nature communications

Article

https://doi.org/10.1038/s41467-025-57687-9

# New insights on angiosperm crown age based on Bayesian node dating and skyline fossilized birth-death approaches

Received: 13 July 2024

Accepted: 25 February 2025

Published online: 07 March 2025

Check for updates

Xiaoya Ma[1], Chi Zhang[2], Lingxiao Yang[1], S. Blair Hedges[3] & Bojian Zhong[1,4]

Despite considerable work in recent years, pinpointing the time when angiosperms originated has been challenging. However, the rapid development of molecular clock methodology has provided new tools to resolve this conundrum. In particular, the fossilized birth-death model establishes a rich interplay between molecules and stratigraphy by incorporating fossils explicitly into dating analyses. In this study, we apply Bayesian node dating and the skyline fossilized birth-death model, which differ in how the calibration is applied, to estimate the crown age of angiosperms. Node dating analyses with different calibration strategies show that the posterior distribution is strongly constrained by the effective prior at the node of crown angiosperms, dominated by the maximum age constraint. Using the skyline fossilized birth-death model, we reveal that assigning different priors for origin time resulted in similar crown ages for angiosperms. Moreover, the oldest fossils play a significant role in time estimates, and the dating results are robust to sampling assumptions of extant taxa. Our dating analyses indicate a largely Triassic crown age (255–202 Ma) for angiosperms, the period when mammals, dinosaurs, and squamate reptiles first appeared, and highlight the potential of morphological data to redefine the timeline of angiosperms.

Angiosperms, also known as flowering plants, are the most diverse, widely distributed and adaptable group of all major lineages of plants. The emergence of angiosperms was one of the most important evolutionary events in the history of the Earth, forming the structural and energetic basis of the vast majority of extant terrestrial biomes and reshaping major ecosystems around the globe. In particular, the diversification of angiosperms is an important driver of the decline or expansion of other plant groups such as ferns and conifers[1,2]. The time of origin of angiosperms also has a direct bearing on the discussion of their geographical origin and on the understanding of their paleoenvironmental conditions[3–5]. A reliable evolutionary timeline of angiosperms is needed to improve our understanding of how angiosperms came to dominate terrestrial ecosystems, and to explain patterns of diversity in other groups (e.g., insects and amphibians).

It is widely accepted that the earliest unequivocal fossils assignable to angiosperms appear in the early Cretaceous of northern Gondwana, ~133–125 million years ago (Ma)[4,6–8]. However, many molecular clock studies have pushed the origin of living angiosperms (the crown age) back to the early Jurassic or even earlier[9–11], much older than fossil evidence. This age difference between molecular and fossil data was referred as the "Jurassic gap"[10], which hinders our understanding of angiosperm evolution. Importantly, current Bayesian dating analyses based on large numbers of nuclear genes are limited and the taxon sampling is relatively sparse, and there has been no

[1]College of Life Sciences, Nanjing Normal University, Nanjing, China. [2]Institute of Vertebrate Paleontology and Paleoanthropology, Chinese Academy of Sciences, Beijing, China. [3]Center for Biodiversity, Temple University, Philadelphia, PA, USA. [4]Ministry of Education Key Laboratory of NSLSCS, Nanjing Normal University, Nanjing, China. e-mail: bjzhong@gmail.com

consensus on the crown age of angiosperms (e.g., 139–136 Ma[12], 199–167 Ma[13], 247–154 Ma[14], 255–222 Ma[11], 267–187 Ma[10]).

Calibration is the key factor in molecular clock dating, typically based on the fossil record or biogeographic events, converting the distances from molecular data into absolute times and rates[15]. One mainstream approach, referred to as node dating (ND), uses probability distributions (e.g., uniform or Cauchy distribution) to calibrate one or more internal nodes in the tree to estimate the ages of the other nodes[16,17], reflecting our knowledge of the amount of elapsed time. Each calibration density is specified by the user to accommodate uncertainty in the relationship of the node age to the age of the reference fossil, taking into account processes such as fossilization, recovery, and morphological evolution. Although the ND approach has been widely used in estimating angiosperm timescales[11,12,18–21], there are still some concerns about calibrating divergence time. Firstly, only the oldest fossil can be useful in constructing the calibration prior for the node in ND analyses. However, the existence of different morphospecies for different parts of the plant makes it difficult to identify the appropriate oldest fossils, with pollen being an extreme example. Secondly, unlike the minimum bound in the calibration distribution, the maximum bound is much less restrictive. When the maximum age of a lineage divergence has not yet reached consensus, specifying different maximum constraints could produce widely varying ages[22]. Thirdly, the user-specified calibration priors typically deviate from the prior that has effect in dating (called the effective prior or induced prior), due to interactions among calibration densities and the underlying tree and clock priors. Even though the tree prior can be constructed conditional on the calibration densities, the truncation effect resulting from the requirement that ancestral nodes be older than descendant nodes, still contributes to the deviation[23].

An attractive alternative approach, named tip dating or total-evidence dating, directly includes fossils as terminal taxa in the tree[24], thus obviating the need to specify calibration densities that are often assigned based on arbitrary criteria. Unlike ND analyses, tip dating allows for inclusion of all available fossils from a specific group, rather than only referring to the oldest one, and thus can take advantage of abundant time information. As the tree prior in the tip-dating approach, the fossilized birth-death (FBD) model represents a significant improvement over the uniform tree model, since the uniform tree prior has no birth, death, and sampling rate parameters, and it is usually considered a vague or uninformative prior[24–26]. The FBD model provides a coherent mechanistic description of species diversification, fossil preservation, and taxa sampling processes that produced the tree[27,28]. Recent advances extend the FBD model to allow the speciation, extinction, and fossil recovery rates to vary over time in a piecewise-constant manner[26,29], referred as skyline FBD (SFBD) model. Furthermore, the SFBD model was extended to accommodate diversified sampling for larger, species-rich clades[30], which could correct overestimation of divergence times resulting from violations of the random-sampling assumption[26,31]. Moreover, the SFBD model with many time epochs has recently proved to be robust to violations of sampling assumptions[30]. Despite the enthusiastic reception of the FBD process[25,32–35], no studies have so far used the FBD model to estimate the origin time of angiosperms.

In this work, we expand the taxon sampling to use more angiosperm fossils to infer the timeline of angiosperms. Specifically, we focus on two questions: (1) the longstanding debate on the crown age of angiosperms; and (2) determining differences in divergence time estimation between the SFBD and ND approaches. The SFBD approach we use here differs from total evidence dating (TED) because no morphological characters are used. Instead, it relies on a fixed tree topology including fossils, and the effective prior on divergence times is induced by the SFBD process. To investigate key factors on divergence time estimation, we explore several calibrating strategies in ND analyses; while in SFBD analyses, we evaluate the impacts of fossil sampling, sampling strategy, density of extant species, and priors on the origin time. Our study not only provides the foundation for a better understanding of the patterns of angiosperm diversification and the evolution of key traits, but also yields valuable insights for estimating divergence times using the FBD model.

## Results

We reconstructed a well-supported phylogenetic tree using a concatenated supermatrix including 324 nuclear genes from 670 angiosperm taxa and 78 species of gymnosperms (Supplementary Fig. 1). Along the backbone of the phylogeny, Amborellales was strongly supported as the sister lineage to all other angiosperms, followed successively by Nymphaeales and Austrobaileyales (Supplementary Fig. 1). We found eudicots and magnoliids to be sister groups with full support, and Chloranthales was the closest relative of eudicots-magnoliids (Supplementary Fig. 1). A sister relationship between the Ceratophyllales and monocots also had high support (Supplementary Fig. 1).

As the full molecular dataset is computational intractable for the subsequent dating analyses, we subsampled taxa at the genus level of eudicots, monocots, and gymnosperms, but retained the diversity of all sampled families, resulting in 349 angiosperm species and 15 gymnosperm species as an outgroup.

### Impacts of maximum bound on time estimates in the node dating approach

In the ND analyses, we firstly assessed the relative fit of two relaxed-clock models to our data using Bayes factors. The results indicated that the independent-rates model had a higher marginal likelihood than the autocorrelated-rates model (Supplementary Data 1). We then explored whether maximum bounds are responsible for the age disparity of angiosperms using the independent-rates model. Based on the 114 fossil calibrations, we employed five fossil calibration strategies (see Methods for details) to accommodate different maximum bounds of crown angiosperms, mesangiosperms, and eudicots. To have a better understanding of the general impact of priors on posteriors, we compared the user-specified priors, effective (induced) priors, and posteriors for these three nodes.

Strategies 1 and 2 inferred a Carboniferous origin of crown angiosperms (strategy 1: 366–313.2 Ma; strategy 2: 365–304.9 Ma; 95% highest posterior density (HPD) interval), a Carboniferous–Permian diversification of mesangiosperms, and a Permian–Jurassic origin of eudicots (Supplementary Figs. 2 and 3). When reducing the maximum bound to 247.2 Ma for crown angiosperms and mesangiosperms, we obtained younger crown ages of angiosperms (strategy 3: 255.9–219.7 Ma; strategy 4: 292.0–249.5 Ma) and mesangiosperms. We also noted that changing from a narrow uniform distribution (strategy 3) to a more diffuse Cauchy distribution (strategy 4) for crown eudicots not only varied its own age estimate, but also pushed the crown age of angiosperms to be older (Fig. 1, Supplementary Figs. 4 and 5). By further reducing the maximum constraint in strategy 5, we obtained a much younger age of 147.2–140.9 Ma for crown angiosperms (Supplementary Figs. 6 and 7). The HPD intervals were quite narrow for the three nodes, indicating that the calibrations were too restrictive and the maximum constraint might be too young.

