

Evolution of the angiosperms: calibrating the family tree

Niklas Wikström^{1*}, Vincent Savolainen² and Mark W. Chase²

¹*Department of Botany, The Natural History Museum, Cromwell Road, London SW7 5BD, UK*

²*Molecular Systematics Section, Jodrell Laboratory, Royal Botanic Gardens, Kew, Richmond, Surrey TW9 3DS, UK*

Growing evidence of morphological diversity in angiosperm flowers, seeds and pollen from the mid Cretaceous and the presence of derived lineages from increasingly older geological deposits both imply that the timing of early angiosperm cladogenesis is older than fossil-based estimates have indicated. An alternative to fossils for calibrating the phylogeny comes from divergence in DNA sequence data. Here, angiosperm divergence times are estimated using non-parametric rate smoothing and a three-gene dataset covering *ca.* 75% of all angiosperm families recognized in recent classifications. The results provide an initial hypothesis of angiosperm diversification times. Using an internal calibration point, an independent evaluation of angiosperm and eudicot origins is performed. The origin of the crown group of extant angiosperms is indicated to be Early to Middle Jurassic (179–158 Myr), and the origin of eudicots is resolved as Late Jurassic to mid Cretaceous (147–131 Myr). Both estimates, despite a conservative calibration point, are older than current fossil-based estimates.

Keywords: angiosperm; phylogeny; divergence times; non parametric rate smoothing; fossils

1. INTRODUCTION

Flowering plants (angiosperms) comprise an estimated 250 000 species; they completely dominate most terrestrial ecosystems, and in terms of species numbers they represent an overwhelming majority of extant land plants. The morphological, ecological and physiological diversity observed among angiosperms is unparalleled in any other plant group, and this diversity has attracted a significant proportion of plant research. Nevertheless, our understanding of the origin and diversification of angiosperms has been hampered by a number of problems. Relationships among extant lineages have been difficult to resolve, the rooting of the angiosperm clade using morphological criteria has been problematic, and the early fossil record has been comparatively poorly understood and insufficiently known (Crane *et al.* 1995). During the last two decades, significant progress has been made concerning these problems. Phylogenetic analyses of both morphological and molecular (DNA sequence) data have recently resolved major relationships among angiosperm lineages (Donoghue & Doyle 1989; Chase *et al.* 1993, 2000; Doyle *et al.* 1994; Soltis 1997; Nandi *et al.* 1998; Qiu *et al.* 1999; Soltis *et al.* 1999, 2000; Savolainen *et al.* 2000*a,b*), congruent patterns concerning the rooting of the angiosperm clade have emerged (Mathews & Donoghue 1999; Qui *et al.* 1999; Soltis *et al.* 1999), and a renewed interest and focus on the fossil record, particularly on Cretaceous deposits, have yielded a wealth of diverse and well-preserved mummified and charcoalified flowers (see Friis *et al.* 1999).

The earliest fossils generally accepted as angiosperms are pollen records from Valanginian–Hauterivian (141–132 Myr) deposits (Brenner & Bickoff, 1992; Hughes 1994; Brenner 1996; Trevisan 1988), but in the Aptian–Albian (125–97 Myr) of North America and the Barremian–Aptian (132–112 Myr) of Portugal, a rapid expansion of

morphological diversity in flowers, seeds and pollen has recently been documented (Friis *et al.* 1999). Furthermore, fossils considered to be members of derived angiosperm lineages are being documented from increasingly older geological deposits. Crepet & Nixon (1998), for example, documented Clusiaceae from Turonian (90–88 Myr) deposits of New Jersey, Keller *et al.* (1996) and Herendeen *et al.* (1999) documented Actinidiaceae from Campanian (83–74 Myr) and Santonian (87–83 Myr) deposits, Pérez-Hernández *et al.* (1997) documented Phytolaccaceae from the Campanian (83–74 Myr), Herendeen *et al.* (1999) suggested a possible affinity to Araliaceae/Apiaceae for one of their Santonian (87–83 Myr) fossils and Basinger & Dilcher (1984) documented a possible Rhamnaceae/Rosaceae from the early Cenomanian (97–94 Myr) of Nebraska. The full impact of these reports can only be appreciated by considering the emerging patterns of relationships among angiosperm lineages (Soltis *et al.* 1999, 2000). The presence of these derived groups in Cenomanian–Campanian deposits implies either that we have underestimated the rapid and explosive nature of the angiosperm diversification or that cladogenesis in basal angiosperms took place considerably earlier than fossil-based estimates have indicated.

An alternative to fossils for estimating divergence times comes from using divergence in DNA sequence data. Such estimates are, however, known to suffer from problems, some associated with small datasets and stochastic errors (Hillis *et al.* 1996), and others with an inability to correctly infer rate change over the tree (Sanderson 1997, 1998). Furthermore, they have until recently relied on the assumption that sequences evolve roughly at constant rates. A different approach, non-parametric rate smoothing (NPRS), was recently developed (Sanderson 1997). Rather than assuming rate constancy, Sanderson's method allows the rate to change but assumes that such changes are autocorrelated (Sanderson 1997), which supposes that rate change is inherited from an ancestral lineage by their immediate

* Author for correspondence: (N.Wikstrom@nhm.ac.uk).

descendants. Through optimization techniques, the method searches for the solution that minimizes the inferred rate changes. Here we use NPRS to estimate divergence times in angiosperms using a three-gene dataset based on plastid *rbcL* and *atpB* exons and nuclear 18S rDNA that covers 560 angiosperms (Soltis *et al.* 1999, 2000). Our primary aim is to provide an initial hypothesis of angiosperm diversification times based on sequence divergence data that represents a majority of angiosperm families. By using an internal calibration point for relative ages, an independent evaluation of angiosperm and eudicot origins is accomplished. Results are compared with estimates based on fossil information, recently reviewed by Magallón *et al.* (1999), and possible directions for future improvements are discussed.

2. MATERIALS AND METHODS

(a) Dataset

In a collaborative effort, nucleotide sequence data covering a majority of all flowering plant families have over the last decade been assembled for three loci, *rbcL* (Chase *et al.* 1993, 2000; Savolainen *et al.* 2000b) and *atpB* (Savolainen *et al.* 2000a) from the plastid genome, and 18S rDNA (Soltis 1997) from the nucleus. These efforts recently culminated in a three-gene phylogenetic analysis (Soltis *et al.* 1999, 2000) including 560 angiosperm and seven outgroup taxa representing *ca.* 75% of the angiosperm families recognized in the most up-to-date classification (APG 1998). We have used the complete data matrix from this analysis to calculate branch lengths on one of the more than 8000 most parsimonious trees obtained by Soltis *et al.* (1999); the tree used corresponds to that reported in their figs 1B–10B (see § 4).

