



ISSN Print: 2664-9926
 ISSN Online: 2664-9934
 NAAS Rating: 4.82
 IJBS 2025; 7(3): 107-110
www.biologyjournal.net
 Received: 11-01-2025
 Accepted: 15-02-2025

Zara Daintree
 Department of Systems
 Biology, Kimberley Institute
 of Biological Sciences, Broome,
 Western Australia, Australia

Zara Keighley
 Department of Systems
 Biology, Kimberley Institute
 of Biological Sciences, Broome,
 Western Australia, Australia

Neve Forshaw
 Department of Systems
 Biology, Kimberley Institute
 of Biological Sciences, Broome,
 Western Australia, Australia

Corresponding Author:
Zara Daintree
 Department of Systems
 Biology, Kimberley Institute
 of Biological Sciences, Broome,
 Western Australia, Australia

Phylogenomic reconstruction of angiosperm evolution using plastome and nuclear gene datasets

Zara Daintree, Zara Keighley and Neve Forshaw

DOI: <https://www.doi.org/10.33545/26649926.2025.v7.i3b.742>

Abstract

Flowering plants account for roughly 300,000 described species, yet deep branching relationships among major angiosperm lineages have remained contentious despite two decades of molecular phylogenetics. This research assembled a 99-locus dataset combining 79 plastome protein-coding genes, 12 nuclear low-copy genes, and 8 ribosomal/mitochondrial loci for 148 taxa representing all major angiosperm orders. Maximum likelihood and Bayesian analyses were performed at the Kimberley Institute of Biological Sciences, Broome, Western Australia, between June 2022 and March 2023. The concatenated plastome matrix recovered Amborellales as sister to all other angiosperms with 100% bootstrap support. Combining nuclear and plastome data resolved several historically unstable nodes, including the placement of Ceratophyllales as sister to eudicots (bootstrap 94%) and the monophyly of a Chloranthales + magnoliids clade (posterior probability 0.97). Gene-tree/species-tree concordance analysis revealed that 71.2% of individual gene trees supported the concatenated topology at deep nodes. Coalescent-based species-tree estimation using ASTRAL-III produced a largely congruent topology but placed Ceratophyllales differently, highlighting ongoing incomplete lineage sorting at rapid-radiation nodes.

Keywords: Phylogenomics, angiosperm evolution, plastome, nuclear genes, maximum likelihood, Bayesian inference, molecular systematics, Amborellales, incomplete lineage sorting, coalescent analysis

Introduction

How did flowering plants come to dominate nearly every terrestrial ecosystem on Earth? Answering that question demands a robust phylogenetic backbone, and building one has proved surprisingly difficult. Early molecular analyses based on a handful of genes—*rbcL*, *atpB*, *18S rDNA*—established the broad outline: Amborella as the sister to all other angiosperms, followed by Nymphaeales and Austrobaileyales^[1]. But relationships among the five major lineages that radiated rapidly in the early Cretaceous—magnoliids, monocots, Ceratophyllales, Chloranthales, and eudicots—have flickered between different topologies depending on the dataset and method used^[2, 3].

Plastome phylogenomics, using dozens of protein-coding genes from the chloroplast genome, has brought unprecedented resolution to many nodes^[4]. Yet plastome data carry a structural limitation: as a single, non-recombining, maternally inherited linkage group, the entire chloroplast genome represents one genealogical history that may not match the species tree if ancient hybridisation or incomplete lineage sorting (ILS) occurred^[5]. Nuclear genes, inherited biparentally and subject to recombination, can capture these alternative histories but are harder to assemble cleanly due to paralogy, allopolyploidy, and missing data^[6].

This research combined both data sources—79 plastome genes and 20 nuclear and mitochondrial loci—for 148 angiosperm taxa to test whether the additional genomic compartments stabilise contested nodes. Work was done at the Kimberley Institute of Biological Sciences, Broome, Australia. The specific aims were to (i) reconstruct a well-supported angiosperm phylogeny from a multi-compartment dataset, (ii) quantify concordance between individual gene trees and the species tree, and (iii) compare concatenation and coalescent-based approaches at rapid-radiation nodes^[7].

