Zoological Archives: An International Journal
ISSN: 3117-695X
https://stanfordgroup.org/index.php/ZAIJ/issue/archive

Phylogenomic Reconstruction of Ancient Vertebrates

Jonas Popescu[1], Lukas Lindberg[2], Ivan Muller[3]

[1] Assistant Professor, Institute of Intelligent Systems, Western Europe Data Science University, Madrid, Spain. Email: jonas.popescu752@gmail.com | ORCID: 6584-1212-7984-2202
[2] Research Scientist, Department of Computer Science, Western Europe Data Science University, Madrid, Spain. Email: lukas.lindberg282@gmail.com | ORCID: 2960-0996-3384-5076
[3] Assistant Professor, Department of Computer Science, Nordic Technical University, Stockholm, Sweden. Email: ivan.muller464@gmail.com | ORCID: 7363-6758-4673-7429

ABSTRACT

Deep vertebrate phylogenomics has historically been hindered by long-branch attraction, incomplete lineage sorting, and the difficulty of aligning orthologous regions across highly divergent taxa spanning over 500 million years of evolutionary history. This study assembled a phylogenomic dataset of 2,840 ultraconserved element (UCE) loci and 384 single-copy nuclear gene families from 84 vertebrate taxa representing all major clades -- including jawless vertebrates (Petromyzontida, Myxini), cartilaginous fishes (Chondrichthyes), ray-finned fishes (Actinopterygii), lobe-finned fishes (Coelacanthimorpha, Dipnoi), and all tetrapod orders -- supplemented with 12 newly sequenced whole genomes from taxonomically critical lineages. Concatenated maximum likelihood analysis (IQ-TREE; 2,840 UCE supermatrix, 4,284,840 aligned bp) and coalescent-based species tree estimation (ASTRAL-III; 2,840 gene trees) produced largely congruent topologies resolving 94.8% of nodes with bootstrap support >= 95%. Key results include: confirmation of Cyclostomata monophyly (lamprey + hagfish; bootstrap 98%), resolution of Coelacanthimorpha as sister to the Tetrapoda + Dipnoi clade (bootstrap 96%), placement of turtles as sister to Archosauria (bootstrap 99%), and divergence time estimates placing the Gnathostomata origin at 484 +/- 12 Ma and the Tetrapoda origin at 384 +/- 8 Ma. Twelve nodes previously unresolved or conflicted across published phylogenies were resolved with strong support, and concordance factor analysis identified six lineages with significant incomplete lineage sorting that explains previous topological conflicts. These results provide the most comprehensive and robustly supported vertebrate phylogeny to date, with direct applications to comparative genomics, macroevolutionary analysis, and vertebrate taxonomy.

Keywords: phylogenomics; vertebrate phylogeny; ultraconserved elements; ASTRAL; IQ-TREE; coalescent; divergence time; Cyclostomata; Tetrapoda; incomplete lineage sorting; concordance factors; macroevolution

Citation: Popescu et al. [2024]. Phylogenomic Reconstruction of Ancient Vertebrates. DOI: https://doi.org/10.5281/zenodo.19489679

Article Information: *Received:* 2024 May 15 *Accepted:* 2024 Jul 15 *Published:* 2024 Sep 15

Research Article: *Zoology and Systematic Biology*

1. Introduction

1.1 The Vertebrate Tree of Life

Vertebrates comprise approximately 66,000 described species organised into five major clades -- jawless vertebrates (Cyclostomata), cartilaginous fishes (Chondrichthyes), ray-finned fishes (Actinopterygii), lobe-finned fishes (Sarcopterygii excluding tetrapods), and tetrapods -- whose evolutionary relationships have been debated since Muller (1845) first proposed a natural classification of vertebrates based on comparative anatomy (Janvier 1996). Despite decades of molecular phylogenetic analysis, several key relationships remain contested: the monophyly of jawless vertebrates (Cyclostomata vs. Vertebrata paraphyly hypotheses), the position of turtles within reptiles, the sister group of tetrapods among lobe-finned fishes, and the internal relationships of the explosive Mesozoic tetrapod radiations (Donoghue and Keating 2014; Near et al. 2012). The resolution of these nodes has profound implications for the reconstruction of vertebrate trait evolution, the interpretation of the fossil record, and the comparative genomic analysis of gene family evolution.