Moreover, we found that the effective prior has a much older and narrower age range than the calibration range for crown angiosperms (Fig. 1). Particularly in strategies 1–4, the effective prior assigns less than 5% probability of age within 50 Myr of the oldest fossil. In extreme cases, most prior density was in the tail of the specified prior (e.g., strategy 2). These results indicated that without molecular data, the effective priors do not result in a crown age for angiosperms in the early Cretaceous. In other words, the effective prior structure excludes the user-specified prior assigned for the crown node of angiosperms. Analyses with molecular data revealed posterior

https://doi.org/10.1038/s41467-025-57687-9

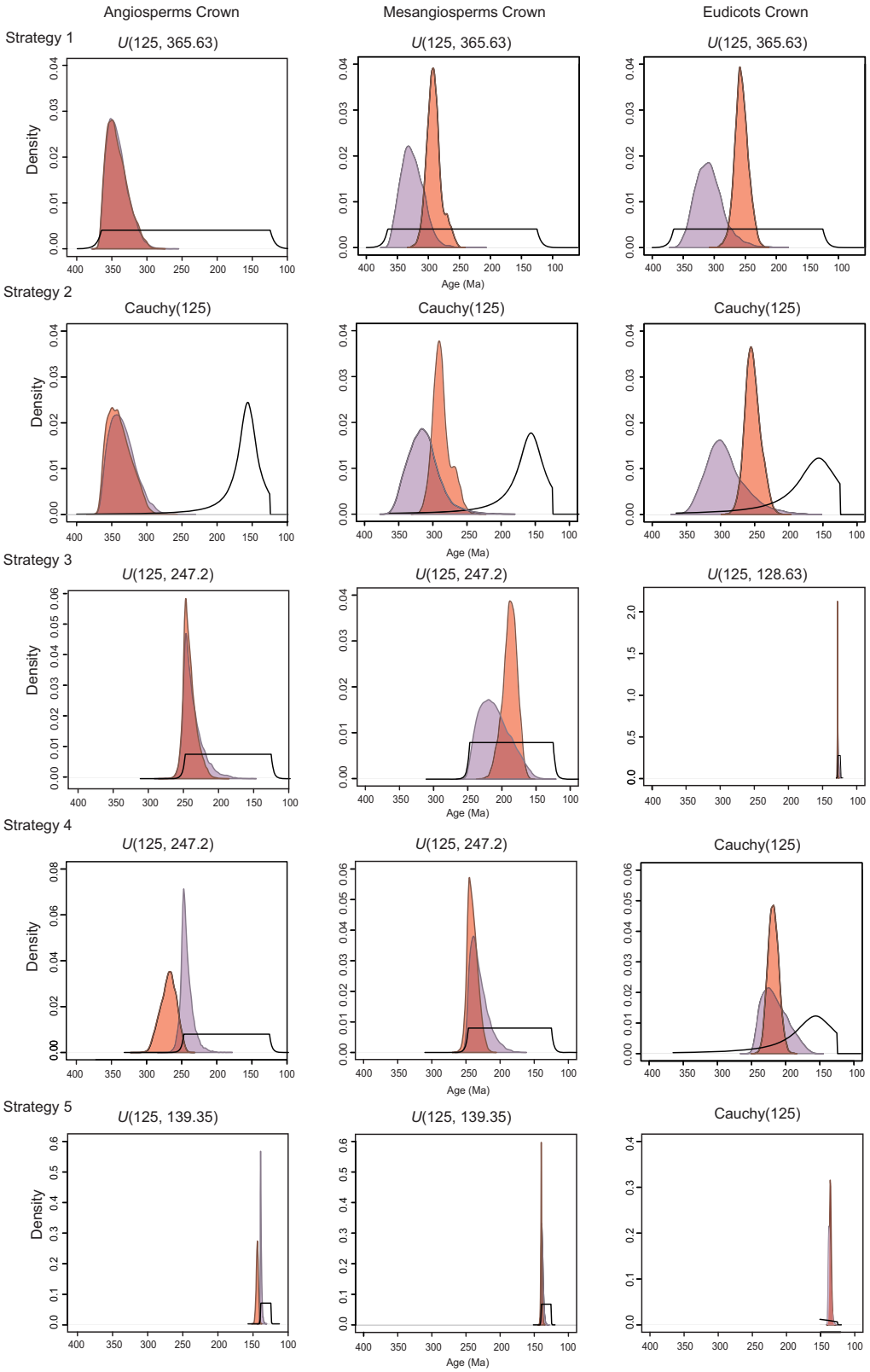


**Fig. 1 | Comparison of the three probability distributions for the age of crown angiosperms, mesangiosperms, and eudicots in the ND approach: user-specified prior (black), effective prior (purple), and posterior (orange).** Strategy 1: Uniform distribution (125, 365.63) was specified for nodes of crown angiosperms, mesangiosperms, and eudicots; strategy 2: Cauchy distribution was used on nodes of crown angiosperms, mesangiosperms, and eudicots; strategy 3: the soft maximum ages were changed from 365.63 Ma to 247.2 Ma for nodes of crown angiosperms and mesangiosperms, and the soft maximum age of 128.63 Ma was used on the node of crown eudicots; strategy 4: the nodes of crown angiosperms and mesangiosperms were assigned 247.2 Ma, and the node of crown eudicots was specified Cauchy distribution; strategy 5: the maximum constraints of crown angiosperms and mesangiosperms were assigned 139.35 Ma, and the node of crown eudicots was specified Cauchy distribution. Source data are provided as a Source Data file.

distributions that nearly overlap with the effective priors for angiosperms (strategies 1–3), implying that there is a strong dependence of the posterior estimates on the effective priors. Interestingly, the posteriors estimated in strategies 4–5 deviated from their effective priors (Fig. 1 and Supplementary Fig. 7). Additionally, we found that varying the birth and death rates in the five strategies did not affect the time estimations (Supplementary Fig. 8). Overall, our results indicated that the divergence times of angiosperm lineages were highly sensitive to the maximum constraints of the calibrated nodes. There is not yet any consensus on the evolutionary history of angiosperms among these calibrating strategies.

Given that the phylogenetic relationships of the five groups (Ceratophyllales, Chloranthales, magnoliids, monocots and eudicots) of mesangiosperms are still uncertain[10,11,20,36], we evaluated the impact of alternative phylogenetic hypotheses based on calibration strategy 4 (Supplementary Fig. 9A). The estimated divergence times of crown angiosperms, mesangiosperms, monocots, and eudicots inferred from different topological relationships were almost identical (Supplementary Fig. 9B). These results indicated that different phylogenetic relationships had little effect on the estimated ages of angiosperms, probably because there is equal support for the various relationships among the genes.

### Impacts of extant and fossil sampling schemes on time estimates under the skyline fossilized birth-death approach

To make direct use of the fossil occurrences rather than relying on controversial node calibrations to date the divergence times, we applied the SFBD approach assuming either random or diversified sampling of extant taxa, using 166 fossils and 349 extant angiosperm species. The results showed that the crown age of angiosperms under the diversified sampling (254.9–201.6 Ma) was similar to the random sampling (253.8–206.6 Ma) (Fig. 2A; Supplementary Figs. 10 and 11). The crown ages of major groups of angiosperms produced similar results, for instance, the mesangiosperms were estimated at 208.9–163.3 Ma under diversified sampling and 208.2–175.4 Ma under random sampling (Fig. 2A). Similarly, for shallow phylogenetic levels, the age intervals across most families under both sampling strategies were consistent (Supplementary Fig. 12). The effective priors under the two sampling strategies were generally older and had wider intervals (random sampling: 277.3–146.1 Ma; diversified sampling: 249.4–132.1 Ma) than posteriors for the crown ages of angiosperms. Meanwhile, we explored the effects of overestimated sampling fraction (at family level) on time estimations under the two sampling schemes, and found the effects were generally limited (Fig. 2B). To investigate whether the observed distribution of fossil ages was consistent with that implied by the SFBD model, we sampled fossil ages given posterior estimates of divergence times and SFBD rates. Overall, the generated fossils had wider age intervals than the observed fossil ages under diversified sampling and random sampling (Fig. 2C). In addition, two sampled stem-group angiosperm fossil ages do not fall within the observed age range, likely because the fossils close to the root are more sensitive to origin time. Despite unavoidable uncertainties, the sampled fossil ages were moderately correlated with the observed data (Fig. 2C), implying that the selected fossils are compatible with the SFBD model in use.

As the SFBD process was conditioned on the origin time ($t_{or}$), we further considered the impact of the choice of different priors for the origin on divergence time estimation. The dating results showed that the crown age of angiosperms was much less sensitive to the maximum age on the $t_{or}$ than under ND approach (Fig. 3). When the maximum

