



This is a repository copy of *Phylogenomics and the rise of the angiosperms*.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/212833/>

Version: Published Version

Article:

Zuntini, A.R. orcid.org/0000-0003-0705-8902, Carruthers, T., Maurin, O. orcid.org/0000-0002-4151-6164 et al. (276 more authors) (2024) Phylogenomics and the rise of the angiosperms. *Nature*, 629. pp. 843-850. ISSN 0028-0836

<https://doi.org/10.1038/s41586-024-07324-0>

Reuse

This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here:

<https://creativecommons.org/licenses/>

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.



eprints@whiterose.ac.uk
<https://eprints.whiterose.ac.uk/>

Phylogenomics and the rise of the angiosperms

<https://doi.org/10.1038/s41586-024-07324-0>

Received: 15 September 2023

Accepted: 15 March 2024

Published online: 24 April 2024

Open access

 Check for updates

Angiosperms are the cornerstone of most terrestrial ecosystems and human livelihoods^{1,2}. A robust understanding of angiosperm evolution is required to explain their rise to ecological dominance. So far, the angiosperm tree of life has been determined primarily by means of analyses of the plastid genome^{3,4}. Many studies have drawn on this foundational work, such as classification and first insights into angiosperm diversification since their Mesozoic origins^{5–7}. However, the limited and biased sampling of both taxa and genomes undermines confidence in the tree and its implications. Here, we build the tree of life for almost 8,000 (about 60%) angiosperm genera using a standardized set of 353 nuclear genes⁸. This 15-fold increase in genus-level sampling relative to comparable nuclear studies⁹ provides a critical test of earlier results and brings notable change to key groups, especially in rosids, while substantiating many previously predicted relationships. Scaling this tree to time using 200 fossils, we discovered that early angiosperm evolution was characterized by high gene tree conflict and explosive diversification, giving rise to more than 80% of extant angiosperm orders. Steady diversification ensued through the remaining Mesozoic Era until rates resurged in the Cenozoic Era, concurrent with decreasing global temperatures and tightly linked with gene tree conflict. Taken together, our extensive sampling combined with advanced phylogenomic methods shows the deep history and full complexity in the evolution of a megadiverse clade.

Flowering plants (angiosperms) represent about 90% of all terrestrial plant species² but, despite their remarkable diversity and ecological importance underpinning almost all main terrestrial ecosystems, their evolutionary history remains incompletely known. Since their Mesozoic origins^{5,10,11}, angiosperms have had a pervasive influence on the biosphere of Earth, shaping climatic changes at global and local scales¹², supporting the structure and assembly of biomes¹³ and influencing the diversification of other organisms, such as insects, fungi and birds¹⁴. The evolution of terrestrial biodiversity is thus inextricably linked with the macroevolution of angiosperms, which can only be shown using a robust and comprehensive tree of life. Reconstructing such a tree, however, is challenging because of the sheer diversity of angiosperms and the complex phylogenetic signal in their genomes.

High-throughput DNA sequencing methods now enable us to reconstruct phylogenetic trees that broadly represent the evolutionary signal across entire genomes. Target sequence capture¹⁵ has revolutionized plant phylogenetics by unlocking herbarium specimens as a source of sequenceable DNA¹⁶, thus removing the chief sampling bottleneck that has obstructed the completion of the tree of life. Although previous work on plants has relied primarily on the widely sequenced plastid genome^{3,4,7}, these technologies now allow us to tap into the evolutionary signal of the much larger and more complex nuclear genome. Universal nuclear probe sets, such as Angiosperms353 (ref. 8), have made target sequence capture consistently applicable across broad taxonomic scales, opening doors to collaboration and data integration¹⁷. As a result, opportunities now present themselves to address fundamental questions in plant evolutionary biology, such as the origin of angiosperms, the tempo and mode of their diversification and the classification of main lineages.

Here, we present a nuclear phylogenomic tree that includes all 64 orders and 416 families of angiosperms recognized by the prevailing classification¹⁸, using the Angiosperms353 (ref. 8) gene panel. Our sampling of 7,923 angiosperm genera (represented by 9,506 species) amounts to a 15-fold increase compared to previous work⁹. Leveraging a dataset of 200 fossil calibrations, we scale the tree to time, effectively capturing evolutionary divergences for all but the most recent 15% of angiosperm history. Although our tree broadly supports relationships predicted by previous studies primarily based on plastid data, it also shows previously unknown relationships and highlights some that remain intractable despite a vast increase in data. Gene tree conflict is tightly linked to diversification across the tree. We find evidence for high levels of conflict associated with an early burst of diversification, which is followed by an extended period of constant diversification rates underpinned by a tapestry of varied lineage-specific patterns. Diversification then increases in the Cenozoic Era, potentially driven by global climatic cooling. Our results highlight the fundamental role of botanical collections in reconstructing the tree of life to illuminate long-standing questions in angiosperm macroevolution.

The angiosperm tree of life

Our phylogenetic tree includes 58% of the approximately 13,600 currently accepted genera of angiosperms (Fig. 1 and Supplementary Table 1; ref. 2). Together, the 7,923 genera encompass 85.7% of total known angiosperm species diversity. We produced data for 6,777 of these genera; before this study, 3,154 of these lacked publicly available genomic data, of which 393 lacked any form of DNA

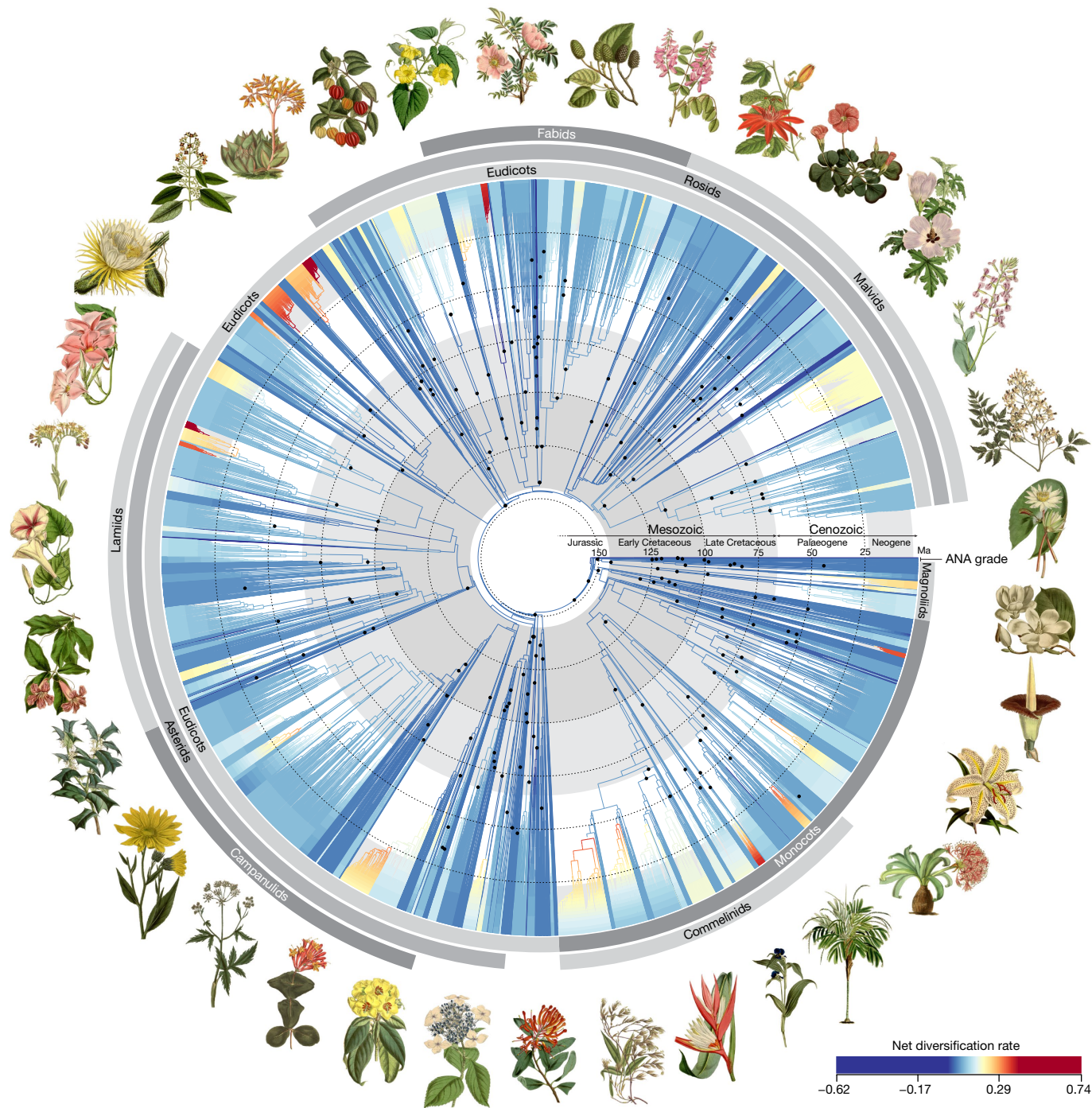


Fig. 1 | Time-calibrated phylogenetic tree for angiosperms based on 353 nuclear genes. All 64 orders, all 416 families and 58% (7,923) of genera are represented. The young tree is illustrated here (maximum constraint at the root node of 154 Ma), with branch colours representing net diversification rates. Black dots at nodes indicate the phylogenetic placement of fossil calibrations based on the updated AngioCal fossil calibration dataset. Note that calibrated nodes can be older than the age of the corresponding fossils owing to the use of minimum age constraints. Arcs around the tree indicate

the main clades of angiosperms as circumscribed in this paper. ANA grade refers to the three consecutively diverging orders Amborellales, Nymphaeales and Austrobaileyales. Plant portraits illustrating key orders were sourced from Curtis’s Botanical Magazine (Biodiversity Heritage Library). These portraits, by S. Edwards, W. H. Fitch, W. J. Hooker, J. McNab and M. Smith, were first published between 1804 and 1916 (for a key to illustrations see Supplementary Table 2). A high-resolution version of this figure can be downloaded from <https://doi.org/10.5281/zenodo.10778206> (ref. 55).

sequence data. For the remaining genera, data were obtained from public repositories. Sampling for this project was possible thanks to the collaborative effort of many biodiversity institutions from around the world, including 163 herbaria in 48 countries. More than one-third of species were sourced directly from herbarium specimens, some dating back nearly 200 years. Many phylogenetically

problematic lineages with unconventional genome evolution were sampled, such as holoparasites, mycoheterotrophs and aquatics. Many of the species included are threatened and four are extinct (or extinct in the wild). The resulting tree of life presented here is one of the largest genomic trees generated yet for angiosperms as a whole.

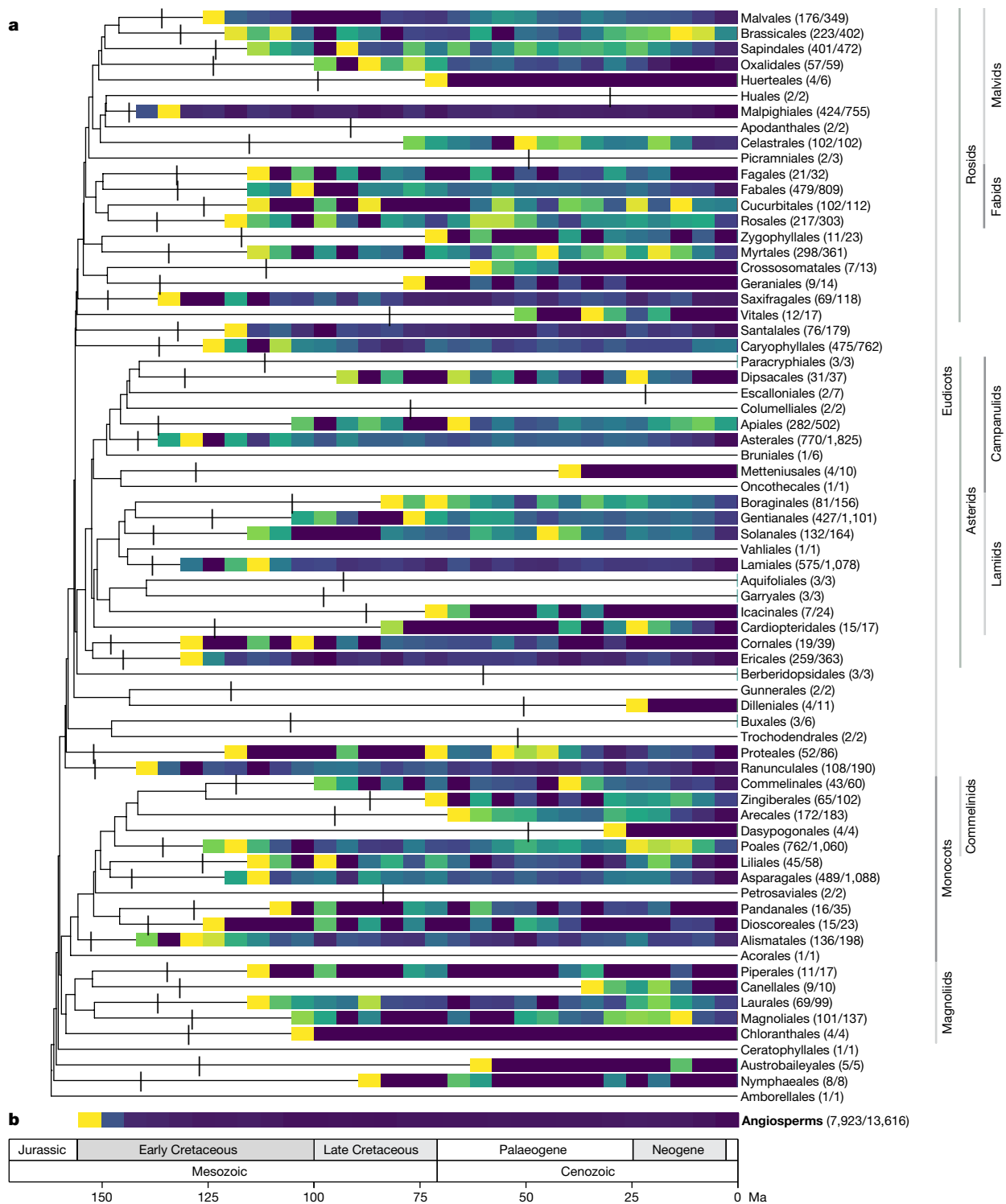


Fig. 2 | Diversification dynamics across angiosperms. The results illustrated are based on the young tree (maximum constraint at the root node of 154 Ma). **a**, Time-calibrated summary phylogenetic tree with LTT plots rendered as heatmaps for all orders with four or more sampled genera. The log-transformed increase in the number of lineages is depicted in 5 Myr intervals, omitting crown

nodes, which disproportionately altered the visualization; crown node locations are indicated by vertical lines. The yellow to blue colour scale represents steep to shallow slopes. For each order, the numbers of sampled and total genera are provided. **b**, A global LTT heatmap for all angiosperms is shown at the bottom of the figure as a whole.

The phylogenomic challenge

Large genomic datasets present challenges to phylogenetic inference. One issue is accurate homology assessment, which proved intractable across the full span of our dataset, even with the most advanced multiple sequence alignment methods. Another challenge

is the efficient search of tree space based on gene matrices that have many more taxa than characters. We overcame both challenges with a divide-and-conquer approach (Supplementary Fig. 1). First, we computed a backbone species tree with sampling limited to five species per family (1,336 (15%) samples in total) and targeted to represent their deepest nodes (Supplementary Fig. 2). We used the

backbone species tree to delimit taxon subsets for the construction of order-level gene alignments, which were then merged into global alignments. We then computed global gene trees from the global alignments, using backbone gene trees (inferred during the estimation of the backbone species tree) as topological constraints to reduce tree space while still letting gene trees differ from each other. The smaller number of samples in the backbone dataset permits a more thorough search of tree space, resulting in greater confidence at deeper nodes than could be achieved in an unconstrained global analysis. This approach allows a trade-off between comprehensive sampling and tree search robustness while accommodating putative discordance among gene trees. Finally, we used the global gene trees to generate a global species tree in a multispecies coalescent framework (Supplementary Fig. 3).

