



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

Effects of taxon sampling on molecular dating for within-genus divergence events, when deep fossils are used for calibration

Citation for published version:

Milne, RI 2009, 'Effects of taxon sampling on molecular dating for within-genus divergence events, when deep fossils are used for calibration', *Journal of Systematics and Evolution*, vol. 47, no. 5, pp. 383-401. <https://doi.org/10.1111/jse.2009.47.issue-5>

Digital Object Identifier (DOI):

[10.1111/jse.2009.47.issue-5](https://doi.org/10.1111/jse.2009.47.issue-5)

Link:

[Link to publication record in Edinburgh Research Explorer](#)

Document Version:

Publisher's PDF, also known as Version of record

Published In:

Journal of Systematics and Evolution

General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.



Effects of taxon sampling on molecular dating for within-genus divergence events, when deep fossils are used for calibration

Richard I. MILNE*

(Institute of Molecular Plant Sciences, The University of Edinburgh, Edinburgh EH9 3JH, UK, and Royal Botanic Garden Edinburgh, Edinburgh EH3 5LR, UK)

Abstract A universal method of molecular dating that can be applied to all families and genera regardless of their fossil records, or lack thereof, is highly desirable. A possible method for eudicots is to use a large phylogeny calibrated using deep fossils including tricolpate pollen as a fixed (124 mya) calibration point. This method was used to calculate node ages within three species-poor disjunct basal eudicot genera, *Caulophyllum*, *Podophyllum* and *Pachysandra*, and sensitivity of these ages to effects such as taxon sampling were then quantified. By deleting from one to three accessions related to each genus in 112 different combinations, a confidence range describing variation due only to taxon sampling was generated. Ranges for *Caulophyllum*, *Podophyllum* and *Pachysandra* were 8.4–10.6, 7.6–20.0, and 17.6–25.0 mya, respectively. However, the confidence ranges calculated using bootstrapping were much wider, at 3–19, 0–32 and 11–32 mya, respectively. Furthermore, deleting 10 adjacent taxa had a large effect in *Pachysandra* only, indicating that undersampling effects are significant among Buxales. Changes to sampling density in neighboring clades, or to the position of the fixed fossil calibration point had small to negligible effects. Non-parametric rate smoothing was more sensitive to taxon sampling effects than was penalized likelihood. The wide range for *Podophyllum*, compared to the other two genera, was probably due to a high degree of rate heterogeneity within this genus. Confidence ranges calculated by this method could be narrowed by sampling more individuals within the genus of interest, and by sequencing multiple DNA regions from all species in the phylogeny.

Key words basal eudicots, fossil calibration, molecular dating, North America–East Asia disjunction, taxon sampling, Tertiary relict, tricolpate pollen.

Molecular dating has become a vital tool in biogeographic research (e.g. Xiang et al., 2000, 2005; Donoghue et al., 2001; Wikström et al., 2001; Renner et al., 2001; Milne & Abbott 2002; Givnish & Renner, 2004; Pennington et al., 2004; Renner, 2004, 2005; Sanmartin & Ronquist, 2004; de Queiroz, 2005) but node ages calculated in this way are liable to vary according to the method used (Sanderson et al., 2004), the choice of molecule examined (Milne & Abbott, 2002; Magallon & Sanderson, 2005), the codon position examined (Sanderson & Doyle, 2001; Yang & Yoder, 2003; Magallon & Sanderson, 2005), the rate smoothing method used (Anderson et al., 2005; Bell & Donoghue, 2005; Linder et al., 2005, Britton et al., 2007), which calibration point is used (Soltis et al., 2002), different phylogeny topologies (Renner, 2005), and the number of taxa sampled from within a clade (Linder et al., 2005). Of all of these, taxon sampling effects have received perhaps the least attention, but error from this source increases with distance from the calibration point (Linder et al., 2005), creating a problem when dating nodes that are not phylogenetically close to any fossil calibration

points. Approaches that do not use fossil calibrations are problematic, as a universal molecular clock cannot be assumed (Gaut, 1998; Kay et al., 2006), and a secondary calibration point introduces large errors into the results (Shaul & Graur, 2002). By contrast, using one or more phylogenetically distant fossils for calibration could be an effective approach, but only if the potential effects of taxon sampling were understood and controlled for.