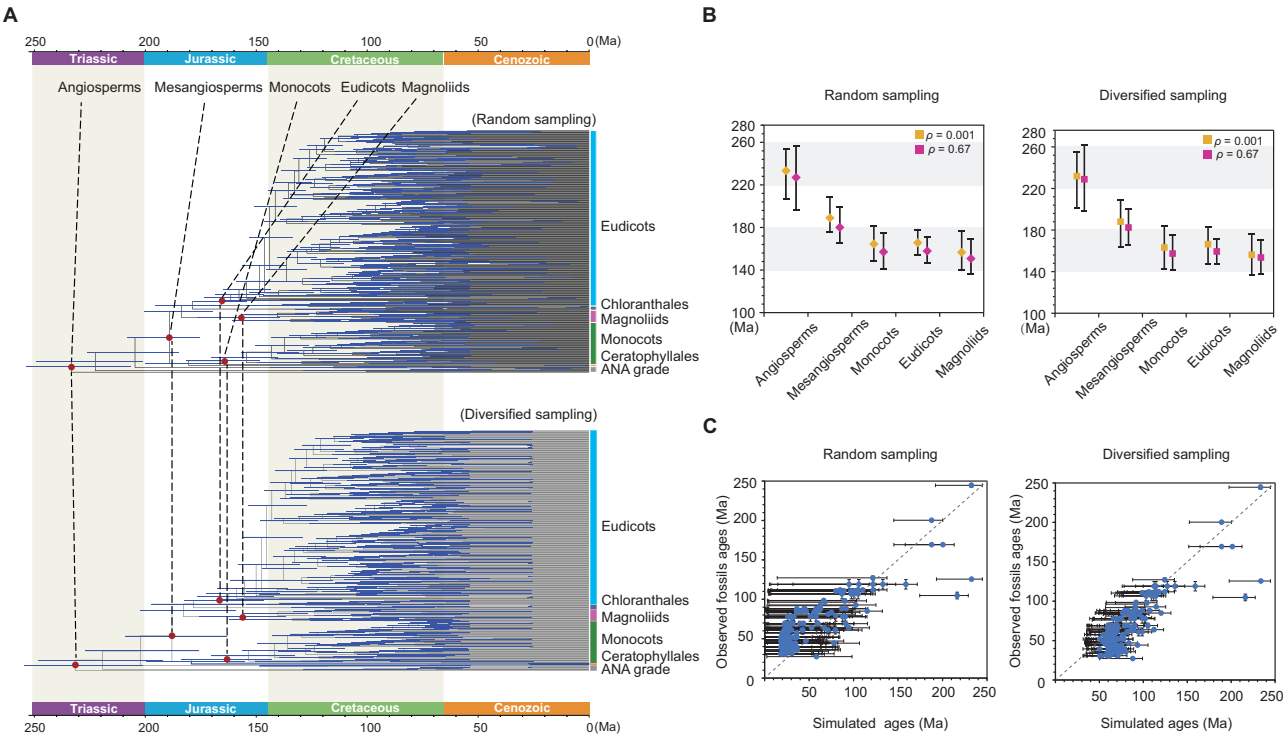


**Fig. 2 | Comparison of divergence times under different sampling schemes and sampling fractions, as well as observed and sampled fossil ages in the SFBD model. A** Time estimates of crown angiosperms and major groups under random and diversified sampling schemes. Node ages are plotted as the posterior medians, with horizontal bars representing 95% highest posterior density intervals. **B** Time estimates of crown angiosperms and major groups inferred under $\rho = 0.001$ (species level, orange) and $\rho = 0.67$ (family level, pink). Each analysis was conducted twice. **C** Simulated fossil ages given posterior medians of divergence times and SFBD rates under the SFBD model. The age was obtained from posterior predictive simulations by fixing the divergence times in the MCC tree and assuming that all fossils were tip fossils. A total of 166 observed fossil ages, plotted against times sampled under the SFBD model. Simulated fossil ages are presented as the estimated 95% HPD intervals and median time. Source data are provided as a Source Data file.

https://doi.org/10.1038/s41467-025-57687-9

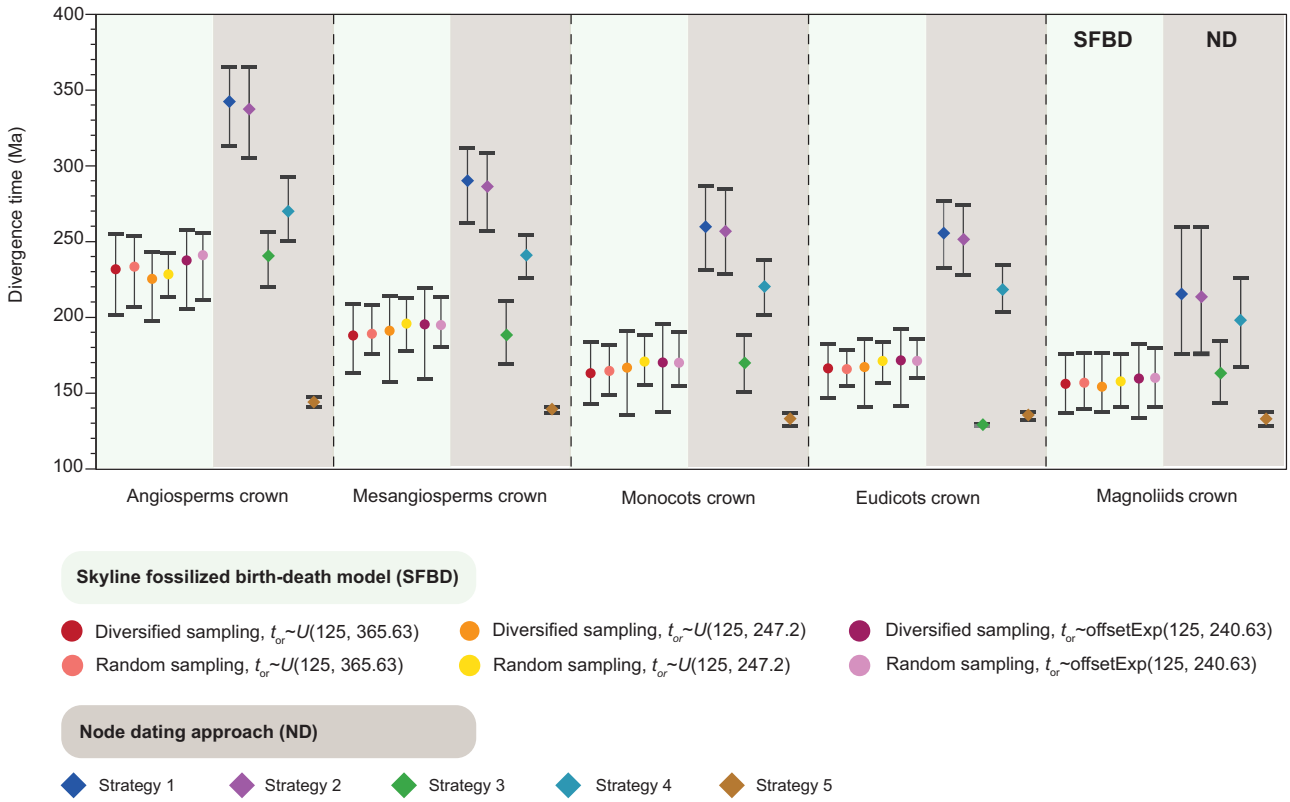


**Fig. 3 | Comparison of divergence time estimates of crown angiosperms and major clades using ND approach and the SFBD model under diversified and random sampling strategies.** In the SFBD analyses, posterior times were inferred from the main analyses under uniform prior (125, 365.63) for origin time $t_{or}$, and from comparative analyses under uniform prior (125, 247.2) and offset-exponential prior (125, 240.63). In the ND analyses, the five calibration strategies settings for the nodes of the crown angiosperms, mesangiosperms, and eudicots are shown in Fig. 1. Circles represent posterior medians in the SFBD model and rhombuses represent posterior means in the ND analyses; lines represent 95% HPD intervals. Each analysis was conducted twice. Source data are provided as a Source Data file.

**Table 1 | Divergence time (median values and 95% highest posterior density intervals) obtained under random and diversified sampling schemes under SFBD model using different strategies**

| Nodes | | Main analyses | Strategy A | Strategy B | Strategy C | Strategy D |
|---|---|---|---|---|---|---|
| Angiosperms crown (Ma) | random | 253.8–206.6 | 239.2–177.0 | 194.9–138.6 | 156.7–151.3 | 277.3–225.3 |
| | diversified | 254.9–201.6 | 253.7–194.1 | 191.3–151.0 | 270.1–210.2 | 281.0–228.5 |
| Mesangiosperms crown (Ma) | random | 208.2–175.4 | 186.5–155.8 | 152.5–121.4 | 150.7–141.6 | 233.8–187.4 |
| | diversified | 208.9–163.3 | 207.2–171.3 | 162.0–123.8 | 233.8–177.1 | 234.6–192.2 |
| Monocots crown (Ma) | random | 181.6–148.4 | 164.1–132.8 | 135.3–105.0 | 140.4–125.0 | 203.4–157.9 |
| | diversified | 183.8–142.9 | 182.5–146.7 | 145.4–107.7 | 205.3–148.7 | 204.9–167.0 |
| Magnoliids crown (Ma) | random | 176.6–139.6 | 162.0–129.3 | 133.2–108.6 | 143.3–128.0 | 190.7–145.1 |
| | diversified | 176.0–137.0 | 178.2–135.6 | 139.7–109.7 | 195.4–141.0 | 196.9–148.8 |
| Eudicots crown (Ma) | random | 178.1–154.3 | 166.8–139.2 | 131.6–110.4 | 141.2–130.7 | 201.5–164.5 |
| | diversified | 182.6–146.9 | 182.4–154.4 | 141.0–109.8 | 207.3–158.3 | 201.3–170.3 |

Main analyses: 166 fossils and 349 extant angiosperm species; Strategy A: 106 fossils represent the oldest fossils in each clade; Strategy B: 60 fossils represent the exclusion of the oldest fossil from each clade; Strategy C: 160 fossils are remaining after the removal of the stem group of angiosperm fossils; Strategy D: inclusion of outgroup for a total of 186 fossils.

value of $t_{or}$ was specified as 247.2 Ma, the crown age of angiosperms was restricted to this interval (diversified sampling: 242.6–197.1 Ma; random sampling: 242.4–213.2 Ma); while using an offset-exponential prior with mean of 240.63 Ma and offset of 125 Ma, the estimated time was similar with the main analyses (i.e., $t_{or}$~$U$(125, 365.63)) (Fig. 3).

Furthermore, we explored the effects of fossil sampling on estimating divergence times using four sampling strategies. Our results showed that when only including the oldest fossils (strategy A), the crown-group angiosperm time is largely in agreement with that estimated using all fossils (random sampling: 239.2–177 Ma; diversified sampling: 253.7–194.1 Ma) (Table 1 and Supplementary Figs. 13 and 14). In contrast, the corresponding age estimate from strategy B was relatively young (random sampling: 194.9–138.6 Ma; diversified sampling: 191.3–151 Ma) (Table 1, Supplementary Figs. 15 and 16). Removing fossils from stem-group angiosperms (strategy C) led to dramatically younger estimates of the crown age of angiosperms (156.7–151.3 Ma) and the major lineages under random sampling, while slightly older estimates under diversified sampling (Table 1, Supplementary Figs. 17 and 18). When the outgroup taxa with fossils were included (strategy D), we observed a much older crown age of angiosperms

(random sampling: 277.3–225.3 Ma; diversified sampling: 281–228.5 Ma) (Table 1, Supplementary Figs. 19 and 20). This is possibly due to different diversification and sampling patterns between the outgroup and angiosperms, which are not accommodated in the SFBD model.

