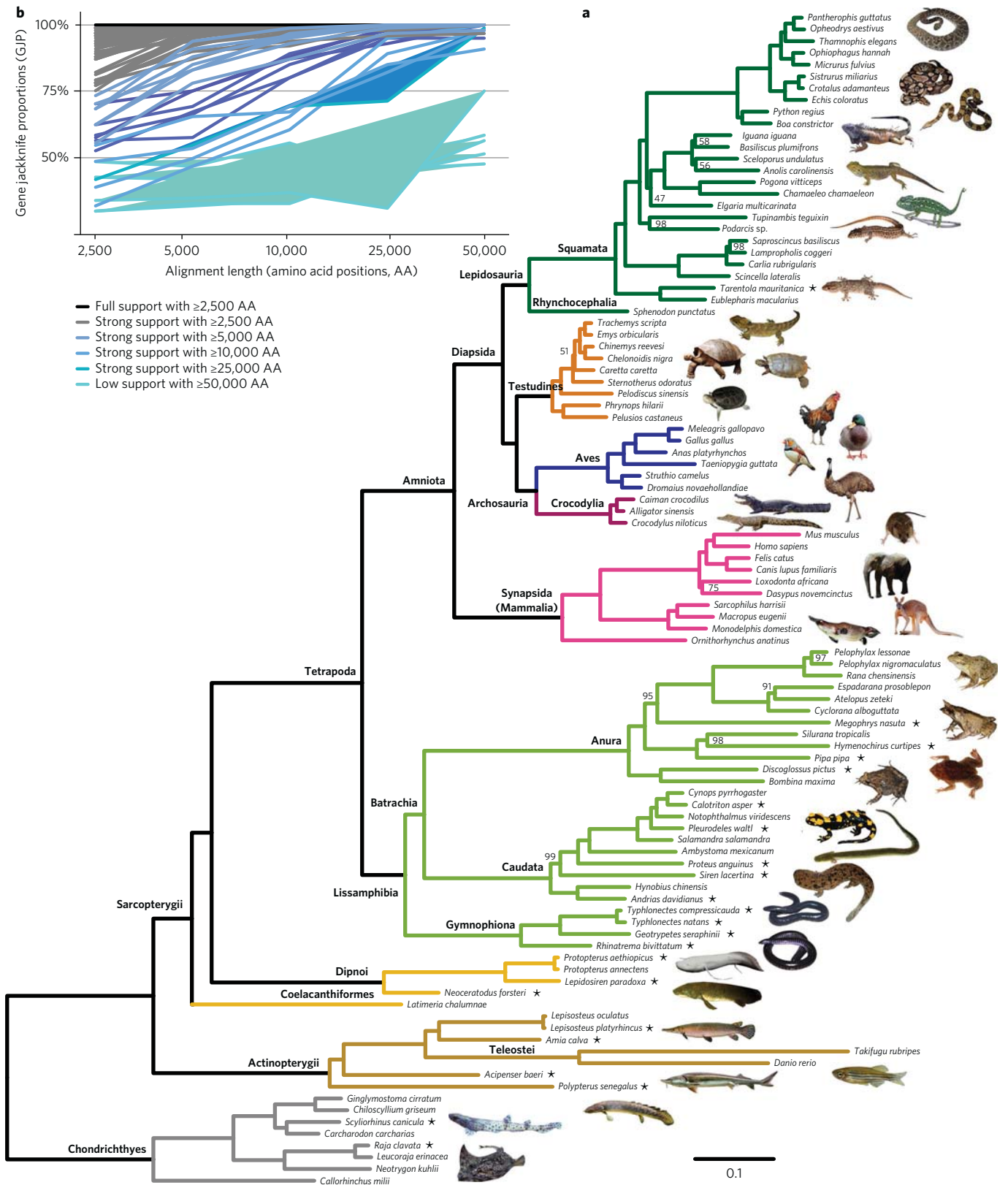


Fig. 2 | Backbone phylogeny of jawed vertebrates. **a**, Bayesian majority-rule consensus tree from 100 independent Markov-chain Monte Carlo chains derived from gene jackknife replicates (~50,000 amino acid (AA) positions each) of the NoDP nuclear dataset, estimated by PhyloBayes under the CAT + Γ model. All nodes received full gene jackknife support (100%), except those displaying the actual value. The scale bar corresponds to the expected number of substitutions per site. Asterisks denote new transcriptomic data generated in this study. **b**, Effect of alignment length on the recovery of single nodes in the phylogeny assessed by gene jackknife proportions derived from the NoDP dataset.