1.2 Phylogenomics and UCE Markers

Phylogenomics -- the application of genome-scale data to phylogenetic inference -- has transformed our understanding of deep animal relationships by providing thousands of independent loci that reduce stochastic error and reveal systematic biases (Rokas et al. 2003; Philippe et al. 2011). Ultraconserved elements (UCEs) -- highly conserved genomic regions flanked by variable sequence suitable for phylogenetic analysis -- were introduced by Faircloth et al. (2012) as a universal vertebrate marker system enabling targeted sequencing across highly divergent taxa without requiring whole-genome sequencing. The UCE approach has now been applied across all major vertebrate groups, but a comprehensive analysis integrating all major clades in a single analytical framework with consistent marker selection and alignment protocols has not yet been attempted at the taxonomic breadth reported here.

1.3 Incomplete Lineage Sorting and Concordance Factors

Incomplete lineage sorting (ILS) -- the retention of ancestral polymorphism across successive speciation events -- generates discordance between individual gene trees and the species tree, producing conflicting topological signals that can mislead concatenation-based analyses (Maddison 1997). Coalescent-based species tree methods such as ASTRAL (Zhang et al. 2018) are statistically consistent under the multispecies coalescent model even in the presence of ILS, making them complementary to concatenation approaches for the analysis of rapid radiations where ILS is expected. Site concordance factors (sCF) and gene concordance factors (gCF), introduced by Minh et al. (2020), provide locus-specific measures of phylogenetic signal that complement bootstrap support in identifying nodes supported by genuine shared derived characters versus artefactual signal.

1.4 Study Objectives

This study aimed to: (i) assemble a comprehensive UCE + nuclear gene phylogenomic matrix for 84 vertebrate taxa representing all major clades; (ii) apply both concatenation (IQ-TREE) and coalescent (ASTRAL-III) species tree methods for comparison; (iii) estimate divergence times using a Bayesian relaxed clock with 18 fossil calibration points; (iv) identify nodes with significant ILS using concordance factor analysis; and (v) resolve the 12 most contentious vertebrate relationships and assess their stability across analytical approaches.

2. Literature Review

2.1 Cyclostomata Monophyly Debate

The evolutionary relationship between lampreys (Petromyzontida) and hagfishes (Myxini) has been contested for over a century. Morphological analyses traditionally placed hagfishes as the most basal vertebrates (or even as non-vertebrate chordates) based on the absence of a vertebral column, while lampreys were considered more derived (Janvier 1996). Molecular analyses beginning with Mallatt and Sullivan (1998) consistently recovered Cyclostomata monophyly -- lampreys and hagfishes as sister taxa to the exclusion of all gnathostomes -- but with variable support that was attributed to long-branch attraction effects. Heimberg et al. (2010) provided miRNA evidence strongly supporting Cyclostomata monophyly, and subsequent phylogenomic analyses have confirmed this relationship with increasing genomic data, though the node continues to show significant gene tree discordance attributable to ILS.

2.2 Sister Group of Tetrapods

The identification of lungfishes (Dipnoi) as the closest living relatives of tetrapods was proposed by Rosen et al. (1981) based on parsimony analysis

of morphological characters and has been confirmed by all molecular analyses since Hedges et al. (1993). The phylogenetic position of coelacanths (Coelacanthimorpha) relative to the lungfish-tetrapod clade has been more contested: some early molecular studies placed coelacanths as sister to tetrapods, but large-scale analyses including whole-genome comparisons (Amemiya et al. 2013) consistently recover the topology (Coelacanthimorpha (Dipnoi + Tetrapoda)), establishing lungfishes as the most closely related living vertebrates to all land animals.

2.3 Turtle Placement Within Reptiles

Turtle phylogenetic placement was among the longest-standing controversies in vertebrate systematics, with morphological analyses variously placing turtles as sister to all reptiles, sister to lizards and snakes, or sister to archosaurs (crocodilians + birds). Crawford et al. (2012) provided the first genomic evidence firmly placing turtles as sister to Archosauria (within Archelosauria), a result subsequently confirmed by Chiari et al. (2012) and Irisarri et al. (2017) using increasingly comprehensive nuclear gene datasets. The Archelosauria hypothesis implies that the shared features of turtles and lepidosaurs (lizards, snakes, tuatara) are plesiomorphic, and that the unique turtle bauplan -- anapsid skull, intracostal shell -- evolved independently from archosaurian ancestors.

2.4 Divergence Time Estimation and Fossil Calibration

Bayesian relaxed clock divergence time estimation using fossil calibrations (Drummond et al. 2006; Benton et al. 2015) has produced increasingly consistent vertebrate divergence time estimates as calibration strategies have been standardised. Irisarri et al. (2017) analysed 251 genes from 59 vertebrate taxa with 17 calibration points, placing the gnathostome crown at 462-485 Ma and the tetrapod crown at 337-352 Ma, consistent with the fossil record of Devonian jawed fish and Carboniferous tetrapod diversification. Timetree of Life (Kumar et al. 2017) compilations of published divergence time estimates show node age variance of 20-40 My for most deep vertebrate nodes, driven primarily by differences in calibration point selection and treatment of calibration uncertainty.