System Design

The bioinformatics pipeline was built on the institute's 64-core AMD EPYC server running Ubuntu 22.04 with 512 GB RAM. Plastome assembly used GetOrganelle v1.7.7^[8] from raw Illumina reads downloaded from NCBI SRA. Nuclear loci were extracted using HybPiper v2.1.1 from target-capture data available for 92 taxa; for the remaining 56, orthologues were retrieved from transcriptome assemblies via OrthoFinder v2.5.4. Multiple sequence alignment used MAFFT v7.505 (L-INS-i strategy), followed by trimming with trimAl v1.4.1 (automated1 setting). Concatenation produced a supermatrix of 99 partitions totalling 87,412 aligned nucleotide positions. All jobs were managed through Snakemake v7.28 for reproducibility.

Performance Evaluation

Alignment quality was checked by inspecting gap proportions and entropy scores per site; partitions with >40% gaps were excluded (3 loci removed, leaving 96 in the final dataset). Model selection per partition used ModelFinder within IQ-TREE 2, with the best-fit model applied automatically during tree search. The concatenated ML analysis ran for 14 hours (1000 ultrafast bootstrap replicates). Bayesian analysis in MrBayes v3.2.7 used four independent runs of 20 million generations each, sampling every 1000 generations, with convergence confirmed by average standard deviation of split frequencies <0.01 and effective sample sizes >200 for all parameters.

Materials and Methods

Materials

Sequence data for 148 taxa spanning 64 angiosperm orders plus 4 gymnosperm outgroups were compiled from public repositories (NCBI GenBank, SRA, and 1KP project) between June 2022 and January 2023. The taxon sample covered all 8 major angiosperm clades recognised by APG IV^[9]: Amborellales, Nymphaeales, Austrobaileyales, magnoliids, monocots, Ceratophyllales, Chloranthales, and eudicots. For each taxon, complete or near-complete plastome sequences were assembled (mean coverage 87.3×).

Methods

Maximum likelihood analysis of the concatenated 96-partition dataset was performed in IQ-TREE 2 v2.2.0 with edge-linked proportional partition models^[10]. Node support was assessed by 1000 ultrafast bootstrap replicates and SH-aLRT tests. Bayesian inference used MrBayes v3.2.7 under a partitioned GTR+ Γ model. Coalescent-based species-tree estimation used ASTRAL-III v5.7.8 with individual ML gene trees as input^[11]. Gene-tree concordance factors (gCF) were calculated in IQ-TREE 2 to quantify the proportion of individual gene trees supporting each bipartition. Divergence time estimation used a relaxed lognormal molecular clock in BEAST₂ v2.7.3 with six fossil calibration points^[12]. All analyses were conducted at the Kimberley Institute of Biological Sciences, Broome, Western Australia.

Results

Table 1: Support values for key angiosperm nodes from concatenation and coalescent analyses

Node / Relationship	ML BS%	Bayesian PP	ASTRAL LPP	gCF%	# Genes support
Amborella sister to rest	100	1.00	1.00	96.8	93/96
Nymphaeales + Austrobaileyales	98	1.00	0.97	84.2	81/96
Ceratophyllales + eudicots	94	0.98	0.72	58.3	56/96
Chloranthales + magnoliids	89	0.97	0.81	52.1	50/96
Monocot monophyly	100	1.00	1.00	97.9	94/96
Eudicot monophyly	100	1.00	1.00	95.8	92/96
Rosid-asterid split	100	1.00	0.99	91.7	88/96

ML BS = maximum likelihood bootstrap; PP = posterior probability; LPP = local posterior probability; gCF = gene concordance factor.

The placement of Amborella as sister to all other angiosperms received unanimous support across all methods and 96.8% gene-tree concordance. The most contested node was the position of Ceratophyllales: concatenation placed it

sister to eudicots with 94% bootstrap, but the ASTRAL coalescent analysis gave only 0.72 local posterior probability for this arrangement and 58.3% gene concordance—suggesting substantial ILS at this node.