A widespread concern in phylogenomic analysis is the presence of undetected gene copies. Our findings are unlikely to be affected by this because we used genes that have been selected to be mostly single-copy across green plants^{8,9}. Although gene duplication cannot be ruled out¹⁹, the methods we used have been shown to be robust to the presence of paralogues²⁰. In addition, a full assessment of orthologues was not computationally tractable but should be undertaken when methods become available to fully unravel the complexity of genome evolution at this scale²¹.

Phylogenetic insights from nuclear data

Our results broadly corroborate the prevailing understanding of angiosperm phylogenetic relationships, which rests on three decades of molecular systematic research largely built on data from the plastid genome^{3,4,18,22}. We recover all main lineages of angiosperms, namely Amborellales, Nymphaeales, Austrobaileyales, Ceratophyllales and the three larger clades, monocots, magnoliids (including Chloranthales) and eudicots (Figs. 1 and 2). Although some of the relationships among those groups, such as the placement of Amborellales as sister group to all other angiosperms, are well-established and confirmed here, others, such as the placement of Ceratophyllales, which have been unstable in previous work^{4,9}, remain inconclusive in our results. Despite the contrasting biological properties of the nuclear and plastid genomes (for example, size, copy number, mode of inheritance, recombination and evolutionary rate), which can lead to conflicting phylogenetic results, our findings largely support the mostly plastid-based phylogenetic classification of the Angiosperm Phylogeny Group¹⁸ (Extended Data Fig. 1). For example, 58 of the 64 now accepted orders and 406 of the 416 families are recovered as monophyletic (excluding artefacts; Supplementary Table 1). The most striking exception is the non-monophyly of Asteraceae, the largest angiosperm family comprising the sunflowers and their relatives. Our tree also confirms 85% of the relationships among families recovered by ref. 4 using plastid genomes (Supplementary Fig. 4).

The overall stability of established relationships is unevenly distributed across the tree, as observed in contrasting patterns in the main eudicot clades, the asterids and rosids, which account for 35% and 29% of angiosperm diversity, respectively². The relationships among main orders of asterids are stable⁹, with a clade comprising Ericales and Cornales sister to all other asterids and the remaining 15 orders divided in two main clades (campanulids and lamiids), both long characterized by their contrasting floral ontogeny²³. Relationships contrasting with the status quo are mostly restricted to small orders, such as the paraphyly of Aquifoliales, Bruniales and Icaciniales. These DNA-defined orders were consistently recovered as highly supported clades in plastome analyses^{4,24} but they lack morphological cohesion. Given their placement in our phylogenetic tree and unique morphologies, these changes, although small, will alter our understanding of the evolution of asterids.

By contrast to asterids, our findings in rosids conflict markedly with plastid-based evidence. First, we resolve Saxifragales, rather than

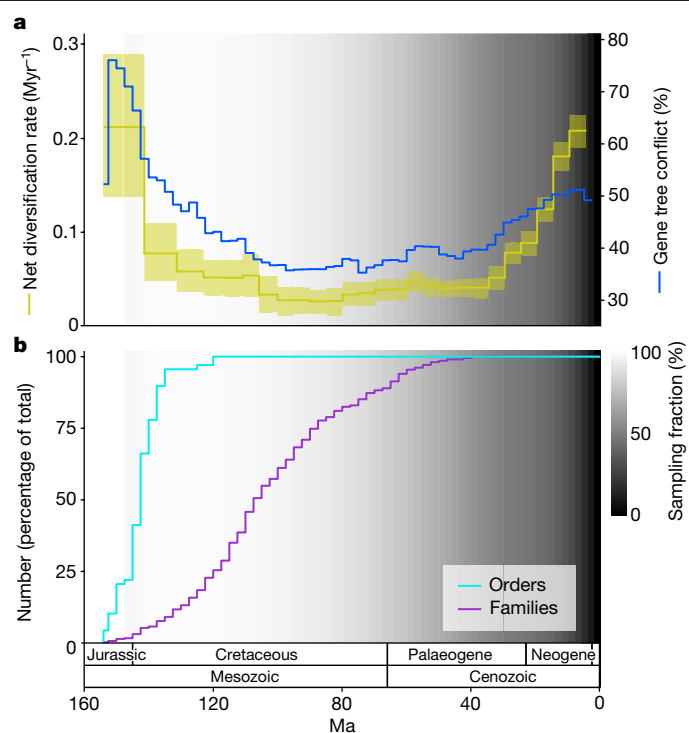


Fig. 3 | Angiosperm-wide diversification and gene tree conflict through time. The results illustrated are based on the young tree (maximum constraint at the root node of 154 Ma). See Extended Data Fig. 5 for results based on the old tree. **a**, Estimated net diversification rate through time (yellow, left y axis) and the level of gene tree conflict through time (blue, right y axis). Net diversification rates are estimated with a model that enables speciation rates to vary between time intervals; the line is the posterior mean and the yellow shaded area is the 95% highest posterior density. Gene tree conflict is calculated from the percentage of gene trees that do not share a congruent bipartition with each species tree branch, with the plotted value being the mean across all species tree branches that cross each 2.5 Myr time slice. **b**, Cumulative percentage of extant orders and families that have originated through time. In both **a** and **b**, the background grey-scale gradient is the estimated percentage of extant lineages represented in the species tree through time (sampling fraction).

Vitales⁴, as sister to the remainder of rosids. In rosids, the fabid and malvid subclades, recovered as reciprocally monophyletic by plastid data^{4,22}, are substantially rearranged into a grade of four orders subtending two well-supported sister clades, which we designate here as the recircumscribed fabids and malvids. The new fabid clade (Cucurbitales, Fabales, Fagales and Rosales) has long been characterized by symbiotic nitrogen fixation²⁵. In the new malvids (Brassicales, Celastrales, Huerteales, Malpighiales, Malvales, Oxalidales, Picramniales and Sapindales), Oxalidales is resolved as two independent lineages, the core emerging closer to Brassicales, Malvales and Sapindales, whereas Huaceae emerges in the position conventionally occupied by Oxalidales, that is, closer to Malpighiales and Celastrales (the former Celastrales–Oxalidales–Malpighiales (COM) clade¹⁸).

Notwithstanding the many well-supported confirmatory and new findings, some key relationships remain contentious and cannot be resolved by our data. These areas of high gene tree conflict often coincide with biological processes that confound phylogenetic inference. For example, the uncertain placements of eudicot orders Caryophyllales, Dilleniales and Gunnerales are probably impacted by key whole genome duplications^{9,26}. The poor support for relationships among magnoliids, monocots, eudicots and Ceratophyllales might be explained by ancient hybridization events, such as that recently proposed for the origin of the monocots²⁷. These examples highlight

the importance of areas of poor resolution as waymarkers to biological events meriting further study.

Time frame for angiosperm macroevolution

Our tree was analysed in combination with a dataset of 200 fossil calibrations (originally described in ref. 5, with modifications) to estimate divergence times and rates of diversification. Because the age of angiosperms is uncertain²⁸, we dated the tree with two different maximum constraints at the angiosperm crown node (154 and 247 million years ago (Ma), termed the young tree and old tree, respectively), which reflect realistic upper and lower bounds for the maximum age of this node^{5,28}. These different constraints affected age estimates across angiosperms (Extended Data Fig. 2, Supplementary Fig. 5 and Supplementary Table 3). For example, in the young tree, stem node age estimates for Nymphaeales, Austrobaileyales and Ceratophyllales were 153, 152 and 152 Ma, respectively, whereas in the old tree the equivalent age estimates were 245, 244 and 243 Ma. Likewise, for larger clades such as magnoliids, monocots and eudicots, crown node age estimates were 151, 149 and 151 Ma in the young tree and 238, 237 and 241 Ma in the old tree. This range in age estimates is consistent with the most comprehensive comparable study⁵ (Extended Data Fig. 3) but our trees provide age estimates for a further 7,000 nodes. In subsequent analyses, we indicate if differing age estimates between the young tree and old tree cause substantially different interpretations of angiosperm diversification.

With our sampling across angiosperms, we ensured that deeper branching events leading to extant lineages are comprehensively represented, while recognizing that extinct lineages are inaccessible to genomic methods. However, our dated trees are sparsely sampled at the species-level, meaning that branching events are incompletely represented towards the present, limiting diversification inferences in that time window. To address this, we developed a simulation-based approach to quantify the sampling fraction through time. For both dated trees, the lineage representation begins to drop substantially (below 75%) around 50 Ma (Supplementary Fig. 6). However, the most dramatic fall in lineage representation occurs in the most recent 20 Myr, in which it falls from around 50% to slightly more than 1% at present. Our investigation of angiosperm diversification should be interpreted with this broader context in mind. In particular, inferences in the most recent 20 Myr may be updated in the future with denser species sampling.

The diversification of angiosperms

Diversification linked to gene conflict

We used our dated trees to reconstruct both diversification and gene tree conflict across a broad range of temporal and phylogenetic scales and investigate the relationship between them. We show that throughout angiosperm macroevolution, elevated gene tree conflict was tightly associated with elevated diversification. At a general level, this relationship is visible by simply comparing estimated diversification rates with gene tree conflict across all angiosperms through time (Fig. 3a). Meanwhile, in a branch-specific analysis using the temporal duration of branches as a proxy for the rate at which branches are diversifying, we also show that conflict and diversification rate are positively correlated (Extended Data Fig. 4) ($P < 0.001$, $r^2 = 0.51$).

To characterize the theoretical basis of this relationship, we simulated species trees with corresponding gene trees under different diversification scenarios in a multispecies coalescent framework. These simulations showed that gene tree conflict is positively correlated with diversification when caused by incomplete lineage sorting, assuming that effective population size is constant (Supplementary Fig. 7). Our empirical results are largely consistent with such a scenario. Other potential causes of gene tree conflict such as whole genome duplication

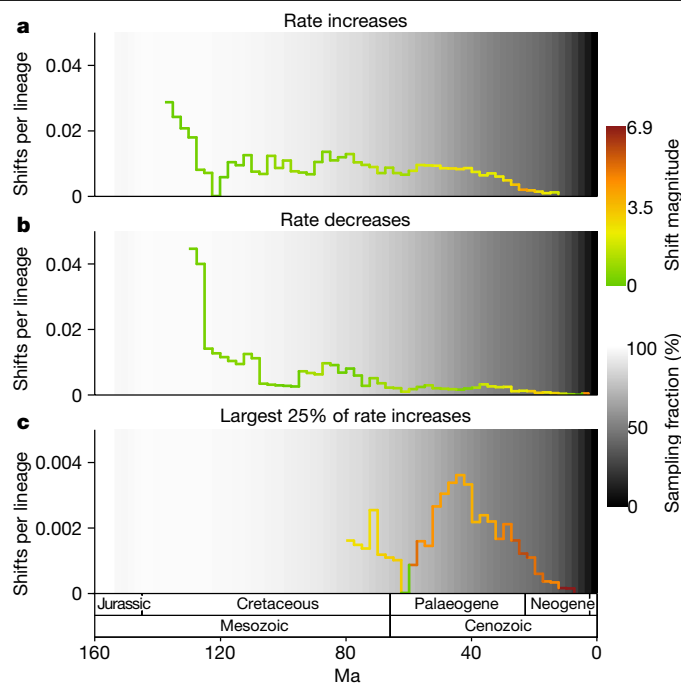


Fig. 4 | Summary of lineage-specific diversification rate shifts estimated by BAMM. The results illustrated are based on the young tree (maximum constraint at the root node of 154 Ma). See Extended Data Fig. 6 for results based on the old tree. **a**, Diversification rate increases per LTT. The colour corresponds to the average magnitude of the rate increases during the time period. **b**, Equivalent to **a** but for rate decreases. **c**, Equivalent to **a** but focusing on the largest 25% of diversification rate increases. In **a**, **b** and **c**, the number of shifts is from the maximum a posteriori shift configuration with the prior for the number of shifts set to 10 and the background grey-scale gradient is the estimated percentage of extant lineages represented in the species tree through time (sampling fraction).

and hybridization may also be associated with rapid diversification and have been recorded extensively throughout angiosperms^{29,30}. Overall, however, gene tree conflict seems to be reliable corroborating evidence for investigating temporal patterns of angiosperm diversification.

Early burst of angiosperm diversification

Our lineage-through-time (LTT) heatmap and diversification rate estimates through time both indicate an explosive early phase of diversification of extant lineages during the Late Jurassic and Early Cretaceous Periods (Fig. 2b and Fig. 3a). An early burst of angiosperm diversification, popularized as ‘Darwin’s abominable mystery’^{31,32}, is expected given the sudden emergence of diverse angiosperm fossils during the Early Cretaceous^{11,33–35}. Phylogenetic studies based on single or few genes have also implied that angiosperms diversified rapidly in the Early Cretaceous^{7,36–38}. Our dated tree corroborates the existence of a distinct early burst of diversification, associated with high levels of gene tree conflict (Fig. 3a and Supplementary Fig. 8), further increasing our confidence in this finding.

More than 80% of extant angiosperm orders originated during the early burst of diversification (Fig. 3b). Although not strictly comparable because of their subjective delimitation, orders represent the main components of angiosperm feature diversity, which have arisen rapidly after the crown node of angiosperms. In the young tree (Fig. 3), the early burst occurs during the Cretaceous, consistent with the hypothesis that a Cretaceous terrestrial revolution was triggered by the establishment of main angiosperm lineages^{14,39,40}. More controversially, the old tree places the early burst in the Triassic Period (Extended Data Fig. 5), which is dramatically at variance with the palaeobotanical record^{33,34},

highlighting that current molecular dating methods are unable to resolve the age of angiosperms²⁸.

A tapestry of lineage-specific histories

Following the early burst, overall rates of diversification across angiosperms continued at a lower, constant pace for at least 80 Myr (Fig. 3a), during which time around three-quarters of all families originated (Fig. 3b). As expected, this phase of slower diversification was associated with lower levels of gene tree conflict. Despite the constancy of overall rates, diversification during this period was underpinned by a complex tapestry of lineage-specific patterns. This is illustrated by the LTT heatmap, which shows profound differences in diversification trajectories among orders (Fig. 2) and by the estimation of around 160 lineage-specific diversification rate shifts in angiosperms, most of which occur during this period. These rate shifts have a widespread phylogenetic distribution, with most orders containing at least one rate shift and many containing several nested shifts (Supplementary Table 4). The importance of nested rate shifts is highlighted extensively in discussions of evolutionary radiation^{41,42} and underpins the continual response of diversification to dynamic extrinsic and intrinsic conditions. However, because these rate shifts are temporally scattered, as also shown by ref. 43, they do not lead to observable global rate shifts across angiosperms.

A Cenozoic diversification surge

A second surge in angiosperm diversification occurred during the Cenozoic Era (Fig. 3a). The occurrence of this surge, despite the already high standing diversity of angiosperms at the time, suggests that diversification was unaffected by diversity-dependent processes, that is, the filling of available niche space as clades diversify⁴⁴. Instead, this finding is consistent with previously proposed positive feedbacks between increased diversity and increased rates of diversification in angiosperms¹⁴, alongside more positive feedbacks, for example, between angiosperm and insect diversification^{45,46}. Alternatively, global climatic cooling during the Cenozoic acting as a driver of angiosperm diversification could explain this finding^{7,47–49}. Importantly, an even larger Cenozoic surge would probably be inferred with increased sampling that addresses the under-representation of branching events in the recent time window. The temporal distribution of lineage-specific diversification rate shifts may offer some insight into the cause of the Cenozoic surge. Many of the largest diversification rate increases occur during the Cenozoic, whereas the number of diversification rate decreases declines markedly during this period (Fig. 4). These large rate increases may underpin the Cenozoic surge. The expansion of taxon sampling should be given priority to confirm these patterns.