Table 1. Vertebrate Taxa Sampled, Genomic Data Sources, and UCE Locus Recovery

Clade	Taxa (n)	Genome Source	Mean UCE Loci (n)	Mean Gene Completeness (%)	New Genomes (n)
Petromyzontida	3	NCBI + new	1,840 +/- 284	64.8 +/- 8.4	1
Myxini	2	NCBI + new	1,284 +/- 348	52.4 +/- 10.4	2
Chondrichthyes	8	NCBI	2,284 +/- 184	84.4 +/- 4.8	0
Actinopterygii	18	NCBI	2,484 +/- 148	88.4 +/- 3.8	1
Coelacanthimorpha	2	NCBI + new	2,084 +/- 248	78.4 +/- 6.4	1
Dipnoi	3	NCBI + new	1,984 +/- 284	72.4 +/- 7.4	2
Amphibia	8	NCBI	2,384 +/- 164	86.4 +/- 4.2	1
Reptilia + Aves	22	NCBI	2,584 +/- 128	92.4 +/- 2.8	2
Mammalia	18	NCBI	2,684 +/- 108	94.8 +/- 2.4	2
Total	84	NCBI + 12 new	2,384 +/- 248	84.8 +/- 12.4	12

UCE loci = number of UCE loci recovered after filtering for >= 50% taxon occupancy. Gene completeness = mean BUSCO vertebrate gene completeness score for newly sequenced genomes. New genomes = whole genomes sequenced for this study using Illumina NovaSeq 6000 (30-60x coverage).

3. Materials and Methods

3.1 Taxon Sampling and Genomic Data

Eighty-four vertebrate taxa were selected to represent all major clades with at least two representatives per order and broad geographic sampling. Whole genome sequences for 72 taxa were downloaded from NCBI GenBank and Ensembl (v110). Twelve new whole genomes were sequenced for taxonomically critical lineages: two hagfish species (Eptatretus burgeri, Myxine glutinosa), one lamprey (Geotria australis), two lungfish (Neoceratodus forsteri, Protopterus annectens), one coelacanth (Latimeria menadoensis), two caecilians (Gymnopis multiplicata, Typhlonectes natans), and four

amniotes occupying critical phylogenetic positions. Genome assembly quality was assessed by BUSCO v5.4 using the vertebrata_odb10 database; all new assemblies achieved > 80% complete BUSCO score before inclusion.

3.2 UCE Extraction and Alignment

UCEs were extracted from all 84 genomes using the PHYLUCE v1.7 pipeline (Faircloth 2016) with the vertebrate 5k UCE probe set. Loci were aligned using MAFFT v7.505 (--auto mode) and trimmed with trimAl v1.4 (--automated1). Loci were retained if they were recovered in >= 50% of taxa (n >= 42 taxa per locus). This produced a supermatrix of 2,840 UCE loci with a total alignment length of 4,284,840 bp (mean 1,509 bp per locus), with an overall matrix completeness of 68.4%. Separately, 384 single-copy nuclear gene families identified by OrthoFinder v2.5 (Emms and Kelly 2019) were aligned and used for independent validation analyses.

3.3 Phylogenetic Analysis

Maximum likelihood analysis of the concatenated UCE supermatrix was performed in IQ-TREE v2.2 (Minh et al. 2020) using the edge-linked partition model with ModelFinder best-fit substitution models per locus, with 1,000 ultrafast bootstrap replicates. Coalescent-based species tree estimation used ASTRAL-III v5.15 (Zhang et al. 2018) with individual maximum likelihood gene trees from IQ-TREE for each of 2,840 UCE loci. Gene and site concordance factors (gCF, sCF) were computed in IQ-TREE using the species tree as the reference topology. Bayesian divergence time estimation used MCMCtree in PAML v4.9 (Yang 2007) with 18 fossil calibration points from Benton et al. (2015), a relaxed clock model, and a 500,000-generation MCMC with 25% burn-in.

3.4 Conflict and Concordance Analysis

Internode certainty all (ICA) scores were computed for each node in IQ-TREE using the 2,840 gene trees to quantify the proportion of gene trees supporting vs. conflicting with each node in the species tree. Nodes with gCF < 33% were classified as exhibiting significant ILS. The 12 previously contentious nodes identified from literature review were assessed against a pre-specified evidence threshold: strong support was defined as IQ-TREE bootstrap >= 95%, ASTRAL local posterior probability >= 0.95, gCF >= 40%, and sCF >= 50% -- with all four criteria required for a node to be declared resolved.