Although the dataset includes three genes that may have different rate dynamics, existing methods cannot combine data and at the same time account for such differences. We have therefore calculated branch lengths on our tree using the combined data, and although there are likely to be different rate dynamics, there are also no compelling reasons that such differences would violate the assumption of autocorrelation. If anything, the use of three different genes with different patterns of molecular evolution would tend to compensate for unusual patterns in any single dataset, as has been argued by Qiu *et al.* (1999). As an explorative measure, we did conduct separate analyses for *rbcL* and *atpB*, and about half of the node dates fell outside the estimated error bounds based on the three genes combined (data not shown). This approach, however, leads to difficulties with short (zero length) branches, which creates severe analytical problems (Sanderson 1997) and also greatly increases the stochastic errors. The seven outgroup taxa used by Soltis *et al.* (1999) were initially included to obtain branch length estimates for the first ingroup branching point but were subsequently removed from the analyses. Branch lengths were estimated with both parsimony methods, accelerated and delayed transformation (ACCTRAN and DELTRAN respectively), and with maximum likelihood methods. The HKY85 model of sequence evolution (Hasegawa *et al.* 1985) was used in the likelihood estimates, and transition/transversion ratios and nucleotide frequencies were estimated from the data. Branch length calculations were made using PAUP 4.0b4a (Swofford 1998).

(b) Non-parametric rate smoothing analyses

NPRS analyses were done using the r8s program (Sanderson 1997). To prevent the algorithm converging on a local optimum,

the searches were started at five different initial time estimates. Local stability of the solutions for each estimate was checked by perturbing them and restarting the search three times. Three consecutive analyses were carried out using the different branch lengths from the ACCTRAN and DELTRAN optimizations and likelihood analysis. No minimum age constraints were enforced.

Errors in age estimates resulting from the stochastic nature of substitution processes were assessed using a bootstrap resampling procedure (Efron & Tibshirani 1993). One hundred bootstrap replicates were constructed using the SEQBOOT program (Felsenstein 1993), and branch lengths were calculated using ACCTRAN optimization for each replicate and input to the r8s program. Bootstrap estimates of standard error for each node were calculated for the age-distribution estimates obtained (Efron & Tibshirani 1993).

(c) Time calibration

To convert the relative ages obtained through the analyses into dates, a single absolute calibration point has to be selected with reference to the fossil record. Important considerations for this choice include: (i) terminal nodes should be avoided to minimize effects of a poor taxon sampling; (ii) the fossil taxon should undisputedly be part of the group defined by the selected node; (iii) the age of the fossil taxon should as closely as possible represent the actual divergence time for the selected node; and (iv) relationships of the selected group to other taxa should be well supported by the bootstrap/jackknife.

Given these criteria, we have chosen to calibrate our tree by fixing the split between Fagales and Cucurbitales in the Late Santonian at 84 Myr based on the occurrence of *Protofagacea* (Herendeen *et al.* 1995) and *Antiquacupula* (Sims *et al.* 1998) in the Campanian and Late Santonian of Georgia. A number of floral features indicate that they are part of the Fagales lineage (Herendeen *et al.* 1995; Sims *et al.* 1998), and both have flowers and fruits born in a typical Fagales cupule.

Evaluating their precise relationships is, however, complicated by uncertainties regarding the origin of the cupule (figure 1). Recent analyses based on both morphological and molecular data indicated that Fagaceae *sensu lato* are paraphyletic, with Nothofagaceae and Fagaceae *sensu stricto* forming two separate lineages (Chase *et al.* 1993; Manos *et al.* 1993; Manos 1997; Manos & Steele 1997; Nixon 1989). The cupule must therefore either have evolved twice or originated once in the Fagales lineage and subsequently been lost in the lineage leading to Betulaceae, Casuarinaceae, Juglandaceae and Myricaceae. By using this conservative estimate (Fagales–Cucurbitales split) we can control the direction of incorporated errors and be confident that we are underestimating the age of our calibration point.

3. RESULTS

Results of the analyses are presented in the form of a chronogram (figure 2) calibrated against the geological time-scale (Harland *et al.* 1990). Additional chronograms (figs 3–13) covering all included taxa are given in electronic Appendix A and can be retrieved from The Royal Society Web Site (<http://www.pubs.royalsoc.ac.uk>). Chronograms presented (figs 2–13) are based on the analysis using parsimony with ACCTRAN optimization for calculating branch lengths. Details of all three analyses, using both ACCTRAN and DELTRAN optimizations and likelihood analyses for calculating branch

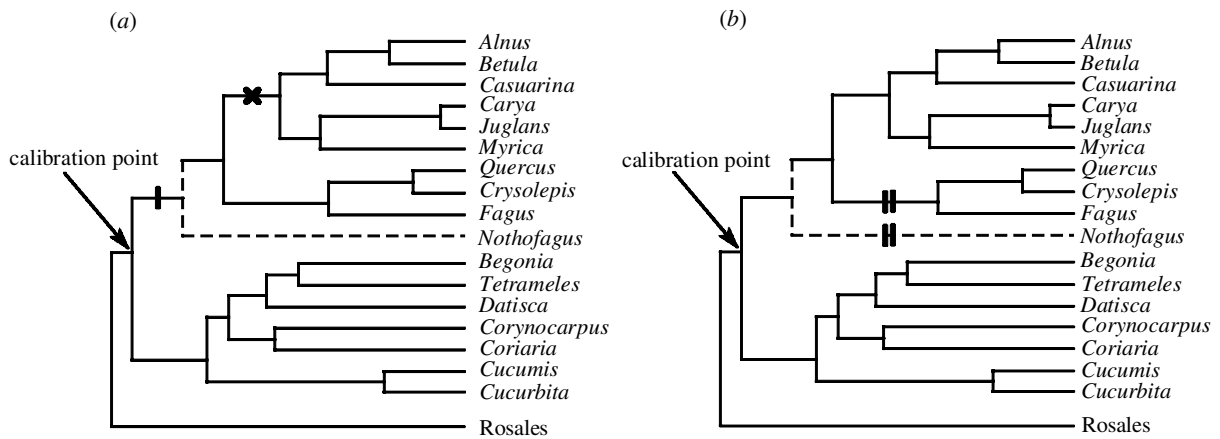


Figure 1. Two alternative possibilities for the origin of the Fagales cupule: (a) it originated once and was subsequently lost in the lineage leading to Betulaceae, Casuarinaceae, Juglandaceae and Myricaceae; (b) it originated twice, once in Fagaceae s. str. and once in Notofagaceae. Evaluating the precise relationships of the fossil taxa *Antiquacupula* and *Protofagacea* is complicated by this uncertainty, and we have therefore chosen a conservative strategy by calibrating our estimates using the split between Cucurbitales and Fagales. The position of Notofagaceae on the tree (Chase *et al.* 1993; Manos *et al.* 1993; Manos 1997; Manos & Steele 1997; Nixon 1989) is marked by a dashed line because the analyses by Soltis *et al.* (1999) did not include members of this group.

lengths, together with estimates of standard errors, are presented in table A1 in electronic Appendix A, which lists age estimates of all nodes.