Synthesis

The nuclear phylogenomic framework presented here is the result of an ongoing initiative to complete the tree of life for all angiosperm genera⁵⁰, a milestone in our understanding of angiosperm evolutionary relationships. This study not only sheds light on much of the deep diversification history of the angiosperms but also lays foundations for future work towards a species-level tree⁵⁰. The standardized panel of nuclear genes in our dataset paves the way for more collaborations and data integration^{17,51}, while the open availability of universal tools to sequence them (that is, Angiosperms353 probes⁸) has made nuclear genomic data more accessible at relatively low cost. The accelerating uptake of this approach^{52–54}, which is readily applicable to herbarium collections¹⁶, indicates that large volumes of data will soon become available for a wide range of applications in plant diversity, systematic and macroevolutionary research.

Our fossil-calibrated, phylogenomic tree enables a range of unique insights into broad-scale diversification dynamics of angiosperms, substantiating the early burst of diversification anticipated by Darwin

while illuminating the complexity and conflict in the lineage histories underlying it. This sets the scene for future research, extending these investigations to shallower phylogenetic scales or digging more deeply into the data to discover the processes driving angiosperm diversification, such as genomic conflict, polyploidy, selection, trait evolution and adaptation. The challenges brought by the scale of this dataset and its ongoing expansion may also catalyse the development of methods which take full advantage of the global proliferation of genomic data.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-024-07324-0>.

- Diamond, J. Evolution, consequences and future of plant and animal domestication. *Nature* **418**, 700–707 (2002).
- Govaerts, R., Nic Lughadha, E., Black, N., Turner, R. & Paton, A. The World Checklist of Vascular Plants, a continuously updated resource for exploring global plant diversity. *Sci. Data* **8**, 215 (2021).
- Chase, M. W. et al. Phylogenetics of seed plants: an analysis of nucleotide sequences from the plastid gene *rbcL*. *Ann. Missouri Bot. Gard.* **80**, 528–580 (1993).
- Li, H.-T. et al. Plastid phylogenomic insights into relationships of all flowering plant families. *BMC Biol.* **19**, 232 (2021).
- Ramírez-Barahona, S., Sauquet, H. & Magallón, S. The delayed and geographically heterogeneous diversification of flowering plant families. *Nat. Ecol. Evol.* **4**, 1232–1238 (2020).
- Li, H.-T. et al. Origin of angiosperms and the puzzle of the Jurassic gap. *Nat. Plants* **5**, 461–470 (2019).
- Dimitrov, D. et al. Diversification of flowering plants in space and time. *Nat. Commun.* **14**, 7609 (2023).
- Johnson, M. G. et al. A universal probe set for targeted sequencing of 353 nuclear genes from any flowering plant designed using k-medoids clustering. *Syst. Biol.* **68**, 594–606 (2019).
- One Thousand Plant Transcriptomes Initiative. One thousand plant transcriptomes and the phylogenomics of green plants. *Nature* **574**, 679–685 (2019).
- Barba-Montoya, J., dos Reis, M., Schneider, H., Donoghue, P. C. J. & Yang, Z. Constraining uncertainty in the timescale of angiosperm evolution and the veracity of a Cretaceous Terrestrial Revolution. *New Phytol.* **218**, 819–834 (2018).
- Doyle, J. A. Molecular and fossil evidence on the origin of angiosperms. *Annu. Rev. Earth Planet. Sci.* **40**, 301–326 (2012).
- Holbourn, A. E. et al. Late Miocene climate cooling and intensification of southeast Asian winter monsoon. *Nat. Commun.* **9**, 1584 (2018).
- Pennington, R. T., Cronk, Q. C. B. & Richardson, J. A. Introduction and synthesis: plant phylogeny and the origin of major biomes. *Philos. Trans. R. Soc. Lond. B* **359**, 1455–1464 (2004).
- Benton, M. J., Wilf, P. & Sauquet, H. The Angiosperm terrestrial revolution and the origins of modern biodiversity. *New Phytol.* **233**, 2017–2035 (2022).
- Dodsworth, S. et al. Hyb-Seq for flowering plant systematics. *Trends Plant Sci.* **24**, 887–891 (2019).
- Brewer, G. E. et al. Factors affecting targeted sequencing of 353 nuclear genes from herbarium specimens spanning the diversity of angiosperms. *Front. Plant Sci.* **10**, 1102 (2019).
- Baker, W. J. et al. Exploring Angiosperms353: an open, community toolkit for collaborative phylogenomic research on flowering plants. *Am. J. Bot.* **108**, 1059–1065 (2021).
- The Angiosperm Phylogeny Group. et al. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV. *Bot. J. Linn. Soc.* **181**, 1–20 (2016).
- Joyce, E. M. et al. Phylogenomic analyses of Sapindales support new family relationships, rapid Mid-Cretaceous Hothouse diversification and heterogeneous histories of gene duplication. *Front. Plant Sci.* **14**, 1063174 (2023).
- Yan, Z., Smith, M. L., Du, P., Hahn, M. W. & Nakhleh, L. Species tree inference methods intended to deal with incomplete lineage sorting are robust to the presence of paralogs. *Syst. Biol.* **71**, 367–381 (2022).
- Watanabe, T., Kure, A. & Horiike, T. OrthoPhy: a program to construct ortholog data sets using taxonomic information. *Genome Biol. Evol.* **15**, evad026 (2023).
- Ruhfel, B. R., Gitzendanner, M. A., Soltis, P. S., Soltis, D. E. & Burleigh, J. G. From algae to angiosperms—inferring the phylogeny of green plants (Viridiplantae) from 360 plastid genomes. *BMC Evol. Biol.* **14**, 23 (2014).
- Endress, P. K. Origins of flower morphology. *J. Exp. Zool.* **291**, 105–115 (2001).
- Stull, G. W., Duno de Stefano, R., Soltis, D. E. & Soltis, P. S. Resolving basal lamiid phylogeny and the circumscription of Icacinaeaceae with a plastome-scale data set. *Am. J. Bot.* **102**, 1794–1813 (2015).
- Soltis, D. E. et al. Chloroplast gene sequence data suggest a single origin of the predisposition for symbiotic nitrogen fixation in angiosperms. *Proc. Natl Acad. Sci. USA* **92**, 2647–2651 (1995).
- Walker, J. F. et al. From cacti to carnivores: improved phylotranscriptomic sampling and hierarchical homology inference provide further insight into the evolution of Caryophyllales. *Am. J. Bot.* **105**, 446–462 (2018).

27. Guo, X. et al. Chloranthus genome provides insights into the early diversification of angiosperms. *Nat. Commun.* **12**, 6930 (2021).
28. Sauquet, H., Ramírez-Barahona, S. & Magallón, S. What is the age of flowering plants? *J. Exp. Bot.* **73**, 3840–3853 (2022).
29. Landis, J. B. et al. Impact of whole-genome duplication events on diversification rates in angiosperms. *Am. J. Bot.* **105**, 348–363 (2018).
30. Tank, D. C. et al. Nested radiations and the pulse of angiosperm diversification: increased diversification rates often follow whole genome duplications. *New Phytol.* **207**, 454–467 (2015).
31. Darwin, C. *The Correspondence of Charles Darwin* (Cambridge Univ. Press, 1879).
32. Buggs, R. J. A. Reconfiguring Darwin's abominable mystery. *Nat. Plants* **8**, 194–195 (2022).
33. Coiro, M., Doyle, J. A. & Hilton, J. How deep is the conflict between molecular and fossil evidence on the age of angiosperms? *New Phytol.* **223**, 83–99 (2019).
34. Herendeen, P. S., Friis, E. M., Pedersen, K. R. & Crane, P. R. Palaeobotanical redux: revisiting the age of the angiosperms. *Nat. Plants* **3**, 17015 (2017).
35. Friis, E. M., Crane, P. R. & Pedersen, K. R. *Early Flowers and Angiosperm Evolution* (Cambridge Univ. Press, 2011).
36. Davies, T. J. et al. Darwin's abominable mystery: insights from a supertree of the angiosperms. *Proc. Natl Acad. Sci. USA* **101**, 1904–1909 (2004).
37. Magallón, S., Gómez-Acevedo, S., Sánchez-Reyes, L. L. & Hernández-Hernández, T. A metacalibrated time-tree documents the early rise of flowering plant phylogenetic diversity. *New Phytol.* **207**, 437–453 (2015).
38. Mathews, S. & Donoghue, M. J. The root of angiosperm phylogeny inferred from duplicate phytochrome genes. *Science* **286**, 947–950 (1999).
39. Dilcher, D. Toward a new synthesis: major evolutionary trends in the angiosperm fossil record. *Proc. Natl Acad. Sci. USA* **97**, 7030–7036 (2000).
40. Meredith, R. W. et al. Impacts of the Cretaceous terrestrial revolution and KPg extinction on mammal diversification. *Science* **334**, 521–524 (2011).
41. Bouchenak-Khelladi, Y., Onstein, R. E., Xing, Y., Schwery, O. & Linder, H. P. On the complexity of triggering evolutionary radiations. *New Phytol.* **207**, 313–326 (2015).
42. Donoghue, M. J. & Sanderson, M. J. Confluence, synnovation and depauperons in plant diversification. *New Phytol.* **207**, 260–274 (2015).
43. Magallón, S., Sánchez-Reyes, L. L. & Gómez-Acevedo, S. L. Thirty clues to the exceptional diversification of flowering plants. *Ann. Bot.* **123**, 491–503 (2019).
44. Rabosky, D. L. Diversity-dependence, ecological speciation and the role of competition in macroevolution. *Annu. Rev. Ecol. Evol. Syst.* **44**, 481–502 (2013).
45. Asar, Y., Ho, S. Y. W. & Sauquet, H. Early diversifications of angiosperms and their insect pollinators: were they unlinked? *Trends Plant Sci.* **27**, 858–869 (2022).
46. Peris, D. & Condamine, F. L. The dual role of the angiosperm radiation on insect diversification. *Nat. Commun.* **15**, 552 (2024).
47. Folk, R. A. et al. Rates of niche and phenotype evolution lag behind diversification in a temperate radiation. *Proc. Natl Acad. Sci. USA* **116**, 10874–10882 (2019).
48. Sun, M. et al. Recent accelerated diversification in rosids occurred outside the tropics. *Nat. Commun.* **11**, 3333 (2020).
49. Soltis, P. S., Folk, R. A. & Soltis, D. E. Darwin review: angiosperm phylogeny and evolutionary radiations. *Proc. R. Soc. B* **286**, 20190099 (2019).
50. Baker, W. J. et al. A comprehensive phylogenomic platform for exploring the angiosperm tree of life. *Syst. Biol.* **71**, 301–319 (2022).
51. McDonnell, A. J. et al. Exploring Angiosperms353: developing and applying a universal toolkit for flowering plant phylogenomics. *Appl. Plant Sci.* **9**, e11443 (2021).
52. Bratzel, F. et al. Target-enrichment sequencing reveals for the first time a well-resolved phylogeny of the core Bromelioideae (family Bromeliaceae). *TAXON* **72**, 47–63 (2023).
53. Gagnon, E. et al. Phylogenomic discordance suggests polytomies along the backbone of the large genus *Solanum*. *Am. J. Bot.* **109**, 580–601 (2022).
54. Murillo-A, J., Valencia-D, J., Orozco, C. I., Parra-O, C. & Neubig, K. M. Incomplete lineage sorting and reticulate evolution mask species relationships in Brunelliaceae, an Andean family with rapid, recent diversification. *Am. J. Bot.* **109**, 1139–1156 (2022).
55. Zuntini, A. R. & Carruthers, T. Phylogenomics and the rise of the angiosperms. Zenodo <https://doi.org/10.5281/zenodo.10778206> (2024).
- Toral Shah¹, Juan Viruel¹, Theodore R. Allnutt²⁰, Gabriel K. Ameka²¹, Rose L. Andrew²², Marc S. Appelhans²³, Montserrat Arista²⁴, María Jesús Ariza²⁵, Juan Arroyo²⁴, Watchara Arthan¹, Julien B. Bachelier²⁶, C. Donovan Bailey²⁷, Helen F. Barnes²⁰, Matthew D. Barrett³, Russell L. Barrett²⁸, Randall J. Bayer²⁹, Michael J. Bayly³⁰, Ed Biffin³¹, Nicky Biggs¹, Joanne L. Birch³⁰, Diego Bogarin^{14,32}, Renata Borosova¹, Alexander M. C. Bowles³³, Peter C. Boyce³⁴, Gemma L. C. Bramley¹, Marie Briggs¹, Linda Broadhurst³⁵, Gillian K. Brown³⁶, Jeremy J. Bruhl³⁷, Anne Bruneau³⁷, Sven Buerki³⁸, Edie Burns¹, Margaret Byrne³⁹, Stuart Cable¹, Ainsley Calladine³¹, Martin W. Callmander⁴⁰, Ángela Cano⁴¹, David J. Cantrill²⁰, Warren M. Cardinal-McTeague⁴², Mónica M. Carlsen⁴⁰, Abigail J. A. Carruthers¹, Alejandra de Castro Mateo²⁴, Mark W. Chase¹⁴⁴, Lars W. Chatrou⁴⁵, Martin Cheek¹, Shilin Chen^{46,47}, Maarten J. M. Christenhusz^{148,49}, Pascal-Antoine Christin⁵⁰, Mark A. Clements³⁵, Skye C. Coffey⁵¹, John G. Conran⁵², Xavier Cornejo⁵³, Thomas L. P. Couvreur⁵⁴, Ian D. Cowie⁵⁵, Laszlo Csiba¹, Iain Darbyshire¹, Gerrit Davidse⁴³, Nina M. J. Davies¹, Aaron P. Davis¹, Kor-jent van Dijk⁵⁶, Stephen R. Downie⁵⁷, Marco F. Duretto⁵⁸, Melvin R. Duvall⁵⁸, Sara L. Edwards¹, Urs Eggli⁵⁹, Roy H. J. Erkens^{14,60,61}, Marcial Escudero²⁴, Manuel de la Estrella⁶², Federico Fabiani⁴⁵, Michael F. Fay¹, Paola de L. Ferreira^{63,64}, Sarah Z. Ficinski¹, Rachael M. Fowler³⁰, Susie Frisby¹, Lin Fu⁶⁵, Tim Fulcher¹, Mercè Galbany-Casals⁶⁶, Elliot M. Gardner⁶⁷, Dmitry A. German⁶⁸, Augusto Giarretta⁶⁹, Marc Gibernau⁷⁰, Lynn J. Gillespie⁷¹, Cynthia C. González⁷², David J. Goyder¹, Sean W. Graham¹⁷, Aurélie Grall¹, Laura Green¹, Bee F. Gunn²⁰, Diego G. Gutiérrez⁷³, Jan Hackett^{1,74}, Thomas Haevermans⁷⁵, Anna Haigh¹, Jocelyn C. Hall⁷⁶, Tony Hall¹, Melissa J. Harrison³, Sebastian A. Hatt¹, Oriane Hidalgo⁷⁷, Trevor R. Hodgkinson⁷⁸, Gareth D. Holmes²⁰, Helen C. F. Hopkins¹, Christopher J. Jackson²⁰, Shelley A. James⁵¹, Richard W. Jobson²⁸, Gudrun Kadereit⁷⁹, Imalka M. Kahandawala¹, Kent Kainulainen⁸⁰, Masahiro Kato⁸¹, Elizabeth A. Kellogg⁸², Graham J. King⁸³, Beata Klejevskaite⁸⁴, Bente B. Klitgaard¹, Ronell R. Klopffer^{85,86}, Sandra Knapp⁸⁷, Marcus A. Koch⁸⁸, James H. Leebens-Mack⁸⁹, Frederic Lens¹⁴, Christine J. Leon¹, Étienne Lévillé-Bourret⁹⁰, Gwilym P. Lewis¹, De-Zhu Li¹⁹, Lan Li⁹¹, Sigrid Liede-Schumann⁹², Tatjana Livshultz^{93,94}, David Lorence⁹⁵, Meng Lu¹, Patricia Lu-Irving²⁸, Jaqueline Luber⁹⁶, Eve J. Lucas¹, Manuel Luján¹, Mabel Lum⁹⁷, Terry D. Macfarlane⁹⁸, Carlos Magdalena¹, Vidal F. Mansano⁹⁶, Lizo E. Masters¹, Simon J. Mayo¹, Kristina McColl²⁸, Angela J. McDonnell⁹⁸, Andrew E. McDougall⁹⁶, Todd G. B. McLay²⁰, Hannah McPherson²⁸, Rosa I. Meneses⁹⁹, Vincent S. F. D. Merckx¹⁴, Fabián A. Michelangeli¹⁰⁰, John D. Mitchell¹⁰⁰, Alexandre K. Monro¹, Michael J. Moore¹⁰¹, Taryn L. Mueller¹⁰², Klaus Mummenhoff¹³, Jérôme Munzinger¹⁰³, Priscilla Muriel¹⁰⁴, Daniel J. Murphy²⁰, Katharina Nargar^{3,35}, Lars Nauheimer³, Francis J. Nge³¹, Reto Nyffeler¹⁰⁵, Andrés Orejuela^{106,107}, Edgardo M. Ortiz¹⁵, Luis Palazzesi⁷³, Ariane Luna Peixoto⁹⁶, Susan K. Pell¹⁰⁸, Jaume Pellicer⁷⁷, Darin S. Penneys¹⁰⁹, Oscar A. Perez-Escobar¹, Claes Persson¹¹⁰, Marc Pignal⁷⁵, Yohan Pillon¹¹¹, José R. Pirani¹⁰, Gregory M. Plunkett¹⁰⁰, Robyn F. Powell¹¹, Ghilleen T. Prance¹, Carmen Puglisi⁴³, Ming Qin⁶⁵, Richard K. Rabaler¹⁸, Paul E. J. Rees¹, Matthew Renner²⁸, Eric H. Roalson¹¹², Michele Rodda¹¹³, Zachary S. Rogers¹¹⁴, Saba Rokni¹, Rolf Rutishauser¹⁰⁵, Miguel F. de Salas¹¹⁵, Hanno Schaefer¹⁵, Rowan J. Schley¹¹⁶, Alexander Schmidt-Lebuhn³⁵, Alison Shapcott¹¹⁷, Ihsan Al-Shehbaz¹³, Kelly A. Shepherd⁵¹, Mark P. Simmons¹¹⁸, André O. Simões¹¹⁹, Ana Rita G. Simões¹, Michelle Siros¹¹²⁰, Eric C. Smidt¹²¹, James F. Smith³⁸, Neil Snow¹²², Douglas E. Soltis¹²³, Pamela S. Soltis¹²³, Robert J. Soreng¹²⁴, Cynthia A. Sothers¹, Julian R. Starr¹²⁵, Peter F. Stevens⁴³, Shannon C. K. Straub¹²⁶, Lena Struwe¹²⁷, Jennifer M. Taylor⁹¹, Ian R. H. Telford¹²², Andrew H. Thornhill^{22,31,52}, Ifeana Tooth²⁸, Anna Trias-Blasi¹, Frank Udovicic²⁰, Timothy M. A. Utteridge¹, Jose C. Del Valle²⁴, G. Anthony Verboom¹²⁸, Helen P. Vonow³¹, Maria S. Vorontsova¹, Jurriaan M. de Vos¹²⁹, Noor Al-Wattar¹, Michelle Waycott^{31,52}, Cassiano A. D. Welker¹³⁰, Adam J. White¹³¹, Jan J. Wieringa¹⁴, Luis T. Williamson³⁶, Trevor C. Wilson²⁸, Sin Yeng Wong¹³², Lisa A. Woods²⁸, Roseina Woods¹, Stuart Worboys³, Martin Xanthos¹, Ya Yang¹³³, Yu-Xiao Zhang¹³⁴, Meng-Yuan Zhou¹⁹, Sue Zmarzty¹, Fernando O. Zuloaga¹³⁵, Alexandre Antonelli^{110,136,137}, Sidonie Bellot¹, Darren M. Crayn³, Olwen M. Grace¹⁰⁶, Paul J. Kersey¹, Ilija J. Leitch¹, Hervé Sauquet²⁸, Stephen A. Smith^{18,139}, Wolf L. Eisehardt^{1,64,139}, Félix Forest¹¹³⁹ & William J. Baker^{1,64,139}✉