### Evolutionary timescale of angiosperms

Using the two dating approaches, our results indicate that divergence times of angiosperms are sensitive to calibrations. However, we found that for the major groups of angiosperms, the width of the HPD intervals was similar between SFBD and ND approaches (~50 Ma) (Fig. 3), implying that the two methods are equally capable of estimating precision. There is no doubt that both approaches are undergoing continual refinement and have their own advantages and weaknesses. It is clear that the crown age of angiosperms depends on underlying assumptions about the fossil record and that there is still no consensus on maximum bounds in the ND approach. Given this, and our finding that assignment of different priors for the origin in the SFBD model resulted in relatively consistent crown ages of angiosperms compared to the ND approach (Fig. 3), we considered using the date estimates across SFBD main analyses under two sampling schemes to interpret the evolutionary history of angiosperms.

The SFBD analyses supported that crown-group angiosperms originated at 254.9–201.6 Ma (Fig. 4 and Supplementary Figs. 21 and 22). The crown group mesangiosperms originated in the late Triassic to late Jurassic (208.9–163.3 Ma). Apart from Amborellales, Nymphaeales and Austrobaileyales (namely ANA grade) diverged at 249–192.3 Ma and 227.2–175.8 Ma, respectively. The estimated timeline suggests that crown-group magnoliids arose in the early Jurassic to early Cretaceous (176.6–137 Ma). The Piperales and Canellales diverged at 162.1–111.1 Ma, Magnoliales and Laurales split at 155.9–126.4 Ma, with all four orders of magnoliids originating in the Cretaceous. We inferred that the crown-group monocots also arose during the early Jurassic to early Cretaceous (183.8–142.9 Ma). Within the monocots, the Acorales were the earliest to diverge, followed by the Alismatales splitting at 150.7–111.2 Ma. Furthermore, several radiations have occurred among the commelinids during the Cretaceous (128.8–103.3 Ma: Poales; 122.7–94.7 Ma: Commelinales and Zingiberales; 98.8–76.4 Ma: Zingiberales) (Fig. 4 and Supplementary Figs. 21 and 22).

We inferred an age of 182.6–146.9 Ma for crown eudicots. The orders corresponding to ~88% (31/35) of the core eudicots originated from the Aptian to the K-Pg (Cretaceous-Paleogene) boundary (125–66 Ma) in the Cretaceous (Fig. 4). This age range corresponds closely with the extensive angiosperm fossil record of this period (Supplementary Information), which implies that the Cretaceous was an important period in shaping the morphological diversity of angiosperms. This period has also been reported as a critical time for the co-evolution of angiosperms with insects, birds, and mammals[37–40], which are key components of contemporary terrestrial biodiversity. Among the nine species-rich families of beetles (>10,000 species), seven of them originated in the Cretaceous and feed on plants or decomposing plant materials and woody tissues[41]. The recent discovery of a remarkably well-preserved short-winged flower beetle from mid-Cretaceous amber offers direct evidence of pollen-feeding in a Cretaceous beetle[42]. This amber was associated with pollen aggregations and coprolites consisting mainly of pollen, confirming that diverse beetle lineages visited early-diverging lineages of angiosperms in the Cretaceous. By examining the evidence for matching dates of origination of crown angiosperms and key insect pollinator lineages, the study highlighted the importance of the K-Pg extinction event in the context of angiosperm and insect co-speciation[37]. These results not only contribute to the elucidation of the origin and evolution of angiosperms, but also provide a theoretical basis for understanding key questions such as the diversification of other lineages that have co-evolved with angiosperms.

## Discussion

### The crown age of angiosperms mostly depends on underlying assumptions about the fossil record in the node dating approach

A puzzling range of angiosperm crown ages, from early Cretaceous to Permian, have been reported across many ND analyses[10–13,21]. In addition to the varying amounts of molecular data, different maximum bounds were used for the calibration of the node of crown-group angiosperms in these studies[12,13,18]. For example, the recent study[43] dated the angiosperm tree with two different maximum constraints at the angiosperm crown node, and showed that these different constraints greatly affected age estimates across angiosperms. In our ND analyses, we showed that different maximum bounds conferred considerable influence on inferring the crown age of angiosperms. For instance, when the maximum bound was relaxed from 139.35 to 365.63 Ma, the crown age of angiosperms changed from Cretaceous to Carboniferous (Strategy 1 and 5 in Fig. 1). Recently, Budd and Mann[44] revealed that the dependence of posterior estimates on the priors and the problems of range truncations are typical of ND approaches in general, especially for the deep nodes. In our study of the crown node of angiosperms here, this truncation is striking: the user-specified prior for all strategies was not the same as the effective prior implemented in the estimation of divergence times (Fig. 1). This deviation from the user-specified prior was described as "pseudo-data"[45]. Thus, we argue that the effective priors cannot accurately reflect the paleontological constraints on divergence time estimation in our dataset.

Multiple analyses have demonstrated a strong correlation between posterior age estimates with effective priors[44,45]. In our study, these similar distributions of effective prior and posterior on the crown node of angiosperms in strategies 1–3 highlighted that the effective prior had an overwhelming influence on the posterior. Because the effective priors, which were actually taking effect in dating, are similar in strategies 1 and 2, the posterior estimates for the ages of these three nodes are similar, despite the different shapes of the specified priors in strategies 1 and 2. We noted that the effective prior at the crown node of the angiosperms was remarkably similar to that observed in Mesangiosperms in strategies 4 and 5. At the crown node of mesangiosperms, the overlap area between the posterior and effective prior distributions is large and their peaks are close. To satisfy the condition that ancestral nodes are older than descendant nodes, the posterior distribution was unsurprisingly pushed to deviate from the effective prior of the crown node of angiosperms.

It is well known that the reliability and precision of fossil calibrations have a significant impact on time estimates even with an infinite amount of data[23]. Thus, the molecular data could not explain or support old crown ages of angiosperms in most fossil-calibrated molecular dating analyses. It has been showed that the time information in "molecular dating" approaches is mostly coming from the fossil maximum calibrations and not the molecules[46]. Neither genes, taxa, nor internal calibrations appear to be the key to resolving the crown age of angiosperms. Underlying assumptions about the fossil record and the assignment of different maximum bounds, in particular, can easily shape the origin of crown angiosperms into different epochs (e.g., Cretaceous, Triassic, Permian, or earlier).

### Further exploration in the FBD approach to refine the timeline of angiosperms

One appealing feature of the FBD model is that it unifies extinct and extant species in a single macroevolutionary model. As for the mystery of angiosperm origins, many studies expect to use tip dating methods to address the time gap between the fossil record and molecular dating estimates[45,46]. In our study, we inferred the age interval for crown angiosperms (essentially, the Triassic Period) using the most extensive and reliable fossil record under the SFBD model. However, there is still plenty of room that could potentially improve age estimates. One