A second problem in molecular dating is that of calculating maximum ages for disjuncts. Fossil calibration points normally only offer minimum ages for a node, so ages calculated for other nodes based on such calibration are also minimum ages (Renner, 2005; Milne, 2006), even beyond any confidence ranges that are calculated. Using multiple calibration points, and/or where fossils are relatively abundant, allows assertions that nodes are probably not much older than the calculated dates (Feng et al., 2005; Renner, 2005; Xiang et al., 2005), and permits upper age limits to be calculated directly from the fossil record (Tavaré et al., 2002; Marshall, 2008). However, this solves the problem only in groups with a better than average fossil record.

Perhaps the most reliable fixed age fossil calibration point for angiosperms is the appearance of tricolpate pollen 124 mya marking the origin of eudicots

Received: 31 November 2008 Accepted: 12 March 2009

* E-mail: r.milne@ed.ac.uk; Tel.: 44-131-6505322; Fax: 44-131-6505392.

(Crane, 1989; Anderson et al., 2005). Pollen is fossilized relatively easily, and tricolpate pollen appears both in southern England (Hughes & McDougall, 1990) and equatorial Africa (Doyle, 1992) at the Barremian–Aptian border (approximately 124 mya); hence this calibration point is less prone to inaccuracy than others (Crane, 1989; Anderson et al., 2005, q.v. for more detailed arguments). The 124 mya age might be a slight underestimate based on tricolpate pollen already being morphologically diverse (Friis et al., 2006), and present on two different continents (Hughes & McDougall, 1990; Doyle, 1992); however, the effect from this can be controlled somewhat by attaching the 124 my age to the crown node rather than the stem node for eudicots (see below). This calibration point has been used with other deep fossils to provide multiple calibration points to date major divergence events among basal eudicots (Anderson et al., 2005) or major angiosperm clades (Moore et al., 2007). They could also be used to date within-genus divergence events, provided that the above-mentioned effects of taxon sampling were controlled for. However, in addition to the effects of taxon undersampling throughout a clade (as examined by Linder et al., 2005), there would also be potential for error due to large imbalances between species numbers in adjacent clades. For example, Buxales is a species-poor order that is sister to all core eudicots (APG 2003; Anderson et al., 2005). Hence the effects of such imbalances would also need to be tested for if such distance calibration points were to be used for dating within-genus nodes.

Accurate divergence time estimates are of particular value among Tertiary relict floras, which comprise genera that occur disjunctly between two or more of E Asia, N America and SW Eurasia (Tiffney, 1985a,b; Wen, 1999; Xiang et al., 2000; Donoghue et al., 2001; Milne & Abbott, 2002; Milne, 2006), for three reasons. First, unlike the southern hemisphere landmasses, there have been multiple land routes between these landmasses over the past 50 my, each with a different lifespan, climate and latitude (Tiffney 1985a,b; Wen 1999; Donoghue et al., 2001; Milne & Abbott 2002), and accurate divergence time data can thus distinguish between these as putative routes between continents. Second, divergence time data has led to the rejection of vicariance explanations for all but a handful of major southern hemisphere disjunctions (Renner et al., 2001; Givnish & Renner, 2004; Pennington et al., 2004; Renner, 2004, 2005; Sanmartin & Ronquist, 2004; de Queiroz, 2005; McGlone, 2005), and because of this Tertiary relict floras might provide the best example of a major disjunction, involving large numbers of genera, that came about by vicariance. Finally, for many N America–E

Asia disjuncts among Tertiary relict floras, calculated divergence times are generally in the range of 3–15 mya (Xiang et al., 2000; Donoghue et al., 2001). Such a date range is compatible with the last connection having been across the Bering land bridge (BLB). However, with more precise dates including a meaningful upper age limit, it would also be possible to determine whether vicariance was caused by climatic cooling in Beringia (10–15 mya; Wolfe, 1994; White et al., 1997; Tiffney & Manchester, 2001; Milne, 2006) or the sundering of the BLB itself (5.5–5.4 mya, Gladenkov et al., 2002).