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024

Alexandre R. Zuntini¹¹³⁸, Tom Carruthers¹¹³⁸, Olivier Maurin¹, Paul C. Bailey¹, Kevin Leempoel¹, Grace E. Brewer¹, Niroshini Epitawalage¹, Elaine Francisco¹², Berta Gallego-Paramo¹, Catherine McGinnie¹, Raquel Negrão¹, Shyamali R. Roy¹, Lalita Simpson³, Eduardo Toledo Romero¹, Vanessa M. A. Barber¹, Laura Botigué⁴, James J. Clarkson¹, Robyn S. Cowan¹, Steven Dodsworth³, Matthew G. Johnson⁹, Jan T. Kim⁷, Lisa Pokorny¹⁸, Norman J. Wickett⁹, Guilherme M. Anta^{10,31}, Lucinda DeBolt¹², Karime Gutierrez¹², Kasper P. Hendriks¹³¹⁴, Alina Hoewener¹⁵, Ai-Qun Hu¹, Elizabeth M. Joyce³¹⁸, Izai A. B. S. Kikuchi¹⁷, Isabel Larridon¹, Drew A. Larson¹⁸, Elton John de Lirio¹⁰, Jing-Xia Liu¹⁹, Panagiota Malakasi¹, Natalia A. S. Przelomska¹⁵,

¹Royal Botanic Gardens, Kew, Richmond, UK. ²Centre for Ecology, Evolution and Behaviour, Department of Biological Sciences, School of Life Sciences and the Environment, Royal Holloway University of London, London, UK. ³Australian Tropical Herbarium, James Cook University, Smithfield, Queensland, Australia. ⁴Centre for Research in Agricultural Genomics (CRAG), CSIC-IRTA-UAB-UB, Campus UAB, Barcelona, Spain. ⁵School of Biological Sciences, University of Portsmouth, Portsmouth, UK. ⁶Texas Tech University, Lubbock, TX, USA. ⁷School of Physics, Engineering and Computer Science, University of Hertfordshire, Hatfield, UK. ⁸Department of Biodiversity and Conservation, Real Jardín Botánico (RJB-CSIC), Madrid, Spain. ⁹Department of Biological Sciences, Clemson University, Clemson, SC, USA. ¹⁰Departamento de Botânica, Instituto de Biociências, Universidade de São Paulo, São Paulo, Brazil. ¹¹Departamento de Ciências Agrárias e Biológicas, Centro Universitário Norte do Espírito Santo, Universidade Federal do Espírito Santo, São Mateus, Brazil. ¹²Smith College, Northampton, MA, USA. ¹³Department of Biology, University of Osnabrück, Osnabrück, Germany. ¹⁴Naturalis Biodiversity Center, Leiden, The Netherlands. ¹⁵Plant Biodiversity, Technical University Munich, Freising, Germany. ¹⁶Systematic, Biodiversity and Evolution of Plants, Ludwig Maximilian University of Munich, Munich, Germany. ¹⁷Department of Botany, University of British Columbia, Vancouver, British Columbia, Canada. ¹⁸Department of Ecology & Evolutionary Biology, University of Michigan, Ann Arbor, MI, USA. ¹⁹Germplasm Bank of Wild Species, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming, China. ²⁰Royal Botanic Gardens Victoria, Melbourne, Victoria, Australia. ²¹Department of Plant and Environmental Biology, University of Ghana, Accra, Ghana. ²²Botany and N.C.W. Beadle Herbarium, University of New England, Armidale, New South Wales, Australia. ²³Department of Systematics, Biodiversity and Evolution of Plants, Albrecht-von-Haller Institute of Plant Sciences, University of Göttingen, Göttingen, Germany. ²⁴Departamento de Biología Vegetal y Ecología, Facultad de Biología, Universidad de Sevilla, Seville, Spain. ²⁵General Research Services, Herbario SEV, CITIUS, Universidad de Sevilla, Seville, Spain. ²⁶Institute of Biology,

Freie Universität, Berlin, Germany. ²⁷Department of Biology, New Mexico State University, Las Cruces, NM, USA. ²⁸National Herbarium of NSW, Botanic Gardens of Sydney, Mount Annan, New South Wales, Australia. ²⁹Department of Biological Sciences, University of Memphis, Memphis, TN, USA. ³⁰School of BioSciences, The University of Melbourne, Parkville, Victoria, Australia. ³¹State Herbarium of South Australia, Botanic Gardens and State Herbarium, Adelaide, South Australia, Australia. ³²Jardín Botánico Lankester, Universidad de Costa Rica, Cartago, Costa Rica. ³³School of Geographical Sciences, University of Bristol, Bristol, UK. ³⁴Centro Studi Erbario Tropicale, Dipartimento di Biologia, University of Florence, Florence, Italy. ³⁵Centre for Australian National Biodiversity Research, National Research Collections Australia, CSIRO, Canberra, Australian Capital Territory, Australia. ³⁶Queensland Herbarium and Biodiversity Science, Brisbane Botanic Gardens, Toowoong, Queensland, Australia. ³⁷Institut de Recherche en Biologie Végétale and Département de Sciences Biologiques, University of Montreal, Montreal, Quebec, Canada. ³⁸Department of Biological Sciences, Boise State University, Boise, ID, USA. ³⁹Biodiversity and Conservation Science, Department of Biodiversity, Conservation and Attractions, Government of Western Australia, Kensington, Western Australia, Australia. ⁴⁰Conservatoire et Jardin Botaniques de Genève, Chambésy, Switzerland. ⁴¹Cambridge University Botanic Garden, Cambridge, UK. ⁴²Department of Forest and Conservation Sciences, University of British Columbia, Vancouver, British Columbia, Canada. ⁴³Missouri Botanical Garden, St. Louis, MO, USA. ⁴⁴Department of Environment and Agriculture, Curtin University, Bentley, Western Australia, Australia. ⁴⁵Department of Biology, Ghent University, Ghent, Belgium. ⁴⁶Institute of Herbgonomics, Chengdu University of Traditional Chinese Medicine, Chengdu, China. ⁴⁷Institute of Medicinal Plant Development, Chinese Academy of Medical Sciences, Beijing, China. ⁴⁸Department of Environment and Agriculture, Curtin University, Perth, Western Australia, Australia. ⁴⁹Plant Gateway, Den Haag, The Netherlands. ⁵⁰Ecology and Evolutionary Biology, School of Biosciences, University of Sheffield, Sheffield, UK. ⁵¹Western Australian Herbarium, Department of Biodiversity, Conservation and Attractions, Government of Western Australia, Kensington, Western Australia, Australia. ⁵²School of Biological Sciences, The University of Adelaide, Adelaide, South Australia, Australia. ⁵³Herbario GUAY, Facultad de Ciencias Naturales, Universidad de Guayaquil, Guayaquil, Ecuador. ⁵⁴DIADÉ, Université Montpellier, CIRAD IRD, Montpellier, France. ⁵⁵Northern Territory Herbarium Department of Environment Parks & Water Security, Northern Territory Government, Palmerston, Northern Territory, Australia. ⁵⁶The University of Adelaide, North Terrace Campus, Adelaide, South Australia, Australia. ⁵⁷Department of Plant Biology, University of Illinois at Urbana-Champaign, Urbana, IL, USA. ⁵⁸Department of Biological Sciences and Institute for the Study of the Environment, Sustainability and Energy, Northern Illinois University, DeKalb, IL, USA. ⁵⁹Sukkulanten-Sammlung Zürich/ Grün Stadt Zürich, Zürich, Switzerland. ⁶⁰Maastricht Science Programme, Maastricht University, Maastricht, The Netherlands. ⁶¹System Earth Science, Maastricht University, Venlo, The Netherlands. ⁶²Departamento de Botánica, Ecología y Fisiología Vegetal, Facultad de Ciencias, Universidad de Córdoba, Córdoba, Spain. ⁶³Departamento de Biología, Faculdade de Ciências e Letras de Ribeirão Preto, Universidade de São Paulo, São Paulo, Brazil. ⁶⁴Department of Biology, Aarhus University, Aarhus, Denmark. ⁶⁵South China Botanical Garden, Chinese Academy of Sciences, Guangzhou, China. ⁶⁶Systematics and Evolution of Vascular Plants (UAB)—Associated Unit to CSIC by IBB, Departament de Biologia Animal, Biologia Vegetal i Ecologia, Facultat de Biociències, Universitat Autònoma de Barcelona, Bellaterra, Spain. ⁶⁷Department of Biology, Case Western Reserve University, Cleveland, OH, USA. ⁶⁸Altai State University, Barnaul, Russia. ⁶⁹Faculdade de Ciências Biológicas e Ambientais, Universidade Federal da Grande Dourados, Dourados, Brazil. ⁷⁰Laboratoire Sciences Pour l'Environnement, Université de Corse, Ajaccio, France. ⁷¹Canadian Museum of Nature, Ottawa, Ontario, Canada. ⁷²Herbario Trelew, Universidad Nacional de la Patagonia San Juan Bosco, Trelew, Argentina. ⁷³Museo Argentino de Ciencias Naturales (MACN-CONICET), Buenos Aires, Argentina. ⁷⁴Department of Biology, Universität Marburg, Marburg, Germany. ⁷⁵Institut de Systématique, Evolution, Biodiversité, Muséum National d'Histoire Naturelle, Paris, France. ⁷⁶Department of Biological Sciences, University of Alberta, Edmonton, Alberta, Canada. ⁷⁷Institut Botànic de Barcelona (IBB CSIC-Ajuntament de Barcelona), Barcelona, Spain. ⁷⁸Botany, School of Natural Sciences, Trinity College Dublin, The University of Dublin,