https://doi.org/10.1038/s41467-025-57687-9

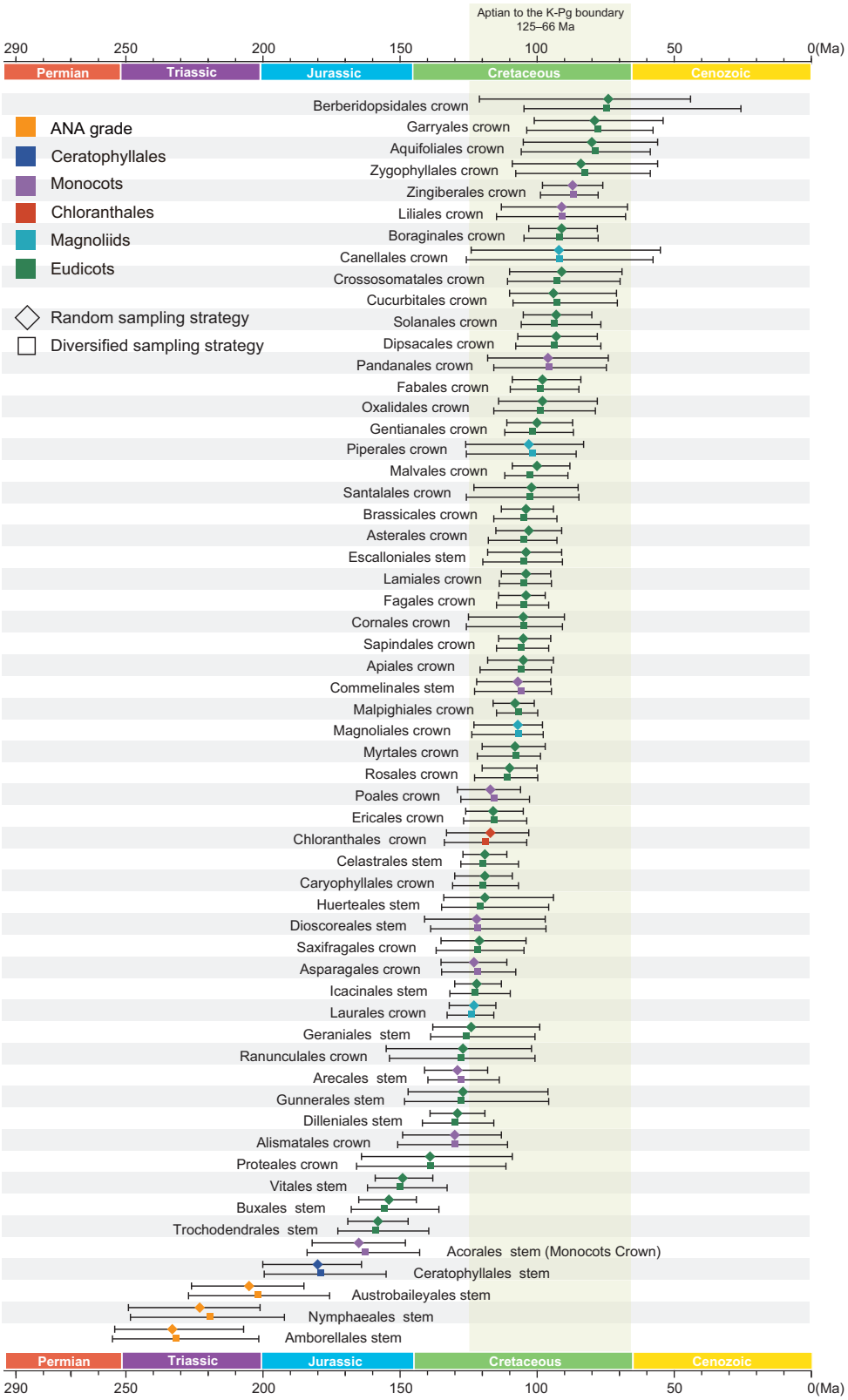


**Fig. 4 | Node age estimates for the major clades of angiosperms under diversified and random samplings using the SFBD model.** The median and 95% highest posterior density interval are depicted for each key node. The Aptian to the K-Pg (Cretaceous-Paleogene) boundary (125–66 Ma) was the period when most of the ordinal and interordinal angiosperm clades diversified. Rhombus box symbols represent random-sampling strategy and square boxes represent diversified-sampling strategy. Each analysis was conducted twice. Source data are provided as a Source Data file.

further effort is coding and including morphological characters in the FBD dating analysis, that is, TED. In the absence of morphological data, incorporating prior knowledge from previous studies to establish accurate topological constraints of fossil placements has important implications for divergence time estimation[47,48]. The dating approach will become refined with TED, whereby the relationship of fossils and extant taxa is instead more inferred from the morphological characters. Recently, a simulation study revealed that incorporating morphological data in TED can improve the accuracy and precision of time estimates[48]. Although constructing effective models of morphological evolution and accounting for the non-random nature of missing characters in fossil data is considerably challenging, incorporation of morphological characters in tip dating is a trend in molecular dating analyses. Another effort would be to include more fossils in the analysis, despite the lack of morphological data. It is worth noting that there are still thousands of angiosperm fossils available and the 166 fossils in our dataset may not be a representative subsample. However, it is typically computationally intractable with large fossil datasets, especially because the placement of fossils needs to be averaged using Markov chain Monte Carlo (MCMC). Future reliable identification of morphological characters and the development of morphological models and efficient computational methods would provide a solid basis to narrow down the range of plausible ages for crown angiosperms.

Our SFBD dating analyses show that the oldest fossils and the stem fossils exert the influence in estimating time, as the prior times are basically informed by the SFBD process. Fossils close to the root or in the stem are particularly valuable as they provide information of early divergence events. We also discovered that including outgroup taxa can result in older ages of the internal groups, such as the crown ages of angiosperms, monocots, and eudicots (Table 1). The SFBD model assumes that the diversification and fossilization rates are the same across lineages (although they can vary among different time intervals), which is unrealistic for angiosperms and gymnosperms. The two groups also have different extant sampling fractions (~0.001 vs. ~0.015). Given the absence of methods that account for variation in diversification and sampling rates among clades in the current FBD model (but see studies when the fossil-sampling rate is constant[49,50]), it is crucial for future studies to acknowledge potential bias when including outgroup sampling.

Although the diversified sampling could somewhat avoid the unrealistic assumption of exhaustive or random sampling of extant taxa in large, species-rich clades[26], we should acknowledge that both random and diversified sampling assumptions are ideal cases and empirical datasets (including this one) typically do not employ either of them exactly as specified by the model. Under diversified sampling, the reconstructed tree has long terminal branches that maximize the tree length of the chronogram of extant taxa, which rely on the high sampling fraction[48]. The angiosperms have ~350,000 extant species and a rich fossil record. According to the assumption of diversified sampling, when the sampling fraction of our empirical analyses was fixed at 0.001, it implied that the ~349 extant species we sampled should represent the oldest diversified clades. In reality, taxa are usually sampled on the basis of their taxonomic ranks and other factors (e.g., specimen availability and a balanced sampling between various groups). These sampled lineages would not necessarily be the same as those defined by a single cutoff time in the chronogram. Therefore, based on our empirically sampled dataset, the cutoff time will be underestimated. It is unlikely that the assumption of diversified sampling can be matched no matter how the cut-off is changed. Thus, it is important to use a model, such as the SFBD model, that is robust to model violations in dating divergence times, at least in our empirical dataset. Nevertheless, it is crucial to explore various aspects that could violate the model and explore their effects on the inference.

### Insights into the puzzle of the origin of angiosperms

The crown age of angiosperms has been one of the most thought-provoking topics in evolutionary biology. In recent years, there has been some progress in the application of molecular dating, such as detecting potential bias due to habit-related molecular rate heterogeneity (herbaceous and woody)[19], considering the uncertainties in dating analyses[9,15] and expanding sampling densities of the genes, species, and fossil calibrations[10,46]. Most of these studies supported the hypothesis that crown-angiosperms originated before the Cretaceous, but inherent biases in the molecular dating methods were recognized[4,45,51]. A case as demonstrated in our ND results, the effective prior pushed the crown age of angiosperms towards much older than expected by the user-specified prior, and various maximum boundaries would determine the crown age of angiosperms. To advance our understanding of angiosperm macroevolution, several studies emphasize the importance of future fossil discoveries and innovative macroevolutionary model development, including tip dating approaches[15,46,52]. Our study inferred a Triassic origin of angiosperms using the SFBD model, which conflicts with the sequence of appearance of fossils[4]. Although our dating results are robust to sampling assumptions of extant taxa, the limited number of fossils and extant species sampled, as well as the effect of the specified prior for $t_{or}$, may have resulted in the pre-Cretaceous origin of angiosperms that we obtained. These factors also may explain the origin of some angiosperm orders that are not entirely compatible with the fossil record. For example, the Vitales, with an inferred Jurassic age, whose oldest reliable fossils are dated from the Late Cretaceous[53]. Additionally, these two dating methods make different assumptions about whether speciation and extinction rates change through time in our study. We observed that the inferred net diversification and turnover rates varied over time in the SFBD analyses, as shown in the plotted skylines (Supplementary Fig. 23). While the implemented ND model assumes birth and death rates which are constant throughout the evolutionary process in the MCMCTREE. Although it has been shown that the most important factor affecting time estimation in ND is fossil calibration, future research could provide a "fair" comparison (i.e., allowing birth and death rates to change through time) between ND and FBD approaches.

Another possible explanation for the absence of Jurassic crown angiosperm fossil evidence, including that of pollen, is that the early representatives were geographically restricted. This could happen if they were confined to a region not represented in the fossil record for that time period[54]. The investigation into the diversification patterns of angiosperm flowers and leaves revealed that angiosperms emerged during the Cretaceous[6,55]. But the origin time of angiosperms can date back even earlier if pre-Cretaceous angiosperms exhibited a low level of morphological diversification[56,57]. A recent study[44] indicated that fossils alone provide strong evidence of origin that biologists are keenly interested in, such as in the bilaterian animals and placental mammals. As the fundamental basis of the origin and early evolution of angiosperms, palaeobotanical records need to be critically investigated. Especially for pre-Cretaceous fossils, an unambiguous documentation of the diagnostic structural features of angiosperms is essential for distinguishing between other extant and extinct seed plants[4,58]. Recently, much attention has been paid to which clades of Mesozoic gymnosperms are possible ancestors or sister branches of flowering plants. It has been reported that the fossil found in Upper Mongolia and the earlier reported fossils with cupules were close relatives of angiosperms[59], which may imply that ancestral taxa of angiosperms emerged as early as about 250 Ma (early Triassic), although those seed-bearing organs might be broadly comparable to extant conifer seed cones[60,61]. The fossil *Xadzigacalix quatsinoensis* was regarded as possibly constituting a new order of seed plants, and perhaps has affinities to gnetophytes or angiosperms[60]. The study modeled the diversity trajectories of angiosperm families based on

evidence from ~15,000 fossils, reporting that flowering plants originated in the Jurassic or even earlier[62]. It has been controversial that this approach does not directly estimate the crown age of angiosperms, but rather extrapolates from the estimated stem ages of an extensive sampling of angiosperm families[63]. Nevertheless, these analyses have implied the possibility of a pre-Cretaceous origin of angiosperms.

Overall, our dating analyses provide valuable insight for the timescale of early angiosperm evolution by integrating a large number of reliable fossils and using two Bayesian dating methods. Our results demonstrate why the crown age of angiosperms inferred using ND varies among different studies and indicate that the crown age of angiosperms is essentially conditioned by the maximum bound in the calibrations and the effective priors. Furthermore, the sensitivity analyses using the SFBD model highlighted the dominant influence of the oldest fossils and the potential bias of time estimates from inclusion of outgroup taxa. Compared to ND analyses, we found that the crown age of angiosperms was much less sensitive to the maximum age on the origin time in the SFBD model. Based on the SFBD analyses, we inferred the crown age of angiosperms to be essentially in the Triassic Period when mammals, dinosaurs, and squamate reptiles first appeared in the fossil record and which immediately followed the largest known mass extinction, the Permian-Triassic extinction event. Considering that there are thousands of angiosperm fossils available and most of them have not been coded with morphological characters, further progress will require combined efforts from paleontologists and evolutionary biologists to build densely sampled trait datasets and integrate paleontological data into the study of angiosperm macroevolution.