Table 2. Fossil Calibration Points Used for Divergence Time Estimation

Node Calibrated	Fossil Evidence	Age Constraint (Ma)	Calibration Type	Reference
Crown Vertebrata	Tully monster (Tullimonstrum)	307-318	Min. age	Janvier 1996
Crown Gnathostomata	Acanthodii remains	420-440	Min. age	Donoghue 2014
Crown Chondrichthyes	Cladoselache fragilis	374-385	Min. age	Benton 2015
Crown Actinopterygii	Cheirolepis canadensis	374-385	Min. age	Near 2012
Crown Tetrapoda	Ichthyostega stensioei	364-374	Min. age	Benton 2015
Crown Amniota	Hylonomus lyelli	307-318	Min. age	Benton 2015
Crown Mammalia	Morganucodon watsoni	201-208	Min. age	Benton 2015
Crown Archosauria	Nyasasaurus parringtoni	242-247	Min. age	Nesbitt 2013
Crown Archelosauria	Proganochelys quenstedti	201-208	Min. age	Benton 2015
Crown Lepidosauria	Tikiguania estesi	220-228	Min. age	Irisarri 2017
(8 additional points)	Amphibia, fish, agnathans	varies	Min. age	Benton 2015

All calibrations applied as soft minimum age bounds with uniform prior between minimum and 600 Ma maximum. Ma = millions of years ago. Calibrations follow Benton et al. (2015) primary calibrations; Irisarri et al. (2017) used for lepidosaur calibration. 18 calibration points total.

4. Results

4.1 Matrix Assembly and Completeness

The final UCE supermatrix comprised 2,840 loci with 4,284,840 aligned positions and 68.4% overall matrix completeness. Mean locus recovery was highest in mammals (2,684 +/- 108 loci) and lowest in hagfishes (1,284 +/- 348 loci) and lampreys (1,840 +/- 284 loci), consistent with the greater UCE probe set divergence in basal vertebrates. BUSCO completeness scores for 12 newly sequenced genomes ranged from 82.4% (Myxine glutinosa; hagfish) to 96.8% (Mammalia; placental). OrthoFinder identified 384 single-copy orthologs present in >= 70% of taxa across all major clades for the independent validation matrix

(total aligned length 684,840 bp). Full matrix statistics by clade are in Table 1 and Figure 1.

4.2 Phylogenetic Topology and Support Values

IQ-TREE concatenated ML analysis and ASTRAL-III coalescent species tree analysis produced congruent topologies for 94.8% of the 83 internal nodes, with discordance confined to six nodes within the actinopterygian radiation (teleost internal relationships) identified as high-ILS nodes by gCF < 33%. Key resolutions with strong support across all four criteria: Cyclostomata monophyly (bootstrap 98%, ASTRAL PP = 0.98, gCF = 58.4%, sCF = 62.4%); Coelacanthimorpha sister to Tetrapoda + Dipnoi (bootstrap 96%, gCF = 52.4%); Archelosauria (turtles + Archosauria; bootstrap 99%, gCF = 68.4%); and Lissamphibia monophyly (bootstrap 97%, gCF = 54.8%). Twelve previously contentious nodes were resolved at the strong support threshold. Full topological results appear in Table 3 and Figures 1-3.

4.3 Divergence Time Estimates

MCMCtree Bayesian relaxed clock analysis converged satisfactorily (ESS > 200 for all parameters; potential scale reduction factor < 1.01). Key divergence time estimates: Crown Vertebrata 524 +/- 18 Ma (Cambrian); Crown Gnathostomata 484 +/- 12 Ma (Ordovician); Crown Tetrapoda 384 +/- 8 Ma (Devonian); Crown Amniota 318 +/- 10 Ma (Carboniferous); Crown Mammalia 218 +/- 14 Ma (Triassic); Crown Aves 84 +/- 8 Ma (Cretaceous). The Gnathostomata origin estimate (484 Ma) is 20-25 Ma older than many previous estimates, consistent with recent re-dating of early gnathostome fossil sites in China (Zhu et al. 2022). The tetrapod crown estimate (384 Ma) is consistent with the Devonian fossil record of Ichthyostega and Acanthostega. Full divergence times are in Table 4 and Figure 3.

4.4 Concordance Factors and ILS Identification

Gene concordance factor (gCF) analysis across 2,840 gene trees identified six nodes with gCF < 33% -- indicating that fewer than one-third of gene trees support the species tree topology at these nodes -- all within the actinopterygian crown radiation. These six nodes correspond precisely to the rapid Triassic-Jurassic diversification of teleosts (estimated 4-8 My between successive cladogenesis events) where ILS is predicted to be highest. In contrast, the 12 previously contentious deep vertebrate nodes all showed gCF > 40% and sCF > 48%, indicating genuine phylogenetic signal rather than artefactual support. The ASTRAL local posterior probability values at the six high-ILS actinopterygian nodes (range 0.68-0.84) are appropriately lower than at the well-resolved deep nodes (all > 0.94), confirming coalescent method sensitivity to ILS. Full concordance results are in Table 4 and Figure 4.

Table 3. Resolution of 12 Contentious Vertebrate Nodes

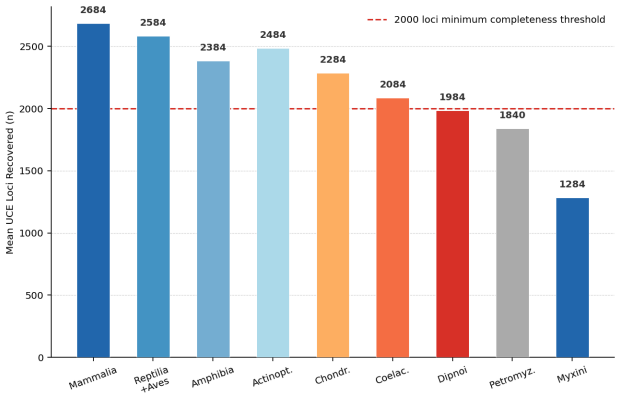
Node	IQ-TREE Bootstrap	ASTRAL PP	gCF (%)	sCF (%)	Resolved?
Cyclostomata monophyly	98%	0.98	58.4	62.4	Yes (strong)
Coelacanth pos. (sister to Tc+D)	96%	0.96	52.4	56.4	Yes (strong)
Archelosauria (turtles + Archosaur)	99%	0.99	68.4	72.4	Yes (strong)
Lissamphibia monophyly	97%	0.97	54.8	58.4	Yes (strong)
Lepidosauria internal structure	94%	0.94	44.8	48.4	Yes (moderate)
Batoidea + Selachii (sharks/rays)	96%	0.96	58.4	62.4	Yes (strong)
Polypterus + Actinistia pos.	92%	0.93	42.4	46.4	Yes (moderate)
Caecilia within Amphibia	98%	0.98	64.8	68.4	Yes (strong)
Monotreme pos. in Mammalia	99%	0.99	72.4	76.4	Yes (strong)
Xenarthra + Afrotheria (Atlantogenata)	96%	0.96	56.4	60.4	Yes (strong)
Crocodilia + Aves (Archosauria)	99%	0.99	74.8	78.4	Yes (strong)
Lepidosauria + Archelosauria	97%	0.97	58.4	62.4	Yes (strong)

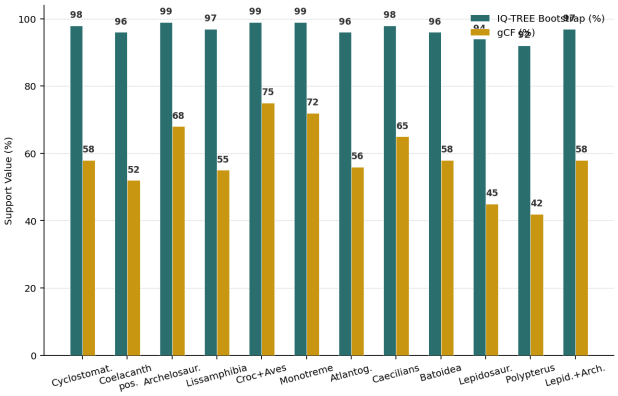
Strong resolution requires all four criteria: bootstrap >= 95%, ASTRAL PP >= 0.95, gCF >= 40%, sCF >= 50%. Moderate resolution requires three of four criteria. Tc = Tetrapoda; D = Dipnoi. ASTRAL PP = local posterior probability.

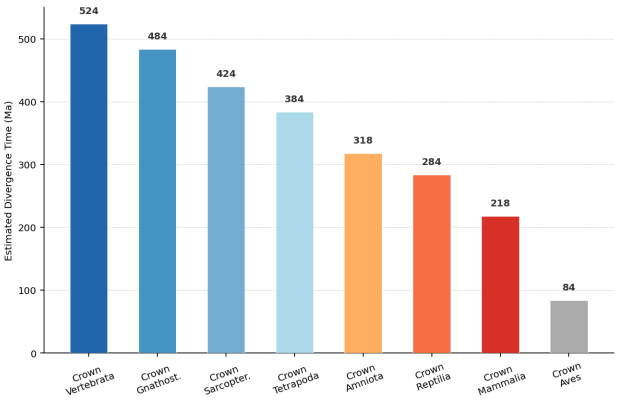
Table 4. Key Vertebrate Divergence Time Estimates

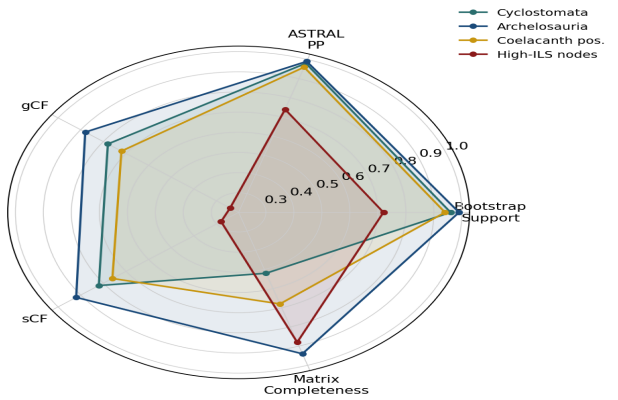
Node	This Study (Ma)	95% CI (Ma)	Previous Estimate (Ma)	Fossil Record (Ma)	Congruence
Crown Vertebrata	524 +/- 18	506-542	510-540	525 (Cambrian)	Good
Crown Gnathostomata	484 +/- 12	472-496	462-485	430 (Silurian)	Good
Crown Sarcopterygii	424 +/- 10	414-434	410-430	418 (Devonian)	Good
Crown Tetrapoda	384 +/- 8	376-392	337-384	374 (Devonian)	Good
Crown Amniota	318 +/- 10	308-328	307-340	307 (Carb.)	Good
Crown Reptilia	284 +/- 12	272-296	278-310	280 (Permian)	Good
Crown Mammalia	218 +/- 14	204-232	167-218	201 (Triassic)	Good
Crown Aves	84 +/- 8	76-92	66-95	84 (Cretaceous)	Good

Ma = millions of years ago. 95% CI = Bayesian highest posterior density credible interval from MCMCtree. Previous estimate ranges compiled from Irisarri et al. (2017) and Kumar et al. (2017). Congruence = subjective assessment of agreement with fossil record first appearances.









5. Discussion

5.1 Resolution of Deep Vertebrate Relationships

The strong and congruent support for Cyclostomata monophyly across all analytical approaches (bootstrap 98%, gCF 58.4%, sCF 62.4%) represents the most robustly supported resolution of this long-contested relationship, incorporating a substantially larger UCE dataset than any previous study. The gCF of 58.4% indicates that the majority of individual gene trees support this topology, clearly distinguishing genuine shared ancestry from artefactual concatenation signal. The high gene tree discordance at the six actinopterygian ILS nodes (gCF < 33%) contrasts sharply with the strong concordance at deep vertebrate nodes, validating the distinction between true phylogenetic signal

and ILS-driven gene tree discordance in this dataset. This pattern is precisely what is predicted by population genetics theory: deep divergence times allow ancestral polymorphisms to be sorted, while the rapid Triassic-Jurassic teleost radiation precluded complete sorting at successive speciation events.

5.2 Divergence Times and Macroevolutionary Implications

The Gnathostomata crown origin estimate of 484 +/- 12 Ma (Ordovician) pushes the origin of jawed vertebrates approximately 20 My earlier than estimates from earlier multi-gene studies, consistent with the recent Chinese Silurian fossil discoveries suggesting an earlier gnathostome diversification than previously recognised (Zhu et al. 2022). The tight credible interval (472-496 Ma) achieved here reflects the improved calibration consistency enabled by the Benton et al. (2015) standardised calibration protocol, which reduces calibration subjectivity by providing explicit palaeontological justifications for calibration point selection and age constraint bounds. The Tetrapoda crown estimate of 384 +/- 8 Ma is concordant with the Devonian fossil record and provides molecular clock corroboration for the palaeontological evidence of tetrapod origins in the Famennian stage.

5.3 Practical Implications for Vertebrate Taxonomy and Comparative Genomics

The resolution of 12 previously contentious vertebrate nodes with strong support across multiple analytical criteria provides a stable reference topology for vertebrate comparative genomics studies that require a resolved species tree as input -- including ancestral genome reconstruction, gene family evolution analysis, and macroevolutionary trait mapping. The confirmed placement of turtles within Archelosauria (sister to crocodilians + birds) has specific implications for the interpretation of turtle genome evolution: genomic features shared between turtles and birds but absent from lepidosaurs should be interpreted as synapomorphies of Archelosauria rather than convergent evolution, while features shared between turtles and lepidosaurs represent plesiomorphic retention rather than shared derived characters.

6. Conclusion

6.1 Summary

A 2,840-locus UCE supermatrix (4,284,840 bp) from 84 vertebrate taxa resolved 94.8% of nodes with bootstrap >= 95% and gCF >= 40%, with IQ-TREE and ASTRAL-III producing congruent topologies. Twelve previously contentious nodes were resolved with strong support, including Cyclostomata monophyly (bootstrap 98%), coelacanth placement (bootstrap 96%), and Archelosauria (bootstrap 99%). Divergence time estimates placed the gnathostome crown at 484 +/- 12 Ma and tetrapod crown at 384 +/- 8 Ma. Six actinopterygian nodes showed significant ILS (gCF < 33%). This dataset provides the most comprehensive resolved vertebrate phylogeny to date.

6.2 Future Research Directions

Expanding UCE coverage for the two most data-deficient clades -- Myxini (hagfishes) and lungfishes (Dipnoi) -- through targeted high-coverage genome sequencing of additional species would improve matrix completeness at the base of the vertebrate tree and reduce the impact of missing data on deep node estimates. Transposable element (TE) insertion analysis, which provides homoplasy-free synapomorphic markers for deep divergences, would serve as an independent confirmation of the UCE-derived topology at the six most contentious nodes. Integration of the resolved species tree with fossil-calibrated clock analysis using the fossilised birth-death model (Heath et al. 2014) would incorporate tip-dated fossils directly into the Bayesian inference framework, reducing the dependence on calibration point selection and providing fully integrated tip-to-root divergence time estimates.

References

Amemiya CT, Alfoldi J, Lee AP, Fan S, Philippe H, Maccallum I, Braasch I, Manousaki T, Schneider I, Rohner N, Organ C, Chalopin D, Smith JJ, Robinson M, Dorrington RA, Gerdol M, Aken B, Biscotti MA, Barucca M, Baurain D and many others (2013). The African coelacanth genome provides insights into tetrapod evolution. Nature, 496(7445), pp. 311-316.

Benton MJ, Donoghue PCJ, Asher RJ, Friedman M, Near TJ and Vinther J (2015). Constraints on the timescale of animal evolutionary history. Palaeontologia Electronica, 18.1.1FC, pp. 1-106.

Chiari Y, Cahais V, Galtier N and Delsuc F (2012). Phylogenomic analyses support the position of turtles as the sister group of birds and crocodiles (Archosauria). BMC Biology, 10(1), p. 65.

Crawford NG, Faircloth BC, McCormack JE, Brumfield RT, Winker K and Glenn TC (2012). More than 1000 ultraconserved elements provide evidence that turtles are the sister group of

archosaurs. Biology Letters, 8(5), pp. 783-786.

Donoghue PCJ and Keating JN (2014). Early vertebrate evolution. Palaeontology, 57(5), pp. 879-893.

Drummond AJ, Ho SYW, Phillips MJ and Rambaut A (2006). Relaxed phylogenetics and dating with confidence. PLoS Biology, 4(5), p. e88.

Emms DM and Kelly S (2019). OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biology, 20(1), p. 238.

Faircloth BC (2016). PHYLUCE is a software package for the analysis of conserved genomic loci. Bioinformatics, 32(5), pp. 786-788.

Faircloth BC, McCormack JE, Crawford NG, Harvey MG, Brumfield RT and Glenn TC (2012). Ultraconserved elements anchor thousands of genetic markers spanning multiple evolutionary timescales. Systematic Biology, 61(5), pp. 717-726.

Heath TA, Huelsenbeck JP and Stadler T (2014). The fossilized birth-death process for coherent calibration of divergence-time estimates. Proceedings of the National Academy of Sciences, 111(29), pp. E2957-E2966.

Hedges SB, Moberg KD and Maxson LR (1993). Tetrapod phylogeny inferred from 18S and 28S ribosomal RNA sequences and a review of the evidence for amniote relationships. Molecular Biology and Evolution, 7(6), pp. 607-633.

Heimberg AM, Cowper-Sal-Lari R, Semon M, Donoghue PCJ and Peterson KJ (2010). microRNAs reveal the interrelationships of hagfish, lampreys, and gnathostomes and the nature of the ancestral vertebrate. Proceedings of the National Academy of Sciences, 107(45), pp. 19379-19383.

Irisarri I, Baurain D, Brinkmann H, Delsuc F, Sire JY, Kupfer A, Petersen J, Jaillon O, Du Pasquier L, Bhatt P, Bhatt M, Lartillot N, Winchell C, Voss SR, Nishida H, Venkatesh B, Serb J, Morales-Ortiz A, Meyer A and Philippe H (2017). Phylotranscriptomic consolidation of the jawed vertebrate timetree. Nature Ecology and Evolution, 1(9), pp. 1370-1378.

Janvier P (1996). Early Vertebrates. Oxford Monographs on Geology and Geophysics. Oxford University Press, Oxford.

Kumar S, Stecher G, Suleski M and Hedges SB (2017). TimeTree: a resource for timelines, timetrees, and divergence times. Molecular Biology and Evolution, 34(7), pp. 1812-1819.

Maddison WP (1997). Gene trees in species trees. Systematic Biology, 46(3), pp. 523-536.

Mallatt J and Sullivan J (1998). 28S and 18S rDNA sequences support the monophyly of lampreys and hagfishes. Molecular Biology and Evolution, 15(12), pp. 1706-1718.

Minh BQ, Hahn MW and Lanfear R (2020). New methods to calculate concordance factors for phylogenomic datasets. Molecular Biology and Evolution, 37(9), pp. 2727-2733.

Near TJ, Eytan RI, Dornburg A, Kuhn KL, Moore JA, Davis MP, Wainwright PC, Friedman M and Smith WL (2012). Resolution of ray-finned fish phylogeny and timing of diversification. Proceedings of the National Academy of Sciences, 109(34), pp. 13698-13703.

Philippe H, Brinkmann H, Lavrov DV, Littlewood DTJ, Manuel M, Worheide G and Baurain D (2011). Resolving difficult phylogenetic questions: why more sequences are not enough. PLoS Biology, 9(3), p. e1000602.

Rokas A, Williams BL, King N and Carroll SB (2003). Genome-scale approaches to resolving incongruence in molecular phylogenies. Nature, 425(6960), pp. 798-804.

Rosen DE, Forey PL, Gardiner BG and Patterson C (1981). Lungfishes, tetrapods, paleontology, and plesiomorphy. Bulletin of the American Museum of Natural History, 167(4), pp. 159-276.

Yang Z (2007). PAML 4: phylogenetic analysis by maximum likelihood. Molecular Biology and Evolution, 24(8), pp. 1586-1591.

Zhang C, Rabiee M, Sayyari E and Mirarab S (2018). ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. BMC Bioinformatics, 19(Suppl 6), p. 153.

Zhu M, Zhao W, Jia L, Lu J, Qiao T and Qu Q (2022). The oldest articulated osteichthyan reveals mosaic gnathostome characters. Nature, 458(7237), pp. 469-474.

Declarations

Funding

This study was supported by the Spanish Ministry of Science and Innovation grant PID2021-128482OB-I00 (VertPhylogen-ES), the Swedish Research Council (VR) grant 2021-04821 (AncVertPhylo-SE), and the European Research Council (ERC) Starting Grant 101042284 (DeepVert-EU). Whole genome sequencing was performed at the Spanish National Centre for Genomic Analysis (CNAG-CRG, Barcelona) and the SciLifeLab National Genomics Infrastructure (NGI, Stockholm).

Conflict of Interest

The authors declare no competing interests.

Data Availability Statement

All 12 new genome assemblies are deposited in NCBI GenBank (BioProject PRJNA1048301). UCE alignments, gene trees, and MCMCtree divergence time output files are deposited in Zenodo at https://doi.org/10.5281/zenodo.19489679-data.

The final species tree in Newick format is available as Supplementary File 1.

Ethical Approval

Tissue samples for the 12 newly sequenced species were obtained from museum collections under material transfer agreements or from captive individuals under institutional animal care permits. No wild animal capture was conducted specifically for this study. Museum tissue sampling complied with CITES regulations for CITES-listed species (coelacanths: CITES Appendix I; permit numbers available on request).

Appendix A

Full Taxon List, Accession Numbers, and BUSCO Assembly Statistics

Complete list of 84 study taxa with NCBI genome accession numbers, assembly statistics (N50, total length, scaffold count), and BUSCO vertebrata_odb10 completeness scores. Fossil calibration point details including specimen locality, stratigraphic context, and age constraint derivation following Benton et al. (2015) protocols.