Dublin, Ireland. ⁷⁹Prinzessin Therese von Bayern-Lehrstuhl für Systematik, Biodiversität & Evolution der Pflanzen, Ludwig-Maximilians-Universität München, Botanische Staatssammlung München, Botanischer Garten München-Nymphenburg, Munich, Germany. ⁸⁰Gothenburg Botanical Garden, Gothenburg, Sweden. ⁸¹National Museum of Nature and Science, Tsukuba, Japan. ⁸²Donald Danforth Plant Science Center, St. Louis, MO, USA. ⁸³Southern Cross University, Lismore, New South Wales, Australia. ⁸⁴Synergy SRG, Luton, UK. ⁸⁵Foundational Biodiversity Science Division, South African National Biodiversity Institute, Pretoria, South Africa. ⁸⁶Department of Plant and Soil Sciences, University of Pretoria, Pretoria, South Africa. ⁸⁷Natural History Museum, London, UK. ⁸⁸Centre for Organismal Studies, Biodiversity and Plant Systematics, Heidelberg University, Heidelberg, Germany. ⁸⁹Department of Plant Biology, University of Georgia, Athens, GA, USA. ⁹⁰Institut de Recherche en Biologie Végétale, University of Montreal, Montreal, Quebec, Canada. ⁹¹CSIRO, Canberra, Australian Capital Territory, Australia. ⁹²Department of Plant Systematics, University of Bayreuth, Bayreuth, Germany. ⁹³Department of Biodiversity, Earth and Environmental Sciences, Drexel University, Philadelphia, PA, USA. ⁹⁴Academy of Natural Science, Drexel University, Philadelphia, PA, USA. ⁹⁵National Tropical Botanical Garden, Kalaheo, HI, USA. ⁹⁶Instituto de Pesquisas Jardim Botânico do Rio de Janeiro, Rio de Janeiro, Brazil. ⁹⁷Bioplatforms Australia Ltd, Sydney, New South Wales, Australia. ⁹⁸Department of Biological Sciences, Saint Cloud State University, Saint Cloud, MN, USA. ⁹⁹Instituto de Arqueología y Antropología, Universidad Católica del Norte, San Pedro de Atacama, Chile. ¹⁰⁰New York Botanical Garden, Bronx, NY, USA. ¹⁰¹Department of Biology, Oberlin College, Oberlin, OH, USA. ¹⁰²Department of Ecology, Evolution & Behavior, University of Minnesota, St. Paul, MN, USA. ¹⁰³AMAP Lab, Université Montpellier, IRD, CIRAD, CNRS INRAE, Montpellier, France. ¹⁰⁴Laboratorio de Ecofisiología, Escuela de Ciencias Biológicas, Pontificia Universidad Católica del Ecuador, Quito, Ecuador. ¹⁰⁵Department of Systematic and Evolutionary Botany, University of Zürich, Zürich, Switzerland. ¹⁰⁶Royal Botanic Garden Edinburgh, Edinburgh, UK. ¹⁰⁷Grupo de Investigación en Recursos Naturales Amazónicos, Instituto Tecnológico del Putumayo, Mocoa, Colombia. ¹⁰⁸US Botanic Garden, Washington, DC, USA. ¹⁰⁹Department of Biology and Marine Biology, University of North Carolina Wilmington, Wilmington, NC, USA. ¹¹⁰Department of Biological and Environmental Sciences, University of Gothenburg, Gothenburg, Sweden. ¹¹¹LSTM Université Montpellier, CIRADIR, Montpellier, France. ¹¹²School of Biological Sciences, Washington State University, Pullman, WA, USA. ¹¹³National Parks Board, Singapore Botanic Gardens, Singapore, Singapore. ¹¹⁴New Mexico State University, Las Cruces, NM, USA. ¹¹⁵Tasmanian Herbarium, University of Tasmania, Sandy Bay, Tasmania, Australia. ¹¹⁶University of Exeter, Exeter, UK. ¹¹⁷School of Science Technology and Engineering, Center for Bioinnovation, University Sunshine Coast, Sippy Downs, Queensland, Australia. ¹¹⁸Department of Biology, Colorado State University, Fort Collins, CO, USA. ¹¹⁹Departamento de Biologia Vegetal, Universidade Estadual de Campinas, Campinas, Brazil. ¹²⁰University of California, San Francisco, San Francisco, CA, USA. ¹²¹Departamento de Botânica, Universidade Federal do Paraná, Curitiba, Brazil. ¹²²Pittsburg State University, Pittsburg, KS, USA. ¹²³Florida Museum of Natural History, University of Florida, Gainesville, FL, USA. ¹²⁴Smithsonian Institution, Washington, DC, USA. ¹²⁵Department of Biology, University of Ottawa, Ottawa, Ontario, Canada. ¹²⁶Hobart and William Smith Colleges, Geneva, NY, USA. ¹²⁷Rutgers University, New Brunswick, NJ, USA. ¹²⁸Department of Biological Sciences and Bolus Herbarium, University of Cape Town, Cape Town, South Africa. ¹²⁹Department of Environmental Sciences—Botany, University of Basel, Basel, Switzerland. ¹³⁰Instituto de Biologia, Universidade Federal de Uberlândia, Uberlândia, Brazil. ¹³¹Australian National Herbarium, Centre for Australian National Biodiversity Research, National Research Collections Australia, CSIRO, Canberra, Australian Capital Territory, Australia. ¹³²Institute of Biodiversity And Environmental Conservation, Universiti Malaysia Sarawak, Samarahan, Malaysia. ¹³³University of Minnesota-Twin Cities, St. Paul, MN, USA. ¹³⁴Southwest Forestry University, Kunming, China. ¹³⁵Instituto de Botánica Darwinion, San Isidro, Argentina. ¹³⁶Gothenburg Global Biodiversity Centre, University of Gothenburg, Gothenburg, Sweden. ¹³⁷Department of Biology, University of Oxford, Oxford, UK. ¹³⁸These authors contributed equally: Alexandre R. Zuntini, Tom Carruthers. ¹³⁹These authors jointly supervised this work: Stephen A. Smith, Wolf L. Eisehardt, Félix Forest, William J. Baker. [✉]e-mail: w.baker@kew.org

Methods

As part of the Plant and Fungal Trees of Life (PAFTOL) Project at the Royal Botanic Gardens, Kew⁵⁰, we assembled a nuclear genomic dataset consisting of newly generated data and data mined from public repositories. Our objective was to sample at least 50% of all angiosperm genera, with genera selected in a phylogenetically representative manner on the basis of published research. To avoid excessive imbalance in the tree, we included only one sample per species and a maximum of three species per genus. When several samples were available for the same species, we selected those with the largest amount of data, that is, more genes and a higher sum of gene length. For genera with several species available, the criterion for selection was primarily phylogenetic representation followed by amount of data. One species of each gymnosperm family was selected to form the outgroup, totalling 12 samples.

We produced target sequence capture data for 7,561 samples using the universal Angiosperms353 probe set⁸ following established laboratory protocols^{50,56}. We complemented our dataset with publicly available data for 2,054 species, sourced from the One Thousand Plant Transcriptomes Initiative⁹ (OneKP; 564 samples), annotated and unannotated genomes (151 samples) and the sequence read archive (SRA; 1,339 samples), the last including transcriptomes (for example, see refs. 57,58) and target capture data (for example, see refs. 59,60). To standardize taxonomy and nomenclature, all species names and families were harmonized with the World Checklist of Vascular Plants² and orders with APG IV if possible¹⁸.

Sequence recovery

Sequence recovery was carried out in two ways, depending on the type of input data. For recovery on the basis of raw reads, that is, Angiosperms353 data or data mined from the SRA, we used HybPiper v.1.31 (ref. 61), embedded in a bespoke pipeline (<https://github.com/baileyp1/PhylogenomicsPipelines>). Raw reads were trimmed using Trimmomatic⁶² to remove low-quality bases and short sequences. In HybPiper, reads were initially binned into genes using BLASTN and an amino acid target file as reference (Supplementary File 1). Individual genes were assembled de novo using SPADRES⁶³ and refined by joining and trimming gene contigs to match coding regions using Exonerate⁶⁴. For genes with paralogue warnings, only the putative orthologue as identified by HybPiper was used. Exclusion of genes with several copies per species has been shown to have negligible impact on species tree inference when it is performed under a multispecies coalescent framework, as described below²⁰. Conversely, the inclusion of several copies per species would have rendered our study computationally intractable. Gene sequences from assembled genomes and OneKP transcriptomes were recovered using custom scripts described in ref. 50. Briefly, the assembled sequences were searched against the target file mentioned above using BLASTN, selecting the best match for each gene and trimming it to the BLAST hit. For a few Angiosperms353 samples that represented the sole accession of their respective families (*Ixonanthes reticulata*, *Mitrastemon matudae* and *Tetracarpaea tasmannica*) and had poor recovery from HybPiper (that is, below 5 kilobase pairs (kb) in total sum of contig length), recovery was undertaken following ref. 50, using less stringent recovery thresholds. The average recovery per order is presented in Supplementary Fig. 9.

Phylogenetic inference

To analyse the dataset, we devised a divide-and-conquer approach. First, we computed a backbone tree, sampling up to five species per family, to test the monophyly of orders and to rigorously explore deep relationships. We used the backbone tree to identify groups (orders or groups of orders) for multiple sequence alignment, with the aim of producing refined subalignments among closely related taxa. Subsequently, the subalignments were merged into global gene alignments and global gene trees were inferred from these using the respective gene

trees from the backbone analysis as constraints. Finally, we inferred a multispecies coalescent tree using the estimated gene trees. The inference pipeline is summarized in Supplementary Fig. 1.

Backbone tree inference. The samples for the backbone were selected so as to represent the crown node and deepest divergences in each family. For families with five or fewer samples (279 families), all samples were included. For those with more than five samples (156), we selected the best sample (most genes and longest sequence) of each consecutively diverging clade (based on published phylogenetic evidence and preliminary analyses of our own data), until five samples were included. To evaluate the extent to which sample selection might affect the backbone tree topology, we inferred 20 backbone replicates, randomly selecting five samples for each family with more than five samples (among the 50% best samples in terms of gene number and gene length recovered). We then summarized the trees to family level and computed Robinson–Foulds distances between the backbone and the 20 replicates (Supplementary Fig. 10).

The phylogenetic reconstruction of the backbone involved up to two iterations of gene alignment and gene tree estimation, with an intermediate step of outlier removal. This was followed by species tree inference in a multispecies coalescent framework. In the first iteration, all sequences for a given gene were aligned using MAFFT v.7.480 (ref. 65) (with `ffnsl` method, that is, `--retree 2 --maxiterate 1000`) and with the direction of the sequence adjusted (`--adjustdirection`). After removing sites with more than 90% missing data with Phyutility⁶⁶, gene trees were estimated using IQ-TREE v.2.2.0-beta⁶⁷, keeping identical sequences in the analysis (`--keep-ident`), setting the substitution model to GTR + G and estimating branch support with 1,000 ultrafast bootstrap replicates⁶⁸. Before the second iteration, we identified long branch outliers using TreeShrink⁶⁹ in ‘all-genes’ mode and rerooting at the centroids of the trees. A second iteration of gene alignment, removal of gappy sites and gene tree estimation was performed for genes with outliers after the removal of outlier sequences. Subsequently, the resulting gene trees were summarized into a species tree using ASTRAL III v.5.7.3, a quartet-based species tree estimation method statistically consistent with the multispecies coalescent model⁷⁰, enabling the full annotation option (`-t 2`), having first collapsed poorly supported nodes (ultrafast bootstrap $\leq 30\%$) in the input gene trees using Newick utilities⁷¹.

Order-level subalignments. For the order-level subalignments, most orders were analysed individually, following the same method described for the backbone. In some cases, smaller orders (fewer than 50 samples) were analysed together with larger ones if they formed monophyletic groups in the backbone. These groups are: (1) Commelinales with Zingiberales, (2) Dioscoreales with Pandanales, (3) Fagales with Fabales, (4) Columelliales, Dipsacales, Escalloniales and Paracryphiales with Apiales, (5) all magnoliids (Canellales, Laurales, Magnoliales and Piperales) and (6) all gymnosperms together (Cycadales, Ephedrales, Gnetales, Ginkgoales and Pinales). Conversely, orders emerging as non-monophyletic in the backbone were split into monophyletic subgroups as follows: (1) Cardiopteridaceae and Stemonuraceae separate from the rest of Aquifoliales, (2) Dasyogonaceae separate from the rest of Arecales, (3) Collumelliaceae separate from the rest of Bruniales, (4) Oncothecaceae separate from the rest of Icaciniales and (5) Huaceae separate from the rest of Oxalidales. The groupings of samples used in the order-level subalignments are provided in Supplementary Table 1. Very small groups, comprising one or two samples (termed orphan sequences), were not included in subalignments and were incorporated directly in global analyses.

Global gene alignments and trees. We produced global gene alignments by merging the order-level subalignments (before removal of gappy sites) and adding the orphan non-aligned sequences using MAFFT⁶⁵, with up to 100 refinement iterations. This approach yields

Article

alignment across the order-level subalignments without disrupting the structure in the subalignments. The final gene alignments were cleaned by removing gappy sites. A summary of the alignments was produced with AMAS⁷² (Supplementary Table 5) and the average occupancy per gene per order is presented in Supplementary Fig. 11.

We then estimated gene trees in Fasttree v.2.1.10 (ref. 73), setting the model to GTR + G, using pseudocounts to avoid biases from fragmentary sequences and increasing search thoroughness (-spr 4, -mlacc 2 and -slownni). We used the gene trees from the backbone analysis to constrain the topology of each respective global gene tree. To avoid propagating error from the backbone analysis to the global analysis, we removed potentially misleading signal from the backbone gene trees before applying them as constraints. First, branches with bootstrap values below 80% were collapsed to avoid enforcing poorly supported relationships. Second, tips placed far from the rest of their order were algorithmically removed (but retained in global gene alignments). Once global gene trees were estimated, outlier long branches were removed using TreeShrink and the set of pruned gene trees was used to compute the global species tree using ASTRAL-MP v.5.15.5 (ref. 74), after collapsing branches with poor support (that is, those with support lower than 10% in the Shimodaira–Hasegawa test).

Divergence time estimation

Divergence times were estimated by penalized likelihood in treePL^{75,76}. This method is computationally efficient for datasets of this scale and typically estimates similar divergence times to more computationally intensive Bayesian analyses. The coalescent species tree topology was used as the input tree with molecular branch lengths estimated in IQ-TREE, on the basis of a concatenated alignment of the top 25 genes selected by SortaDate⁷⁷. Genes were selected by ranking their corresponding gene trees according to the number of congruent bipartitions with the species tree. We selected genes on this basis because high gene tree conflict leads to error in divergence time estimates^{78,79}.

Fossil calibrations were based on the AngioCal fossil calibration dataset described in ref. 5. We used an updated version of this dataset, referred to as AngioCal v.1.1 (Supplementary Table 6 and Supplementary File 2). Assigning fossil calibrations in this dataset to our tree topology led to 200 unique minimum age calibrations at internal nodes (Supplementary Table 7 and Supplementary Fig. 12). A maximum constraint of 154 or 247 Ma was used at the angiosperm crown node. These two values, respectively, represent a young and old constraint for the maximum age of the angiosperm crown node^{5,28}. Both values are nonetheless considerably older than the oldest known crown group angiosperm fossils of around 127.2 Ma (ref. 80). Both maximum constraints, in combination with all the minimum age constraints, were used to time-calibrate the species tree. Depending on the maximum constraint at the root node, these dated phylogenetic trees are referred to as young tree and old tree, respectively. For both the young tree and old tree, four analyses were performed in treePL, using smoothing values of 0.1, 1, 10 or 100. These different smoothing values assume high to low levels of among-branch substitution rate variation.

Sampling extant lineages through time

At 1 Myr intervals from the root age of the dated phylogenetic trees to the present, we calculated how many angiosperm lineages would have been present in a hypothetical tree that sampled 100% of extant angiosperm species diversity. We used this to quantify the proportion of extant lineages incorporated by our phylogenetic trees through time (Supplementary Methods). To do this we simulated unsampled diversity on the dated trees: the species diversity of unsampled genera was simulated as a constant-rate birth–death branching process originating in the crown group of its respective family, whilst unsampled species diversity in sampled genera was simulated as a constant-rate birth–death branching process originating at the stem node of the relevant genus. The extant diversity of each simulated branching process was determined using

the World Checklist of Vascular Plants². At each time interval, we then calculated the proportional difference between the number of lineages in our dated phylogenetic tree and the hypothetical fully sampled tree.

Diversification rate estimation

Dated trees estimated with alternative smoothing values were very similar (Extended Data Fig. 2 and Supplementary Fig. 5), so diversification rate estimates were only performed with the dated trees estimated with a smoothing value of 10. By contrast, age estimates in the young and old trees differed markedly. Diversification rate estimates were therefore performed for both these dated trees. In each case, the dated trees were pruned such that there was a maximum of one tip for each genus.

An initial analysis of diversification rates was performed by generating LTT plots as heatmaps for angiosperms as a whole, as well as for each order, with colours representing the steepness of each LTT curve at 5 Myr intervals. To calculate the steepness of the curve, we calculated the running difference between logarithmic corrected cumulative sums of lineages and applied Tukey's running median smoothing to avoid excessive noise. For order plots, the cumulative sum starts at the first branching point, that is, order crown nodes.