## Methods

### Extant taxon sampling

We sampled 670 angiosperm taxa (representing 669 angiosperm species) including 504 eudicots, 121 monocots, 29 magnoliids, four Chloranthales (five transcriptomes of Chloranthales representing four species), two Ceratophyllales, and nine early-diverging angiosperms, representing 58 orders and 278 families. Another 78 gymnosperm taxa were used as an outgroup. Our data of the 144 genomes and 579 transcriptomes were obtained from the public databases, and other 25 transcriptomic data were downloaded from the NCBI SRA database and assembled by Trinity v2.4.0[64] with default settings (min_kmer_cov = 2). The complete list of data information is provided in Supplementary Data 2.

### Orthology inference

The orthology inference was based on identifying homologs followed by cleaning, aligning, and cutting homolog trees[65]. First, we selected the complete nuclear genomes of 19 representative species (*Ginkgo biloba*, *Papaver somniferum*, *Amborella trichopoda*, *Nymphaea colorata*, *Ceratophyllum demersum*, *Vitis vinifera*, *Actinidia eriantha*, *Coffea canephora*, *Helianthus annuus*, *Arabidopsis thaliana*, *Eucalyptus grandis*, *Prunus persica*, *Musa schizocarpa*, *Oryza sativa*, *Asparagus officinalis*, *Zostera marina*, *Cinnamomum kanehirae*, *Liriodendron chinense*, *Piper nigrum*) to generate putative orthologous groups (OGs). The CD-HIT (-c 0.995 -n 5) was used to reduce amino acid redundancy[66]. Initial homology search was carried out using all-by-all BLASTP of 19 complete genomes with an *E* value cutoff of 10 and -max_target_seqs 1000. We used a hit_fraction cutoff of 0.5 and set "IGNORE_INTRASPECIFIC_HITS = True" to filter raw blast output. The filtered hits were used as input for MCL (MCL v14-137)[67], with the *E* value cutoff set to $10^{-5}$ and an inflation value of 1.5. Clusters containing fewer than 19 species were further removed. Then, we built the 3503 homolog trees by performing the tree estimation, trimming, masking, and cutting deep paralogs (keep the parameter default). Finally, we used homologs that had no duplicated taxon and are one-to-one orthologs to prune all homologous gene trees and inferred a set of 902 orthologs. The transcriptomic and genomic data of the remaining 728 taxa were then incorporated into 902 OGs through Hmmsearch (HMMER v3.2.1)[68].

### Phylogenetic analyses

All 902 OGs were aligned using MAFFT v7.310[69] with the option"-localpair -maxiterate 1000" and then trimmed using Gblocks 0.91b[70], allowing 50% gap positions. To remove the short sequences, trimAL v1.4[71] was used for pruning with the option resoverlap 0.5 -seqoverlap 50. All alignments were discarded if the length was below 100 amino acids or species coverage was lower than 90%. Finally, 324 OGs from a diverse selection of 748 taxa were obtained for phylogenetic inference. We reconstructed phylogenetic relationships using the concatenated supermatrix that was partitioned by gene in IQ-TREE v2.0.5[72]. The best-fit substitution model for each partition was determined and branch support was assessed with 1000 ultrafast bootstrap replicates.

### Divergence time estimation

**Molecular dataset and fossil occurrences**. Our divergence time analyses were conducted using two Bayesian methods: node-dating (ND) and SFBD approaches. We performed preliminary analyses of the concatenated data set (324 OGs for 748 taxa), but such a large dataset is computationally intractable. Previous study showed that the time estimates were largely consistent in full compared to reduced data sets, and suggested selecting a subset of informative genes for molecular dating analyses[9]. "Clock-like" genes were defined as those evolving in a clock-like manner[73], which can minimize errors associated with model mis-specification and have been used for molecular dating[9,22,74]. Therefore, we used the SortaDate[74] to identify the clock-likeliness of each gene. According to the three criteria sequentially: "1 = minimal root-to-tip variance, 3 = least topological conflict against the species tree, and 2 = tree length (discernible information content)", the top 100 genes of 324 OGs were selected for molecular-clock dating analyses. In addition, we reduced taxa at the genus-level of eudicots, monocots, and gymnosperms, but retained the diversity of all sampled families. The final dataset included 27,326 amino acids representing 100 nuclear genes from 364 taxa (349 angiosperm species and 15 gymnosperm species as an outgroup) for the dating analyses.

Following best practice principles of fossil relationships[75], we selected 186 fossils that could reliably indicate the occurrence of seed plant clades down to family level. These contained 160 angiosperm fossils, 6 stem angiosperm lineage fossils, 11 gymnosperm fossils, and 9 extinct seed lineage fossils. The detailed fossil information including fossil ages, locality, fossil type, and relationships with taxa in the tree, are provided in Supplementary Data 3 and Supplementary Information. Most of (not all) the fossils were cited from the previous study[14].

**Node dating analyses**. For the ND approach, divergence times were estimated using the program MCMCTree in the PAML v4.9j[76], based on a fixed tree topology of the 364 extant taxa. The 27,326 amino acids were treated as a single partition, and the substitution model used was LG + F + Γ4.

To evaluate the relative fit of two relaxed-clock models (i.e., autocorrelated-rates model and independent-rates model) to the datasets, we selected 32 *b* values to estimate the marginal likelihood of a given model using a "stepping-stone" algorithm in MCMCTree. The output of MCMCTree was parsed and the marginal likelihood for the chosen relaxed-clock model was calculated using the R package *mcmc3r*[77].

We consistently used a time unit of 100 MY (million years) across the ND analyses. For the evolutionary rate, we used the better-fit independent-rates model which assumes the branch rates follow the same lognormal distribution. To specify a sensible prior on the mean rate, we calculated the genetic distance between *Arabis alpina* and *Ginkgo biloba* based on LG + F + Γ4 model using CODEML[76]. Fossil evidence suggested that the divergence time of the two species is

~307 Ma[78], which indicated that the mean substitution rate was assigned to a gamma hyperprior $G(2, 26)$ with 2/26 = 0.076 amino acid substitutions per site per 100 MY. The variance (rate-drift) parameter $\sigma^2$ was assigned $G(1, 10)$.

The tree prior follows a birth-death process. As MCMCTree does not support estimating the speciation and extinction rates, we had to use fixed values for them. Based on the diversification rate estimates from previous sudy[79], we set the birth rate ($\lambda$) at 9.88 and the death rate ($\mu$) at 2.77. In addition, we assessed the impact of varying birth and death rates ($\lambda = 0.988$, $\mu = 0.277$) on the time estimates. The sampling fraction of extant species ($\rho$) was set to 0.001 (364 species compared with the 352,000 extant angiosperms species) (https://www.mobot.org/MOBOT/research/APweb/). The root age was constrained to be between 307 and 365.63 Ma, based on the oldest fossil of Acrogymnospermae[78] and the first record of seeds in the form of preovules that satisfy the criteria of the seed habit[13].

We applied 114 fossil calibrations. In most cases, the uppermost boundary of the stratigraphic interval of the fossil attribution can be used as the minimum age for the calibration distribution. Nodes 1–9 were set with uniform distributions with soft bounds ($pL = pU = 0.025$). Truncated Cauchy distributions with hard minimum bounds ($pL = 1e^{-300}$) were used at nodes 11–12, 14–38 and 40–114 (Supplementary Data 3, Supplementary Figs. 2–6). We implemented five strategies to test the effects of different maximum bounds on nodes 10 (crown angiosperms), 13 (crown mesangiosperms), and 39 (crown eudicots). In strategy 1, we used uniform distribution with 365.63 Ma as the maximum age for three nodes, which is based on the first record of seeds in the form of preovules that satisfy the criteria of the seed habit[13]. In strategy 2, Cauchy distributions truncated at 125 Ma were used on nodes of crown angiosperms, mesangiosperms, and eudicots. In strategy 3, we changed the maximum age in strategy 1 from 365.63 Ma to 247.2 Ma for nodes of crown angiosperms and mesangiosperms, based on sediments devoid of angiosperm-like pollen below their first report in the middle Triassic[80]; and applied the maximum age of 128.63 Ma for node of crown eudicots, based on the maximum age of the oldest potential age of tricolpate pollen[80]. In strategy 4, the nodes of crown angiosperms and mesangiosperms have the same specified priors as in strategy 3, while crown node of eudicots was assigned a Cauchy distribution. In strategy 5, following previous study[12], we used we used 139.35 Ma for nodes of crown angiosperms and mesangiosperms, and Cauchy distribution for node of crown eudicots. Furthermore, to explore the effects of different topologies on the estimation of angiosperm crown age, we selected four topologies published in recent years[10,11,20,36] and estimated the timeline of angiosperms based on strategy 4.

The MCMC analyses were run for 5 million generations sampled every 500 generations after discarding 1 million generations as burn-in. We conducted each analysis twice to check consistency and confirmed the effective sample sizes (ESS) of all parameters were larger than 200 by Tracer v.1.7.1[81].

**Skyline fossilized birth-death dating analyses.** For the SFBD approach, we used the BDSKY v1.5.0 package for BEAST v2.7.5[29,82] to infer the divergence times.