crown-group iguanians and turtles, this applies to the sister-group between anguimorph (*Elgaria*) and iguanian lizards which was sensitive to the use of alternative models (GTR+ Γ and LG+ Γ +F in ML; Supplementary Figs. 3, 4) or the inclusion of deep paralogy (BI on the 1DP and 2DP datasets; Supplementary Figs. 6, 7), which recovered anguimorphs as the sister group of snakes (rejected, however, by topology tests; Supplementary Table 5). In agreement with Fig. 2a, coalescent-based analyses reconstructed an anguimorph + iguanian clade, which was robust to the inclusion of deep paralogy (Supplementary Figs. 5, 8, 9). In addition, only moderate support (75% GJP) was recovered for the controversial position^{14,17} of armadillo (*Xenarthra*) plus elephant (*Afrotheria*) sister to the remaining placental mammals (*Atlantogenata*^{13,16}), in agreement with coalescent analyses (Supplementary Fig. 5), and the two alternative resolutions were rejected by topology tests (Supplementary Table 5). These problematic nodes correspond to fast radiations whose resolution requires extended taxon sampling in addition to accounting for incomplete lineage sorting. Our study minimized the possibility of model misspecification by using also complex evolutionary models and assessing the stability of tree topology to the effect of gene sampling and deep paralogy. For definitively resolving the above nodes, we argue for a careful exploration using suitable methodology and increased taxon sampling.

Robustness to gene sampling: size of gene jackknife replicates and gene trees. The use of gene jackknifing (100 replicates of ~50,000 amino acids each) allowed the recovery of an almost fully supported tree and the resolution of a number of controversial relationships. To explore the stability of the nodes in our tree and assess the amount of data required to recover them, we further analysed four sets of 100 gene jackknife replicates of increasing total length (~2,500, ~5,000, ~10,000 and ~25,000 amino acids) under ML. Relatively short replicates (~2,500 amino acids) recovered 33% and 76% of the nodes with full and high GJP, respectively (Fig. 2b). Increasing alignment length to 25,000 amino acids led to an increase of 47% of fully supported nodes (Fig. 2b; Supplementary Table 3). The relationships among the earliest-branching salamander lineages (*Andrias*, *Hynobius*, *Siren*) were particularly unstable and required long replicates (~50,000 amino acids) to be recovered with strong support. Gene length positively correlated with the proportion of final-tree bipartitions, more strongly for deep (>150 Ma; $r=0.21$, $P<2.2\times 10^{-16}$) than for recent (<150 Ma; $r=0.13$, $P<2.2\times 10^{-16}$) relationships, suggesting that longer genes correctly resolve more ancient nodes.

Mitogenomes and limits to phylogenetic resolution. To assess the phylogenetic resolution power of mitogenomes, we assembled a mitogenomic dataset matching the species in our nuclear datasets. Mitogenomic trees inferred by both ML and BI (Supplementary Figs. 10–17) correctly recovered some major clades with strong support, but failed to recover well-established relationships such as the monophyly of ray-finned and lobe-finned fishes or the sister-group position of platypus to all other mammals⁴, even after excluding the fastest-evolving taxa and using complex mixture models to minimize long-branch attraction artefacts (Supplementary Figs. 14,15). Besides stochastic error due to limited alignment length, these incongruences most probably originate from long-branch attraction (despite using sophisticated models such as CAT-GTR), suggesting that mitogenomes are inadequate for resolving ancient divergences (>400 Ma) using currently available models of sequence evolution. The correlation between nuclear and mitochondrial rates is low ($r=0.35$; $P<2.49\times 10^{-5}$; Supplementary Fig. 18) but still higher than expected from random datasets ($r=0.13\pm 0.08$; $P>0.05$ averaged for 100 replicates). Hence, commonly assumed determinants of substitution rates, such as demography (population size changes, bottlenecks) or life-history traits (body size, metabolic rate, generation time or genome size), have to some extent influenced both

genomes similarly, but other factors must be invoked to explain the observed rate disparity between the two genomes at many branches (Supplementary Fig. 18). These might include clade-specific variation in mitochondrial effective population sizes, genome-specific mutation rate, or acceleration of mitochondrial genes due to selection shifts in respiratory function.

No general relationship among evolutionary rates, species diversity and genome size. Comparing 44 main clades of jawed vertebrates of ages >150 Ma confirmed enormous differences in species diversity, from 1 to 31,826 species (Supplementary Table 6). Species diversity was not overall correlated with substitution rate ($r=0.18$, $P=0.25$) nor were higher rates significantly associated with higher species diversity in a sister-group approach (Sign test, $P=0.13$). Our dataset includes the entire range of genome sizes in vertebrates (from 0.4 pg in pufferfish to 109 pg in lungfishes). Yet we found no association of genome size with evolutionary rate or species diversity ($r=-0.28$, $P=0.061$, and $r=-0.13$, $P=0.44$, respectively). Previous studies have also suggested that genome size might be associated with indels in coding regions³⁸, but we detected no significant correlation, neither within conserved ($r=0.1983$, $P=0.0722$) nor within variable coding regions ($r=0.0533$, $P=0.6325$) as defined by the software BMGE (Block Mapping and Gathering with Entropy). The results of these correlation analyses were confirmed by a Bayesian joint modelling of the above traits with parameters of the evolutionary process at the sequence level (see Supplementary Table 7).