Time-dependent diversification parameters (speciation and extinction rates) were also explicitly estimated across all angiosperms. These analyses were performed in RevBayes with the dnEpisodicBirthDeath function⁸¹. The smallest time windows in which rates were estimated were 5 Ma. However, larger windows were used toward the root of the tree such that there were at least 50 branching events in each time window. Three different models were used: equal rates of speciation and extinction across all windows; variable rates of speciation across windows but equal rates of extinction; and equal rates of speciation across windows but variable rates of extinction. Bayes factor comparison was used to compare models and offered strong support for the variable rate models but could not distinguish between the two variable rate models (Supplementary Information), indicating that they are probably from the same congruent set of models for the species tree⁸². In subsequent discussion we primarily refer to results from the variable speciation rate model (for justification see Supplementary Information), although both variable rate models estimate similar patterns of net diversification rates through time (Supplementary Information).

Lineage-specific diversification rate estimation was performed in BAMM⁸³ and RevBayes. For analyses in BAMM, the setBammPriors function from the R package BAMMtools⁸⁴ was used to define appropriate priors. Different sets of analyses were performed with the prior for the expected number of shifts set to either 10 or 100. These different prior settings had minimal effect on parameter estimates. Clade-specific sampling fractions were specified for each sampled family and a backbone sampling fraction of 1 was used. We therefore accounted for incomplete sampling within families alongside comprehensive sampling of the backbone of the tree. For analyses in RevBayes, the dnCDBDP function was used and the prior for the total number of rate shifts was set to either 10 or 100. Clade-specific sampling fractions cannot be specified with this function. Therefore, the sampling fraction was set to 1 meaning that estimates will become inaccurate toward the present because of unsampled within-family diversity.

Simulations on gene tree conflict

Simulations were based on a multispecies coalescent process. Each species tree contained 100 tips and was simulated as a birth–death branching process with time-dependent rates of speciation and extinction. In experiment 1, the extinction rate was always 0. The speciation rate was 0.75 species Myr⁻¹ at times over 6 Ma, between 6 and 2 Ma the speciation rate was 0.075 species Myr⁻¹ and less than 2 Ma the speciation rate was 0.75 species Myr⁻¹. In experiment 2, the net diversification rates were the same as in experiment 1; however, in this case changes to the extinction rate led to the net diversification rate shifts. Therefore, for all time intervals, the speciation rate was 0.75 species Myr⁻¹. At times

over 6 Ma the extinction rate was 0 species Myr⁻¹, between 6 and 2 Ma the extinction rate was 0.675 species Myr⁻¹ and at times less than 2 Ma the extinction rate was 0 species Myr⁻¹.

Species trees with extinct lineages have extra complexities: first, changes in the extinction rate have a less direct impact on the duration of extant lineages in the species tree compared to changes in the speciation rate (Supplementary Information); and second, the effect of extinction is reduced at times close to the present. This causes shorter branches in the species tree, leading to the so-called ‘pull of the present’. We therefore performed a further analysis that was similar to experiment 2 but with no decrease in the extinction rate at the present. This offered insight into the effect of the ‘pull of the present’ on inferences of gene tree conflict and diversification rates and the relationship between these variables and the timing of rate shifts.

One-hundred gene trees were simulated along the branches of the birth–death branching processes according to a multispecies coalescent process. For most experiments, the effective population size was 5,000. In one further experiment, which was otherwise the same as experiment 1, the effective population size was 50,000. For each simulated dataset, the degree to which the simulated gene trees exhibited conflicting topologies with the species tree was plotted through time (Supplementary Information). This enabled characterization of the relationship between gene tree conflict caused by incomplete lineage sorting and shifts in speciation and extinction rates in the species tree.

More methods, results and discussion are available (Supplementary Information; Supplementary Figs. 13–24 and Supplementary Table 8).

Inclusion and ethics statement

The research described here results from a highly inclusive, large-scale, international collaboration, which has actively encouraged the participation of many individuals from around the world. The authorship comprises many nationalities and is representative in terms of gender, career stage and career path. A total of 163 herbaria from 48 countries provided samples and/or house herbarium vouchers related to samples used in the study (see Acknowledgements). These samples originated from many countries, including Indigenous lands. We recognize the complex histories underlying all natural history collections and the global challenge we face in acknowledging them. We gave priority to recently collected samples and, as a result, most (85%) date from the postcolonial era (estimated here as 1970 onward). To share the benefits of our research, all data generated through this collaboration have been made publicly available before the submission of this work in several data releases, starting in 2019 (see Data availability).

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

All raw DNA sequence data generated for this study are deposited in the European Nucleotide Archive under the following bioprojects PRJNA478314, PRJEB35285, PRJEB49212 and PRJNA678873. All analysed data and metadata are available in Zenodo at <https://doi.org/10.5281/zenodo.10778206> (ref. 55). The resulting trees and metadata are also available in GBIF (<https://doi.org/10.15468/4nfn8b>) and Open Tree of Life (https://tree.opentreeoflife.org/curator/study/view/ot_2304). The names used in this work match the World Checklist of Vascular Plants (<https://doi.org/10.34885/jdh2-dr22>).

Code availability

The code used and developed to perform analyses is available in GitHub at <https://github.com/RBGKew/AngiospermPhylogenomics> and Zenodo at <https://doi.org/10.5281/zenodo.10778206> (ref. 55).

56. Hendriks, K. P. et al. Global Brassicaceae phylogeny based on filtering of 1,000-gene dataset. *Curr. Biol.* **33**, 4052–4068 (2023).
57. Chen, L.-Y. et al. Phylogenomic analyses of Alismatales shed light into adaptations to aquatic environments. *Mol. Biol. Evol.* **39**, msac079 (2022).
58. Timilsena, P. R. et al. Phylogenomic resolution of order- and family-level monocot relationships using 602 single-copy nuclear genes and 1375 BUSCO genes. *Front. Plant Sci.* **13**, 876779 (2022).
59. Ogutcen, E. et al. Phylogenomics of Gesneriaceae using targeted capture of nuclear genes. *Mol. Phylogenet. Evol.* **157**, 107068 (2021).
60. Yardeni, G. et al. Taxon-specific or universal? Using target capture to study the evolutionary history of rapid radiations. *Mol. Ecol. Resour.* **22**, 927–945 (2022).
61. Johnson, M. G. et al. HybPiper: extracting coding sequence and introns for phylogenetics from high-throughput sequencing reads using target enrichment. *Appl. Plant Sci.* **4**, 1600016 (2016).
62. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
63. Bankevich, A. et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* **19**, 455–477 (2012).
64. Slater, G. S. C. & Birney, E. Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics* **6**, 31 (2005).
65. Katoh, K. & Standley, D. M. MAFFT Multiple Sequence Alignment Software Version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
66. Smith, S. A. & Dunn, C. W. Phylutility: a phyloinformatics tool for trees, alignments and molecular data. *Bioinformatics* **24**, 715–716 (2008).
67. Minh, B. Q. et al. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* **37**, 1530–1534 (2020).
68. Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B. Q. & Vinh, L. S. UFBoot2: improving the ultrafast bootstrap approximation. *Mol. Biol. Evol.* **35**, 518–522 (2018).
69. Mai, U. & Mirarab, S. TreeShrink: fast and accurate detection of outlier long branches in collections of phylogenetic trees. *BMC Genomics* **19**, 272 (2018).
70. Zhang, C., Rabiee, M., Sayyari, E. & Mirarab, S. ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinformatics* **19**, 153 (2018).
71. Junier, T. & Zdobnov, E. M. The Newick utilities: high-throughput phylogenetic tree processing in the Unix shell. *Bioinformatics* **26**, 1669–1670 (2010).
72. Borowiec, M. L. AMAS: a fast tool for alignment manipulation and computing of summary statistics. *PeerJ* **4**, e1660 (2016).
73. Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree 2—approximately maximum-likelihood trees for large alignments. *PLoS ONE* **5**, e9490 (2010).
74. Yin, J., Zhang, C. & Mirarab, S. ASTRAL-MP: scaling ASTRAL to very large datasets using randomization and parallelization. *Bioinformatics* **35**, 3961–3969 (2019).
75. Sanderson, M. J. Estimating absolute rates of molecular evolution and divergence times: a penalized likelihood approach. *Mol. Biol. Evol.* **19**, 101–109 (2002).
76. Smith, S. A. & O’Meara, B. C. treePL: divergence time estimation using penalized likelihood for large phylogenies. *Bioinformatics* **28**, 2689–2690 (2012).
77. Smith, S. A., Brown, J. W. & Walker, J. F. So many genes, so little time: a practical approach to divergence-time estimation in the genomic era. *PLoS ONE* **13**, e0197433 (2018).
78. Britton, T. Estimating divergence times in phylogenetic trees without a molecular clock. *Syst. Biol.* **54**, 500–507 (2005).
79. Carruthers, T. et al. The implications of incongruence between gene tree and species tree topologies for divergence time estimation. *Syst. Biol.* **71**, 1124–1146 (2022).
80. Gomez, B., Daviero-Gomez, V., Coiffard, C., Martin-Closas, C. & Dilcher, D. L. Montsechia, an ancient aquatic angiosperm. *Proc. Natl Acad. Sci. USA* **112**, 10985–10988 (2015).
81. Höhna, S. et al. RevBayes: Bayesian phylogenetic inference using graphical models and an interactive model-specification language. *Syst. Biol.* **65**, 726–736 (2016).
82. Louca, S. & Pennell, M. W. Extant timetrees are consistent with a myriad of diversification histories. *Nature* **580**, 502–505 (2020).
83. Rabosky, D. L. Automatic detection of key innovations, rate shifts and diversity-dependence on phylogenetic trees. *PLoS ONE* **9**, e89543 (2014).
84. Rabosky, D. L. et al. BAMMtools: an R package for the analysis of evolutionary dynamics on phylogenetic trees. *Methods Ecol. Evol.* **5**, 701–707 (2014).

Acknowledgements The PAFTOL project was funded by grants from the Calvea Foundation to the Royal Botanic Gardens, Kew. Data were also contributed by the Genomics for Australian Plants Framework Initiative consortium funded by Bioplatforms Australia (enabled by the National Collaborative Research Infrastructure Strategy) and partner organizations. The work was further supported by research grants from VILLUM FONDEN (grant no. 00025354) and the Aarhus University Research Foundation (grant no. AUFF-E-2017-7-19) to W.L.E. and from grant nos NSF DBI 1930030 and DEB 1917146 to S.A.S. Computational resources and technical support were provided by the Research/Scientific Computing teams at The James Hutton Institute and the National Institute of Agricultural Botany (NIAB) through the ‘UK’s Crop Diversity Bioinformatics HPC’ (BBSRC grant no. BB/S019669/1). The following provided technical assistance to the project at various stages: O. Berry, N. Black, M. Corcoran, S. Dequiere, I. Fairlie, L. Frankel, T. Freeth, A. Gilbert, B. Lepschi, D. Lewis, L. May, A. McArdle, E. O’Loughlin, S. Phillips, T. Sarkinen, L. Simmons, N. Walsh and M.-H. Weech. We thank all institutions who made their biological collections available and the many botanists and co-workers in the field who have collected, identified and curated the specimens used in this project. Specifically, we thank the following herbaria and their staff for providing samples for genomic analysis and/or for housing voucher specimens associated with analysed samples: A, ABH, AD, ALTB, APSC, B, BA, BC, BCN, BCRU, BG, BH, BHCB, BISH, BJFC, BKF, BM, BNRH, BOL, BONN, BR, BRI, BRIT, BRLU, BRUN, C, CAN, CANB, CAS, CBG, CNS, COL, CONC, CORD, CS, CTES, CUVK, DNA, E, EA, F, FI, FLAS, FMB, FTG, G, GB, GC, GENT, GH, GOET, GUAY, GZU, HAW, HEID, HITBC, HNG, HO, HPUJ, HRCB, HTW, HUA, HUAL, HUAZ, HUB, HUEFS, HUFU, IBSO, IBUG, ICN, IEB, INB, INPA, JBB, JBL, JRAU, K, KAS, KLU, KRB, KUN, L, LE, LISG, LP, LPB, LYJB, M, MA, MAU, MBA, MBML, MEDEL, MEL, MELU, MHA, MICH, MIN, MJG, MO, MSUN, MT, MY, N, NBS, NCU, NCV, NE, NH, NHM, NHMR, NMNL, NOU, NSW, NU, NY, OS, OSBU, P, PERTH, PG, PH, PRE, PTBG, QBG, QCA, QRS, RB, REU, S, SALA, SAR, SEV, SGO, SI, SING, SP, SPF, SPFR, SUVA, TCD,

Article

TEX, TNS, TUH, TUM, U, UAPC, UB, UBT, UDW, UEC, UPBC, UPR, UPS, UPTC, US, USM, W, WAG, WS, WTU, YA and ZSS; acronyms follow Index Herbariorum (<https://sweetgum.nybg.org/science/ih/>). We also thank the Millennium Seed Bank Partnership for supporting access to samples. We acknowledge all national, state and regional authorities who authorized and facilitated the sourcing of these specimens. See also extended acknowledgements in the Supplementary Information.

Author contributions A.R.Z., T.C., A.A., S. Bellot, D.M.C., O.M.G., P.J.K., I.J.L., H. Sauquet, S.A.S., W.L.E., F. Forest and W.J.B. were involved in conceptualization of this work. A.R.Z., T.C., A.A., S. Bellot, D.M.C., O.M.G., H. Sauquet, S.A.S., W.L.E., F. Forest and W.J.B. contributed to the methodology. A.R.Z., O.M., E.F., C. McGinnie, S.R.R., L. Simpson, J.J.C., R.S.C., S.D., L. Pokorny, G.M.A., K.G., K.P.H., A. Hoewener, A.-Q.H., E.M.J., I.A.B.S.K., I.L., D.A.L., E. J. Lirio, J.-X.L., P. Malakasi, N.A.S.P., T.S., J.V., G.K.A., R.L.A., M.S.A., M.A., M.J.A., J.A., W.A., J.B.B., C.D.B., H.F.B., M.D.B., R.L.B., R.J.B., M.J.B., E. Biffin, N.B., J.L.B., D.B., R.B., A.M.C.B., P. C. Boyce, G.L.C.B., M. Briggs, L. Broadhurst, G.K.B., J.J.B., A.B., S. Buerki, E. Burns, M. Byrne, S. Cable, A.C., M. W. Callmander, Á.C., D.J.C., W.M.C.-M., M.M.C., A.J.A.C., A.C.M., M. W. Chase, L.W.C., M.C., S. Chen, M.J.M.C., P.-A.C., M.A.C., S.C.C., J.G.C., X.C., T.L.P.C., I.D.C., L.C., I.D., G.D., N.M.J.D., A.P.D., K.-J.D., S.R.D., M.F.D., M.R.D., S.L.E., U.E., R.H.J.E., M. Escudero, M. Estrella, F. Fabriani, M.F.F., P.L.F., S.Z.F., R.M.F., S.F., L.F., T.F., M.G.-C., E.M.G., D.A.G., A. Giaretta, M.G., L.J.G., C.C.G., D.J.G., S.W.G., A. Grall, L.G., B.F.G., D.G.G., J.H., T. Haevermans, A. Haigh, J.C.H., T. Hall, M.J.H., S.A.H., O.H., T.R.H., G.D.H., H.C.F.H., C.J.J., S.A.J., R.W.J., G.K., I.M.K., K.K., M.K., E.A.K., G.J.K., B.K., B.B.K., R.R.K., S.K., M.A.K., J.H.L.-M., F.L., C.J.L., É.L.-B., G.P.L., D.-Z.L., L.L., S.L.-S., T.L., D.L., M. Lu, P.L.-I., J.L., E. J. Lucas, M. Luján, M. Lum, T.D.M., C. Magdalena, V.F.M., L.E.M., S.J.M., K. McColl, A.J.M., A.E.M., T.G.B.M., H.M., R.I.M., V.S.F.T.M., F.A.M., J.D.M., A.K.M., M.J.M., T.L.M., K. Mummenhoff, J.M., P. Muriel, D.J.M., K.N., L.N., F.J.N., R. Nyffeler, A.O., E.M.O., L. Palazzesi, A.L.P., S.K.P., J.P., D.S.P., O.A.P.-E., C. Persson, M.P., Y.P., J.R.P., G.M.P., R.F.P., G.T.P., C. Puglisi, M.Q., R.K.R., P.E.J.R., M. Renner, E.H.R., M. Rodda, Z.S.R., S.R., R.R., M.F.S., H. Schaefer, R. J. Schley, A.S.-L., A.S., I.S., K.A.S., M.P.S., A.O.S., A.R.G.S., M.S., E.C.S., J.F.S., N.S., D.E.S., P.S.S., R. J. Soreng, C.A.S., J.R.S., P.F.S., S.C.K.S., L. Struwe, J.M.T., I.R.H.T., A.H.T., I.T., A.T.-B., F.U., T.M.A.U., J.C.V., G.A.V., H.P.V., M.S.V., J.M.V., N.W., M.W., C.A.D.W., A.J.W., J.J.W., L.T.W., T.C.W., S.Y.W., L.A.W.,