4. DISCUSSION

(a) *Origins of angiosperms and eudicots*

The crown group of extant angiosperms is resolved to have originated in the Early–Middle Jurassic (179–158 Myr), and eudicots are indicated as Late Jurassic–mid Cretaceous (Gallic) (147–131 Myr). Despite the conservative age estimate for our calibration point, these estimates are older than nearly all previous fossil-based estimates.

Claims of a pre-Cretaceous crown group diversification of angiosperms have been made before, based both on fossil evidence (Cornet & Habib 1992; Cornet 1993) and molecular clock estimates (Ramshaw *et al.* 1972; Martin *et al.* 1989, 1993; Wolfe *et al.* 1989; Brandl *et al.* 1992; Goremykin *et al.* 1997); however, from a palaeobotanical perspective, the appearances of angiosperms in the Valanginian (through putative magnolid pollen), eudicots around the Barremian–Aptian boundary (through their triaperturate pollen), and rosids and hamamelids in the Early Cenomanian, has been described as an orderly sequence, and one that such pre-Cretaceous claims must confront (Crane *et al.* 1995). It is, however, not the sequence of appearance that poses a problem, but the ages themselves. The fossil evidence indicates that the time-intervals separating basal branches are short and that major angiosperm lineages diverged within a comparatively short time-span (Hickey & Doyle 1977; Lidgard & Crane 1988; Crane & Lidgard 1989; Taylor & Hickey 1990; Crane *et al.* 1995). Nevertheless, we see a substantial amount of nucleotide change on those branches, and in our molecular-based estimates, angiosperm and eudicot origins are pushed back in time. Our results, in this respect, corroborate previous molecular estimates in placing the origin of extant angiosperms in the Early–Middle Jurassic and the origin of extant eudicots in the Late Jurassic–mid Cretaceous.

If claims of a pre-Cretaceous angiosperm diversification need to confront the orderly sequence of appearances seen in the fossil record, claims of a Cretaceous diversification need to confront the long branch lengths seen in the molecular phylogenetic trees (Soltis *et al.* 1999, 2000). There are, of course, alternative explanations for those branch lengths. They might be incorrectly inferred, and true branch lengths might be considerably shorter. An explanation such as this would, however, have serious consequences for phylogenetic analyses, indicating that the support for basal cladogenesis in angiosperms is based on spurious and incorrectly inferred evidence. However, other non-molecular lines of evidence corroborating these phylogenetic relationships are substantial (Nandi *et al.* 1998). Alternatively, if these branch lengths are correct, then the inferred rates may not be correct, resulting in time-intervals between cladogenic events that are too large. This implies that rapid morphological diversification of early angiosperms was accompanied by equally rapid molecular change. Such a pattern would contrast starkly with the patterns seen in groups that have diversified more recently such as Asterales and Lamiales, and it thus seems illogical that more recent patterns would be qualitatively different from older ones. This issue of correlated or non-correlated change of morphological and molecular characters was addressed by Bateman (1999) by looking at ‘architectural radiations’ on volcanic islands, and this island approach may provide a way to address this question at a more general level.

Pushing the origins of angiosperms and eudicots back in time implies that there is a gap in the fossil record, and from a palaeobotanical perspective such a gap may seem unlikely to be real (Crane *et al.* 1995). It is somewhat difficult to evaluate the extent to which there are inconsistencies within the fossil-based estimates themselves. Palaeobotanical work often uses unresolved and collapsed phylogenies (Crane *et al.* 1995; Magallón *et al.* 1999), and any inconsistency might become clearer within a more rigorous hierarchical framework. Within the zoological community, two methods have been