Divergence times of major lineages of jawed vertebrates. Genome-scale datasets have been shown to produce more precise and accurate divergence time estimates³⁹, but this ultimately depends on the use of realistic evolutionary and clock models that appropriately account for among-lineage heterogeneities⁴⁰ and multiple calibration intervals whose uncertainty and internal consistency is accounted for^{41,42}. We applied an autocorrelated log-normal relaxed clock model and best-fitting sequence evolution model (CAT-GTR) to estimate genome-wide divergence times, averaged over 100 gene jackknife replicates. We used a conservative approach to setting calibrations, starting from multiple well-established calibrations with solid palaeontological evidence, and used conservative intervals to account for dating and phylogenetic uncertainty⁴³ (Supplementary Table 8). On top of that, the internal congruence among these calibrations was verified through extensive cross-validation procedures to remove any poorly performing calibration, either examining the performance of single calibrations⁴¹ or removing one calibration at a time to check the congruence between estimated ages and priors⁴². The performance of each calibration scheme (named C16 and C30) derived from the above cross-validation strategies was assessed in independent dating analyses with a test dataset (a subset of the 14,352 most complete amino acid positions from the NoDP dataset that was computationally tractable with PhyloBayes). Both schemes produced largely congruent divergence times, but C16 yielded more reasonable dates within turtles, frogs, neoavian birds, modern frogs (overestimated in C30 if compared with previous data; www.time-tree.org), or iguanian squamates and snakes (underestimated in C30) (Supplementary Table 9).

To estimate genome-wide divergence times, we calculated averaged divergence times (and conservative 95% credibility intervals; CrI) across 100 timetrees based of jackknife sampling of ~15,000 positions and the more stringently cross-validated C16 calibration scheme (Fig. 3). The genome-averaged timetree places divergences among cartilaginous, ray-finned and lobe-finned fishes in the Ordovician period, between 458 (CrI: 465–438) and 449 (462–431) Ma. The first split within lobe-finned fishes occurred in the Silurian period c. 427 (444–413) Ma, and lungfishes separated from tetrapods in the early Devonian c. 412 (419–408) Ma. The split between amphibians and amniotes occurred in the early Carboniferous

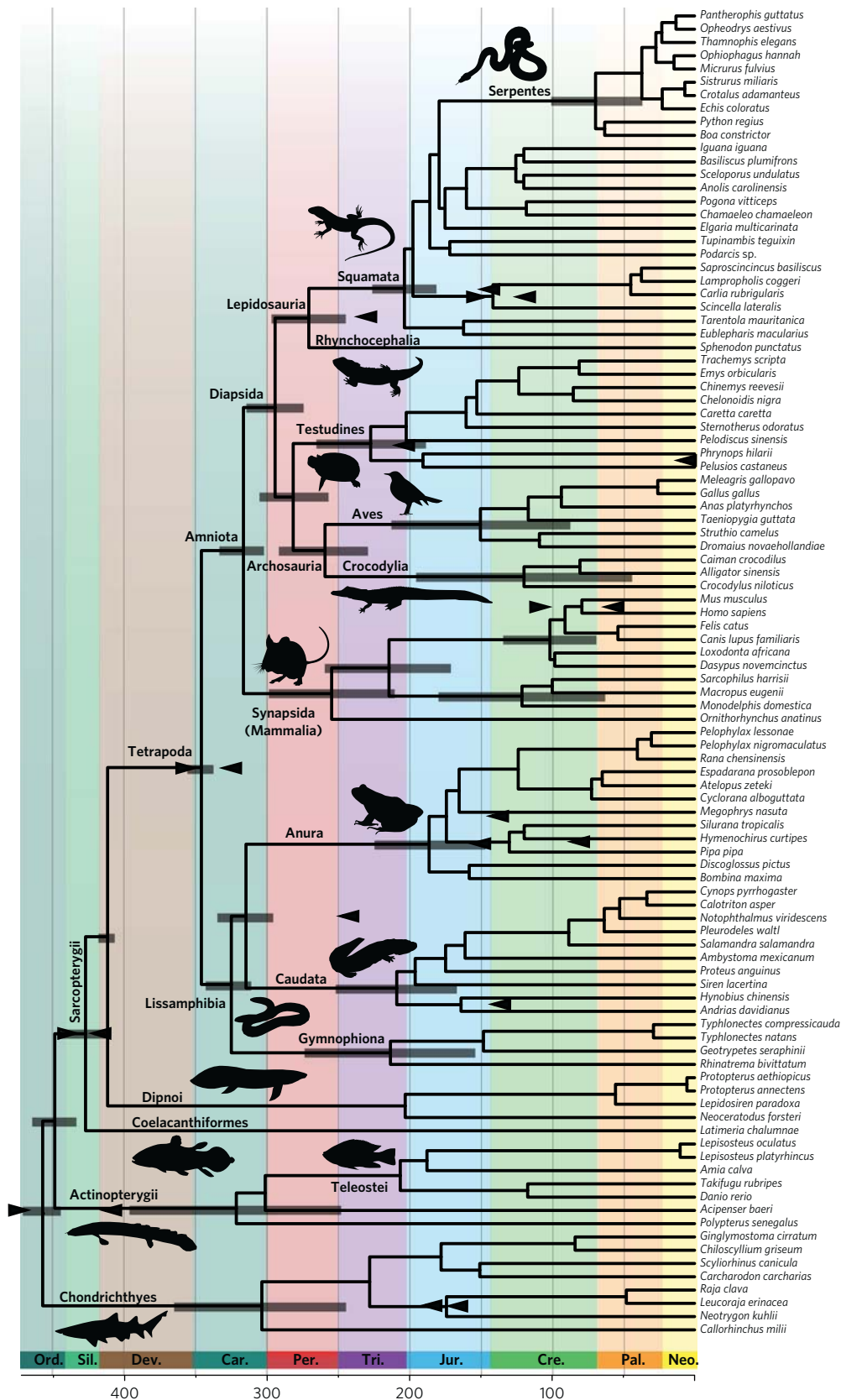


Fig. 3 | Time-calibrated phylogeny of jawed vertebrates. Divergences have been averaged across 100 timetrees estimated from independent gene jackknife replicates in PhyloBayes, using the subset of most congruent calibrations (C16; marked by arrows) and best-fit evolutionary (CAT-GTR+ Γ) and relaxed clock (autocorrelated log-normal) models. Credibility intervals (CrI) are calculated as the absolute maximum and minimum values of 95% confidence intervals across 100 timetrees (only displayed for key nodes; see Supplementary Table 9 for detailed results). The dimension of the scale is given in million years, and main geological periods are highlighted.

period c. 346 (351–333) Ma, and the three amphibian orders separated during the Carboniferous from 325 (338–307) to 315 (332–293) Ma, as did synapsids (mammals) and diapsids (turtles, archosaurs and lepidosaurs) c. 317 (330–299) Ma. The origins of the main sauropsid groups—turtles, crocodiles, birds, squamates and tuatara (*Sphenodon*)—took place in the Permian period from 294 (313–273) to 259 (288–226) Ma. The crown diversification of extant frogs, salamanders and caecilians occurred in the late Triassic to early Jurassic periods between 213 (270–151) and 186 (231–153) Ma, almost simultaneously with the crown splits within squamates c. 204 (228–183) Ma, cryptodiran turtles c. 202 (243–159) Ma, pleurodiran turtles c. 191 (248–116) Ma and therian mammals c. 214 (257–169) Ma.

Estimated divergences are generally in line with previous time-calibrated phylogenies using different dating methodologies, molecular data and calibrations, particularly for the deepest splits in the backbone^{44,45}, as well as divergences within amphibians⁶, squamates⁴⁶, snakes⁴⁷ and placental mammals⁸. Estimated ages for crown-groups of cartilaginous and ray-finned fishes are younger than in previous analyses^{48,49}, which is probably caused by the removal of incongruent calibrations in the C16 scheme (the C30 scheme produced estimates more similar to previous studies for these groups; see Supplementary Table 9). The younger age of cartilaginous fishes, however, is consistent with recent palaeontological analyses⁵⁰. Compared with previous time-calibrated phylogenies, we obtain older divergences for turtles⁵¹ and birds¹⁵, but our estimates are in line with the ages of recently discovered fossils of stem turtles³⁶ and an ornithuromorph bird that pushes back the origin of the group to at least 130.7 Ma⁵². The Cretaceous/Palaeogene boundary (67 Ma) in our tree is not associated with a notable concentration of divergences, but our dataset does not capture the crown diversification of several species-rich taxa that might have occurred in this period, such as spiny-ray fishes, modern birds (Neoaves), boreoeutherian mammals, ranoid frogs, gekkonid geckos or skinks. We support a diversification of placental mammals prior to the Cretaceous/Palaeogene boundary c. 102 (139–73) Ma, in agreement with most previous molecular and macroevolutionary studies^{8,39}.

Reliability of phylogenomic analyses. Inferring phylogenies can be difficult, particularly in the presence of ancient or closely spaced speciation events, and the use of genome-scale datasets poses additional challenges related to poor data quality and more importantly systematic error²⁵. In principle, the jawed vertebrate phylogeny is a solvable problem, being devoid of excessively old divergences and mostly long internal branches (Fig. 2a). It thus represents a good benchmark to test the abovementioned challenges. Yet, poor data quality¹⁸ (Supplementary Fig. 1) can lead to incorrect results (for example non-monophyly of amphibians, misplacement of turtles). We adopt a phylotranscriptomic approach to assemble an alignment of >7,000 genes for 100 species with rigorous quality controls. The quality and resolving power of our NoDP dataset are higher than those of previous studies, including ref.¹⁸ and the most comprehensive dataset analysed so far¹², with 70% versus 17% and 61% mean congruence respectively for the two datasets, measured as the proportion of final-tree bipartitions recovered by single genes. The higher congruence of NoDP persisted after correcting for gene length (Supplementary Fig. 1).

We thus confirm that RNA-Seq is a cost-effective method to anchor phylogenomic analyses, which can result in robust fossil-dated trees, provided that careful data curation and appropriate analytical methods are used. Moreover, we show that gene jackknifing allows stringent testing of phylogenetic relationships and makes it possible to overcome the limitations and possible biases of small datasets that aim to represent the entire genome, and that phylogenomics is resilient to limited levels of deep paralogy, provided that a large number of genes (>1,000) are used and internal

branches are relatively long. In such cases, realistic models allow recovery of correct phylogenetic hypotheses, even in the presence of extreme among-lineage evolutionary rate variation. In contrast, resolving closely spaced radiations, which were not targeted in this work, could require a detailed study with specific gene and taxon sampling¹⁴ and testing the robustness to model misspecification³². Overall, our results highlight the importance of data quality in phylogenomics, as well as the application of realistic evolutionary and clock models, and the validation of calibrations in timetree estimation, both a priori (based on palaeontological data) and a posteriori (cross-validation).

Methods

An extended description of our bioinformatic pipeline and detailed Methods are available as Supplementary Information.

Ethics statement. Animal experiments conformed to the European Parliament and council of 22 September 2010 (Directive 2010/63/EU) and the French Rural Code (Articles R214-87 to R214-137, decree no. 2013-118 of 1 February 2013). Experiments performed in France were authorized by certificate no. 75-600.

Assembly of phylogenomic datasets. New RNA-Seq data was generated for 23 gnathostome species using Illumina MiSeq (2 × 250 bp) and HiSeq2000 (2 × 50 bp, 2 × 100 bp) technologies. Available RNA-Seq data were downloaded from NCBI SRA. Transcriptomes were assembled *de novo* with Trinity or MIRA. Species names and accession numbers are available in Supplementary Table 10.

Nuclear datasets were assembled using a new pipeline summarized in Supplementary Fig. 2. Briefly, proteomes of 21 vertebrate genomes (ENSEMBL) were grouped into orthologue clusters, and those not containing data for all major jawed vertebrate lineages were discarded. The resulting 11,656 protein clusters were aligned and positions of unreliable homology removed. To identify and resolve paralogy issues, we implemented a paralogue-splitting pipeline based on gene trees. The 9,852 orthologue clusters obtained were complemented with new genomes and transcriptomes using the software Forty-Two (<https://bitbucket.org/dbaurain/42/>). Several decontamination steps were carried out. Any sequence contamination from non-vertebrates and human was detected by BLAST and eliminated. We searched for cross-contamination that can arise during library preparation using gene trees, and removed contaminants based on expression data. After eliminating overlapping redundant sequences that were too divergent, we filtered out incomplete or short sequences and alignments, leading to 7,687 genes. The paralogy splitting procedure was repeated to resolve any paralogy caused by the addition of new species, and gene alignments were classified into three datasets that contained zero (NoDP), one (1DP) and two or more (2DP) deep paralogues. Sequence stretches with unusually low similarity (usually due to frame shifts) were masked with HMM-cleaner (R. Pujol, unpublished), and alignments were trimmed. For each gene, we used ScaFoS³⁰ to merge conspecific sequences and resolve putative remaining paralogy. A third decontamination step used extremely long branches estimated on a fixed reference tree as proxy for contamination.

Mitochondrial datasets were assembled from mitogenomes available at NCBI with a taxon sampling mirroring the nuclear datasets plus a few additional species to reduce long-branch attraction artefacts expected in mitogenomic trees (Supplementary Table 11). The resulting alignments consisted of 106 species (2,773 amino acid positions) and 95 species (2,866 amino acid positions) after removing the fastest-evolving species.