R.W., S.W., M.X., Y.Y., Y.-X.Z., M.-Y.Z., S.Z., F.O.Z., S. Bellot, D.M.C., O.M.G., H. Sauquet, W.L.E., F. Forest and W.J.B. provided resources. A.R.Z., T.C., O.M., P. C. Bailey, K.L., G.E.B., N.E., E.F., B.G.-P., C. McGinnie, S.R.R., L. Simpson, L. Botigué, J.J.C., R.S.C., S.D., M.G.J., J.T.K., L. Pokorny, N.J.W., G.M.A., L.D., K.G., K.P.H., A. Hoewener, A.-Q.H., E.M.J., I.A.B.S.K., I.L., D.A.L., E. J. Lirio, J.-X.L., P. Malakasi, N.A.S.P., T.S., J.V. and S. Bellot carried out the investigations. A.R.Z., T.C., O.M., G.E.B., N.E., E.F., B.G.-P., C. McGinnie, R. Negrão, S.R.R., L. Simpson, E.T.R., V.M.A.B., K.P.H., J.V., T.R.A. and H. Sauquet were responsible for data curation. A.R.Z., T.C., P. C. Bailey and K.L. conducted the formal analysis. A.R.Z., T.C., P. C. Bailey, K.L., M.G.J. and J.T.K. developed the software. A.R.Z., T.C. and R. Negrão prepared the visualizations. A.R.Z., T.C., W.L.E., F. Forest and W.J.B. wrote the original manuscript with support from A.A., S. Bellot, D.M.C., O.M.G., H. Sauquet and S.A.S. K.L., E.F., J.J.C., J.T.K., L. Pokorny, N.J.W., A.-Q.H., E.M.J., I.L., J.V., M.S.A., J.B.B., M.D.B., R.L.B., A.M.C.B., L. Broadhurst, A.B., D.J.C., M.M.C., M. W. Chase, L.W.C., M.J.M.C., P.-A.C., T.L.P.C., U.E., R.H.J.E., M. Estrella, M.F.F., P.L.F., M.G., L.J.G., S.W.G., J.H., T. Haevermans, J.C.H., O.H., T.R.H., K.K., E.A.K., M.A.K., F.L., C.J.L., D.-Z.L., S.L.-S., T.D.M., V.S.F.T.M., F.A.M., A.K.M., M.J.M., D.J.M., F.J.N., L. Palazzesi, J.P., D.S.P., O.A.P.-E., Y.P., G.M.P., R.K.R., R.R., H. Schaefer, A.S.-L., M.P.S., A.R.G.S., N.S., D.E.S., P.S.S., R. J. Soreng, P.F.S., S.C.K.S., A.H.T., T.M.A.U., J.M.V., J.J.W., T.C.W., Y.Y., S.Z. and I.J.L. reviewed the final manuscript. S.A.S., W.L.E., F. Forest and W.J.B. undertook supervision. P.J.K., I.J.L., F. Forest and W.J.B. acquired funding. V.M.A.B., P.J.K., I.J.L., F. Forest and W.J.B. were responsible for project administration.

Competing interests The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41586-024-07324-0>.

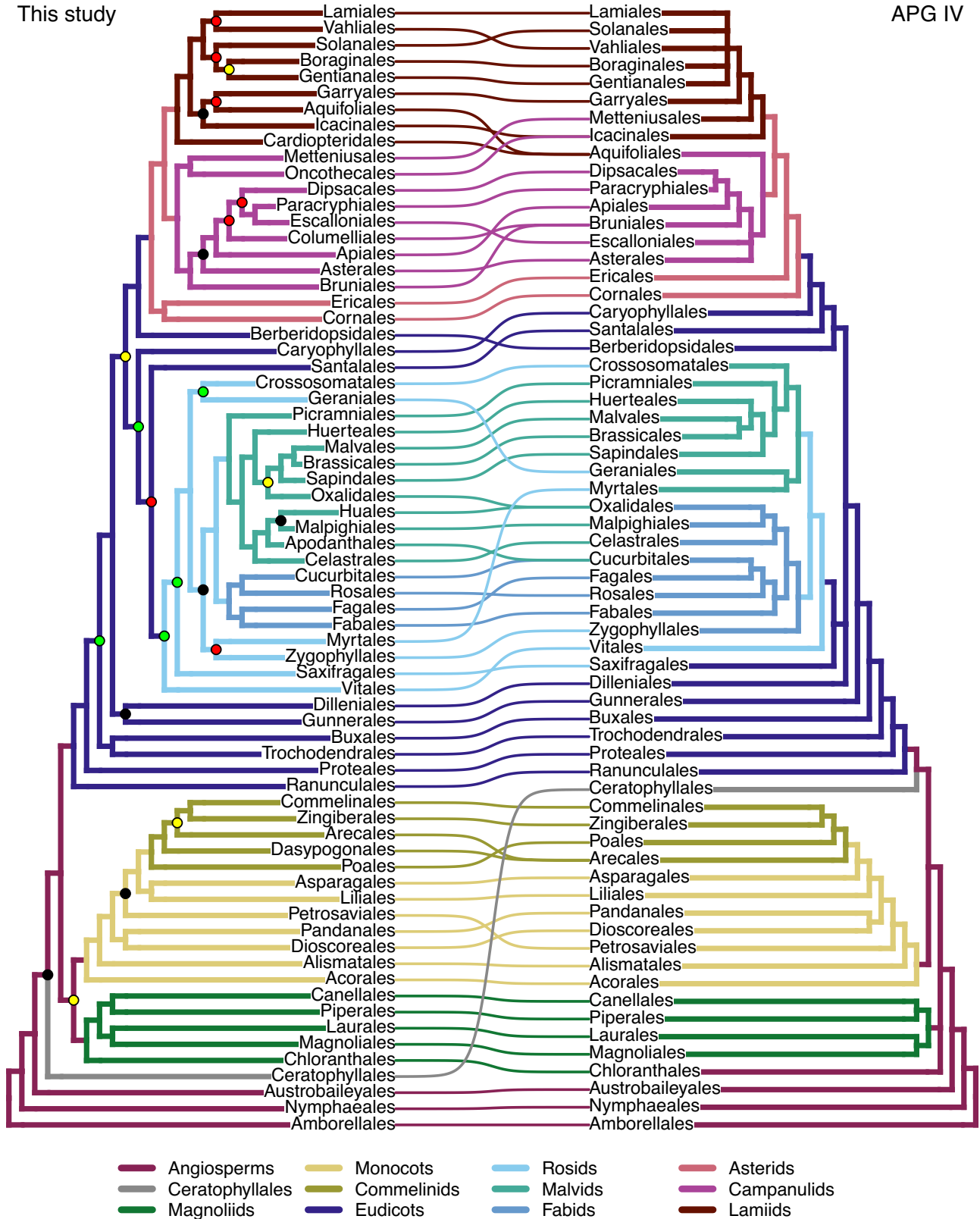
Correspondence and requests for materials should be addressed to William J. Baker.

Peer review information *Nature* thanks Craig Barrett, Elena Kramer, Patrick Wincker and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permissions information is available at <http://www.nature.com/reprints>.

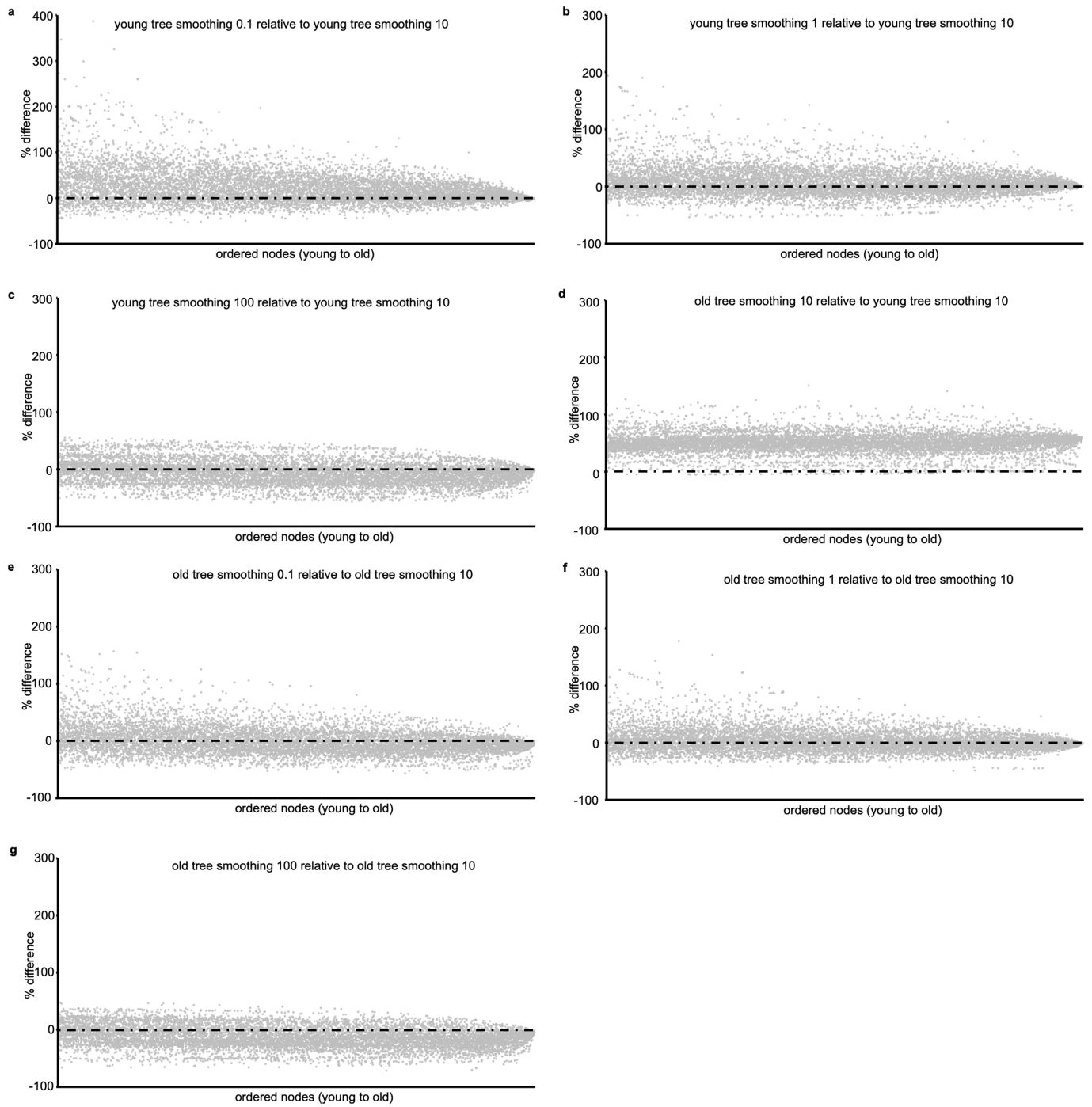
This study

APG IV

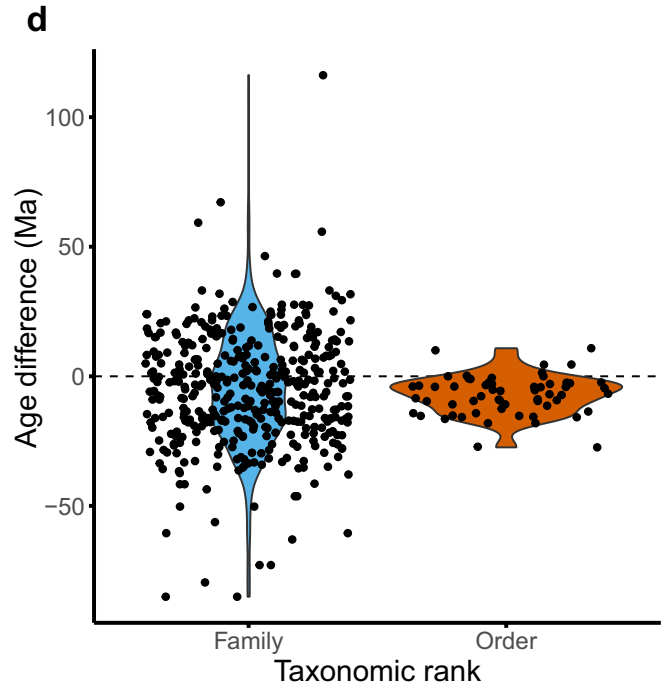
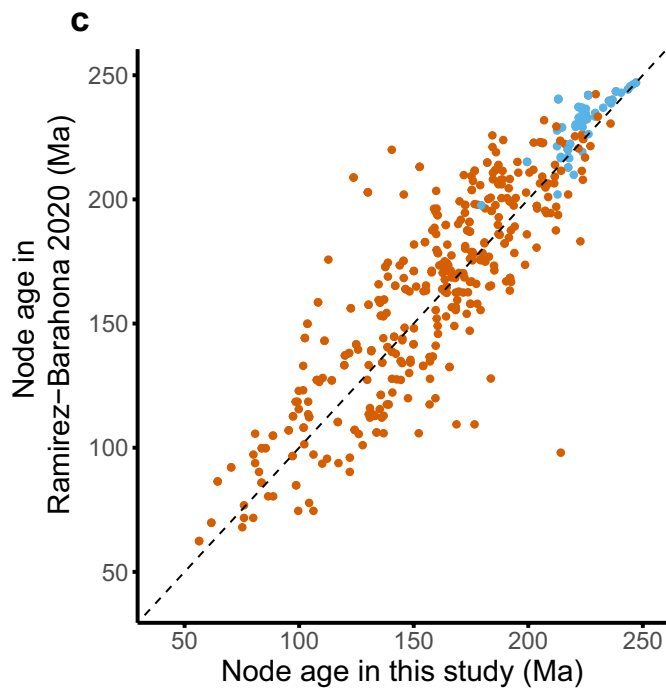
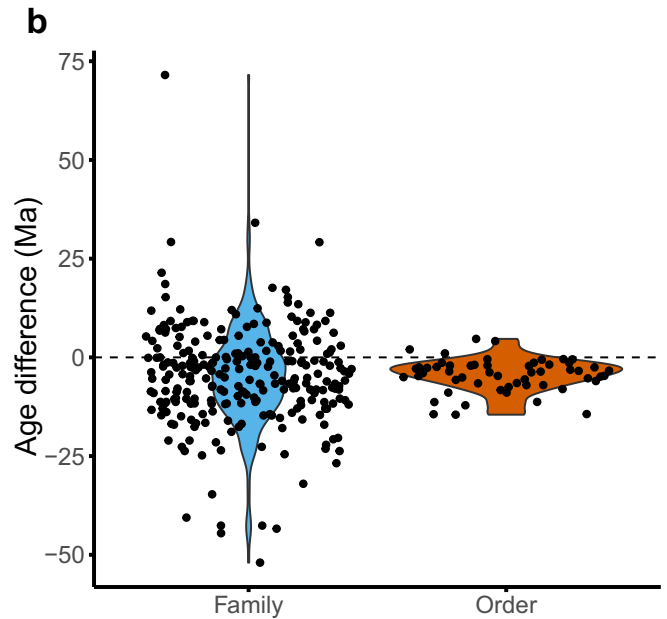
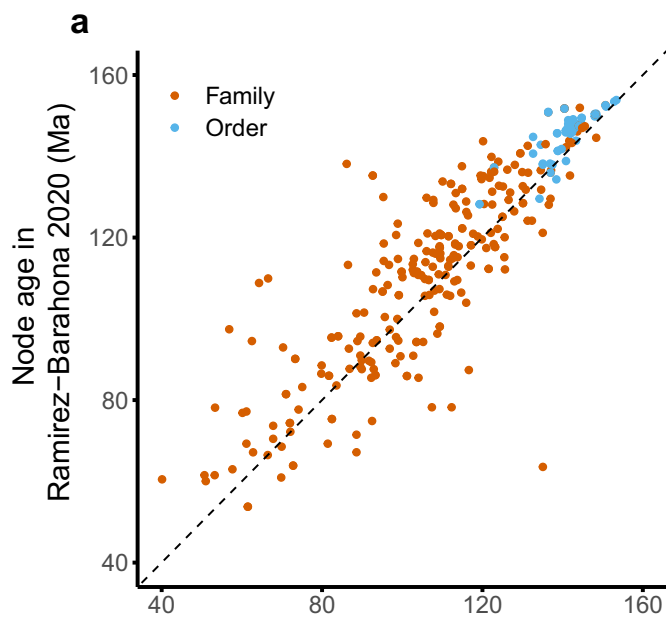