When morphological characteristics of extant species and fossils are unavailable, setting both accurate and precise topological constraints of fossil placements should be beneficial for analyses on FBD trees[48]. Following best practice principles[75], we scrutinized the placement of fossils used in our study. Instead of integrating over the uncertainties of fossil placements through MCMC, which is computationally intractable in our case, we fixed the tree topology, as well as the position of the fossils in the tree. Detailed information on the placement of the fossils in the tree were available in the Supplementary Information. The fossil ages were given uniform distributions according to their stratigraphic intervals (Supplementary Data 3) and were sampled as part of the divergence time estimation. As the gymnosperms likely have different diversification and sampling patterns from the angiosperms, which would violate the SFBD model assumption, we used 349 angiosperm species and 166 fossils (including 6 stem group of angiosperm fossils) in the main analyses. Time was divided into ten intervals equally spaced from the root to the present. The speciation, extinction, and fossil-sampling rates were all assumed independent across intervals. We specified diffuse Beta(1,1) prior for the turnover ($\mu/\lambda$) and fossil sampling proportion ($\psi/(\psi+\mu)$), and Exp(1) prior for the diversification rate ($\lambda - \mu$) in each interval. The sampling fraction of extant species ($\rho$) was fixed to 0.001 as in the ND analyses, and the sampling strategy was assumed either random or diversified[26,30]. But we also set $\rho = 0.67$ (sampling fraction at family level) to explore the effects of overestimated $\rho$ on time estimation. Considering the relationships of many stem branches of angiosperms remain highly uncertain and should not be relied upon[46], we conservatively conditioned the SFBD process on the origin time ($t_{or}$) using the minimum age constraint at 125 Ma based on the oldest fossil of angiosperms. The maximum age constraint at 365.63 Ma for $t_{or}$ is according to the first records of seeds in the form of preovules that satisfy the criteria of the seed habit, which represents a soft maximum constraint on the divergence of crown seed plants[13]. Since the true origin of the clade must have taken place before the oldest fossil date from within the clade, but we do not know how much earlier the origin was than this fossil. Therefore, we assigned the $t_{or}$ a uniform distribution (125, 365, 63), assuming that the origin time has the same probability in this interval.

To explore the impact of the choice of different priors for the $t_{or}$ on divergence time estimation, we performed two additional analyses: (1) the origin time was assigned to an offset-exponential prior with offset at 125 Ma and mean of 240.63 Ma; (2) the origin time was set to a uniform prior (125, 247.2). Furthermore, we employed four strategies to explore the effects of fossil sampling on estimating divergence times in the SFBD approach. In strategy A, we selected the oldest fossils from the clades containing fossils, forming a subset of 106 fossils. In strategy B, we excluded the oldest fossil from each clade and the remaining 60 fossils were used to perform the dating analyses. In strategy C, we excluded the 6 stem angiosperm fossils and used the remaining 160 fossils. In strategy D, the 11 gymnosperm fossils, 9 extinct seed lineages fossils, and 15 gymnosperm outgroup species (186 fossils and 364 extant species) were all used for dating. The prior for the origin time was changed to $U(307, 365.63)$ to accommodate the inclusion of the outgroup.

Still, the amino acids were treated as a single partition, and the substitution model used was LG + Γ4 (BEAST does not support +F). We also used the independent lognormal relaxed clock model[83] for the evolutionary rate, and the implementation has been optimized to improve its computational performance[84]. We also calculated the amino acid pairwise distance between *Amborella trichopoda* and *Actinidia chinensis* with LG + Γ4 + F model by the package CODEML in PAML package v4.9j. According to *Tricolpate pollen grain* fossil placement, the divergence time between the two species was about 125 Ma, which indicated that the mean rate was assigned to $G(2, 0.00085)$ with $2 \times 0.00085 = 0.0017$ substitutions per site per time unit (1 My) (note that the change of time scale and that BEAST used shape-scale parameterization while MCMCTree used shape-rate parameterization in which rate = 1/scale). The standard deviation of the rate was given $G(2, 0.05)$ prior with mean 0.1.

For each run, the MCMC was executed for 400 million generations with sampling every 10,000 generations and a discarded burn-in fraction of 0.2. We ran each analysis twice and checked consistency and ESS > 200 for important parameters with Tracer v.1.7.1. Maximum-clade-credibility (MCC) trees were summarized from the combined samples using TreeAnnotator v2.6.0. To investigate the dating results and for better comparison with the ND results, we

pruned the fossils from the MCC trees to produce annotated trees containing only extant taxa.

For all the analyses, we explored the effective priors for the divergence times induced by the SFBD process (by fixing the likelihood to 1.0 in the MCMC). Besides, we also performed a simple predictive check for the fossil ages in the main analyses. Specifically, we fixed the divergence times in the MCC tree, the speciation, extinction, and fossil-sampling rates in each time interval were all to their posterior medians. All fossils were given $U(0, 500)$ distributions for their ages, and were all assumed as tips (this treatment is because the MCMC sampler does not support moving fossil ancestors along the tree). The fossil ages were then sampled using MCMC and were compared with the actual ages used in the dating inference.

## Data availability

The sequence alignments, inferred phylogeny and divergence time trees have been deposited in the Figshare: https://doi.org/10.6084/m9.figshare.25333804. Supplementary Information and Source data are provided in this study. Taxon sampling used in the study is provided in Supplementary Data 2, and fossil information is described in Supplementary Data 3. Source data are provided with this paper.

## References


1. Condamine, F. L., Silvestro, D., Koppelhus, E. B. & Antonelli, A. The rise of angiosperms pushed conifers to decline during global cooling. *Proc. Natl. Acad. Sci. USA* **117**, 28867 (2020).
2. Schneider, H. et al. Ferns diversified in the shadow of angiosperms. *Nature* **428**, 553–557 (2004).
3. Buerki, S., Forest, F. & Alvarez, N. Proto-South-East Asia as a trigger of early angiosperm diversification. *Bot. J. Linn. Soc.* **174**, 326–333 (2014).
4. 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).
5. Feild, T. S., Arens, N. C., Doyle, J. A., Dawson, T. E. & Donoghue, M. J. Dark and disturbed: a new image of early angiosperm ecology. *Paleobiology* **30**, 82–107 (2004).
6. Friis, E. M., Crane, P. R. & Pedersen, K. R. *Early Flowers and Angiosperm Evolution* (Cambridge University Press, 2011).
7. Hughes, N. F. & McDougall, A. B. Records of angiospermid pollen entry into the english early cretaceous succession. *Rev. Palaeobot. Palyno.* **50**, 255–272 (1987).
8. Hughes, N. F. & McDougall, A. B. Barremian-Aptian angiospermid pollen records from southern England. *Rev. Palaeobot. Palyno.* **65**, 145–151 (1990).
9. Foster, C. S. P. et al. Evaluating the impact of genomic data and priors on Bayesian estimates of the angiosperm evolutionary timescale. *Syst. Biol.* **66**, 338–351 (2017).
10. Li, H. et al. Origin of angiosperms and the puzzle of the Jurassic gap. *Nat. Plants* **5**, 461–470 (2019).
11. Yang, L. et al. Phylogenomic insights into deep phylogeny of angiosperms based on broad nuclear gene sampling. *Plant Commun.* **1**, 100027 (2020).
12. 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).
13. Bell, C. D., Soltis, D. E. & Soltis, P. S. The age and diversification of the angiosperms re-revisited. *Am. J. Bot.* **97**, 1296–1303 (2010).
14. Ramírez-Barahona, S., Sauquet, H. & Magallón, S. The delayed and geographically heterogeneous diversification of flowering plant families. *Nat. Ecol. Evol.* **4**, 1232–1238 (2020).
15. Barba-Montoya, J., Dos Reis, M. & Yang, Z. Comparison of different strategies for using fossil calibrations to generate the time prior in Bayesian molecular clock dating. *Mol. Biol. Evol.* **114**, 386–400 (2017).
16. Donoghue, P. C. J. & Yang, Z. The evolution of methods for establishing evolutionary timescales. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **371**, 20160020 (2016).
17. Ho, S. Y. W. & Phillips, M. J. Accounting for calibration uncertainty in phylogenetic estimation of evolutionary divergence times. *Syst. Biol.* **58**, 367–380 (2009).
18. 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).
19. Beaulieu, J. M., O'Meara, B. C., Crane, P. & Donoghue, M. J. Heterogeneous rates of molecular evolution and diversification could explain the Triassic age estimate for angiosperms. *Syst. Biol.* **64**, 869–878 (2015).
20. Zeng, L. et al. Resolution of deep angiosperm phylogeny using conserved nuclear genes and estimates of early divergence times. *Nat. Commun.* **5**, 4956 (2014).
21. Zhang, Z. et al. Origin and evolution of green plants in the light of key evolutionary events. *J. Integr. Plant Biol.* **64**, 516–535 (2022).
22. Su, D. et al. Large-scale phylogenomic analyses reveal the monophyly of bryophytes and Neoproterozoic origin of land plants. *Mol. Biol. Evol.* **38**, 3332–3344 (2021).
23. Yang, Z. & Rannala, B. Bayesian estimation of species divergence times under a molecular clock using multiple fossil calibrations with soft bounds. *Mol. Biol. Evol.* **23**, 212–226 (2006).
24. Ronquist, F. et al. A total-evidence approach to dating with fossils, applied to the early radiation of the hymenoptera. *Syst. Biol.* **61**, 973–999 (2012).
25. May, M. R. et al. Inferring the total-evidence timescale of marattialean fern evolution in the face of model sensitivity. *Syst. Biol.* **70**, 1232–1255 (2021).
26. Zhang, C., Stadler, T., Klopfstein, S., Heath, T. A. & Ronquist, F. Total-evidence dating under the fossilized birth–death process. *Syst. Biol.* **65**, 228–249 (2016).
27. Heath, T. A., Huelsenbeck, J. P. & Stadler, T. The fossilized birth–death process for coherent calibration of divergence-time estimates. *Proc. Natl. Acad. Sci. USA* **111**, E2957 (2014).
28. Stadler, T. Sampling-through-time in birth–death trees. *J. Theor. Biol.* **267**, 396–404 (2010).
29. Gavryushkina, A., Welch, D., Stadler, T. & Drummond, A. J. Bayesian inference of sampled ancestor trees for epidemiology and fossil calibration. *PLoS Comput. Biol.* **10**, e1003919 (2014).
30. Zhang, C., Ronquist, F. & Stadler, T. Skyline fossilized birth–death model is robust to violations of sampling assumptions in total-evidence dating. *Syst. Biol.* **72**, 1316–1336 (2023).
31. Ronquist, F., Lartillot, N. & Phillips, M. J. Closing the gap between rocks and clocks using total-evidence dating. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **371**, 20150136 (2016).
32. Bossert, S. et al. Phylogeny, biogeography and diversification of the mining bee family Andrenidae. *Syst. Entomol.* **47**, 283–302 (2022).
33. Card, D. et al. Phylogeographic and population genetic analyses reveal multiple species of Boa and independent origins of insular dwarfism. *Mol. Phylogenet. Evol.* **102**, 104–106 (2016).
34. O'Reilly, J. E. & Donoghue, P. C. J. The effect of fossil sampling on the estimation of divergence times with the fossilized birth–death Process. *Syst. Biol.* **69**, 124–138 (2020).
35. Saladin, B. et al. Fossils matter: improved estimates of divergence times in *Pinus* reveal older diversification. *BMC Evol. Biol.* **17**, 95 (2017).
36. Leebens-Mack, J. H. et al. One thousand plant transcriptomes and the phylogenomics of green plants. *Nature* **574**, 679–685 (2019).
37. Asar, Y., Ho, S. Y. W. & Sauquet, H. Early diversifications of angiosperms and their insect pollinators: were they unlinked? *Trends Plant Sci.* **27**, 8589–8869 (2022).