Phylogenetic inference. Concatenated nuclear gene sets (NoDP, 1DP and 2DP) were analysed separately using ML with RAxML v.8⁵³ under LG + F + Γ and GTR + Γ models and BI with PhyloBayes MPI v1.5⁵⁴ under the better-fitting CAT + Γ model (selected after 10-fold cross-validation). The Bayesian consensus tree was calculated from 100 post-burn-in tree collections, each from gene jackknife replicates of ~50,000 amino acid positions. Convergence was verified with the diagnostic tools of PhyloBayes. Branch support was computed from 100 bootstrap pseudo-replicates in ML and from gene jackknife proportions (GJP) in BI. To assess the robustness to gene sampling, we analysed by ML gene jackknife replicates of c. 2,500, 5,000, 10,000 and 25,000 aligned positions under the LG + Γ model. Coalescent analyses were run in ASTRAL-II v.4.10.12 using ML gene trees as input (estimated under best-fit models in RAxML), and node stability was assessed as local posterior support and 100 replicates of multi-locus bootstrapping.

The mitochondrial datasets were analysed by ML under MTREV + Γ and GTR + Γ models, and by BI under CAT + Γ and CAT-GTR + Γ models.

Molecular dating. Divergence times were estimated in PhyloBayes v.4.1 using best-fit CAT-GTR + Γ and autocorrelated log-normal clock models (selected after 10-fold cross-validation), a birth–death prior on divergence times, and 30 calibration points with uniform priors and soft bounds (see Supplementary Table 8). After cross-validation procedures (see SI Materials and Methods), we applied the C16 and C30 calibration sets to compute timetrees based on a subset of 14,352 amino

acid positions from NoDP (two independent chains). To estimate genome-wide divergence times, we estimated 100 timetrees from 100 gene jackknife replicates of ~15,000 amino acids from the NoDP dataset, using the most stringent C16 calibration scheme. Divergence times were averaged and conservative 95% credibility intervals (CrI) calculated as the absolute maximum and minimum values of 95% confidence intervals across 100 timetrees.

Nuclear and mitochondrial rates. Substitution rates were measured as branch lengths optimized under CAT + Γ and a reference tree (Fig. 2a) in PhyloBayes, independently for the nuclear (NoDP) and mitochondrial datasets, both pruned to a common subset of 78 species. Correlation between mitochondrial and nuclear rates was assessed by Pearson's correlation among all pairs of internal and terminal branches. We simulated 100 random alignments characterized by the amino acid proportions of either mitochondrial or nuclear datasets. Branch lengths were then optimized on a reference tree and rates correlated as above.

Association of life-history traits and molecular features. We estimated Pearson's correlation after correcting for phylogenetic non-independence among the following life-history traits and molecular features: (i) genome size (retrieved from www.genomesize.com) versus number of gaps in either conserved or variable gene regions (defined by BMGE on untrimmed gene alignments), (ii) genome size versus nuclear substitution rate, and (iii) substitution rate versus species diversity (tabulated from the literature), for 44 lineages divided by an ad hoc cut-off of >150 Ma defined to capture sister groups characterized by obvious differences in species diversity. Nuclear substitution rates and species diversity were also compared in a sister-group approach, assessing by non-parametric Sign test whether higher substitution rates (tested by relative-rate tests) were associated with higher species diversity. We further used Bayesian joint modelling to study the correlation between substitution rates, genome size and the number of gaps in conserved and variable gene regions.