Extended Data Fig. 1 | Tanglegram at ordinal level between this work (left) and the APG IV schematic tree (right). Branches colours represent the clades according to the composition proposed in each work. Posterior probability is presented only for nodes without maximum support. Coloured circles in the

left tree represent the posterior probability of each node as: maximum (absent), between 1 and 0.95 (green), between 0.95 and 0.75 (yellow), between 0.75 and 0.5 (red), below 0.5 (black).

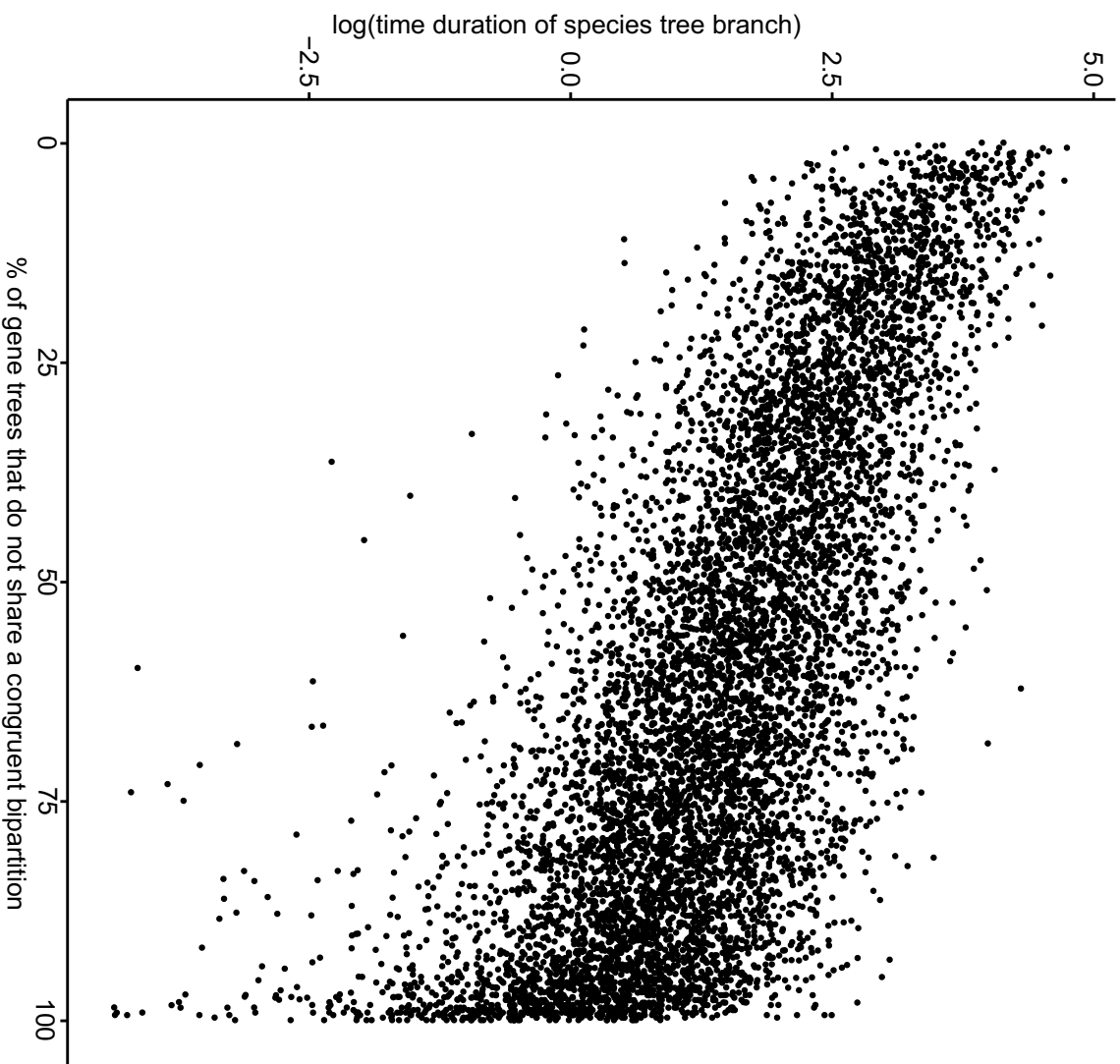


Extended Data Fig. 2 | Comparison of node age estimates in the eight time-calibrated phylogenetic trees. Each point represents a node and corresponds to the percentage difference in age estimates for that node between the two trees that are compared in each plot.

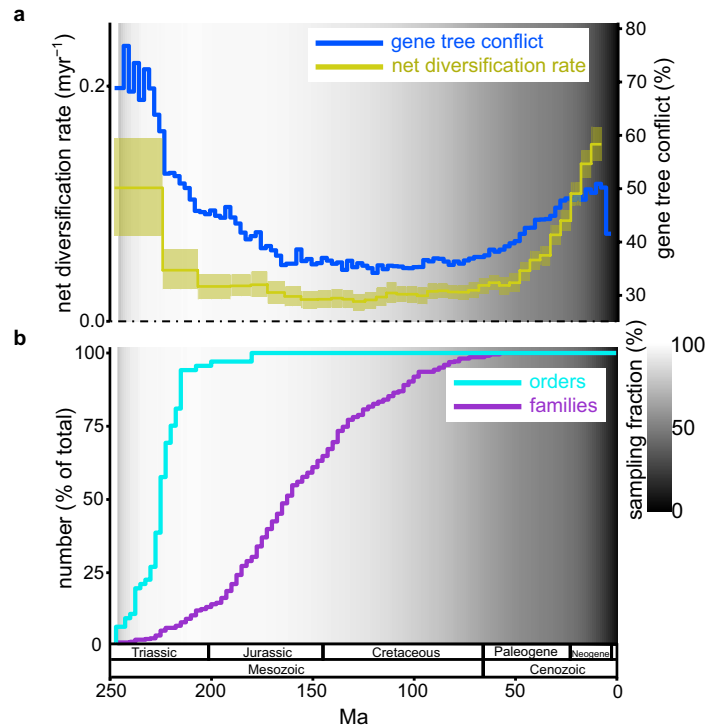


Extended Data Fig. 3 | Comparison of stem ages of families and orders inferred in this study and Ramirez-Barahona et al.⁵. **a** and **b**, Stem age comparison between our young tree (maximum constraint at the root node of 154 Ma) and the dataset CC_complete of Ramirez-Barahona et al.⁵. **a**, Ages in each study, coloured according to taxonomic rank and **b**, Age differences, calculated as age

in this study minus age in Ramirez-Barahona et al.⁵ **c** and **d**, Stem ages comparison between our old tree (maximum constraint at the root node of 247 Ma) and the dataset UC_complete from Ramirez-Barahona et al.⁵ **c**, Ages in each study, coloured according to taxonomic rank and **d**, Age differences, calculated as age in this study minus age in Ramirez-Barahona et al.⁵.

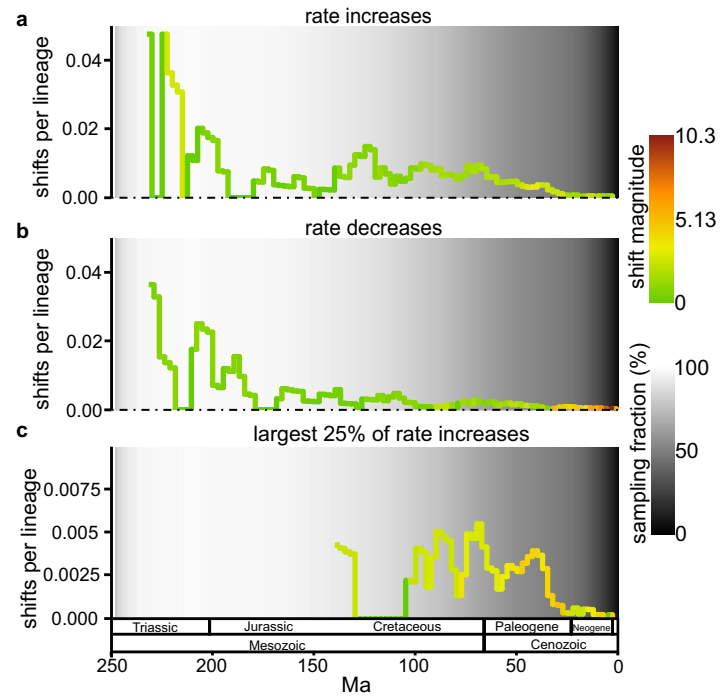


Extended Data Fig. 4 | Correlation between branch time duration and percentage of gene trees that do not share a congruent bipartition for the branch. The results are based on the young tree (maximum constraint at the root node of 1.54 Ma). For each branch in the young tree, the percentage of gene trees that do not share a congruent bipartition with the species tree branch is plotted against the logarithm of the time duration for the branch.



Extended Data Fig. 5 | Angiosperm-wide diversification and gene tree conflict through time. This is equivalent to Fig. 3 but for the old tree (maximum constraint at the root node of 247 Ma). **a**, Estimated net diversification rate through time (yellow, left y-axis) and the level of gene tree conflict through time (blue, right y-axis). Net diversification rates are estimated with a model that enables speciation rates to vary between time intervals; the line is the posterior mean and the yellow shaded area is the 95% highest posterior density.

Gene tree conflict is calculated from the percentage of gene trees that do not share a congruent bipartition with each species tree branch, with the plotted value being the mean across all species tree branches that cross each 2.5 Myr time slice. **b**, Cumulative percentage of extant orders and families that have originated through time. In both **a** and **b**, the background grey-scale gradient is the estimated percentage of extant lineages represented in the species tree through time (“sampling fraction”).



Extended Data Fig. 6 | Summary of lineage-specific diversification rate shifts estimated by BAMM. This is equivalent to Fig. 4, but for the old tree (maximum constraint at the root node of 247 Ma). **a**, Diversification rate increases per lineage through time. The colour corresponds to the average magnitude of the rate increases during the time period. **b**, Equivalent to a, but for rate decreases.

c, Equivalent to a, but focusing on the largest 25% of diversification rate increases. In **a**, **b** and **c**, the number of shifts is extracted from the maximum a posteriori shift configuration, the prior for the number of shifts is set to 10 and the background grey-scale gradient is the estimated percentage of extant lineages represented in the species tree through time (“sampling fraction”).

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection

Data analysis https://doi.org/10.5281/zenodo.10778206).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

All raw DNA sequence data generated for this study are deposited in the European Nucleotide Archive under the following bioprojects PRJNA478314, PRJEB35285, PRJEB49212 and PRJNA678873. All analysed data and metadata are available in Zenodo (<https://doi.org/10.5281/zenodo.10778206>). The resulting trees and metadata are also available in GBIF (<https://doi.org/10.15468/4njin8b>) and Open Tree of Life (https://tree.opentreeoflife.org/curator/study/view/ot_2304). The names used in this work match the World Checklist of Vascular Plants (<https://doi.org/10.34885/jdh2-dr22>).

Research involving human participants, their data, or biological material

Policy information about studies with [human participants or human data](#). See also policy information about [sex, gender \(identity/presentation\), and sexual orientation](#) and [race, ethnicity and racism](#).

Reporting on sex and gender	Not applicable
Reporting on race, ethnicity, or other socially relevant groupings	Not applicable
Population characteristics	Not applicable
Recruitment	Not applicable
Ethics oversight	Not applicable

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

- Life sciences Behavioural & social sciences Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	We sampled nearly 8,000 genera and 353 genes to infer a phylogenetic tree for angiosperms, dated and calibrated it with a dataset of 200 fossils, and used this evolutionary time frame to study the diversification of the group.
Research sample	We aimed to produce a dataset with at least one species per genus for >50% of the ca. 13,600 currently angiosperms genera. For genera with multiple species available, we retained up to three species, selecting primarily by phylogenetic representation followed by amount of data (number of genes and total length of recovery). Twelve species of gymnosperms were used as outgroups, one from each family. The final dataset contains 9,506 species, 7,923 genera, 416 families and 64 orders of angiosperms. It comprises 7,561 samples with data produced by this project and 1,963 samples mined from public repositories.
Sampling strategy	We sourced samples from herbarium, living, DNA bank and and tissue bank collections.
Data collection	We produced target sequence capture data using the Angiosperms353 probe kit. Total DNA was extracted using CTAB protocol and quantified by fluorometry. Average fragment size was assessed with 1% agarose gel. DNA extracts were diluted to 4 ng/ul. For extracts with high molecular weight, total DNA was fragmented using Covaris M220 Focused ultrasonicator. Genomic DNA libraries were prepared using NEBNext Ultra II DNA Library Prep Kit, following manufacturer's protocol at half volume, and with dual indexing. Libraries were normalised to 10 nM and pooled in equimolar amounts according to the average fragment size and taxonomic groups. Pools included, on average, 20-24 libraries. The pools were hybridised with Angiosperms353 probe kit. Hybridised pools were normalised to 7nM and combined for sequencing. Sequencing was carried out in Illumina platforms MiSeq and HiSeq.
Timing and spatial scale	Data production started in May 2016 and ended in December 2021.
Data exclusions	We excluded samples that failed both phylogenetic and barcode validation, as described in Baker (2022). Syst. Biol. 71: 301-319. We

Data exclusions	also excluded samples if more than three accessions were available for the same genus.
Reproducibility	All raw data and intermediate files are provided, making every step reproducible. All accessions used are listed in the Supplementary Table 1.
Randomization	Not applicable - randomization was not required.
Blinding	Not applicable - this work does not involve trials or controlled experiments.
Did the study involve field work?	<input type="checkbox"/> Yes <input checked="" type="checkbox"/> No

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involvement in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> Antibodies
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology and archaeology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data
<input checked="" type="checkbox"/>	<input type="checkbox"/> Dual use research of concern
<input type="checkbox"/>	<input checked="" type="checkbox"/> Plants

Methods

n/a	Involvement in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging