

A metacalibrated time-tree documents the early rise of flowering plant phylogenetic diversity

Susana Magallón¹, Sandra Gómez-Acevedo¹, Luna L. Sánchez-Reyes^{1,2} and Tania Hernández-Hernández³

¹Instituto de Biología, Universidad Nacional Autónoma de México, Mexico City, Mexico; ²Posgrado en Ciencias Biológicas, Universidad Nacional Autónoma de México, Mexico City, Mexico;

³Departamento de Biología Evolutiva, Instituto de Ecología A.C., Xalapa, Veracruz, México

Summary

Author for correspondence:

Susana Magallón

Tel: +52 55 5622 9087

Email: s.magallon@ib.unam.mx

Received: 1 August 2014

Accepted: 21 November 2014

New Phytologist (2015) **207**: 437–453

doi: 10.1111/nph.13264

Key words: calibrations, constraints, diversification, fossil record, maximum age, radiations, relaxed molecular clocks.

- The establishment of modern terrestrial life is indissociable from angiosperm evolution. While available molecular clock estimates of angiosperm age range from the Paleozoic to the Late Cretaceous, the fossil record is consistent with angiosperm diversification in the Early Cretaceous.
- The time-frame of angiosperm evolution is here estimated using a sample representing 87% of families and sequences of five plastid and nuclear markers, implementing penalized likelihood and Bayesian relaxed clocks. A literature-based review of the palaeontological record yielded calibrations for 137 phylogenetic nodes. The angiosperm crown age was bound within a confidence interval calculated with a method that considers the fossil record of the group.
- An Early Cretaceous crown angiosperm age was estimated with high confidence. Magnoliidae, Monocotyledoneae and Eudicotyledoneae diversified synchronously 135–130 million yr ago (Ma); Pentapetalae is 126–121 Ma; and Rosidae (123–115 Ma) preceded Asteridae (119–110 Ma). Family stem ages are continuously distributed between *c.* 140 and 20 Ma.
- This time-frame documents an early phylogenetic proliferation that led to the establishment of major angiosperm lineages, and the origin of over half of extant families, in the Cretaceous. While substantial amounts of angiosperm morphological and functional diversity have deep evolutionary roots, extant species richness was probably acquired later.

Introduction

Life on Earth today is critically linked to flowering plants (angiosperms). Angiosperms are primary producers and fundamental structural components in modern terrestrial ecosystems, and contribute vast diversity in terms of species richness and functional innovations. Many biological lineages have flourished in association with angiosperms, or the biomes they created, depending on them for food, shelter, or symbiotic partnering (e.g. Wikström & Kenrick, 2001; Schneider *et al.*, 2004; McKenna *et al.*, 2009; Cardinal & Danforth, 2013). An understanding of the evolutionary establishment of modern terrestrial biomes is indissociable from angiosperm origin, diversification and rise to ecological predominance.

The timing of angiosperm diversification is among the classic questions in evolutionary biology (Friedman, 2009). Several elements to investigate it effectively have recently become available. Increasingly powerful relaxed clock methods have been incorporated into the general toolbox of modern phylogenetic biology (e.g. Baum & Smith, 2013). The ubiquitous application of relaxed clocks to all major branches of the tree of life (e.g. Hunt *et al.*, 2007; Hibbett & Matheny, 2009; Jetz *et al.*, 2012; Wahlberg *et al.*, 2013; Ericson *et al.*, 2014) has highlighted

critical issues that affect the accuracy of age estimation. The crucial relevance of independent calibrations in relaxed clock analyses has been recognized (e.g. Aris-Brosou & Yang, 2003; Smith *et al.*, 2006; Yang & Rannala, 2006; Donoghue & Benton, 2007; Ho, 2007; Rannala & Yang, 2007; Wilkinson *et al.*, 2011), and also that relaxed clock models may insufficiently capture the degree of molecular variation in empirical phylogenies, leading to incorrect estimation of absolute rates and divergence times (e.g. Dornburg *et al.*, 2012; Wertheim *et al.*, 2012). Simultaneously, promising new avenues are being developed, for example, renewed implementations of local clocks (e.g. Drummond & Suchard, 2010; Ronquist *et al.*, 2012a), and highly parametric approaches to implement prior distributions (e.g. Heath, 2012).

On another front, significant fossil findings – including structurally preserved reproductive organs from Early Cretaceous sediments (Heimhofer *et al.*, 2007; Friis *et al.*, 2011) – document the minimum time of lineage origin and morphological evolution. These findings are especially relevant in the context of the increasingly solid molecular-based picture of angiosperm relationships at all phylogenetic levels (e.g. Soltis *et al.*, 2011), and the as yet few, but increasing numbers of studies explicitly investigating their phylogenetic relationships (e.g. Doyle & Endress, 2000, 2010; Magallón, 2007; Martínez-Millán *et al.*, 2009;

Sauquet *et al.*, 2012). Currently available methods allow one to estimate the phylogenetic position of fossils using maximum likelihood (ML) and Bayesian approaches, and to include fossils as terminals in phylogenetic dating analyses (Pyron, 2011; Ronquist *et al.*, 2012b).

Relaxed clock analyses applied to angiosperms have combined these methodological and palaeontological advances. However,

most of them lack a firm temporal constraint on the onset of extant angiosperm diversification and, consequently, provide different estimates of the age of angiosperms as a whole (Fig. 1) and of the clades within them. Molecular clock estimates of crown angiosperm age range between 300 Ma or older (e.g. Ramshaw *et al.*, 1972; Brandl *et al.*, 1992; Magallón, 2010) and 86 Ma (Sanderson & Doyle, 2001), with many recent estimates lying

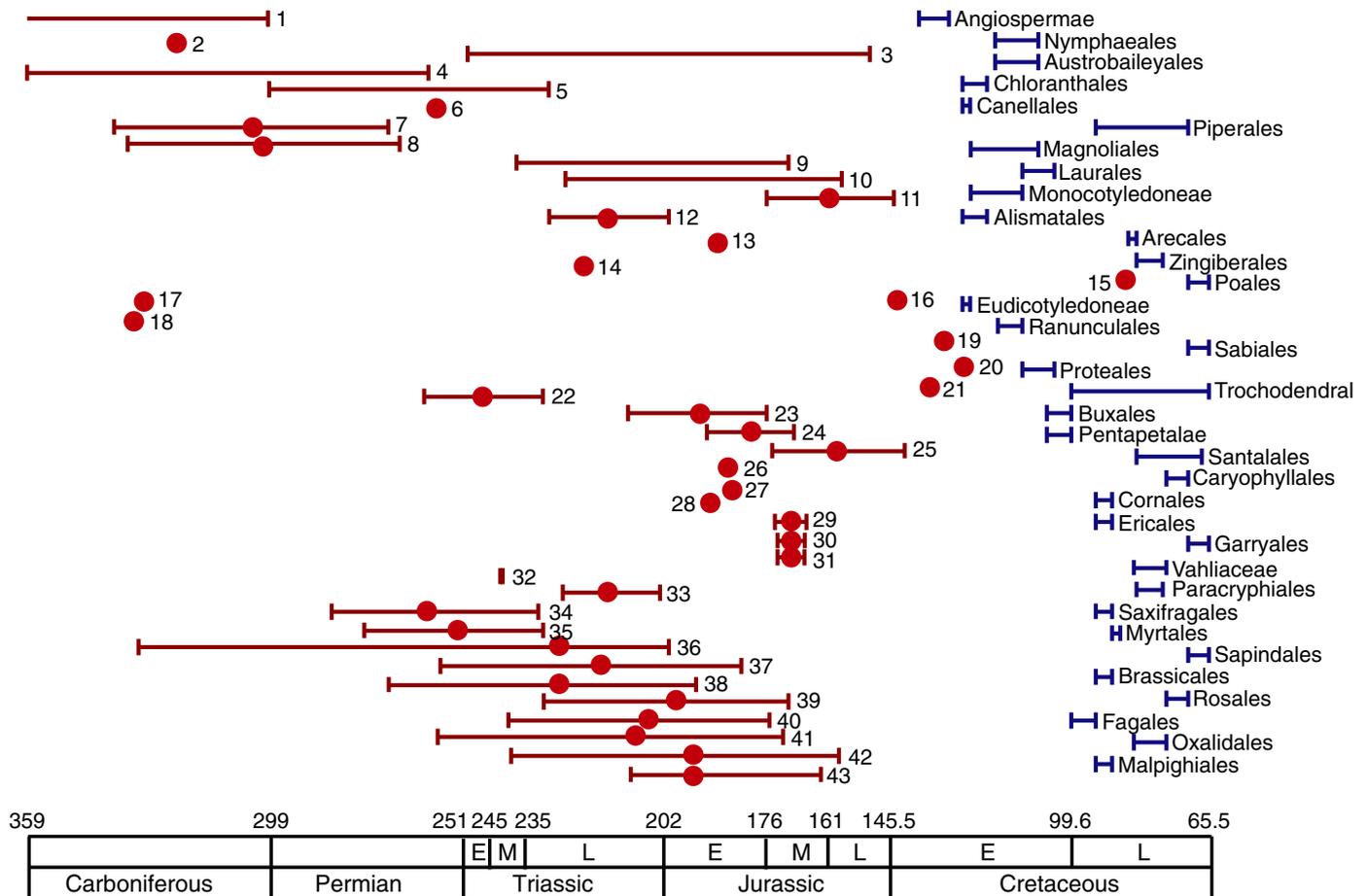


Fig. 1 Molecular and fossil-based estimates of angiosperm age. Blue bars indicate the stratigraphic interval from which the oldest fossil belonging to angiosperm orders and major clades is known, obtained by selecting the oldest representative fossil in Supporting Information Methods S1. Orders and major clades with oldest Tertiary fossils are not shown. The name of the order or major clade is shown next to each bar. Red dots and bars indicate the age and/or range of angiosperm crown age estimated in molecular clock studies. The numbers next to each red bar or dot correspond to estimates in published analyses indicated below. While the fossil record is consistent with an onset of angiosperm crown diversification in the Early Cretaceous, molecular estimates provide a generally older, but disparate picture of the age of the angiosperm crown group. 1, Ramshaw *et al.* (1972); 2, Martin *et al.* (1989); 3, Wolfe *et al.* (1989); 4, Brandl *et al.* (1992). Maize–wheat divergence calibration. 5, Brandl *et al.* (1992). Bryophyte–tracheophyte divergence calibration. 6, Brandl *et al.* (1992). Plant–animal divergence calibration. 7, Martin *et al.* (1993); method of Li & Graur (1991); 8, Martin *et al.* (1993); method of Li & Tanimura (1987); 9, Laroche *et al.* (1995). Maize–wheat divergence calibration. 10, Laroche *et al.* (1995). Viciae–Phaseolinae divergence calibration. 11, Goremykin *et al.* (1997); 12, Sanderson (1997); 13, Sanderson & Doyle (2001). 18S. 14, Sanderson & Doyle (2001). *rbcl* 1st + 2nd. 15, Sanderson & Doyle (2001). *rbcl* 3rd. 16, Sanderson & Doyle (2001). *rbcl* all. 17, Soltis *et al.* (2002). Calibration 12: *Dicksonia/Plagiogyria/Cyathea*. 18, Soltis *et al.* (2002). Calibration 19: *Angiopteris/Marattia*. 19, Soltis *et al.* (2002). Calibration 25: gymnosperms. 20, Soltis *et al.* (2002). Calibration 29: lycopsids at 377.4. 21, Soltis *et al.* (2002). Calibration 19: lycopsids at 400. 22, Schneider *et al.* (2004); 23, Bell *et al.* (2005). MD with three minimum age constraints. 24, Bell *et al.* (2005). BLs estimated separately for each data partition. 25, Bell *et al.* (2005). Penalized likelihood (PL) with three minimum age constraints. 26, Magallón & Sanderson (2005). Four genes, 1st + 2nd. 27, Magallón & Sanderson (2005). Four genes, 3rd. 28, Magallón & Sanderson (2005). Four genes, all. 29, Moore *et al.* (2007). Unconstrained. 30, Moore *et al.* (2007). One hundred and twenty-five million years ago (Ma) minimum age to stem eudicots. 31, Moore *et al.* (2007). One hundred and twenty-five Ma minimum age to crown eudicots. 32, Magallón & Castillo (2009). Unconstrained. 33, Magallón (2010). PLBP. 34, Magallón (2010). PLFB. 35, Magallón (2010). MD. 36, Magallón (2010). Uncorrelated lognormal (UCLN). 37, Smith *et al.* (2010). With eudicot calibration. 38, Smith *et al.* (2010). Without eudicot calibration. 39, Clarke *et al.* (2011). Embryophyta at 509 Ma. 40, Clarke *et al.* (2011). Embryophyta at 1042 Ma. 41, Magallón *et al.* (2013). UCLN *atpB*, *psaA*, *psbB* and *rbcl*. 42, Magallón *et al.* (2013). UCLN *matK*. 43, Magallón *et al.* (2013). UCLN all genes.

between *c.* 190 and 150 Ma (e.g. Magallón, 2010; Smith *et al.*, 2010; Clarke *et al.*, 2011; Magallón *et al.*, 2013; Fig. 1).

The fossil record is consistent with an onset of angiosperm crown diversification in the Early Cretaceous, as shown by the increasing diversity and abundance of angiosperms in local outcrops and at a global level; increasing morphological complexity of leaves, pollen and flowers in agreement with expectations based on plant morphology; and congruence between the appearance of lineages in stratigraphic sequences and the branching order in molecular phylogenies (Doyle, 2012; Magallón, 2014). The fossil record is consistent with a proliferation of angiosperm lineages during the Cretaceous (145.5–65.5 million yr ago (Ma)), as indicated by the first stratigraphic occurrence of 35 orders and major clades during this period (Fig. 1; Supporting Information Methods S1). The observed distribution of the angiosperm fossil record as a whole may provide consistent landmarks to aid relaxed clock estimations, and for calculation of the age of the group as a whole.

The goal of this study is to estimate a time-frame of angiosperm evolution, including the origin and diversification of its major clades. It is based on a comprehensive sample of nearly 800 placeholders for 87% of angiosperm families, and the sequences of five molecular markers from the plastid and nuclear genomes. Divergence times were estimated with penalized likelihood and an uncorrelated Bayesian method. Two factors set this study apart from previous attempts. First, relaxed clock analyses incorporate a large number of fossil-derived age calibrations across the angiosperms, which provide landmarks for molecular rates and minimum ages across the tree. The phylogenetic assignment and age of each calibrating fossil are critically justified. Secondly, a confidence interval on the age of the angiosperm crown node, calculated with a method that considers the overall fossil record of the group (Marshall, 2008), was introduced in relaxed clock analyses. The resulting time-trees provide a solid framework for investigating angiosperm evolution. We discuss technical aspects, including the differential reliability of estimated family stem and crown ages; and biological implications, including the timing of angiosperm crown diversification and the origin of its major clades. The obtained time-frame documents profuse phylogenetic branching early in angiosperm history leading to the establishment of clades that contain major proportions of extant diversity, and, particularly, the early rise of many angiosperm families.

Materials and Methods

Taxonomic sample, molecular data and phylogenetic analyses

Phylogenetic, dating and diversification analyses were based on a data set including 792 angiosperm, six gymnosperm and one fern species. The gymnosperms belong to six families representing Cycadophyta, Ginkgophyta, Gnetophyta and Coniferae (*sensu* Cantino *et al.*, 2007). The angiosperms belong to 374 family-level clades corresponding to 87% of those recognized on the Angiosperm Phylogeny Website in April 2013 (Stevens, 2013). The represented families encompass 99% of angiosperm species.

All angiosperm orders (plus four families unassigned at order level) are represented.

The data are the nucleotide sequences of three protein-coding plastid genes (*atpB*, *rbcL* and *matK*), and two nuclear markers (18S and 26S nuclear ribosomal DNA). Sequences were obtained through searches in GenBank, aiming to represent all angiosperm families with a complete sample of the five molecular markers. As first choice, species for which the five markers were available were selected. When one or more markers were unavailable, the missing markers were sampled from different species within the same genus. Markers that were unavailable at the genus level were left as missing data. Sequences of each marker were aligned with MUSCLE v3.7 (Edgar, 2004) followed by manual refinements using BIOEDIT v7.0.9.0 (Hall, 1999). The sampled species, represented families and orders, and GenBank accessions are listed in Table S1. The molecular data set is available in the DRYAD Digital Repository (<http://doi.org/10.5061/dryad.k4227>).

Phylogenetic analyses were conducted with maximum likelihood (ML) using RAxML v7.2.8. (Stamatakis, 2006a), implementing a GTRCAT model with 1000 bootstrap replicates (Stamatakis, 2006b, 2008). Substitution parameters were estimated independently for four data partitions: first plus second codon positions of *atpB* plus *rbcL*; third codon positions of *atpB* plus *rbcL*; *matK*; and 18S plus 26S. The fern *Ophioglossum* was specified as the outgroup. A topological constraint specifying relationships among major angiosperm clades according to a recent angiosperm-wide phylogenetic analysis based on a larger molecular data set (Soltis *et al.*, 2011) was implemented. Subsequently, a RAxML analysis to obtain 100 bootstrap trees with optimized branch lengths and model parameters was conducted with the same data partitions and constraints as above, but implementing the GTRGAMMA model.

Fossil calibrations

To obtain temporal calibrations for relaxed clock analyses, we conducted an intensive literature-based search of angiosperm fossils. Several search approaches were combined, including reviewing the palaeobotanical articles published during the last 20 yr, and extracting records from palaeofloristic works (e.g. Collinson, 1983; Knobloch & Mai, 1986) or from palaeobotanical compilations (e.g. Manchester, 1999; Martínez-Millán, 2010; Friis *et al.*, 2011) from which primary references were traced. We assembled a large data set in which fossils were grouped by their presumed affinity with angiosperm family-level or order-level clades, by combining the taxonomic assignment of the authors of the palaeobotanical description and our current knowledge of angiosperm phylogenetic relationships (Stevens, 2013). This data set includes over 3500 entries.

From this large data set, we selected fossils that could reliably represent the oldest record of angiosperm clades down to family (or lower) level. Accurate clade recognition was favoured over older ages (Sauquet *et al.*, 2012) by selecting only those fossils that could be identified as members of a clade with certainty. We gave preference to fossils of flowers, fruits and seeds over vegetative remains or pollen. In several instances we relegated the

putative oldest fossil of a clade because it was a vegetative structure of insecure affinity in favour of the oldest reproductive structure of the same clade. We also preferred whole-plant reconstructions and structurally preserved fossils over compression/impression remains.

To identify the relationship of selected fossils with the taxa in the tree and postulate calibration nodes, we combined intuitive, apomorphy-based and phylogenetic approaches. We relied as much as possible on the description and illustrations, and on the discussion of their taxonomic assignment in original publications. When available, we gave preference to a phylogenetic result (e.g. Doyle & Endress, 2010) over intuitive or apomorphy-based assignments. The absolute age assigned to each fossil was equal to the uppermost boundary of the narrowest stratigraphic interval to which the fossil could be assigned. Ages of stratigraphic boundaries were obtained from Walker & Geissman (2009). The assignment of each fossil to a particular node, as well as the absolute minimum age calibration, are discussed in detail in Methods S1.

Confidence interval on angiosperm age

A confidence interval that contains the true age of the angiosperm crown node was calculated with a method that considers the number of branches in a phylogenetic tree that are represented in the fossil record (Marshall, 2008). This method derives from quantitative palaeobiology approaches to calculate confidence intervals that contain the true time of origin and extinction of lineages based on local or global stratigraphic sequences (Strauss & Sadler, 1989; Marshall, 1990, 1994, 1997). The goal of Marshall's (2008) method is to date a molecular phylogenetic tree by using an absolute time-scale extrapolated from a confidence interval that contains the true age of the lineage in the tree that has the most temporally complete fossil record (i.e. the calibration lineage). In brief, the method has three components. The first is to identify the calibration lineage, which is achieved by finding the branch with the greatest overlap between the age of its oldest fossil and its node-to-tip length in an ultrametric tree estimated without any reference to the fossil record. The second step is to calculate a confidence interval that contains the true age of the calibration lineage. This calculation only requires the number of branches in the tree that are represented in the fossil record, the average number of fossil localities from which each branch represented in the fossil record is known, and the age of the oldest fossil of the lineage. The third step is to date the ultrametric tree by directly transforming its branch lengths into time by using the confidence interval of the calibration lineage as an absolute time-scale.

Because our aim is to calculate a confidence interval that contains the true age of crown angiosperms, we only implemented the second step of Marshall's (2008) method in the angiosperms as a whole. In a collateral study, described in Methods S2, we conducted step 1 and identified the calibration lineage of the angiosperms. This application of the method is justifiable because our intention is to estimate the maximal age of angiosperms, and not to temporally calibrate

an ultrametric tree (C. R. Marshall, pers. comm.). The minimum (i.e. youngest) age of the confidence interval is directly given by the oldest known fossil(s) of the lineage. We consider angiosperm pollen grains from Valanginian to Hauterivian sediments (Early Cretaceous; Hughes & McDougall, 1987; Hughes *et al.*, 1991; Brenner, 1996) as the oldest fossils of the angiosperm crown group on the basis of their morphological and ultrastructural attributes; the increasing abundance, diversity and geographical distribution of angiosperm fossils starting in immediately younger sediments; and the congruence in morphological evolution and sequence of lineage appearance between the fossil record and expectations derived from morphological studies and molecular phylogenies, respectively (Methods S1). We used the Valanginian–Hauterivian boundary, corresponding to 136 Ma (Walker & Geissman, 2009), as the minimum age.

The maximum (i.e. oldest) age of the confidence interval is calculated with eqn 14 from Marshall (2008):

$$FA_c = \frac{FA_{cal}}{\sqrt[nH]{1-C}}$$

FA_c is the maximum age of the confidence interval; FA_{cal} is the age of the oldest fossil of the lineage (corresponding to the minimum age of the interval), in this case, 136 Ma; n is the number of branches in the phylogenetic tree represented in the fossil record; H is the average number of fossil localities from which each branch represented in the fossil record is known; and C is the desired confidence level associated with the interval. To calculate n , we used a tree in which each angiosperm family included in the main phylogenetic and dating analyses was represented by a single terminal (Methods S2). Based on the fossils used to calibrate internal nodes in the main dating analyses (Methods S1), we counted the number of branches in the family-level angiosperm tree represented in the fossil record. The average number of fossil localities from which each branch in the tree represented in the fossil record is known (H) is difficult to calculate. We chose to consider $H=1$, implying that each branch with a fossil record is known from a single locality. Because the maximum age of the confidence interval decreases as H increases (Marshall, 2008), assuming $H=1$ is a conservative approach that will bias the maximum age of angiosperms towards older ages. FA_c was calculated with confidence levels (C) of 0.5, 0.95 and 0.99. The calculated confidence interval was then implemented as a constraint in relaxed clock analyses.

Relaxed clock analyses

Dating analyses were conducted with two relaxed clock methods: penalized likelihood (PL; Sanderson, 2002) and the uncorrelated lognormal (UCLN) Bayesian method available in BEAST (Drummond *et al.*, 2006). Penalized likelihood analyses were conducted combining the softwares r8s (Sanderson, 1997, 2004) and TREE-PL (Smith & O'Meara, 2012). PL analyses were based on the ML phylogram obtained with RAXML after excluding the

outgroup (*Ophioglossum*); hence the seed plant crown node became the new root. To identify the appropriate level of rate heterogeneity in the phylogram, a data-driven cross-validation was conducted with TREEPL. The cross-validation tested nine smoothing values (λ) separated by one order of magnitude, starting at 1×10^{-7} . The age of the root was fixed at 330 Ma, based on previous estimates for crown seed plants (Magallón *et al.*, 2013). The angiosperm crown node was bracketed between 136 and 140 Ma (see the section 'Confidence interval on angiosperm age'), and 136 nodes within angiosperms were constrained with fossil-derived minimum ages (see the Materials and Methods section, and Methods S1).

Penalized likelihood age estimation was conducted with TREEPL and with r8s on the ML phylogram. The identified optimal smoothing value, the root node calibration, the bracket on crown angiosperm age (136–139.35 Ma in r8s; see the Results section), and the 136 minimum age constraints were implemented as indicated above. One hundred ML bootstrap phylograms were also dated with r8s, using the optimal smoothing magnitude identified with TREEPL. Age statistics of internal nodes were summarized with TREEANNOTATOR v1.7.5 (Drummond *et al.*, 2006).

Bayesian age estimation was conducted with the UCLN model in BEAST v1.7.5 (Drummond *et al.*, 2006). The data were the nucleotide sequences of the five molecular markers used in phylogeny estimation concatenated in a single alignment, including only seed plants. Data were partitioned into plastid (*atpB*, *rbcl* and *matK*) and nuclear components (18S and 26S nrDNA). Nucleotide substitution was under a GTR+I+ Γ model, allowing independent estimation of parameters for each partition. Independent uncorrelated relaxed clock models were allowed between partitions, and the tree prior was under a Birth-Death model. The root was calibrated with a uniform distribution between 314 and 350 Ma, corresponding to the credibility interval (95% highest posterior density (HPD)) of the age of this node estimated in an independent study (Magallón *et al.*, 2013). The angiosperm root node was calibrated with a uniform distribution between 136 and 139.35 Ma (see the Results section). The prior ages of 136 nodes within angiosperms were obtained from lognormal distributions with mean equal to the fossil age plus 10%, to place the bulk of the distribution at ages older than the fossil, and a standard deviation of 1 (Methods S1). We considered assigning different standard deviation magnitudes depending on our confidence on each calibration, but, because we were unable to rigorously quantify our perception, we decided to assign the same magnitude to all. The chronogram obtained in the r8s analysis on the ML tree was used as a starting tree, and estimators of tree topology were unselected. Eight independent Markov Chain Monte Carlo (MCMC) runs of different lengths, but under the same estimation conditions, were conducted, for a total of 170×10^6 generations. Each MCMC was sampled every 5000 steps. The initial 600 trees sampled in each run were removed as burn-in and, in all cases, the post-burn-in trees were in the stable part of the chain. Analyses were conducted in the CIPRES Science Gateway (Miller *et al.*,

2010). Log outputs of the BEAST analyses were jointly evaluated with TRACER v1.5 (Rambaut & Drummond, 2009). Effective sample sizes of estimated parameters were in most cases > 200 , and always > 100 . As a consequence of usage restrictions in CIPRES, we were unable to conduct further BEAST analyses. Files containing the sampled trees of each MCMC run were combined using LOGCOMBINER v1.7.5, annotated using TREEANNOTATOR v1.7.5 (Drummond *et al.*, 2006), and visualized using FIGTREE v1.4.0 (<http://tree.bio.ed.ac.uk/software/figtree/>).

Results

Molecular data set and phylogenetic tree

The concatenated alignment of nucleotide sequences of the five molecular markers (cp *atpB*, *rbcl* and *matK*; nu 18S and 26S) is 9089 base pairs long. From the total number of five markers for 799 taxa, 432 are missing; hence the molecular data set is 89% complete. The ML tree is consistent with current understanding of angiosperm relationships and most branches are supported with bootstrap values $\geq 95\%$. Phylogenetic relationships are shown in Figs 2 and 3.

Fossil calibrations

Based on the large data set of angiosperm fossils, 151 fossils that can reliably indicate the oldest occurrence of angiosperm clades down to family (sometimes lower) level were selected. Because of sister group relationships or sampling density, 20 nodes were each calibrated by two fossils; and two nodes by three fossils of the same age. In total, 136 internal nodes were calibrated with fossil-derived minimum ages (calibrations 2–137; Methods S1). An additional record, corresponding to the oldest remains of angiosperms, was used as the minimum bound of the confidence interval on crown angiosperm age (calibration 1; Methods S1). Of the selected fossils, eight provide calibrations for genera, seven for intrafamilial clades, 112 for families, seven for clades between families and orders, 12 for orders, and five for clades above the order level (Stevens, 2013). Twenty calibrations were assigned to nodes based on phylogenetic results, and the rest were placed intuitively or based on apomorphies. Except in one case (Martínez-Millán *et al.*, 2009), the phylogenetic position of fossils matched previous intuitive assignments. Detailed discussions of the 137 fossil-based calibrations, including formal names, authors and references, stratigraphic ranges or radiometric dates, justification of node assignment, and absolute age, are provided in Methods S1.

Confidence interval on angiosperm age

The confidence interval of angiosperm crown age was calculated considering the fossil record of the entire group (Marshall, 2008) in the context of a family-level tree. The number of branches represented in the fossil record (n) is 123. The minimum bound of the confidence interval (FA_{ca}) is given by the age of the oldest fossil of the clade, that is, 136 Ma (Hughes

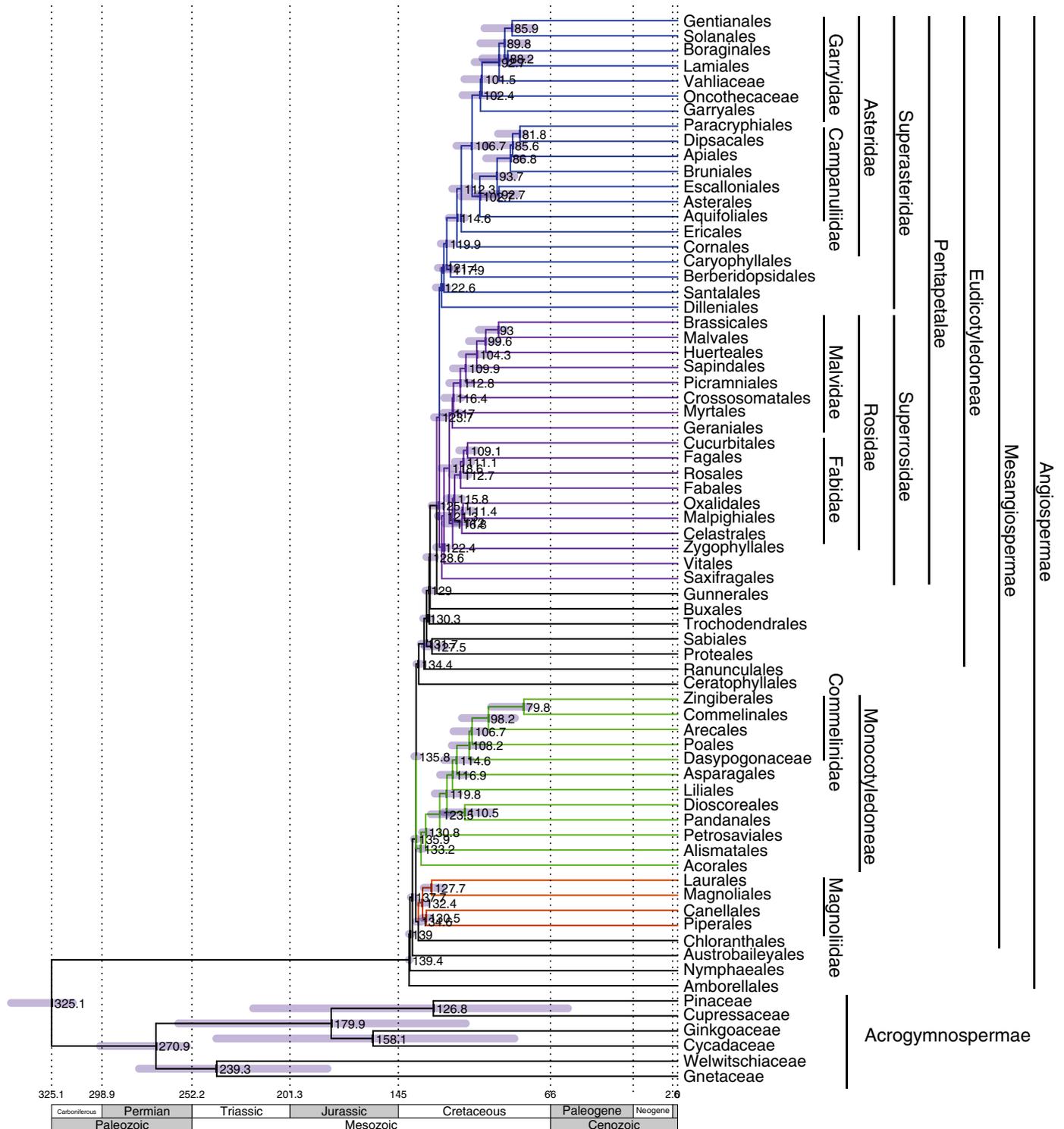


Fig. 2 Angiosperm time-tree estimated using the uncorrelated lognormal method in BEAST, with terminals collapsed to represent orders. Numbers next to nodes indicate the median age, and blue bars correspond to the 95% highest posterior density (HPD).

& McDougall, 1987; Hughes *et al.*, 1991; Brenner, 1996; Methods S1). The maximum bound of the confidence interval (FA), calculated with confidence levels (C) of 0.5, 0.95 and 0.99, is 136.77, 139.35 and 141.19 Ma, respectively. Hence, the 95% confidence interval of angiosperm age is between 136 and 139.35 Ma.

Relaxed clock analyses

Stem and crown ages of 14 major angiosperm clades and 62 orders, and stem ages of 374 families are provided in Table S2. Fig. 2 shows the UNCL time-tree at the level of angiosperm orders, and Fig. 3(a–e) shows the family-level UNCL angiosperm

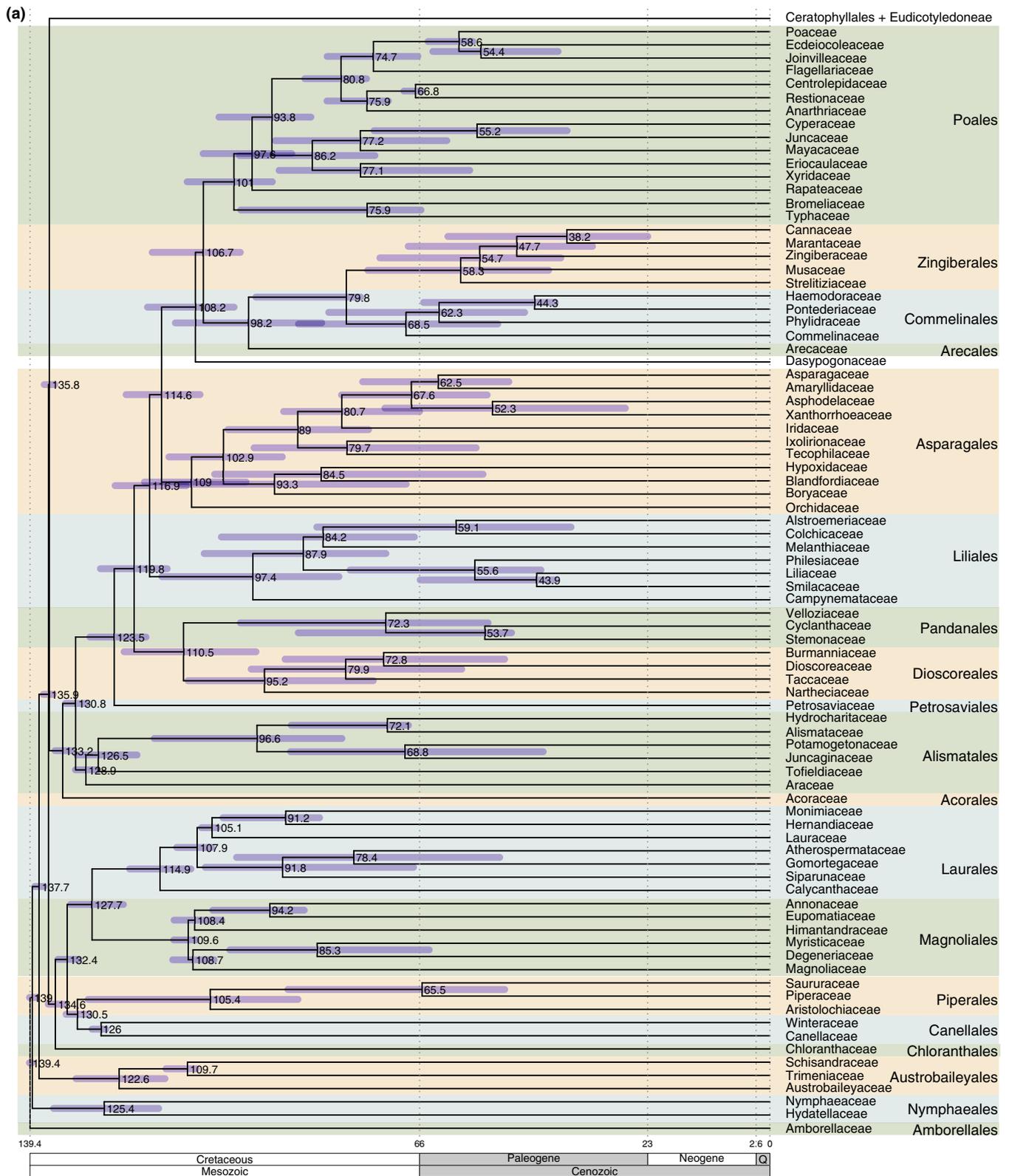


Fig. 3 Angiosperm time-tree estimated using the uncorrelated lognormal method in *BEAST*, with terminals collapsed to represent families. Numbers next to nodes indicate the median age, and blue bars correspond to the 95% highest posterior density (HPD). (a) Amborellales to Poales (Monocotyledoneae). (b) Ceratophyllales to Ericaceae (Asteridae, Eudicotyledoneae). (c) Garryidae to Campanulidae (Asteridae, Eudicotyledoneae). (d) Saxifragales to Brassicales (Malvidae, Rosidae, Eudicotyledoneae). (e) Zygophyllales to Malpighiales (Fabidae, Rosidae, Eudicotyledoneae). Ages of nodes are provided in Supporting Information Table S2.

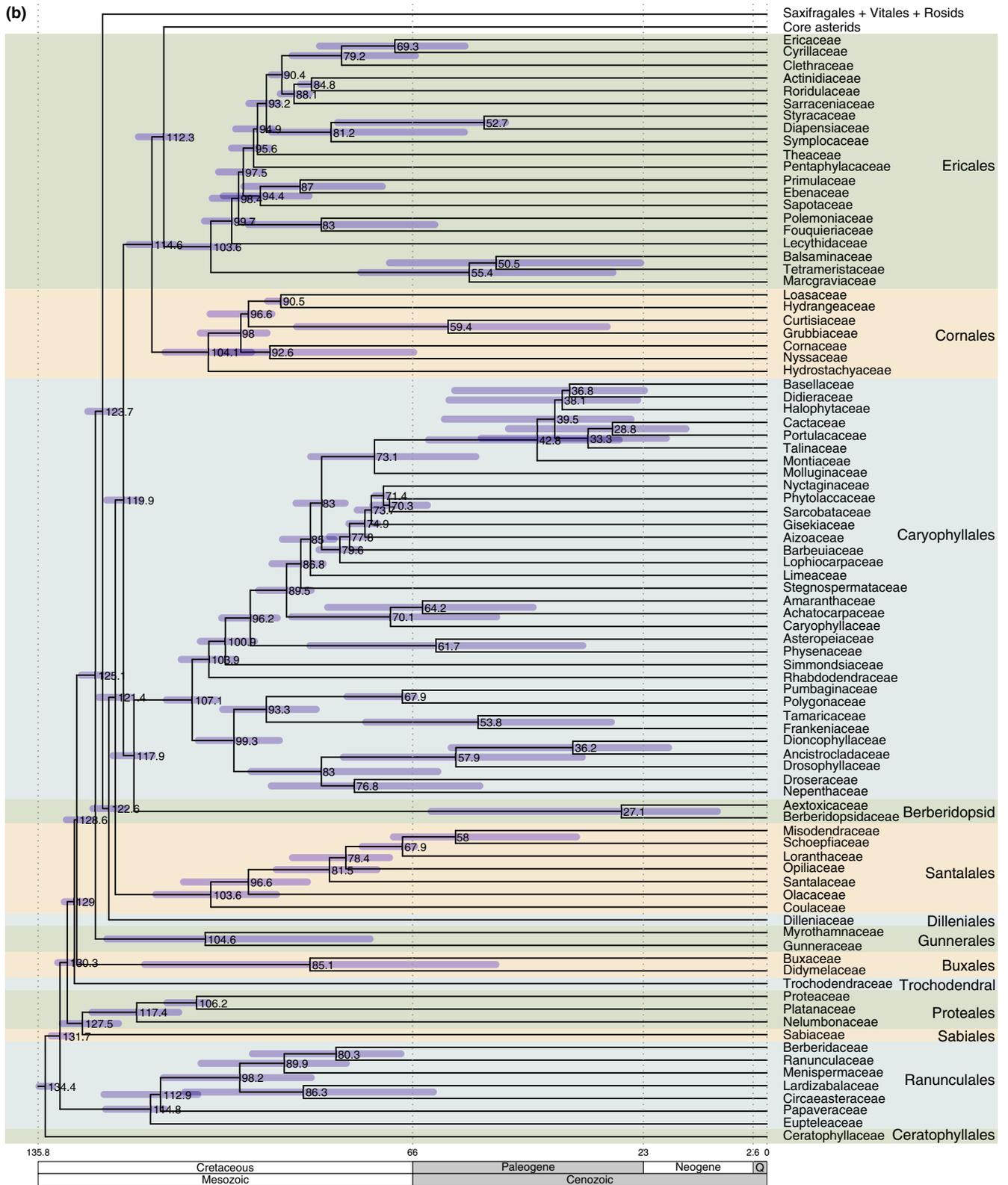


Fig. 3 Continued

time-tree. The full dated trees in NEXUS format are available in the DRYAD Digital Repository (<http://doi.org/10.5061/dryad.k4227>). Ages estimated with PL tend to be older than those estimated with the UCLN method (Table S2). As expected,

associated errors on ages are narrower in the PL time-tree (i.e. average range of values in bootstrap ML trees = 9.25 Ma) than in the UNCL time-tree (i.e. average magnitude of 95% HPD = 27.03 Ma).

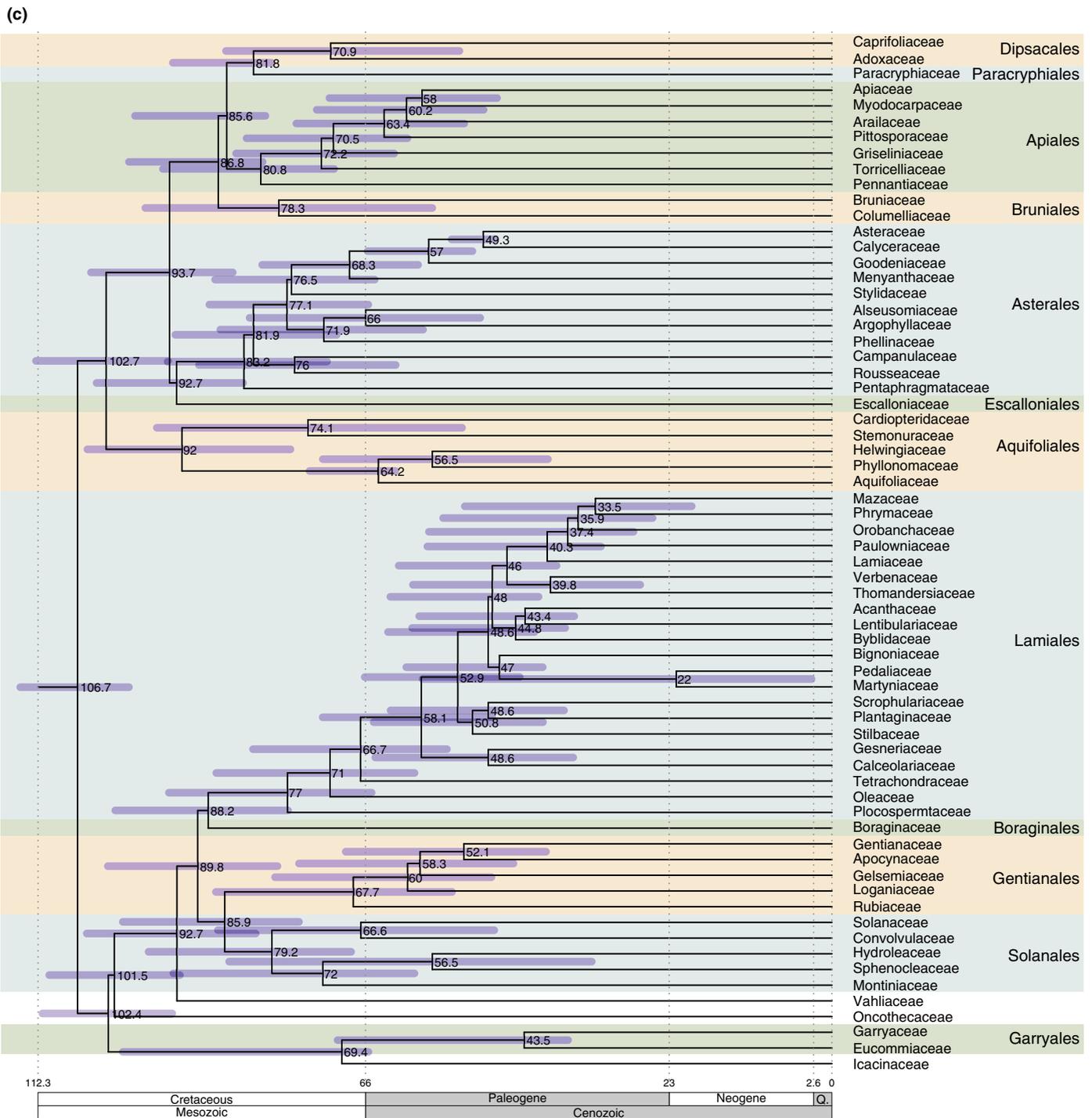


Fig. 3 Continued

Discussion

Estimated dates

The dates obtained depend on the correctness of relationships and branch lengths estimated in phylogenetic and dating analyses. Estimates of some family crown ages are probably too young for the following reasons. (1) The taxonomic sample may not represent the crown node of each family. The crown age of a

clade can only be estimated if at least one member of each of the two sister branches derived from the deepest phylogenetic split in the clade is included. Although we aimed to sample representatives from both sides of the deepest split of each family, this was not consistently achieved because of insufficient knowledge of intrafamilial phylogenetic relationships, or unavailability of molecular markers for the required taxa. (2) Fossils selected as calibrations may not be the oldest fossil members of a clade. To calibrate, we selected fossils with greater chances of correctly

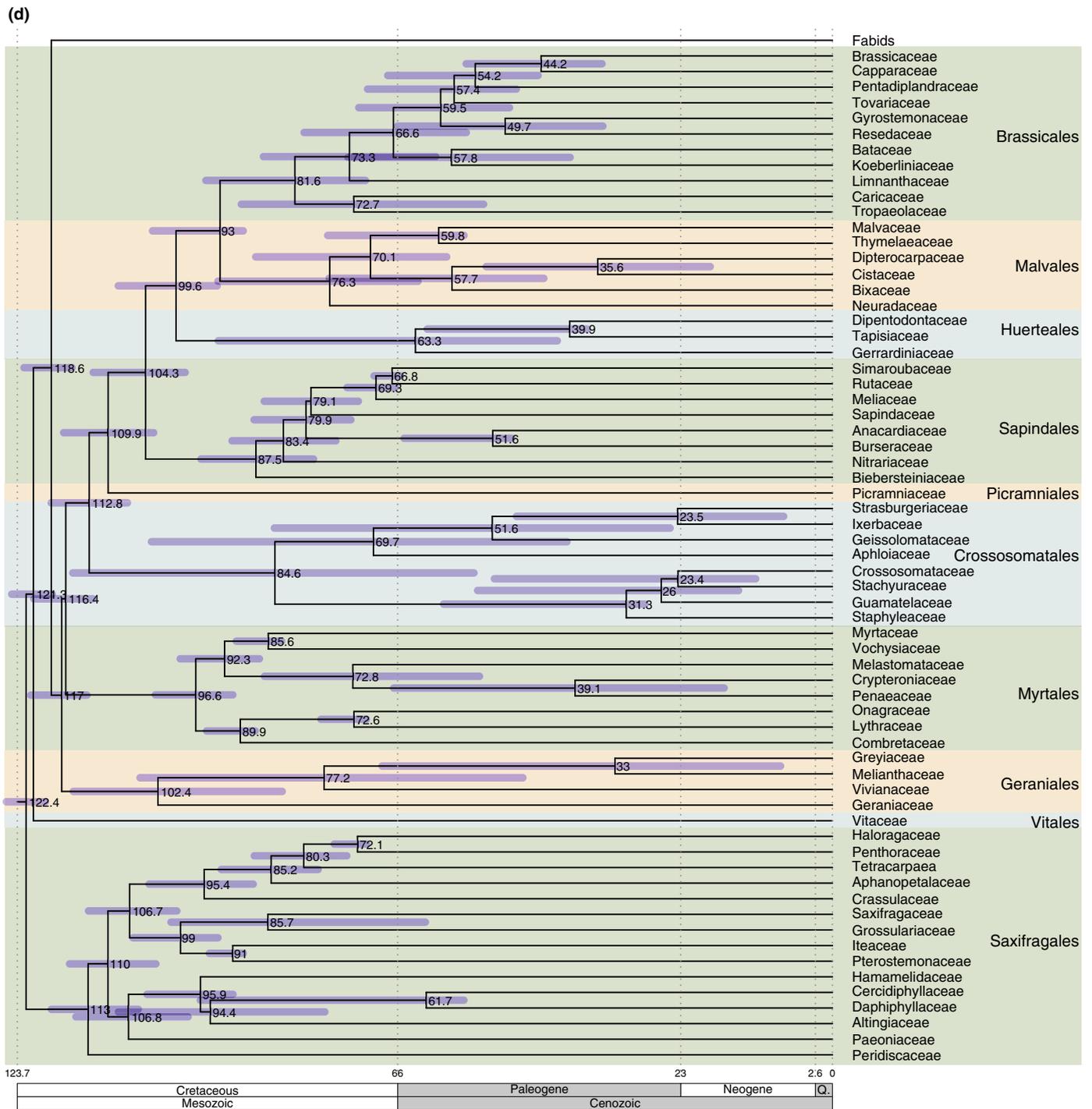


Fig. 3 Continued

reflecting clade membership over fossils whose membership in the clade is equivocal, thus favouring 'safe but late' over 'early but risky' fossils (Sauquet *et al.*, 2012). Consequently, the minimum ages provided by some calibrations may be too young. (3) The assignment of calibrations to nodes on the tree was done conservatively. Most assignments were based on the presence of an apomorphy, or intuitive criteria (Sauquet *et al.*, 2012). In these cases, a fossil was assigned to a more inclusive clade that could securely contain it (e.g. a stem group) than to a less inclusive clade where

its membership was dubious (e.g. a crown group). Consequently, minimum ages that are too young (thus uninformative) might have been applied to stem nodes, rather than minimum ages that are too old (thus incorrect and misleading) to crown nodes. The taxonomic sample in this study represents angiosperm major clades and their relationships; hence the factors mentioned above are unlikely to affect age estimation at deeper phylogenetic levels.

Because previous molecular estimates have provided disparate estimates of angiosperm age (Fig. 1), we implemented a

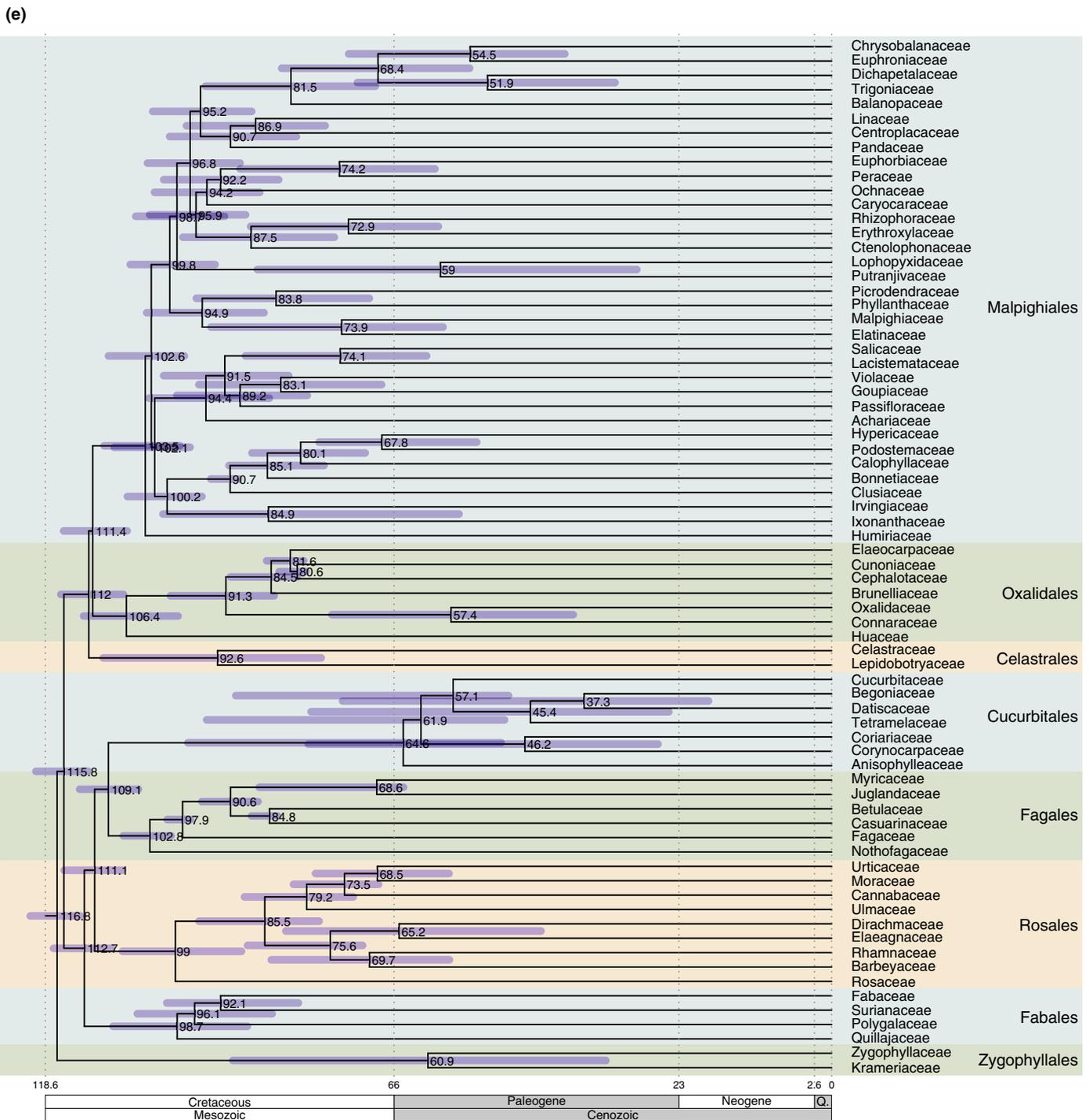


Fig. 3 Continued

confidence interval on the age of the angiosperm crown node. The minimum age of the interval can be obtained from the oldest fossils of the group, but the maximum age is difficult to obtain. Marshall's (2008) method allows calculation of the maximum age by quantifying the density of the fossil record of the group as a whole. Marshall's method has one caveat – reliance on a single calibration lineage, albeit selected from the complete fossil record of the group – and relies on several assumptions: that the affinity

and absolute age of fossils are known with certainty; that the topology and relative branch lengths of the uncalibrated, ultrametric tree are accurate; that fossilization is random; and that the value of H (eqn 14) is close to the true average number of fossil localities from which each lineage represented in the fossil record is known (Marshall, 2008). However, only the last two assumptions are relevant to calculating the confidence interval around the true age of a lineage.

Nonrandom fossilization and, specifically, a lower fossilization probability early in the history of a lineage may underestimate its true time of origin. However, only if all the lineages within the group suffer similarly from a long period of dramatically decreased initial preservation potential will the method fail to bracket the true time of origin (Marshall, 2008). We considered reticulate semitectate pollen grains with columellate sexine from the Valanginian–Hauterivian (Hughes & McDougall, 1987; Brenner, 1996) as belonging, or being close, to the angiosperm crown group, based on their morphological attributes (Doyle, 2012) and, importantly, because their stratigraphic appearance is followed by a continuous and increasingly dense and diverse angiosperm record in immediately younger sediments (Fig. 1). Nevertheless, the possibility of a cryptic pre-Cretaceous angiosperm history has been discussed (e.g. Axelrod, 1952, 1970). Specifically, angiosperm-like pollen grains similar to those from Early Cretaceous sediments have been reported from the Middle Triassic Germanic Basin and other pre-Cretaceous localities (Hochuli & Feist-Burkhardt, 2013). These Triassic pollen grains resemble the *Retimonocolpites* morphotypes, known from late Hauterivian and younger sediments (Phases 1–4 of Hughes, 1994), and not the *Clavatipollenites* morphotypes, to which some of the early Hauterivian angiosperm pollen grains belong (Phase 0 of Hughes, 1994; Hochuli & Feist-Burkhardt, 2013). We agree with Hochuli & Feist-Burkhardt (2013, p. 11) in interpreting the pre-Cretaceous angiosperm-like pollen grains as possible angiosperm stem relatives, based on morphological differences from the earliest Early Cretaceous pollen grains, and a 100 Myr gap in the fossil record before the angiosperm radiation.

If the true average number of fossil localities from which each lineage with a fossil record is known is much larger than the value of H (Marshall, 2008), the calculated maximum bound of the confidence interval will substantially overestimate the true age of the lineage. We considered $H = 1$, which, in the case of angiosperms, is a strong underestimate. Any value $H > 1$ would have resulted in a younger maximum bound of the confidence interval of angiosperm age. The angiosperm age interval was therefore calculated under two biases with opposite effects. It is unknown if the potential underestimation caused by nonrandom fossilization and the overestimation caused by assuming $H = 1$ cancel each other.

We emphasize that the relaxed clock-estimated ages are strongly contingent on the confidence interval placed on the age of the angiosperm crown node. Preliminary analyses conducted under the same conditions as described in the UCLN analysis above, but excluding the confidence interval on the angiosperm crown node, estimated a substantially older age for the angiosperm crown node (219.9 Ma; 160.0–255.8 95% HPD; results available from the authors). Interestingly, internal nodes were not much older than when the confidence interval was used.

Origin of major angiosperm clades

The estimated time-trees indicate the timing of phylogenetic branching that gave rise to major angiosperm clades, and provide a reliable basis for understanding the onset and dynamics of accumulation of different components of extant angiosperm diversity.

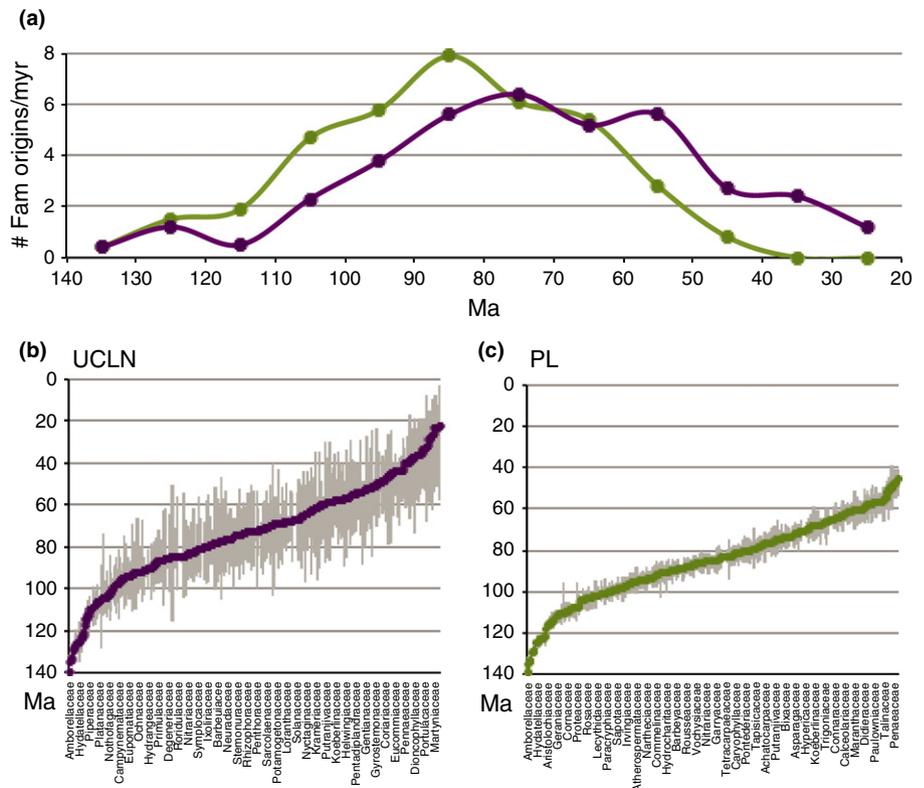
PL-estimated ages are typically older than UCLN ages, but their associated intervals are overlapping, and provide congruent time-frames of the stem ages of families. The following discussion is based on 95% HPDs for crown ages obtained in the UCLN analysis. Mesangiospermae – which contains the vast majority of angiosperm diversity (c. 99.96% of extant species richness) – began to diversify between 137 and 135 Ma, soon after the crown diversification of angiosperms. The clades that include most angiosperm diversity started to diversify almost simultaneously: Magnoliidae (3.61% of extant richness) between 134.1 and 130.2 Ma; Monocotyledoneae (monocots; 23.32% of extant richness) between 134.7 and 131.6 Ma; and Eudicotyledoneae (eudicots; 73% of extant richness) between 133.4 and 129.7 Ma.

The finding that crown eudicots are approximately contemporaneous with crown Magnoliidae and monocots is a noteworthy difference from most previous molecular clock studies (e.g. Soltis *et al.*, 2002; Bell *et al.*, 2005, 2010; Magallón & Sanderson, 2005; Magallón & Castillo, 2009; Magallón *et al.*, 2013), where they were estimated to be younger. Eudicots are morphologically characterized by tricolpate pollen (or derived from this condition; Walker & Walker, 1984; Donoghue & Doyle, 1989), which probably evolved on the stem lineage of this clade. Tricolpate pollen grains are morphologically distinctive, can easily become preserved as fossils, and unequivocally indicate membership to a single clade, thus providing an exceptionally good calibration. Tricolpate pollen has been previously used to calibrate eudicots with a fixed or maximum age of c. 125 Ma, derived from the Barremian–early Aptian age of its oldest fossils (Doyle *et al.*, 1977; Hughes & McDougall, 1990; Doyle & Hotton, 1991). While we recognize the superior potential of tricolpate pollen grains for calibration, here we treated them in the same way as any other fossil used for calibration, and applied their oldest stratigraphic age as a minimum constraint on the eudicot stem node. The eudicots are here estimated as being nearly contemporaneous with Magnoliidae and monocots; therefore, the major components of angiosperm extant diversity began to diversify by the Hauterivian, between 136 and 130 Ma.

Within eudicots, Pentapetalae (70.7% of extant angiosperm richness), characterized by flowers with a five-part organization and distinct calyx and corolla, began to diversify between 126.5 and 120.9 Ma. A very large proportion of species diversity within Pentapetalae is contained in two large clades, Rosidae and Asteridae (29.15% and 35.16% of extant angiosperm richness, respectively), both of which are important components of modern terrestrial biomes. The initial diversification of Rosidae took place between 122.7 and 115.4 Ma, and apparently preceded that of Asteridae, which is estimated to have taken place between 118.8 and 110.4 Ma. Nevertheless, there is a substantial overlap between the two.

Current phylogenetic reconstructions (e.g. Wang *et al.*, 2009; Soltis *et al.*, 2011; but see Qiu *et al.*, 2010 and Zhang *et al.*, 2012) indicate that Rosidae consists of a pair of sister clades, Malvaceae and Fabaceae (10.68% and 18.47% of extant richness, respectively). The onset of Malvaceae diversification is estimated to have occurred between 121.7 and 113.2 Ma, and Fabaceae began to diversify between 120.9 and 113.3 Ma. The crown

Fig. 4 Temporal distribution of family origins. (a) Number of family origins per million year in 10-Myr sliding windows. Green dots: penalized likelihood (PL) estimates; purple dots: Bayesian uncorrelated lognormal (UCLN) method estimates. (b) UCLN-estimated stem ages of families sorted from oldest to youngest. Grey bars correspond to 95% highest posterior density associated with each estimate. Family origins range from *c.* 140 to 20 million yr ago (Ma). (c) PL-estimated stem ages of families sorted from oldest to youngest. Grey bars correspond to confidence interval derived from a sample of dated maximum likelihood bootstrap trees associated with each estimate. Family origins range from *c.* 140 to 40 Ma. According to both UCLN and PL, family origins are constantly distributed, and periods in which family origins are markedly concentrated are not observed. There are fewer family origins at the beginning and end of the respective ranges, and they are more abundant between *c.* 100 and 50–40 Ma. These results are congruent with the number of family origins per Myr shown in (a), where the highest rates are also found between *c.* 100 and 50 Ma.



diversifications of Fabidae and Malvidae took place almost simultaneously, soon after their differentiation within Rosidae.

Asteridae contains a ‘core asterid’ clade consisting of the sister pair Garryidae (18.29% of extant richness) and Campanulidae (12.37% of extant richness). The diversifications of Garryidae and Campanulidae took place almost simultaneously, between 111.7 and 93.4 Ma, and 112.6 and 93.9 Ma, respectively. Hence, the two major clades within Rosidae and within Asteridae each diversified synchronously, but the first pair did so *c.* 10 Myr earlier than the second pair.

The estimated time-frame indicates an early proliferation of major clades in angiosperm history (Figs 2, 3). The major lineages Magnoliidae, Monocotyledoneae and Eudicotyledoneae had originated and started to diversify by the Hauterivian. By the Aptian, the clades that contain a substantial proportion of extant angiosperm species richness, and are major components of extant biomes, had started to radiate: Malvidae and Fabidae in the early Aptian, and Garryidae and Campanulidae in the late Aptian, all during the Early Cretaceous.

The early rise of extant angiosperm families

According to the UNCL time-tree, angiosperm families originated between the Valanginian (Early Cretaceous) and the Miocene (Tertiary), but in the PL time-tree the range is shorter: from the Hauterivian to the Middle Eocene (Tertiary). Considering the oldest and youngest families in the UCLN and PL time-trees, their respective average number of family origins per Myr is 3.18 and 3.96. The number of family origins per Myr calculated in 10-Myr sliding windows indicates the highest rates between 100

and 60 Myr for PL, and between 90 and 50 Myr for UCLN (Fig. 4a).

Plots of UCLN and PL stem ages sorted from oldest to youngest (Fig. 4b,c) show a continuous origin of families from the Early Cretaceous (Hauterivian) to the Tertiary (early Miocene according to UCLN, and Middle Eocene according to PL), with fewer family origins immediately after the onset of angiosperm diversification (between *c.* 140 and 100 Ma), and as the present is approached (between *c.* 55 and 20 Ma in UCLN, and 60–40 Ma in PL). The time between *c.* 100 and 50–40 Ma shows the highest accumulation (Fig. 4b,c). These findings are congruent with the number of family origins per Myr in 10-Myr windows (Fig. 4a), which show lower rates at the beginning and end of the range and higher rates in the middle. The UCLN and PL time-trees show that the number of Cretaceous family origins substantially exceeds the Tertiary number (62.9% and 82.3%, respectively), showing that well over half of extant families have deep evolutionary roots.

Do families originate? The evolutionary significance of clades above the species level is currently being investigated (e.g. Barraclough, 2010; Humphreys & Barraclough, 2014). Processes that influence the generation of new species have been shown to be relevant above the species level, and to lead to higher evolutionarily significant units (Humphreys & Barraclough, 2014). Here, we consider that the origin of a family corresponds to a speciation event in which at least one of the descendants has acquired (or will acquire through its evolutionary trajectory) some type of distinctiveness (genetic, phenotypic, functional, or ecological) that (in hindsight) will allow taxonomists to postulate that species or its descendants as a family. Distinctiveness may be associated with

the phylogenetic differentiation and early stem evolution of the lineage, involving the establishment of new niches. Species richness (including the crown group) would be acquired subsequently. This implies that an adaptive radiation took place early within angiosperms, leading to the establishment of the major morphological and functional attributes that characterize its major lineages. The rapid construction of morphospace early in evolutionary radiations has been documented (Erwin, 2007). Alternatively, distinctiveness may be associated with the acquisition of species richness within a lineage, possibly associated with the diversification of its crown group. This scenario implies independent radiations in separate angiosperm lineages, possibly taking place at different times. These alternatives are not mutually exclusive, specifically considering an early adaptive radiation within angiosperms involving phylogenetic branching associated with large-level differentiation, and subsequent differentiation within separate lineages involving the generation of species richness.

The lineages that contain a very substantial amount of today's angiosperm species richness and their morphological, functional, ecological and genetic diversity were distinct very early in angiosperm history. Nevertheless, because of sampling density, the time-trees cannot show the time of acquisition of species richness. Species richness may be dissociated from the proliferation of major phylogenetic branches. Species that exist in the present (i.e. crown groups) may have originated soon after the differentiation of the family that contains them (short fuse) or substantially postdate it (long fuse), including the possibility that extant species originated recently.

The period in which most family origins are concentrated roughly corresponds to the onset of the Late Cretaceous to the end of the Middle Eocene. We note that this period was a time of pronounced tectonic and geological activity, and high global temperatures (Zachos *et al.*, 2001; Willis & McElwain, 2002). Is there a link between these global events and the increased origin of angiosperm families? Some studies have discussed the possible relationship between high environmental energy and high species richness (Davies *et al.*, 2004; Jansson & Davies, 2008). However, as previously discussed, the origin of a major clade (e.g. a family) implies differences from the increase in speciation, namely, the association with some type of substantial distinctiveness that will allow taxonomists to recognize that lineage as evolutionarily cohesive. There is no conclusive evidence of a relationship between the time of global tectonic change and high temperatures and the major concentration of family origins, and it is not clear what (if any) are the links between higher temperatures and enhanced morphological and functional evolution.

Conclusions

A large number of fossil-derived calibrations and a confidence interval on angiosperm age have been combined in relaxed clock analyses to provide a time-frame of angiosperm evolution. The maximum age of the onset of diversification of angiosperms into their living diversity has been calculated with high confidence to lie in the Early Cretaceous. This estimate was obtained in the context of a bias to estimate an old maximum age, derived from a

strong underestimation of the average number of fossil localities from which each angiosperm family is known. An independent evaluation of the numerically estimated maximum angiosperm age provided here could be conducted with a recently available method that integrates the fossil record of the group in the context of a birth–death process (Heath *et al.*, 2014).

Why many molecular clocks have estimated substantially older ages for the angiosperm crown node (Fig. 1) requires to be investigated. Whereas relaxed clocks rely on increasingly powerful models to estimate divergence times, it has been shown that under complex estimation conditions or molecular model misspecification, absolute molecular rates and divergence times may be erroneously estimated (e.g. Jansa *et al.*, 2006; Hugall *et al.*, 2007; Lepage *et al.*, 2007; Brandley *et al.*, 2011). Relaxed clocks have been found to underestimate the magnitude of rate heterogeneity in trees with extensive rate variation (e.g. Wertheim *et al.*, 2012), or a single parametric distribution has been found to be insufficient to capture the variation in molecular rates found in some empirical trees (e.g. Dornburg *et al.*, 2012), in both cases leading to age overestimation.

This study documents the early rise of angiosperm phylogenetic diversity, including the early origin of more than half of extant angiosperm families. The estimated time-trees represent a solid framework for further investigating angiosperm evolution, for example, the rate of morphological and molecular change, biogeographical history, diversification dynamics, ancestral character reconstruction and state-dependent diversification; as well as coevolution with other biological lineages, correlations between diversification and the physical environment, and the evolution of modern terrestrial biomes. Nevertheless, many questions about the processes involved in early angiosperm evolution remain. To name only a few: What is the relationship between the detected early phylogenetic proliferation and the acquisition of distinctive (e.g. phenotypic or ecological) attributes? What is the relationship between the early rise of angiosperm major clades, including numerous families, and the acquisition of species richness, particularly extant species richness? Is there a relationship between global high temperatures and the origin of major angiosperm clades? If so, how do high temperatures influence the rates of phylogenetic branching and the evolution of phenotypic and functional attributes?

Acknowledgements

We thank Colin Hughes, Peter Linder and Reto Nyffeler for organizing the Radiations Conference, and inviting us to contribute an article to this special issue. We thank C. R. Marshall for guidance in calculating the confidence interval on angiosperm age; J. A. Doyle for information on angiosperm pollen; M. J. Sanderson and J. Schenk for suggestions on dating analyses; H. Sauquet for relevant observations; and P. R. Crane for helpful advice. C. Bell and two anonymous reviewers made many relevant observations. L. Eguiarte, A. Delgado-Salinas, G. Ortega-Leite, R. Hernández-Gutiérrez and James Fouzi provided comments and critical help. P. Linder provided useful comments. Dating analyses were conducted in the CIPRES Science Gateway.

The Coordinación de la Investigación Científica, Universidad Nacional Autónoma de México (UNAM) provided postdoctoral funding to S.G-A. CONACyT (grant no. 410511/262540) provides funding to L.S-R. L.S-R. thanks the Posgrado en Ciencias Biológicas, UNAM.

References

- Aris-Brosou S, Yang Z. 2003. Bayesian models of episodic evolution support a late Precambrian explosive diversification of the Metazoa. *Molecular Biology and Evolution* 20: 1947–1954.
- Axelrod D. 1952. A theory of angiosperm evolution. *Evolution* 6: 29–60.
- Axelrod D. 1970. Mesozoic paleogeography and early angiosperm history. *Botanical Review* 36: 277–319.
- Barracough T. 2010. Evolving entities: towards a unified framework for understanding diversity at the species and higher levels. *Philosophical Transactions of the Royal Society B: Biological Sciences* 365: 1801–1813.
- Baum D, Smith S. 2013. *Tree thinking: an introduction to phylogenetic biology*. Greenwood Village, CO, USA: Roberts & Company.
- Bell C, Soltis D, Soltis P. 2005. The age of the angiosperms: a molecular timescale without a clock. *Evolution* 59: 1245–1258.
- Bell C, Soltis D, Soltis P. 2010. The age and diversification of the angiosperms re-visited. *American Journal of Botany* 97: 1296–1303.
- Brandl R, Mann W, Sprinzl M. 1992. Estimation of the monocot-dicot age through rRNA sequences from the chloroplast. *Proceedings of the Royal Society B* 249: 13–17.
- Brandley M, Wang Y, Guo X, Montes-de-Oca A, Feriá-Ortiz M, Hikida T, Ota H. 2011. Accommodating heterogeneous rates of evolution in molecular divergence dating methods: an example using intercontinental dispersal of Pleistiodon (Eumeces). *Systematic Biology* 60: 3–15.
- Brenner G. 1996. Evidence for the earliest stage of angiosperm pollen evolution: a paleo-equatorial section from Israel. In: Taylor D, Hickey L, eds. *Flowering plant origin, evolution and phylogeny*. New York, NY, USA: Chapman & Hall, 91–115.
- Cantino P, Doyle J, Graham S, Judd W, Olmstead R, Soltis D, Soltis P, Donoghue M. 2007. Towards a phylogenetic nomenclature of Tracheophyta. *Taxon* 56: 1599–1613.
- Cardinal S, Danforth B. 2013. Bees diversified in the age of eudicots. *Proceedings of the Royal Society B* 280: 2012–2686.
- Clarke J, Warnock R, Donoghue P. 2011. Establishing a time-scale for plant evolution. *New Phytologist* 192: 266–301.
- Collinson M. 1983. *Fossils plants of the London Clay*. London, UK: The Palaeontological Association.
- Davies TJ, Savolainen V, Chase MW, Moat J, Barracough TG. 2004. Environmental energy and evolutionary rates in flowering plants. *Proceedings of the Royal Society B* 271: 2195–2200.
- Donoghue MJ, Doyle JA. 1989. Phylogenetic analysis of angiosperms and the relationships of Hamamelidae. In: Crane PR, Blackmore S, eds. *Evolution, systematics, and fossil history of the Hamamelidae*, vol 40A. Oxford, UK: Clarendon Press, 17–45.
- Donoghue P, Benton M. 2007. Rocks and clocks: calibrating the Tree of Life using fossils and molecules. *Trends in Ecology & Evolution* 22: 424–431.
- Dornburg A, Brandley M, McGowen M, Near T. 2012. Relaxed clocks and inferences of heterogeneous patterns of nucleotide substitution and divergence time estimates across whales and dolphins (Mammalia: Cetacea). *Molecular Biology and Evolution* 29: 721–736.
- Doyle J. 2012. Molecular and fossil evidence on the origin of angiosperms. *Annual Review of Earth and Planetary Sciences* 40: 301–326.
- Doyle J, Biens P, Doerenkamp A, Jardiné S. 1977. Angiosperm pollen from the pre-Albian Lower Cretaceous of equatorial Africa. *Bulletin des Centres de Recherches Exploration-Production Elf-Aquitaine* 1: 451–473.
- Doyle J, Endress P. 2000. Morphological phylogenetic analysis of basal angiosperms: comparison and combination with molecular data. *International Journal of Plant Sciences* 161: S121–S153.
- Doyle J, Endress P. 2010. Integrating Early Cretaceous fossils into the phylogeny of living angiosperms: Magnoliidae and eudicots. *Journal of Systematics and Evolution* 48: 1–35.
- Doyle J, Hotton C. 1991. Diversification of early angiosperm pollen in a cladistic context. In: Blackmore S, Barnes SH, eds. *Pollen and spores: patterns of diversification*. Oxford, UK: Clarendon, 169–195.
- Drummond A, Ho S, Phillips M, Rambaut A. 2006. Relaxed phylogenetics and dating with confidence. *PLoS Biology* 4: e88.
- Drummond A, Suchard M. 2010. Bayesian random local clocks, or one rate to rule them all. *BMC Biology* 8: 114.
- Edgar R. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* 32: 1792–1797.
- Ericson P, Klopstein S, Irestedt M, Nguyen J, Nylander J. 2014. Dating the diversification of the major lineages of Passeriformes (Aves). *BMC Evolutionary Biology* 14: 8.
- Erwin D. 2007. Disparity: morphological pattern and developmental context. *Palaeontology* 50: 57–73.
- Friedman W. 2009. The meaning of Darwin's "abominable mystery". *American Journal of Botany* 96: 5–21.
- Friis E, Crane P, Pedersen K. 2011. *Early flowers and angiosperm evolution*. Cambridge, UK: Cambridge University Press.
- Goremykin V, Hansmann S, Martin WF. 1997. Evolutionary analysis of 58 proteins encoded in six completely sequenced chloroplast genomes: revised estimates of two seed plant divergence times. *Plant Systematics and Evolution* 206: 337–351.
- Hall T. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series* 41: 95–98.
- Heath TA. 2012. A hierarchical Bayesian model for calibrating estimates of species divergence times. *Systematic Biology* 61: 793–809.
- Heath TA, Huelsenbeck JP, Stadler T. 2014. The fossilized birth–death process for coherent calibration of divergence-time estimates. *Proceedings of the National Academy of Sciences, USA* 111: E2957–E2966.
- Heimhofer U, Hochuli P, Burla S, Weissert H. 2007. New records of Early Cretaceous angiosperm pollen from Portuguese coastal deposits: implications for the timing of the early angiosperm radiation. *Review of Palaeobotany and Palynology* 144: 39–76.
- Hibbett D, Matheny P. 2009. The relative ages of ectomycorrhizal mushrooms and their plant hosts estimated using Bayesian relaxed molecular clock analyses. *BMC Biology* 7: 13.
- Ho S. 2007. Calibrating molecular estimates of substitution rates and divergence times in birds. *Journal of Avian Biology* 38: 409–414.
- Hochuli PA, Feist-Burkhardt S. 2013. Angiosperm-like pollen and Afropollis from the Middle Triassic (Anisian) of the Germanic Basin (Northern Switzerland). *Frontiers in Plant Science* 4: 344.
- Hugall A, Foster R, Lee M. 2007. Calibration choice, rate smoothing, and the pattern of tetrapod diversification according to the long nuclear gene RAG-1. *Systematic Biology* 56: 543–563.
- Hughes NF. 1994. *The enigma of angiosperm origins*. Cambridge, UK: Cambridge University Press.
- Hughes N, McDougall A. 1987. Records of angiosperm pollen entry into the English Early Cretaceous succession. *Review of Palaeobotany and Palynology* 50: 255–272.
- Hughes N, McDougall A. 1990. Barremian–Aptian angiosperm pollen records from southern England. *Review of Palaeobotany and Palynology* 65: 145–151.
- Hughes N, McDougall A, Chapman J. 1991. Exceptional new record of Cretaceous Hauerivian angiosperm pollen from southern England. *Journal of Micropalaeontology* 10: 75–82.
- Humphreys AM, Barracough TG. 2014. The evolutionary reality of higher taxa in mammals. *Proceedings of the Royal Society B* 281: 20132750.
- Hunt T, Bergsten J, Levkancova Z, Papadopoulou A, St. John O, Wild R, Hammond P, Ahrens D, Balke M, Caterino M *et al.* 2007. A comprehensive phylogeny of beetles reveals the evolutionary origins of a superradiation. *Science* 318: 1913–1916.
- Jansa S, Barker F, Heaney L. 2006. The pattern and timing of diversification of Philippine endemic rodents: evidence from mitochondrial and nuclear gene sequences. *Systematic Biology* 55: 73–88.
- Jansson R, Davies TJ. 2008. Global variation in diversification rates of flowering plants: energy vs. climate change. *Ecology Letters* 11: 173–183.

- Jetz W, Thomas G, Joy J, Hartmann K, Mooers A. 2012. The global diversity of birds in space and time. *Nature* 491: 444–448.
- Knobloch E, Mai D. 1986. Monographie der Früchte und Samen in der Kreide von Mitteleuropa. *Rozprawy Ústředního Ústavu Geologického* 47: 1–219.
- Laroche J, Li P, Bousquet J. 1995. Mitochondrial DNA nd monocot-dicot divergence time. *Molecular Biology and Evolution* 12: 1151–1156.
- Lepage T, Bryant D, Philippe H, Latrillot N. 2007. A general comparison of relaxed molecular clock models. *Molecular Biology and Evolution* 24: 2669–2680.
- Li W-H, Graur D. 1991. *Fundamentals of molecular evolution*. Sunderland, MA: Sinauer.
- Li W-H, Tanimura M. 1987. The molecular clock runs more slowly in man than in apes and monkeys. *Nature* 326: 93–96.
- Magallón S. 2007. From fossils to molecules: phylogeny and the core eudicot floral groundplan in Hamamelidoideae (Hamamelidaceae, Saxifragales). *Systematic Botany* 32: 317–347.
- Magallón S. 2010. Using fossils to break long branches in molecular dating: a comparison of relaxed clocks applied to the origin of angiosperms. *Systematic Biology* 59: 384–399.
- Magallón S. 2014. A review of the effect of relaxed clock method, long branches, genes, and calibrations in the estimation of angiosperm age. *Botanical Sciences* 92: 1–22.
- Magallón S, Castillo A. 2009. Angiosperm diversification through time. *American Journal of Botany* 96: 349–365.
- Magallón S, Hilu K, Quandt D. 2013. Land plant evolutionary timeline: gene effects are secondary to fossil constraints in relaxed clock estimation of age and substitution rates. *American Journal of Botany* 100: 556–573.
- Magallón S, Sanderson M. 2005. Angiosperm divergence times: the effect of genes, codon positions, and time constraints. *Evolution* 59: 1653–1670.
- Manchester S. 1999. Biogeographical relationships of North American Tertiary Floras. *Annals of the Missouri Botanical Garden* 86: 472–522.
- Marshall C. 1990. Confidence intervals on stratigraphic ranges. *Paleobiology* 16: 1–10.
- Marshall C. 1994. Confidence intervals on stratigraphic ranges: partial relaxation of the assumption of randomly distributed fossil horizons. *Paleobiology* 20: 459–469.
- Marshall C. 1997. Confidence intervals on stratigraphic ranges with nonrandom distributions of fossil horizons. *Paleobiology* 23: 165–173.
- Marshall C. 2008. A simple method for bracketing absolute divergence times on molecular phylogenies using multiple fossil calibration points. *The American Naturalist* 171: 726–742.
- Martin W, Gierl A, Saedler H. 1989. Molecular evidence for pre-Cretaceous angiosperm origins. *Nature* 339: 46–48.
- Martin W, Lydiate D, Brinkmann H, Forkmann G, Saedler H, Cerff R. 1993. Molecular phylogenies in angiosperm evolution. *Molecular Biology and Evolution* 10: 140–162.
- Martínez-Millán M. 2010. Fossil record and age of the Asteridae. *The Botanical Review* 76: 83–135.
- Martínez-Millán M, Crepet W, Nixon K. 2009. *Pentapetalum trifasciculandricus* gen. et sp. nov., a thealean fossil flower from the Raritan Formation, New Jersey, USA (Turonian, Late Cretaceous). *American Journal of Botany* 96: 933–949.
- McKenna D, Sequeira A, Marvaldi A, Farrell B. 2009. Temporal lags and overlap in the diversification of weevils and flowering plants. *Proceedings of the National Academy of Sciences, USA* 106: 7083–7088.
- Miller M, Pfeiffer W, Schwartz T. 2010. *Creating the CIPRES Science Gateway for inference of large phylogenetic trees*. New Orleans, LA, USA: Gateway Computing Environments Workshop (GCE).
- Moore M, Bell C, Soltis P, Soltis D. 2007. Using plastid genome-scale data to resolve enigmatic relationships among basal angiosperms. *Proceedings of the National Academy of Sciences, USA* 104: 19363–19368.
- Pyron R. 2011. Divergence time estimation using fossils as terminal taxa and the origins of Lissamphibia. *Systematic Biology* 60: 466–481.
- Qiu Y-L, Li L, Wang B, Xue J-Y, Hendry TA, Li RQ, Brown JW, Liu Y, Hudson GT, Chen ZD. 2010. Angiosperm phylogeny inferred from sequences of four mitochondrial genes. *Journal of Systematics and Evolution* 48: 391–425.
- Rambaut A, Drummond A. 2009. *Tracer version 1.5.0*. [WWW document] URL <http://tree.bio.ed.ac.uk/software/tracer/> [accessed 1 July 2013].
- Ramshaw J, Richardson D, Meatyard B, Brown R, Richardson M, Thompson E, Boulter D. 1972. The time of origin of the flowering plants determined by using amino acid sequence data of cytochrome c. *New Phytologist* 71: 773–779.
- Rannala B, Yang Z. 2007. Inferring speciation times under an episodic molecular clock. *Systematic Biology* 56: 453–466.
- Ronquist F, Klopfstein S, Vilhelmsen L, Schulmeister S, Murray D, Rasnitsyn A. 2012b. A total-evidence approach to dating with fossils, applied to the early radiation of the Hymenoptera. *Systematic Biology* 61: 973–999.
- Ronquist F, Teslenko M, van der Mark P, Ayres D, Darling A, Höhna S, Larget B, Liu L, Suchard M, Huelsenbeck J. 2012a. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology* 61: 539–542.
- Sanderson M. 1997. A nonparametric approach to estimating divergence times in the absence of rate constancy. *Molecular Biology and Evolution* 14: 1218–1231.
- Sanderson M. 2002. Estimating absolute rates of molecular evolution and divergence times: a penalized likelihood approach. *Molecular Biology and Evolution* 19: 101–109.
- Sanderson M. 2004. *r8s, version 1.70 user's manual*. [WWW document] URL <http://loco.biosci.arizona.edu/r8s/> [accessed 1 April 2013].
- Sanderson M, Doyle J. 2001. Sources of error and confidence intervals in estimating the age of angiosperms from rbcL and 18S rDNA data. *American Journal of Botany* 88: 1499–1516.
- Sauquet H, Ho S, Gandolfo M, Jordan G, Wilf P, Cantrill D, Bayly M, Bromham L, Brown G, Carpenter R. 2012. Testing the impact of calibration on molecular divergence times using a fossil-rich group: the case of *Nothofagus* (Fagales). *Systematic Biology* 61: 298–313.
- Schneider H, Schuettelpelz E, Pryer K, Cranfill R, Magallón S, Lupia R. 2004. Ferns diversified in the shadow of angiosperms. *Nature* 428: 553–557.
- Smith A, Pisani D, Mackenzie-Dodds J, Stockley B, Webster B, Littlewood D. 2006. Testing the molecular clock: molecular and paleontological estimates of divergence times in the Echinoidea (Echinodermata). *Molecular Biology and Evolution* 23: 1832–1851.
- Smith S, Beaulieu J, Donoghue M. 2010. An uncorrelated relaxed-clock analysis suggests an earlier origin for flowering plants. *Proceedings of the National Academy of Sciences, USA* 107: 5897–5902.
- Smith S, O'Meara B. 2012. treePL: divergence time estimation using penalized likelihood for large phylogenies. *Bioinformatics* 28: 2689–2690.
- Soltis D, Smith S, Cellinese N, Wurdack K, Tank D, Brockington S, Refulio-Rodriguez N, Walker J, Moore M, Carlswald B *et al.* 2011. Angiosperm phylogeny: 17 genes, 640 taxa. *American Journal of Botany* 98: 704–730.
- Soltis P, Soltis D, Savolainen V, Crane P, Barraclough T. 2002. Rate heterogeneity among lineages of tracheophytes: integration of molecular and fossil data and evidence for molecular living fossils. *Proceedings of the National Academy of Sciences, USA* 99: 4430–4435.
- Stamatakis A. 2006a. Phylogenetic models of rate heterogeneity: a high performance computing perspective. *Proceedings of 20th IEEE International Parallel and Distributed Processing Symposium (IPDPS2006)*, Rhodes, Greece.
- Stamatakis A. 2006b. RAXML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22: 2688–2690.
- Stamatakis A. 2008. A rapid bootstrap algorithm for the RAXML web servers. *Systematic Biology* 57: 758–771.
- Stevens P. 2013. *Angiosperm phylogeny website. Version 12, July 2012 [and more or less continuously updated since]*. [WWW document] URL <http://www.mobot.org/MOBOT/Research/APweb/welcome.html> [accessed 1 April 2013].
- Strauss D, Sadler P. 1989. Stochastic models for the completeness of stratigraphic sections. *Mathematical Geology* 21: 37–59.
- Wahlberg N, Wheat C, Peña C. 2013. Timing and patterns in the taxonomic diversification of Lepidoptera (butterflies and moths). *PLoS One* 8: e80875.
- Walker J, Geissman J. 2009. Geologic time scale. *GSA Today* 19: 60–61.
- Walker JW, Walker AG. 1984. Ultrastructure of Lower Cretaceous angiosperm pollen and the origin and early evolution of flowering plants. *Annals of the Missouri Botanical Garden* 71: 464–521.

- Wang H, Moore M, Soltis P, Bell C, Brockington S, Alexandre R, Davis C, Latvis M, Manchester S, Soltis D. 2009. Rosid radiation and the rapid rise of angiosperm-dominated forests. *Proceedings of the National Academy of Sciences, USA* 106: 3853–3858.
- Wertheim J, Fourment M, Kosakovsky Pond S. 2012. Inconsistencies in estimating the age of HIV-1 subtypes due to heterotachy. *Molecular Biology and Evolution* 29: 451–456.
- Wikström N, Kenrick P. 2001. Evolution of Lycopodiaceae (Lycopsidea): estimating divergence times from *rbcL* gene sequences by use of nonparametric rate smoothing. *Molecular Phylogenetics and Evolution* 19: 177–186.
- Wilkinson R, Steiper M, Soligo C, Martin R, Yang Z, Tavaré S. 2011. Dating primate divergences through an integrated analysis of palaeontological and molecular data. *Systematic Biology* 60: 16–31.
- Willis KJ, McElwain JC. 2002. *The evolution of plants*. Oxford, UK: Oxford University Press.
- Wolfe KH, Gouy M, Yang YW, Sharp PM, Li WH. 1989. Date of the monocot-divergence estimated from chloroplast DNA sequence data. *Proceedings of the National Academy of Sciences, USA* 86: 6201–6205.
- Yang Z, Rannala B. 2006. Bayesian estimation of species divergence times under a molecular clock using multiple fossil calibrations with soft bounds. *Molecular Biology and Evolution* 23: 212–226.
- Zachos J, Pagani M, Sloan L, Thomas E, Billups K. 2001. Trends, rhythms, and aberrations in global climate 65 Ma to present. *Science* 292: 686–693.
- Zhang N, Zeng L, Shan H, Ma H. 2012. Highly conserved low-copy nuclear genes as effective markers for phylogenetic analyses in angiosperms. *New Phytologist* 195: 923–937.

Supporting Information

Additional supporting information may be found in the online version of this article.

Table S1 List of species and GenBank accessions

Table S2 Ages of major angiosperm clades

Methods S1 Fossil-derived calibrations.

Methods S2 Angiosperm branches with the highest empirical scaling factor.

Please note: Wiley Blackwell are not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the *New Phytologist* Central Office.



About New Phytologist

- *New Phytologist* is an electronic (online-only) journal owned by the New Phytologist Trust, a **not-for-profit organization** dedicated to the promotion of plant science, facilitating projects from symposia to free access for our Tansley reviews.
- Regular papers, Letters, Research reviews, Rapid reports and both Modelling/Theory and Methods papers are encouraged. We are committed to rapid processing, from online submission through to publication 'as ready' via *Early View* – our average time to decision is <26 days. There are **no page or colour charges** and a PDF version will be provided for each article.
- The journal is available online at Wiley Online Library. Visit **www.newphytologist.com** to search the articles and register for table of contents email alerts.
- If you have any questions, do get in touch with Central Office (np-centraloffice@lancaster.ac.uk) or, if it is more convenient, our USA Office (np-usaoffice@lancaster.ac.uk)
- For submission instructions, subscription and all the latest information visit **www.newphytologist.com**