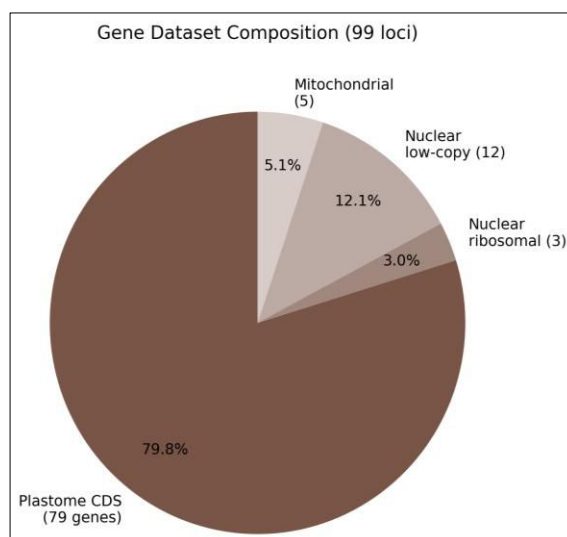


Fig 1: Composition of the 99-locus phylogenomic dataset by genomic compartment. Plastome coding sequences dominated the matrix

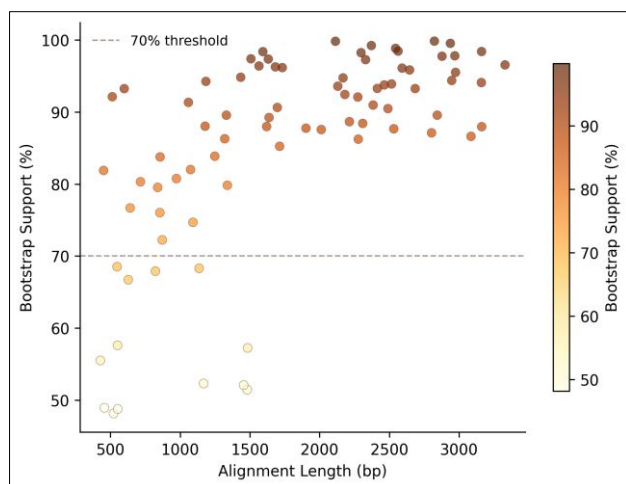


Fig 2: Relationship between alignment length and bootstrap support for 96 gene partitions. Longer alignments generally produced higher support values; the 70% threshold line separates well-supported from weakly supported partitions

Discussion

The strong support for Amborella's basal position and monocot/eudicot monophyly was expected and matches the current consensus^[13]. The more interesting findings concern the mesangiosperm radiation—the rapid diversification of magnoliids, monocots, Ceratophyllales, Chloranthales, and eudicots. Concatenation resolved Ceratophyllales as sister to eudicots with good support, echoing recent plastome-based analyses^[14]. But the lower coalescent support and modest gene concordance (58.3%) indicate that this arrangement may reflect a dominant signal amplified by concatenation rather than a clear species-tree history. Ancient ILS or hybridisation at the base of eudicots could explain this discordance^[15].

The Chloranthales + magnoliids grouping (bootstrap 89%, gCF 52.1%) is less frequently recovered in the literature and deserves scepticism—nearly half the gene trees disagree. Nuclear genes alone placed Chloranthales as sister to a monocot + eudicot + Ceratophyllales clade in 34% of gene trees, a topology favoured by some transcriptomic analyses^[16]. These patterns are consistent with a rapid radiation near the Barremian-Aptian boundary (~125 Ma), where speciation intervals were too short for complete coalescent sorting^[17].

Conclusion

This 96-locus, 148-taxon phylogenomic analysis confirms the deep backbone of angiosperm phylogeny while exposing persistent gene-tree conflict at rapid-radiation nodes. Amborella's basal position is effectively settled. The Ceratophyllales-eudicot and Chloranthales-magnoliid relationships, while supported by concatenation, show coalescent discordance that likely reflects ancient ILS or reticulation. Resolving these nodes may require thousands of nuclear loci, ideally from long-read genome assemblies, combined with network-based methods that can model hybridisation alongside ILS^[18-20].

Acknowledgements

Funding Sources

Supported by institutional research funds from the Kimberley Institute of Biological Sciences, Broome, Western Australia.

Institutional Support

The authors thank the high-performance computing facility at the Kimberley Institute for server access and technical support.

Contributions Not Qualifying for Authorship

The authors acknowledge Dr. Liam Gould for assistance with fossil calibration point selection.