References

- Brown, J. M. & Thomson, R. C. Bayes factors unmask highly variable information content, bias, and extreme influence in phylogenomic analyses. *Syst. Biol.* **66**, 517–530 (2017).
- Jeffroy, O., Brinkmann, H., Delsuc, F. & Hervé, P. Phylogenomics: the beginning of incongruence? *Trends Genet.* **22**, 225–231 (2006).
- Zardoya, R. & Meyer, A. The evolutionary position of turtles revised. *Naturwissenschaften* **88**, 193–200 (2001).
- Janke, A., Erpenbeck, D., Nilsson, M. & Arnason, U. The mitochondrial genomes of the iguana (*Iguana iguana*) and the caiman (*Caiman crocodylus*): implications for amniote phylogeny. *Proc. R. Soc. B Biol. Sci.* **268**, 623–631 (2001).
- Near, T. J. et al. Phylogeny and tempo of diversification in the superradiation of spiny-rayed fishes. *Proc. Natl Acad. Sci. USA* **110**, 12738–12743 (2013).
- Irisarri, I. et al. The origin of modern frogs (Neobatrachia) was accompanied by acceleration in mitochondrial and nuclear substitution rates. *BMC Genom.* **13**, 626 (2012).
- 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).
- Meredith, R. W. et al. Impacts of the Cretaceous terrestrial revolution and KPg extinction on mammal diversification. *Science* **334**, 521–524 (2011).
- Jetz, W., Thomas, G. H., Joy, J. B., Hartmann, K. & Mooers, A. O. The global diversity of birds in space and time. *Nature* **491**, 444–448 (2012).
- Chiari, Y., Cahais, V., Galtier, N. & Delsuc, F. Phylogenomic analyses support the position of turtles as the sister group of birds and crocodiles (Archosauria). *BMC Biol.* **10**, 65 (2012).
- Crawford, N. G. et al. More than 1000 ultraconserved elements provide evidence that turtles are the sister group of archosaurs. *Biol. Lett.* **8**, 783–786 (2012).
- 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).
- Tarver, J. E. et al. The interrelationships of placental mammals and the limits of phylogenetic inference. *Genome Biol. Evol.* **8**, 330–344 (2016).
- Romiguier, J., Ranwez, V., Delsuc, F., Galtier, N. & Douzery, E. J. P. Less is more in mammalian phylogenomics: AT-rich genes minimize tree conflicts and unravel the root of placental mammals. *Mol. Biol. Evol.* **30**, 2134–2144 (2013).
- Jarvis, E. D. et al. Whole-genome analyses resolve early branches in the tree of life of modern birds. *Science* **346**, 1320–1331 (2014).
- Song, S., Liu, L., Edwards, S. V. & Wu, S. Resolving conflict in eutherian mammal phylogeny using phylogenomics and the multispecies coalescent model. *Proc. Natl Acad. Sci. USA* **109**, 14942–14947 (2012).
- Morgan, C. C. et al. Heterogeneous models place the root of the placental mammal phylogeny. *Mol. Biol. Evol.* **30**, 2145–2156 (2013).
- 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).
- Lyson, T. R. et al. MicroRNAs support a turtle + lizard clade. *Biol. Lett.* **8**, 104–107 (2012).
- Gauthier, J. A., Kearney, M., Maisano, J. A., Rieppel, O. & Behlke, A. D. B. Assembling the squamate tree of life: perspectives from the phenotype and the fossil record. *Bull. Peabody Mus. Nat. Hist.* **53**, 3–308 (2012).
- Townsend, T. M., Larson, A., Louis, E., Macey, J. R. & Sites, J. Molecular phylogenetics of Squamata: the position of snakes, amphisbaenians, and dibamids, and the root of the squamate tree. *Syst. Biol.* **53**, 735–757 (2004).
- Zheng, Y. & Wiens, J. J. Combining phylogenomic and supermatrix approaches, and a time-calibrated phylogeny for squamate reptiles (lizards and snakes) based on 52 genes and 4162 species. *Mol. Phylogenet. Evol. Part B* **94**, 537–547 (2016).
- Irisarri, I., Vences, M., San Mauro, D., Glaw, F. & Zardoya, R. Reversal to air-driven sound production revealed by a molecular phylogeny of tongueless frogs, family Pipidae. *BMC Evol. Biol.* **11**, 114 (2011).
- Bewick, A. J., Chain, F. J. J., Heled, J. & Evans, B. J. The pipid root. *Syst. Biol.* **61**, 913–926 (2012).
- Philippe, H. et al. Resolving difficult phylogenetic questions: why more sequences are not enough. *PLoS Biol.* **9**, e1000602 (2011).
- Bininda-Emonds, O. R. P. et al. The delayed rise of present-day mammals. *Nature* **446**, 507–512 (2007).
- Springer, M. S. et al. Waking the undead: implications of a soft explosive model for the timing of placental mammal diversification. *Mol. Phylogenet. Evol.* **106**, 86–102 (2017).
- Phillips, M. J. Geomolecular dating and the origin of placental mammals. *Syst. Biol.* **65**, 546–557 (2016).
- Hara, Y. et al. Optimizing and benchmarking *de novo* transcriptome sequencing: from library preparation to assembly evaluation. *BMC Genom.* **16**, 977 (2015).
- 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).
- Delsuc, F., Tsagkogeorga, G., Lartillot, N. & Philippe, H. Additional molecular support for the new chordate phylogeny. *Genesis* **46**, 592–604 (2008).
- 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).
- Zhu, M. & Schultze, H.-P. in *Major Events in Early Vertebrate Evolution: Paleontology, Phylogeny and Development* (ed. Ahlberg, P. E.) 289–314 (Taylor & Francis, 2001).
- Pyron, R. A. & Wiens, J. J. A large-scale phylogeny of Amphibia including over 2800 species, and a revised classification of extant frogs, salamanders, and caecilians. *Mol. Phylogenet. Evol.* **61**, 543–583 (2011).
- San Mauro, D., Vences, M., Alcobendas, M., Zardoya, R. & Meyer, A. Initial diversification of living amphibians predated the breakup of Pangaea. *Am. Nat.* **165**, 590–599 (2005).
- Schoch, R. R. & Sues, H.-D. A Middle Triassic stem-turtle and the evolution of the turtle body plan. *Nature* **523**, 584–587 (2015).
- Zhang, P. & Wake, D. B. Higher-level salamander relationships and divergence dates inferred from complete mitochondrial genomes. *Mol. Phylogenet. Evol.* **53**, 492–508 (2009).
- Gregory, T. R. Insertion–deletion biases and the evolution of genome size. *Gene* **324**, 15–34 (2004).
- Dos Reis, M. et al. Phylogenomic datasets provide both precision and accuracy in estimating the timescale of placental mammal phylogeny. *Proc. R. Soc. B Biol. Sci.* **279**, 3491–3500 (2012).
- Lepage, T., Bryant, D., Philippe, H. & Lartillot, N. A general comparison of relaxed molecular clock models. *Mol. Biol. Evol.* **24**, 2669–2680 (2007).
- Near, T. J., Meylan, P. A. & Shaffer, H. B. Assessing concordance of fossil calibration points in molecular clock studies: an example using turtles. *Am. Nat.* **165**, 137–146 (2005).
- Sanders, K. L. & Lee, M. S. Y. Evaluating molecular clock calibrations using Bayesian analyses with soft and hard bounds. *Biol. Lett.* **3**, 275–279 (2007).
- Warnock, R. C. M., Parham, J. F., Joyce, W. G., Lyson, T. R. & Donoghue, P. C. J. Calibration uncertainty in molecular dating analyses: there is no substitute for the prior evaluation of time priors. *Proc. R. Soc. B Biol. Sci.* **282**, 20141013 (2014).
- King, B. L. Bayesian morphological clock methods resurrect placoderm monophyly and reveal rapid early evolution in jawed vertebrates. *Syst. Biol.* **66**, 499–516 (2017).

45. Alfaro, M. E. et al. Nine exceptional radiations plus high turnover explain species diversity in jawed vertebrates. *Proc. Natl Acad. Sci. USA* **106**, 13410–13414 (2009).
46. 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).
47. Hsiang, A. Y. et al. The origin of snakes: revealing the ecology, behavior, and evolutionary history of early snakes using genomics, phenomics, and the fossil record. *BMC Evol. Biol.* **15**, 87 (2015).
48. Inoue, J. G. et al. Evolutionary origin and phylogeny of the modern holocephalans (Chondrichthyes: Chimaeriformes): a mitogenomic perspective. *Mol. Biol. Evol.* **27**, 2576–2586 (2010).
49. Dornburg, A., Townsend, J., Friedman, M. & Near, T. J. Phylogenetic informativeness reconciles ray-finned fish molecular divergence times. *BMC Evol. Biol.* **14**, 169 (2014).
50. Coates, M. L., Gess, R. W., Finarelli, J. A., Criswell, K. E. & Tietjen, K. A. symmoriiform chondrichthyan braincase and the origin of chimaeroid fishes. *Nature* **541**, 208–211 (2017).
51. Lourenço, J. M., Claude, J., Galtier, N. & Chiari, Y. Dating cryptodiran nodes: origin and diversification of the turtle superfamily Testudinoidea. *Mol. Phylogenet. Evol.* **62**, 496–507 (2012).
52. Wang, M. et al. The oldest record of Ornithuromorpha from the early Cretaceous of China. *Nat. Commun.* **6**, 6987 (2015).
53. Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).
54. 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).

Acknowledgements

We thank the following people for tissue samples: T. Ziegler (*Andrias*), M. Hasselmann (*Neoceratodus*) and O. Guillaume (*Calotriton*, *Proteus*). We thank V. Michael for

technical assistance. N. Galtier kindly provided access to the *Salamandra* transcriptome. We acknowledge the use of computational resources from the University of Konstanz (HPC2), the Institute of Physics of Cantabria IFCA-CSIC (Altamira), the Fédération Wallonie-Bruxelles (Tier-1; funded by Walloon Region, grant no. 1117545), the Consortium des Équipements de Calcul Intensif (CÉCI; funded by FRS-FNRS, grant no. 2.5020.11) and the Réseau Québécois de Calcul de Haute Performance. I.I. was supported by the Alexander von Humboldt Foundation (1150725) and the European Molecular Biology Organization (EMBO ALTF 440-2013). Further support came from the University of Konstanz (I.I.), the Deutsche Forschungsgemeinschaft (DFG) (A.M.), the European Research Council (ERC Advanced Grant ‘Genadapt’ no. 273900 to A.M.) and the Agence Nationale de la Recherche (TULIP Laboratory of Excellence ANR-10-LABX-41 to H.P. and ANR-12-BSV7-020 project ‘Jaws’ to F.D. and J.-Y.S.). This is contribution ISEM 2017-106 of the Institut des Sciences de l’Evolution de Montpellier.

Author contributions

M.V., H.P. and I.I. designed the research. I.I., F.D., J.-Y.S., A.K., M.J., A.M. and M.V. contributed new data. I.I., D.B., H.B., M.V. and H.P. performed analyses. I.I., M.V. and H.P. drafted the manuscript, and all authors read and approved the final manuscript.

Competing interests

The authors declare no competing financial interests.