38. Cardinal, S. & Danforth, B. N. Bees diversified in the age of eudicots. *Proc. Biol. Sci.* **280**, 20122686 (2013).
39. Moreau, C., Bell, C., Vila, R., Archibald, S. & Pierce, N. Phylogeny of the ants: diversification in the age of angiosperms. *Science* **312**, 101–104 (2006).
40. Wang, H. et al. Rosid radiation and the rapid rise of angiosperm-dominated forests. *Proc. Natl. Acad. Sci. USA* **106**, 3853 (2009).
41. Zhang, S. et al. Evolutionary history of Coleoptera revealed by extensive sampling of genes and species. *Nat. Commun.* **9**, 205 (2018).
42. Tihelka, E. et al. Angiosperm pollinivory in a Cretaceous beetle. *Nat. Plants* **7**, 445–451 (2021).
43. Zuntini, A. R. et al. Phylogenomics and the rise of the angiosperms. *Nature* **629**, 843–850 (2024).
44. Budd, G. E. & Mann, R. P. Two notorious nodes: a critical examination of relaxed molecular clock age estimates of the Bilaterian animals and placental mammals. *Syst. Biol.* **73**, 223–234 (2023).
45. Brown, J. W. & Smith, S. A. The past sure is tense: on interpreting phylogenetic divergence time estimates. *Syst. Biol.* **67**, 340–353 (2018).
46. Sauquet, H., Ramírez-Barahona, S. & Magallón, S. What is the age of flowering plants? *J. Exp. Bot.* **73**, 3840–3853 (2022).
47. Luo, A., Zhang, C., Zhou, Q., Ho, S. Y. W. & Zhu, C. Impacts of taxon-sampling schemes on Bayesian tip dating under the fossilized birth-death process. *Syst. Biol.* **72**, 781–801 (2023).
48. Luo, A., Duchêne, D. A., Zhang, C., Zhu, C. & Ho, S. Y. W. A simulation-based evaluation of tip-dating under the fossilized birth–death process. *Syst. Biol.* **69**, 325–344 (2020).
49. Barido-Sottani, J., Vaughan, T. G. & Stadler, T. Detection of HIV transmission clusters from phylogenetic trees using a multi-state birth–death model. *J. R. Soc. Interface* **15**, 20180512 (2018).
50. Mitchell, J. S., Etienne, R. S. & Rabosky, D. L. Inferring diversification rate variation from phylogenies with fossils. *Syst. Biol.* **68**, 1–18 (2019).
51. Bromham, L. et al. Bayesian molecular dating: opening up the black box. *Biol. Rev.* **93**, 1165–1191 (2018).
52. Sauquet, H. & Magallón, S. Key questions and challenges in angiosperm macroevolution. *New Phytol.* **219**, 1170–1187 (2018).
53. Manchester, S. R., Kapgate, D. K. & Wen, J. Oldest fruits of the grape family (Vitaceae) from the Late Cretaceous Deccan Cherts of India. *Am. J. Bot.* **100**, 1849–1859 (2013).
54. Looy, C., Kerp, H., Duijnstee, I. & Dimichele, W. The late Paleozoic ecological-evolutionary laboratory, a land-plant fossil record perspective. *Sediment. Rec.* **12**, 1–10 (2014).
55. Hickey, L. & Doyle, J. Early Cretaceous fossil evidence for angiosperm evolution. *Bot. Rev.* **43**, 3–104 (1977).
56. Doyle, J. Cretaceous angiosperm pollen of the Atlantic Coastal Plain and its evolutionary significance. *J. Arnold Arbor.* **50**, 1–35 (1969).
57. Muller, J. Palynological evidence on early differentiation of angiosperms. *Biol. Rev. Camb. Philosophic. Soc.* **45**, 417–450 (1970).
58. 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).
59. Shi, G., Herrera, F., Herendeen, P. S., Clark, E. G. & Crane, P. R. Mesozoic cupules and the origin of the angiosperm second integument. *Nature* **594**, 223–226 (2021).
60. Klymiuk, A. A., Rothwell, G. W. & Stockey, R. A. A novel cupulate seed plant, *Xadzigacalix quatsinoensis* gen. et sp. nov., provides new insight into the Mesozoic radiation of gymnosperms. *Am. J. Bot.* **109**, 966–985 (2022).
61. Rothwell, G. W. & Stockey, R. A. Phylogenetic diversification of Early Cretaceous seed plants: the compound seed cone of *Doylea tetrahedrasperma*. *Am. J. Bot.* **103**, 923–937 (2016).
62. Silvestro, D. et al. Fossil data support a pre-Cretaceous origin of flowering plants. *Nat. Ecol. Evol.* **5**, 449–457 (2021).
63. Budd, G. E., Mann, R. P., Doyle, J. A., Coiro, M. & Hilton, J. Fossil data do not support a long pre-cretaceous history of flowering plants. *BioRxiv*. https://doi.org/10.1101/2021.02.16.431478
64. Grabherr, M. G. et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* **29**, 644–652 (2011).
65. Yang, Y. & Smith, S. A. Orthology inference in nonmodel organisms using transcriptomes and low-coverage genomes: improving accuracy and matrix occupancy for phylogenomics. *Mol. Biol. Evol.* **31**, 3081–3092 (2014).
66. Fu, L., Niu, B., Zhu, Z., Wu, S. & Li, W. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* **28**, 3150–3152 (2012).
67. Van Dongen, S. Graph clustering via a discrete uncoupling process. *SIAM J. Matrix Anal. Appl.* **30**, 121–141 (2008).
68. Eddy, S. R. Accelerated profile HMM searches. *PLoS Comput. Biol.* **7**, e1002195 (2011).
69. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
70. Castresana, J. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol. Biol. Evol.* **17**, 540–552 (2000).
71. Capella-Gutiérrez, S., Silla-Martínez, J. M. & Gabaldón, T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972–1973 (2009).
72. Nguyen, L., Schmidt, H. A., von Haeseler, A. & Minh, B. Q. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **32**, 268–274 (2015).
73. Jarvis, E. D. et al. Whole-genome analyses resolve early branches in the tree of life of modern birds. *Science* **346**, 1320 (2014).
74. 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).
75. Parham, J. F. et al. Best practices for justifying fossil calibrations. *Syst. Biol.* **61**, 346–359 (2012).
76. Yang, Z. & Yang, Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**, 1586–1591 (2007).
77. Reis, M. D. et al. Using phylogenomic data to explore the effects of relaxed clocks and calibration strategies on divergence time estimation: primates as a test case. *Syst. Biol.* **67**, 594–615 (2018).
78. Mao, K. et al. Distribution of living Cupressaceae reflects the breakup of Pangea. *Proc. Natl. Acad. Sci. USA* **109**, 7793–7798 (2012).
79. 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).
80. Morris, J. L. et al. The timescale of early land plant evolution. *Proc. Natl. Acad. Sci. USA* **115**, E2274 (2018).
81. Rambaut, A., Drummond, A. J., Xie, D., Baele, G. & Suchard, M. A. Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Syst. Biol.* **67**, 901–904 (2018).
82. Bouckaert, R. et al. BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS Comput. Biol.* **10**, e1003537 (2014).
83. Drummond, A. J., Ho, S. Y. W., Phillips, M. J. & Rambaut, A. Relaxed phylogenetics and dating with confidence. *PLoS Biol* **4**, e88 (2006).
84. Douglas, J., Zhang, R. & Bouckaert, R. Adaptive dating and fast proposals: revisiting the phylogenetic relaxed clock model. *PLoS Comput. Biol.* **17**, e1008322 (2021).


## Acknowledgements

We thank Simon Y.W. Ho and Arong Luo for helpful suggestions. This work was supported by the National Natural Science Foundation of China (32122010 and 32370228) (B.Z.); the Collaborative Innovation Center for Modern Crop Production co-sponsored by Province and

Ministry (B.Z.); the Priority Academic Program Development of Jiangsu Higher Education Institutions (PAPD) (B.Z.); the Postdoctoral Fellowship Program of CPSF (GZC20240730) (X.M.); Jiangsu Funding Program for Excellent Postdoctoral Talent (2024ZB023) (X.M.); the National Natural Science Foundation of China (42172006) and the Hundred Young Talents Program of the Chinese Academy of Sciences (Y902061) (C.Z.); the U.S. National Science Foundation (DBI 2318917) (S.B.H.).

## Author contributions

B.Z. conceived and designed the research; X.M., L.Y. performed the experiments and analyzed the data; X.M. drafted the papers with input from all authors; C.Z., S.B.H., and B.Z. revised and edited the paper. All of the authors read and approved the final paper.

## 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/s41467-025-57687-9.

**Correspondence** and requests for materials should be addressed to Bojian Zhong.

**Peer review information** *Nature Communications* thanks the anonymous reviewer(s) for their contribution to the peer review of this work. A peer review file is available.