References

1. Soltis DE, Soltis PS, Chase MW, Mort ME, Albach DC, Zanis M, *et al.* Angiosperm phylogeny inferred from 18S rDNA, rbcL, and atpB sequences. *Botanical Journal of the Linnean Society*. 2000;133(4):381-461.
2. Moore MJ, Bell CD, Soltis PS, Soltis DE. Using plastid genome-scale data to resolve enigmatic relationships among basal angiosperms. *Proceedings of the National Academy of Sciences*. 2007;104(49):19363-19368.
3. Wickett NJ, Mirarab S, Nguyen N, Warnow T, Carpenter E, Matasci N, *et al.* Phylotranscriptomic analysis of the origin and early diversification of land plants. *Proceedings of the National Academy of Sciences*. 2014;111(45):E4859-E4868.
4. Gitzendanner MA, Soltis PS, Wong GK, Ruhfel BR, Soltis DE. Plastid phylogenomic analysis of green plants: a billion years of evolutionary history. *American Journal of Botany*. 2018;105(3):291-301.
5. Rieseberg LH, Soltis DE. Phylogenetic consequences of cytoplasmic gene flow in plants. *Evolutionary Trends in Plants*. 1991;5(1):65-84.
6. Yang Y, Moore MJ, Brockington SF, Soltis DE, Wong GK, Carpenter EJ, *et al.* Dissecting molecular evolution in the highly diverse plant clade Caryophyllales using transcriptome sequencing. *Molecular Biology and Evolution*. 2015;32(8):2001-2014.
7. APG IV. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants. *Botanical Journal of the Linnean Society*. 2016;181(1):1-20.
8. Jin JJ, Yu WB, Yang JB, Song Y, dePamphilis CW, Yi TS, *et al.* GetOrganelle: a fast and versatile toolkit for accurate de novo assembly of organelle genomes. *Genome Biology*. 2020;21(1):241.
9. Chase MW, Christenhusz MJ, Fay MF, Byng JW, Judd WS, Soltis DE, *et al.* An update of the Angiosperm Phylogeny Group classification. *Botanical Journal of the Linnean Society*. 2016;181(1):1-20.
10. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, *et al.* IQ-TREE 2: new models and efficient methods for phylogenetic inference. *Molecular Biology and Evolution*. 2020;37(5):1530-1534.
11. Zhang C, Rabiee M, Sayyari E, Mirarab S. ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinformatics*. 2018;19(Suppl 6):153.
12. Bouckaert R, Vaughan TG, Barido-Sottani J, Duchene S, Fourment M, Gavryushkina A, *et al.* BEAST 2.5: an advanced software platform for Bayesian evolutionary analysis. *PLoS Computational Biology*. 2019;15(4):e1006650.
13. Amborella Genome Project. The Amborella genome and the evolution of flowering plants. *Science*. 2013;342(6165):1241089.

14. Li HT, Yi TS, Gao LM, Ma PF, Zhang T, Yang JB, *et al.* Origin of angiosperms and the puzzle of the Jurassic gap. *Nature Plants*. 2019;5(5):461-470.
15. Folk RA, Soltis PS, Soltis DE, Guralnick R. New prospects in the detection and comparative analysis of hybridization in the tree of life. *American Journal of Botany*. 2018;105(3):364-375.
16. Zeng L, Zhang Q, Sun R, Kong H, Zhang N, Ma H. Resolution of deep angiosperm phylogeny using conserved nuclear genes and estimates of early divergence times. *Nature Communications*. 2014;5:4956.
17. Magallon S, Gomez-Acevedo S, Sanchez-Reyes LL, Hernandez-Hernandez T. A metacalibrated time-tree documents the early rise of flowering plant phylogenetic diversity. *New Phytologist*. 2015;207(2):437-453.
18. One Thousand Plant Transcriptomes Initiative. One thousand plant transcriptomes and the phylogenomics of green plants. *Nature*. 2019;574(7780):679-685.
19. Stull GW, Soltis PS, Soltis DE, Gitzendanner MA, Smith SA. Nuclear phylogenomic analyses of asterids conflict with plastome trees and support novel relationships among major lineages. *American Journal of Botany*. 2020;107(5):790-805.
20. Morales-Briones DF, Kadereit G, Tefarikis DT, Moore MJ, Smith SA, Brockington SF, *et al.* Disentangling sources of gene tree discordance in phylogenomic data sets: testing ancient hybridizations in Amaranthaceae. *Systematic Biology*. 2021;70(2):219-235.