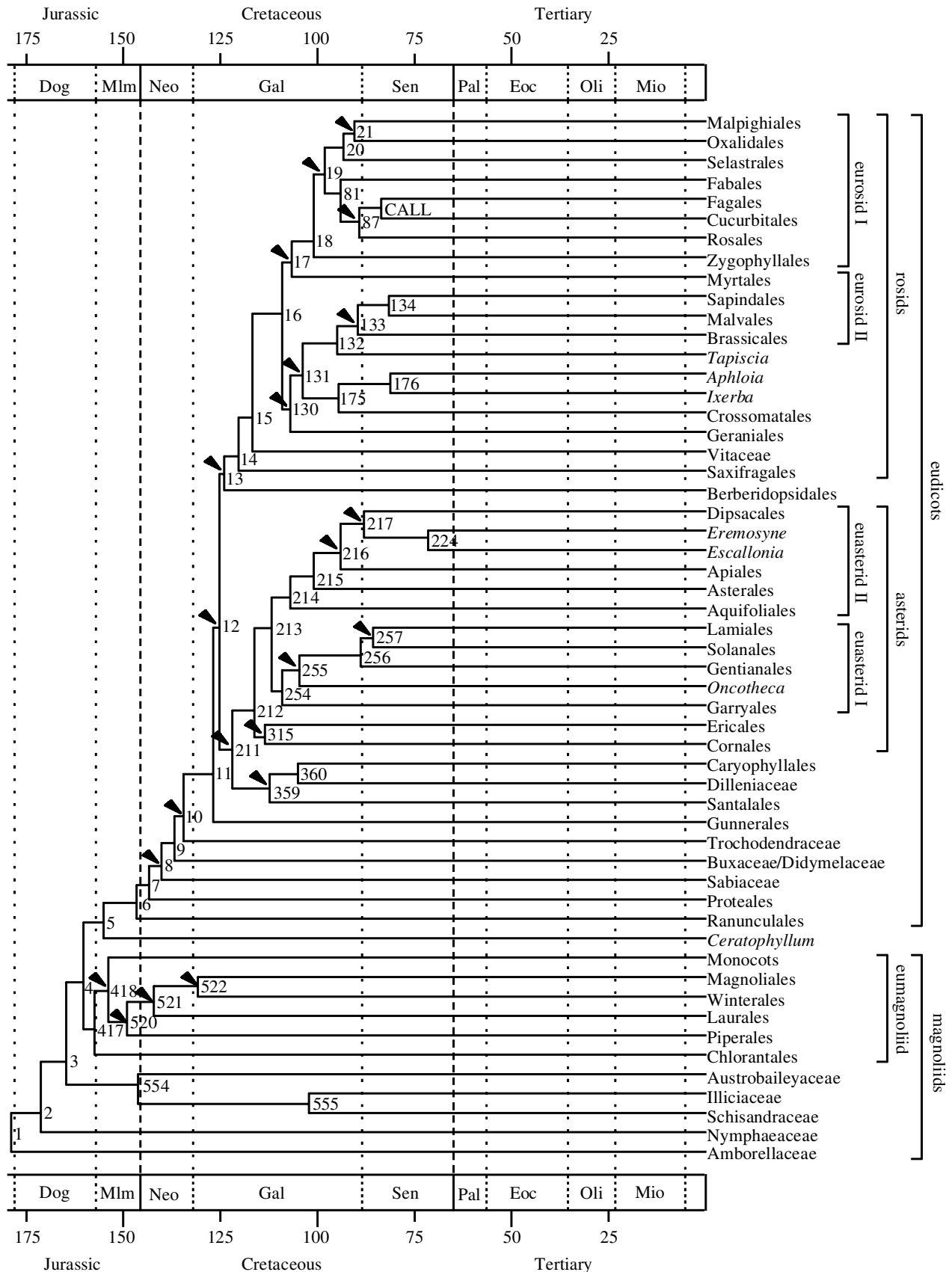


Figure 2. Chronogram calibrated against the geological time-scale (Harland *et al.* 1990) focusing on early cladogenic events within angiosperms. The chronogram is based on the analysis using ACCTRAN optimization for resolving character change ambiguities. Arrows indicate nodes with less than 50% jackknife support (Soltis *et al.* 1999, 2000). Node numbers correspond to those given in electronic Appendix A (<http://www.pubs.royalsoc.ac.uk>) in table 2 and supplementary chronograms (figures 3–13) which show divergence times for all 560 included taxa.

Table 1. Comparison between fossil-based estimates extracted from Magallón *et al.* (1999) and our analyses.

(The fossil-based estimates are, contrary to their use, assumed to provide a minimum age for the split between the taxon in question and its sister group (Doyle & Donoghue 1993). Column 1 lists the taxa according to the usage by Magallón *et al.* (1999), columns 2 and 3 list our age estimates (node numbers correspond to those on the chronograms, figures 2–13). The age span given in columns 2 and 3 results from the three consecutive analyses using ACCTAN and DELTRAN optimization and maximum likelihood for calculating branch lengths. Column 4 lists the specific age estimate given by Magallón *et al.* (1999) and also indicates on what taxon the age estimate was based; column 5 lists instances when the topology implies that some other fossil provides a better minimum age estimate for a node. In addition, further comparisons are made whenever the Magallón *et al.* (1999) estimate was based on a less inclusive taxon. The age for the ranunculid clade, for example, was based on Menispermaceae (Magallón *et al.* 1999). We have therefore included our estimate for both the ranunculid and Menispermaceae clades. MYBP, million years before present.)

taxon	estimated age	(MYBP)	specific fossil-based age (MYBP)	implied fossil-based age (MYBP)
Ranunculid clade	Barremian–Tithonian (node 6)	131–147	69 (Menispermaceae)	118 (Trochodendrales ^a)
Menispermaceae	Albian–Aptian (node 407)	103–113	69	
Nelumbonaceae	Barremian–Valanginian (node 396)	125–137	100	108 (Platanaceae)
Platanaceae	Albian–Aptian (node 397)	108–117	108	
Proteaceae	Albian–Aptian (node 397)	108–117	97	108 (Platanaceae)
Sabiaceae	Barremian–Valanginian (node 8)	128–140	69	118 (Trochodendrales ^a)
Buxaceae	Aptian (node 393)	113–124	104	
Trochodendrales ^a	Aptian–Hauterivian (node 10)	123–135	118 (Tetracentraceae)	
Tetracentraceae ^a	Cenomanian–Albian (node 392)	95–106	18	
Caryophyllid clade	Albian (node 360)	104–111	83 (Amaranthaceae)	
Amaranthaceae	Chattian–Bartonian (node 382)	28–40	83	
Saxifragoids	Albian–Aptian (node 14)	111–121	89 (saxifragaleans)	
Saxifragaleans	Campanian–Cenomanian (node 191)	78–91	89	
Geraniaceae	Maastrichtian–Santonian (node 182)	71–85	8	
expanded Capparales	Santonian–Turonian (node 133)	85–90	89 (Capparales)	
Capparales	Ypresian (node 167)	52–54	89	
Sapindales	Campanian–Santonian (node 134)	80–84	67 (Rutaceae Aceraceae)	69 (Malvales)
Rutaceae	Lutetian (node 141)	45–47	67	
expanded Malvales	Campanian–Santonian (node 134)	80–84	69 (Bombacaceae)	
Bombacaceae	Chattian–Rupelian (node 156)	28–31	69	
Myrtales	Albian (node 17)	100–107	84 (Combretaceae)	
Combretaceae	Campanian (node 122)	75–79	84	
Cucurbitales	our calibration point	84	58 (Cucurbitaceae)	84 (Fagales)
Cucurbitaceae	Maastrichtian (node 97)	65–66	58	
Urticales	Maastrichtian (node 104)	65–67	69 (Celtidoideae)	
Celtidoideae	Ypresian–Thanetian (node 110)	55–57	69	
Rosaceae	Campanian (node 103)	76	44 (Prunoideae)	69 (Urticales)
Prunoideae	Chattian–Rupelian (node 119)	29–35	44	
higher Hamamelididae	our calibration point	84	84 (<i>Normapolles</i>)	
<i>Normapolles</i> clade	Thanetian–Danian (node 89)	60–61	84	
Polygalaceae	Maastrichtian (node 85)	66–68	68	
Fabaceae	Campanian (node 82)	74–79	56–65	68 (Polygalaceae)
expanded Cunoniaceae	Coniacian–Cenomanian (node 21)	88–91	58 (Eleocarpaceae)	
Eleocarpaceae	Danian–Maastrichtian (node 67)	64–66	58	
Malphigiales	Coniacian–Cenomanian (node 21)	88–91	58 (Euphorbiaceae)	

continued

Table 1. *continued*

taxon	estimated age	(MYBP)	specific fossil-based age (MYBP)	implied fossil-based age (MYBP)
Euphorbiaceae	Maastrichtian (node 63)	69–71	58	
Cornalean clade	Albian–Aptian (node 315)	106–114	69 (mastixioid taxa)	89 (ericalean clade)
Mastixioid taxa	Coniacian–Cenomanian (node 348)	87–92	69	
Ericalean clade	Albian–Aptian (node 315)	106–114	89 (Ericaceae s. lato)	
Ericaceae s. lato	Ypresian (node 323)	50–56	89	
<i>Ilex</i> clade	Albian (node 214)	99–107	69 (<i>Ilex</i>)	
<i>Ilex</i>	Lutetian–Ypresian (node 253)	49–55	69	
Apiales	Santonian–Turonian (node 225)	85–90	69 (Araliaceae)	
Araliaceae	Bartonian–Lutetian (node 229)	41–45	69	
Dipsacales	Santonian–Turonian (node 217)	85–90	53 (Caprifoliaceae)	
Caprifoliaceae	Ypresian–Thanetian (node 219)	54–58	53	
Asterales	Cenomanian–Albian (node 215)	94–101	29 (Menyanthaceae, Goodeniaceae)	69 (Araliaceae)
Menyanthaceae	Maastrichtian (node 238)	65–69	29	
Goodeniaceae	Bartonian (node 244)	39–42	29	
<i>Garrya</i> clade	Albian (node 254)	100–109	46 (<i>Eucommia</i>)	53 (Boraginales)
<i>Eucommia</i>	Campanian–Santonian (node 313)	80–84	46	
Boraginales	Campanian (node 258)	77–81	53	
Solanales	Campanian–Santonian (node 257)	82–86	53 (Convolvulaceae)	
Convolvulaceae	Maastrichtian (node 295)	65–66	53	
Gentianales	Santonian–Turonian (node 256)	83–89	53 (Apocynaceae, Rubiaceae)	
Apocynaceae	Lutetian–Ypresian (node 306)	45–53	53	
Rubiaceae	Danian (node 304)	61–64	53	
Lamiales	Maastrichtian (node 259)	71–74	37 (Oleaceae)	
Oleaceae	Ypresian–Danian (node 260)	55–64	37	
Santalales	Albian–Aptian (node 359)	111–118	53 (Olacaceae)	83 (Caryophyllid clade)
Olacaceae	Santonian–Cenomanian (node 385)	85–97	53	
Dilleniaceae	Albian (node 360)	104–111	53	83 (Caryophyllid clade)
<i>Vitis</i> –Leeaceae	Albian–Aptian (node 15)	108–117	58	84 (Fagales)
Gunneraceae	Albian–Aptian (node 391)	108–118	89	

^a Magallón *et al.* (1999) indicated an Aptian (124.5–112 Myr) occurrence for *Populus potomacensis*, but the original documentation of this taxon indicated the Albian (112–97 Myr) (Doyle & Hickey 1976).

suggested for evaluating whether the fossil record is complete enough to disregard the existence of such gaps (Marchall 1998; Foot *et al.* 1999). Bleiweiss (1998) and Benton (1999) adopted the ‘gap analyses’ from Marchall (1998) to test if fossil gaps of birds and mammals, implied by molecular estimates for these groups, could be real, whereas Foot *et al.* (1999) looked at sampling intensity using a likelihood approach, which they applied to the ‘fossil gap’ of mammals. It would be worthwhile to apply this kind of analysis to the early fossil record of angiosperms.

(b) Fossil-based estimates within eudicots

(i) Deep-level nodes

A comprehensive effort to compile and summarize evidence of early eudicot diversification times from the fossil record was recently published by Magallón *et al.*

(1999), providing an opportunity to compare our molecular estimates with fossil-based estimates. A complication, however, is their use of crown groups versus stem lineages. The earliest appearance of a taxon in the fossil record simply provides a minimum age for the split between that taxon and its sister group (Doyle & Donoghue 1993), yet Magallón *et al.* (1999) assumed that fossil taxa correctly assigned to extant groups are members of that crown group, not the stem lineage. In many of their estimates this is a reasonable assumption, but in others it is more likely to represent a static view of taxa and how they relate in terms of morphological similarity to their extant relatives. The fact that fossil ‘Platanaceae’ from the mid Cretaceous share similarities with living species of the family does not indicate that the fossil taxa are part of the crown group of living species. In fact, the original documentation of the earliest fossil Platanaceae indicates

a stem group position (Crane *et al.* 1993). The only reasonable comparison we can make is that of the fossil-based estimate to the stem lineage leading to extant Platanaceae.

The comparisons are summarized in table 1. In addition to the inclusive groups we have also included comparisons for the estimates of Magallón *et al.* (1999) that were based on a less inclusive taxon. Their ranunculid clade estimate, for example, was based on characteristic endocarps of Menispermaceae from the Maastrichtian of Europe (Magallón *et al.* 1999). We have thus included both our estimate for the split between the ranunculid clade and remaining eudicots, and our estimate for the split between Menispermaceae and its sister within the ranunculid clade. We have also extracted data from their analyses that, given our topology, implies older ages for a clade. Proteaceae, for example, are documented from the mid Cretaceous (97 Myr), but given our topology, the occurrence of their sister group, Platanaceae, in the Early Albian (108 Myr) implies that Platanaceae provide a better minimum age estimate for the stem group Proteaceae.

A general pattern (table 1) is that our analyses indicate older divergence times for most clades. This is particularly so if we do not consider what other fossils imply, and if we only compare our estimates with the more inclusive groups (ranunculid clade rather than Menispermaceae). However, if we consider what ages other fossils imply and use the fossil information in a less conservative way, comparing with the less inclusive groups (Menispermaceae, etc.), we see considerably more congruence. This indicates that the Magallón *et al.* (1999) use of the fossil-based age estimates is far too conservative and underestimates the 'true' ages. Instead of our estimates all being older, our estimates are sometimes older and sometimes younger, particularly if we consider the less inclusive taxa. These differences are probably caused by errors in both our and in the fossil-based estimates. The fossil-based Santonian and Campanian estimate of Amaranthaceae, for example, was listed by Collinson *et al.* (1993) with reference to a personal comment by Friis. E. M. Friis (personal communication) has since confirmed that this fossil is not Amaranthaceae.

(ii) Terminal nodes

A general pattern is that the analyses underestimate the ages for more terminal nodes in the tree. This is true if we compare our estimates with the fossil-based estimates in table 1 (Rutaceae, Bombacaceae, Celtidoideae, Prunoideae and Araliaceae) and also if we extend the comparison to other more terminal nodes with reliable fossil-based estimates. Examples include Poaceae, Moraceae, Salicaceae and Aceraceae (Collinson *et al.* 1993), and the list could no doubt be expanded through a more comprehensive comparison.

A partial explanation for this general pattern relates to the resolution of homoplasy and how this resolution is effected by taxon sampling (Sanderson 1990). For homoplastic characters, parsimony only provides a lower bound on the number of changes, and the inferred positions and numbers of those changes are affected by the thoroughness of taxon sampling. Sanderson (1990) demonstrated that decreased taxon sampling often leads

to a dramatic decrease in the estimates of branch lengths. How different resolutions of homoplastic characters affect our age estimates is shown by the two parsimony-based analyses (ACCTRAN versus DELTRAN for branch length calculations), and the effect of a less dense taxon sampling is illustrated by increased differences towards terminal nodes (table A1 in electronic Appendix A). Terminal nodes are generally resolved to be older when DELTRAN optimization is used, but the most terminal nodes (nodes with the most limited sample) are most highly affected. This behaviour is consistent with the findings of Sanderson (1990), and we would expect an extended sample of groups such as Poaceae and Aceraceae to have the effect of pushing the inferred ages closer in line with the fossil-based estimates. This phenomenon becomes less important deeper in the tree, as is observed with smaller differences between ACCTRAN and DELTRAN estimates. Lineage sampling is more thorough at these levels, and thus a more consistent and probably more accurate estimate of change is obtained.

(c) Errors in age estimates

Errors affecting the accuracy of the estimated times arise from several sources. Specific sources include: (i) the calibration point obtained from the fossil record; (ii) noise introduced from stochastic processes of substitution; (iii) rate variations that invalidate the assumptions of the method (see Hillis *et al.* (1996) and Sanderson (1998) for discussions); and (iv) use of an incorrect tree.

(i) Calibration

The results from unconstrained NPRS analyses are a set of relative ages that can only be converted to absolute geological times by choosing a single fixed calibration point with reference to the fossil record. This calibration has no effect on the actual results: it simply converts the relative ages output by the analyses into geological times. Whatever calibration point we choose, there will be errors associated with it that involve two different problems: one has to do with identifying and correctly inferring the fossil's age and relationships to other taxa; the other is the fact that fossils, even though correctly identified, only provide minimum ages (see Doyle & Donoghue (1993) for a discussion of the latter). Both problems may affect our analyses. The uncertainties surrounding the precise relationships of *Protofagacea* and *Antiquacupula* within the Fagales clade (figure 1) probably incorporate errors into our estimates. Secondly, *Normapolles* type pollen is usually associated with taxa within the Fagales lineage, and this pollen type has an extensive fossil record that possibly extends into older geological deposits (Sims *et al.* 1999). However, the conservative approach adopted ensures that we are underestimating the true age for the Fagales–Cucurbitales split. If *Antiquacupula* and *Protofagacea* were shown to have a more derived position within the Fagales clade, or if the older *Normapolles* type pollen records were accepted as part of the Fagales lineage, all our estimates would become older.

(ii) Noise

Errors introduced through the stochastic nature of substitution can be estimated, and Sanderson (1997, 1998) suggested that this could be done through a bootstrap

resampling procedure (Efron & Tibshirani 1993). As seen from table 2 in electronic Appendix A, bootstrap estimates of standard errors are comparatively small for our estimates (on average *ca.* 5 Myr), indicating that stochastic errors can be reduced by including sufficient data. This is one of the benefits of using three genes combined rather than one, even though the models used cannot accommodate differing patterns among the three genes.

(iii) *Rate assumptions*

An inability to infer shifts in the rate of substitution correctly is perhaps the most problematic source of errors. If any amount or any type of changes is allowed and considered, the estimates from an NPRS analysis will be associated with large errors (Sanderson 1997). There is simply no way to avoid making assumptions about the nature of both rate changes and the rates themselves. The NPRS approach allows substitution rates to change but assumes that these are autocorrelated, which means that substitution rates are assumed to be inherited, and if correct, branch lengths and branch-length variation should have a high degree of lineage specificity. There is, however, not much empirical evidence to support these assumptions (but see Harvey *et al.* (1991) for a discussion of autocorrelation and heritability of cladogenesis). A thorough evaluation would require knowledge of absolute rates, which itself requires knowledge of absolute divergence times (Springer 1995). The assumption is, however, intuitively reasonable, and a different examination of its validity could perhaps be accomplished by looking at how rate changes are inferred and trying to corroborate these changes through further and different kinds of analyses.

(iv) *Topology*

How the timing of a group's divergence is ultimately resolved depends on correctly inferring its relationship to other groups. This is most probably not the case for some of the groups in our phylogenetic analysis. The tree used is just one out of more than 8000 most parsimonious trees (Soltis *et al.* 1999), but there is no reasonable way to evaluate the amount of uncertainty this places on all our estimates. This will simply have to be calculated on a group-by-group basis. It is worth noting, though, that the great majority of groups are consistently resolved and receive ample jackknife support; in particular, the spine and major clades of the tree are clear. Relationships within some of the more derived groups such as Malpighiales and Lamiales, are by necessity resolved in the tree used here, but many of these receive less than 50% jackknife support (Soltis *et al.* 1999, 2000). Resolving the relationships differently within these groups will, however, have limited consequences on timings for the more inclusive groups (Malpighiales, Lamiales); we have also indicated on the figures nodes that receive less than 50% jackknife support. The nodes discussed above are consistent in all 8000 trees, and most are well supported by the jackknife (i.e. greater than 85%).

(d) *Future directions*

The calibrated phylogeny presented here is a working hypothesis and should be viewed as such. The analyses are unconstrained, including no fossil-based minimum age constraints; this permits us to evaluate how the molecular

data on their own resolve angiosperm diversification. The type of analysis conducted, however, allows for fossil-based minimum age constraints to be enforced during the analysis. Although after such an inclusion we can no longer independently evaluate the fossil estimates, such an approach may provide ways to improve the actual estimates. Such an analysis will require a detailed and critical evaluation of the available fossil information, which is clearly beyond the scope of this work.

Within existing methods, there is no way to combine data (necessary if stochastic errors are to be reduced) and at the same time to take different rate characteristics into account, much as early likelihood models used for phylogeny reconstruction were all simple and without such capabilities. We hope that work such as this will promote not only an evaluation of the assumptions used in NPRS analyses but also further developments, so that we can look forward to corresponding improvements in age estimation analyses such as those we have seen in the development of likelihood models for phylogeny reconstruction. By using the available fossil information, analyses of this kind would have the advantage of providing ways to estimate the time of origin for groups without a good fossil record. They might also force our estimates into a more rigorous hierarchical framework. Without such a framework, the full implications of documenting derived lineages from successively older geological records become less clear. Judged from the results presented here, the report of Phytolaccaceae (Pérez-Hernández *et al.* 1997) from the Campanian (83–74 Myr), for example, must be incorrect, and reports of Clusiaceae from the Turonian (Crepet & Nixon 1998), Actinidiaceae from the Campanian and Santonian (Keller *et al.* 1996; Herendeen *et al.* 1999), and Apiaceae/Araliaceae from the Santonian (Herendeen *et al.* 1999) all imply that we are still underestimating the timing for early angiosperm diversification. This study is, to our knowledge, the first to attempt calibration of nodes on such a broad-scale phylogenetic tree, and this effort will benefit from yet larger phylogenetic analyses.

We acknowledge the support of our institutions and also thank the Systematics Journal Club where the idea for this paper was spawned.

REFERENCES

- APG 1998 An ordinal classification for the families of flowering plants. *Ann. Missouri Bot. Garden* **85**, 531–553.
- Basinger, J. F. & Dilcher, D. L. 1984 Ancient bisexual flowers. *Science* **224**, 511–513.
- Bateman, R. 1999 Architectural radiations cannot be optimally interpreted without morphological and molecular phylogenies. In *The evolution of plant architecture* (ed. M. H. Kurmann & A. R. Hemsley), pp. 221–250. Kew: Royal Botanic Gardens.
- Benton, M. J. 1999 Early origins of modern birds and mammals: molecules vs. morphology. *BioEssays* **21**, 1044–1052.
- Bleiwiss, R. 1998 Fossil gap analysis supports early Tertiary origin of trophically diverse avian orders. *Geology* **26**, 323–326.
- Brandl, R., Mann, W. & Sprintz, M. 1992 Estimation of the monocot–dicot age through tRNA sequences from the chloroplast. *Proc. R. Soc. Lond.* **B249**, 13–17.
- Brenner, G. J. 1996 Evidence for the earliest stage of angiosperm pollen evolution: a paleoecological section from Israel. In

- Flowering plant origin, evolution and phylogeny* (ed. D. W. Taylor & L. J. Hickey), pp. 91–115. New York: Chapman & Hall.
- Brenner, G. J. & Bickoff, I. S. 1992 Palynology and the age of Lower Cretaceous basal Kurnub Group from the coastal plain to the northern Negev of Israel. *Palynology* **16**, 137–185.
- Chase, M. W. (and 41 others) 1993 Phylogenetics of seed plants: an analysis of nucleotide sequences from the plastid gene *rbcL*. *Ann. Missouri Bot. Garden* **80**, 528–580.
- Chase, M. W. (and 12 others) 2000 Higher-level systematics of the monocotyledons: an assessment of current knowledge and a new classification. In *Monocots: systematics and evolution* (ed. K. L. Wilson & D. A. Morrison), pp. 3–16. Collingwood, Australia: Commonwealth Scientific and Industrial Research Organization.
- Collinson, M. E., Boulter, M. C. & Holmes, P. L. 1993 Magnoliophyta ('Angiospermae'). In *The fossil record 2* (ed. M. J. Benton), chapter 45, pp. 809–841. London: Chapman & Hall.
- Cornet, B. 1993 Dicot-like leaf and flowers from the Late Triassic tropical Newark Supergroup rift zone, U.S.A. *Mod. Geol.* **19**, 81–99.
- Cornet, B. & Habib, D. 1992 Angiosperm-like pollen from the ammonite-dated Oxfordian (Upper Jurassic) of France. *Rev. Palaeobot. Palynol.* **71**, 269–294.
- Crane, P. R. & Lidgard, S. 1989 Angiosperm diversification and paleolatitudinal gradients in Cretaceous floristic diversity. *Science* **246**, 675–678.
- Crane, P. R., Friis, E. M., Pedersen, K. R. & Drinnan, A. N. 1993 Early Cretaceous (Early to Middle Albian) platanoid inflorescences associated with *Sapindopsis* leaves from the Potomac Group of eastern North America. *Syst. Bot.* **18**, 328–344.
- Crane, P. R., Friis, E. M. & Pedersen, K. R. 1995 The origin and early diversification of angiosperms. *Nature* **374**, 27–33.
- Crepet, W. L. & Nixon, K. C. 1998 Fossil Clusiaceae from the Late Cretaceous (Turonian) of New Jersey and implications regarding the history of bee pollination. *Am. J. Bot.* **85**, 1122–1133.
- Donoghue, M. J. & Doyle, J. A. 1989 Phylogenetic studies of seed plants and angiosperms based on morphological characters. In *The hierarchy of life: molecules and morphology in phylogenetic analysis* (ed. B. Fernholm, K. Bremer & H. Jörnvall), pp. 181–193. Amsterdam: Elsevier Science.
- Doyle, J. A. & Hickey, L. J. 1976 Pollen and leaves from the mid-Cretaceous Potomac Group and their bearing on the early angiosperm evolution. In *Origin and early evolution of angiosperms* (ed. C. B. Beck), pp. 139–206. Columbia University Press.
- Doyle, J. A. & Donoghue, M. J. 1993 Phylogenies and angiosperm diversification. *Paleobiology* **19**, 141–167.
- Doyle, J. A., Donoghue, M. J. & Zimmer, E. A. 1994 Integration of morphological and ribosomal RNA data on the origin of angiosperms. *Ann. Missouri Bot. Garden* **81**, 419–450.
- Efron, B. & Tibshirani, R. J. 1993 *An introduction to the bootstrap*. New York: Chapman and Hall.
- Felsenstein, J. 1993 *PHYLIP (Phylogeny Inference Package) version 3.5c*. Distributed by the author, Department of Genetics, University of Washington, Seattle.
- Foot, M., Hunter, J. P., Janis, C. M. & Sepkoski, J. J. 1999 Evolutionary and preservational constraints of origins of biologic groups: divergence times of eutherian mammals. *Science* **283**, 1310–1314.
- Friis, E. M., Crane, P. R. & Pedersen, K. R. 1999 Early angiosperm diversification: the diversity of pollen associated with angiosperm reproductive structures in Early Cretaceous floras from Portugal. *Ann. Missouri Bot. Garden* **86**, 259–296.
- Goremykin, V. V., Hansman, S. & Martin, W. F. 1997 Evolutionary analysis of 58 proteins encoded in six completely sequenced chloroplast genomes: revised molecular estimates of two seed plant divergence times. *Plant Syst. Evol.* **206**, 337–351.
- Harland, W. B., Armstrong, R. L., Cox, A. V., Craig, L. E., Smith, A. G. & Smith, D. G. 1990 *A geologic time scale 1989*. Cambridge University Press.
- Harvey, P. H., Nee, S., Mooers, A. O. & Partridge, L. 1991 These hierarchical views of life: phylogenies and metapopulations. In *Genes in Ecology: 33rd Symp. of the British Ecological Society* (ed. R. J. Berry, T. J. Crawford & G. M. Hewitt), pp. 123–137. Oxford: Blackwell.
- Hasegawa, M., Kishino, H. & Yano, T. 1985 Dating the human–ape splitting by a molecular clock of mitochondrial DNA. *J. Mol. Evol.* **21**, 160–174.
- Herendeen, P. S., Crane, P. R. & Drinnan, A. N. 1995 Fagaceous flowers, fruits, and capsules from the Campanian (Late Cretaceous) of central Georgia, U.S.A. *Int. J. Plant Sci.* **156**, 93–116.
- Herendeen, P. S., Magallón-Puebla, S., Lupia, R., Crane, P. R. & Kobylinska, J. 1999 A preliminary conspectus of the Allon Flora from the Late Cretaceous (Late Santonian) of central Georgia, U.S.A. *Ann. Missouri Bot. Garden*, **86**, 407–471.
- Hickey, L. J. & Doyle, J. A. 1977 Early Cretaceous fossil evidence for angiosperm evolution. *Bot. Rev.* **43**, 3–104.
- Hillis, D. M., Mable, B. K. & Moritz, C., 1996 Applications of molecular systematics: the state of the field and a look to the future. In *Molecular systematics*, 2nd edn. (ed. D. M. Hillis, C. Moritz & B. K. Mable), pp. 515–543. Sunderland, MA: Sinauer Associates.
- Hughes, N. F. 1994 *The enigma of angiosperm origins*. Cambridge University Press.
- Keller, J. A., Herendeen, P. S. & Crane, P. R. 1996 Fossil flowers of the Actinidiaceae from the Campanian (Late Cretaceous) of Georgia. *Am. J. Bot.* **83**, 528–541.
- Lidgard, S. & Crane, P. R. 1988 Quantitative analyses of the early angiosperm radiation. *Nature* **331**, 344–346.
- Magallón, S., Crane, P. R. & Herendeen, P. S. 1999 Phylogenetic pattern, diversity, and diversification of eudicots. *Ann. Missouri Bot. Garden* **86**, 297–372.
- Manos, P. S. 1997 Systematics of *Nothofagus* (Nothofagaceae) based on rDNA spacer sequences (ITS): taxonomic congruence with morphology and plastid sequences. *Am. J. Bot.* **84**, 1137–1155.
- Manos, P. S. & Steele, K. P. 1997 Phylogenetic analyses of “higher” Hamamelididae based on plastid sequence data. *Am. J. Bot.* **84**, 1407–1419.
- Manos, P. S., Nixon, K. C. & Doyle, J. J. 1993 Cladistic analyses of restriction site variation within the chloroplast DNA inverted repeat region of selected Hamamelididae. *Syst. Bot.* **18**, 551–562.
- Marchall, C. R. 1998 Determining stratigraphic ranges. In *The adequacy of the fossil record* (ed. S. K. Donovan & C. R. C. Paul), pp. 23–54. Chichester: John Wiley.
- Martin, W., Gierl, A. & Saedler, H. 1989 Molecular evidence for pre-Cretaceous angiosperm origin. *Nature* **339**, 46–48.
- Martin, W., Lydiate, D., Brinkmann, H., Forkmann, G., Saedler, H. & Cerff, R. 1993 Molecular phylogenies in angiosperm evolution. *Mol. Biol. Evol.* **10**, 140–162.
- Mathews, S. & Donoghue, M. J. 1999 The root of angiosperm phylogeny inferred from duplicate phytochrome genes. *Science* **286**, 947–950.
- Nandi, W. I., Chase, M. W. & Endress, P. K. 1998 A combined cladistic analysis of angiosperms using *rbcL* and non-molecular data sets. *Ann. Missouri Bot. Garden* **85**, 137–212.
- Nixon, K. C. 1989 Origins of Fagaceae. In *Evolution, systematics and fossil history of the Hamamelididae* vol. 2, “Higher” Hamamelididae (ed. P. R. Crane & S. Blackmore), pp. 23–43. Oxford University Press.
- Pérez-Hernández, B. R., Rodríguez-de la Rosa, R. A. & Cevallos-Ferriz, S. R. S. 1997 Permineralized infructescence

- from the Cerro del Pueblo formation (Campanian), near Saltillo, Coahuila, Mexico: Phytolaccaceae. *Am. J. Bot.* **84**(Suppl.), 139.
- Qiu, Y.-L., Lee, J., Bernasconi-Quadroni, F., Soltis, D. E., Soltis, P. S., Zanis, M., Chen, Z., Savolainen, V. & Chase, M. W. 1999 The earliest angiosperms: evidence from mitochondrial, plastid and nuclear genomes. *Nature* **402**, 404–407.
- Ramshaw, J. A. M., Richardson, D. L., Meatyrd, B. T., Brown, R. H., Richardson, M., Thompson, E. W. & Boulter, D. 1972 The time of origin of the flowering plants determined by using amino acid sequence data of cytochrome *c*. *New Phytol.* **71**, 773–779.
- Sanderson, M. J. 1990 Estimating rates of speciation and evolution: a bias due to homoplasy. *Cladistics* **6**, 387–391.
- Sanderson, M. J. 1997 A nonparametric approach to estimating divergence times in the absence of rate constancy. *Mol. Biol. Evol.* **14**, 1218–1231.
- Sanderson, M. J. 1998 Estimating rate and time in molecular phylogenies: beyond the molecular clock. In *Molecular systematics of plants II: DNA sequencing* (ed. D. E. Soltis, P. S. Soltis & J. J. Doyle), pp. 242–264. Norwell, MA: Kluwer Academic.
- Savolainen, V., Chase, M. W., Morton, C. M., Hoot, S. B., Soltis, D. E., Bayer, C., Fay, M. F., de Bruijn, A., Sullivan, S. & Qiu, Y.-L. 2000a Phylogenetics of flowering plants based upon a combined analysis of plastid *atpB* and *rbcL* gene sequences. *Syst. Biol.* **49**, 306–362.
- Savolainen, V. (and 16 others) 2000b Phylogeny of the eudicots: a nearly complete familial analysis based on *rbcL* gene sequences. *Kew Bull.* **55**, 257–309.
- Sims, H. J., Herendeen, P. S. & Crane, P. R. 1998 New genus of fossil Fagaceae from the Santonian (Late Cretaceous) of central Georgia, U.S.A. *Int. J. Plant Sci.* **159**, 391–404.
- Sims, H. J., Herendeen, P. S., Lupia, R. A., Christopher, R. A. & Crane, P. R. 1999 Fossil flowers with *Normapolles* pollen from the Late Cretaceous of southeastern North America. *Rev. Palaeobot. Palynol.* **106**, 131–151.
- Soltis, D. E. 1997 Angiosperm phylogeny inferred from 18S ribosomal DNA sequences. *Ann. Missouri Bot. Garden* **84**, 1–49.
- Soltis, P. S., Soltis, D. E. & Chase, M. W. 1999 Angiosperm phylogeny inferred from multiple genes: a research tool for comparative biology. *Nature* **402**, 402–404.
- Soltis, D. E. (and 13 others) 2000 Angiosperm phylogeny inferred from a combined dataset of 18S rDNA, *rbcL*, and *atpB* sequences. *Bot. J. Linnean Soc.* **133**, 381–461.
- Springer, M. 1995 Molecular clocks and the incompleteness of the fossil record. *J. Mol. Evol.* **41**, 531–538.
- Swofford, D. L., 1998 *PAUP**. *Phylogenetic analyses using parsimony (*and other methods)*. Version 4. Sunderland, MA: Sinauer Associates.
- Taylor, D. W. & Hickey, L. J. 1990 An Aptian plant with attached leaves and flowers: implications for angiosperm origin. *Science* **247**, 702–704.
- Trevisan, L., 1988 Angiospermous pollen (monosulcate–trichotomosulcate phase) from very early Lower Cretaceous of southern Tuscany (Italy): Some aspects. In *Abstracts Seventh Int. Palynol. Conf.* p. 165. Brisbane.
- Wolfe, K. H., Gouy, M., Yang, Y.-W., Sharp, P. M. & Li, W.-H. 1989 Date of the monocot–dicot divergence estimated from chloroplast DNA sequence data. *Proc. Nat. Acad. Sciences USA* **86**, 6201–6205.

As this paper exceeds the maximum length normally permitted, the authors have agreed to contribute to production costs.

An electronic appendix to this paper can be found at (<http://www.pubs.royalsoc.ac.uk>).