Many Tertiary relict disjuncts belong to species-poor genera (e.g. see Xiang et al., 2000) and make ideal subjects for testing the sensitivity of node ages to taxon sampling effects when distant calibration points are used. Therefore, three basal eudicot genera were chosen that have a species each in North America and Eurasia: *Caulophyllum* Michx., *Podophyllum* L. (both Berberidaceae) and *Pachysandra* Michx. (Buxaceae). *Podophyllum* is here taken to include *Sinopodophyllum* T.S. Ying. but excludes other genera occasionally united with it (*Dysosma*, *Diphylleia*), for simplicity. A phylogeny based on *rbcL* for basal eudicots was used for this exercise. Although multiple DNA regions and/or a faster evolving region would have been desirable, only *rbcL* had the necessary breadth of sampling across basal eudicot families for this study. Using multiple calibration points and rate smoothing by penalized likelihood (PL), initial ages and bootstrap confidence ranges were calculated for the basal node within each genus. These values were then used to quantify the effects of the following: (i) slight changes in taxon sampling close to the node to be dated; (ii) reduction of sampling density around the node to be dated; (iii) increase in sampling density in single clades that did not contain nodes to be dated; (iv) changing the topology of the phylogeny by switching the positions of major clades; (v) changing the node to which the fixed age of 124 mya is assigned, from crown to stem node of eudicots; and (vi) using non-parametric rate smoothing (NPRS) to smooth branch lengths instead of PL.

1 Material and methods

1.1 Data matrix construction

All complete *rbcL* sequences available for “stem eudicotyledons” on Genbank in August 2005 were downloaded, as well as three outgroups (one *Chloranthus* Sw. and two *Ceratophyllum* L. accessions) and three representatives of the basal clade among the remaining eudicots. The latter, *Gunnera* L. (two accessions) and *Myrothamnus* Welw. (one accession), served

as placeholders for core eudicots. Three accessions in the dataset failed to group with the correct family in an initial phylogenetic analysis and were therefore discarded as being possibly sequenced from wrongly labelled material. These were *Cocculus trilobus* D85696 (which grouped within *Caulophyllum*), *Podophyllum emodi* AF203487 (which differed by >50 mutations from any sequence in the set and >90 from any other *Podophyllum*), and *Platanus occidentalis* L01943 (which was identical to *Trochodendron* L01958 and differed by 55 mutations from any other *Platanus* L. sequence). In addition, the positions of *Oceanopapaver* Guillaumin and *Hainania* Merr. (both Papaveraceae according to Genbank) were unclear in this phylogeny, so these too were removed for simplicity. Sequences with poly-N stretches were discarded, and initially only one accession per genus was retained, except in the three target genera, as well as three genera that were the sole available representatives of their families, that is, *Ceratophyllum*, *Gunnera* and *Nelumbo* Adans. *Buxus* L. was also sampled twice, to increase representation of Buxaceae, as was *Meliosma*, to increase sampling close to Buxaceae. Finally, families that had a large number of accessions relative to others in the dataset (Ranunculaceae and Proteaceae) had some accessions removed, to reduce imbalance in clade size in the initial analysis, and to allow a test of the effects of varying the size of individual clades. The data matrix for the initial analysis therefore comprised 77 *rbcL* sequences, with 10 and 12 extra sequences from Ranunculaceae and Proteaceae, respectively, that were added in specific analyses (Table 1).

Among the target genera, the data set contained two accessions each of the E Asian *Caulophyllum robustum* Maxim. (AF190441 and D85695), and the N American *Caulophyllum thalictroides* Michx. (L08760 and AF190442); one accession of the E Asian *Podophyllum hexandrum* Royle (AF079455) (= *Podophyllum emodi* Wall., = *Sinopodophyllum hexandrum* (Royle) T.S.Ying) and four of the N American *Podophyllum peltatum* L. (AF093716, AF139877, AF197591, and AF203488); one accession of the E Asian *Pachysandra axillaris* Franch. (AF203485), and three of the N American *Pachysandra procumbens* Michx. (AF061993, AF093718 and AF203486). Other than these, the dataset comprised 10 accessions from Berberidaceae, 4 from Buxaceae, 13 from Proteaceae, 7 from Ranunculaceae, 4 from Menispermaceae, 3 each from Lardizabalaceae and Sabiaceae, 2 each from Fumariaceae, Gunneraceae, Nelumbonaceae and Papaveraceae, and 1 each from Circaeasteraceae, Didymelaceae, Eupteleaceae, Kingdoniaceae, Myrothamnaceae, Platanaceae, Tetracentraceae, and Trochodendraceae. It also included,

from outside of the eudicots, two of *Ceratophyllum* and one of *Chloranthus* (Table 1).

1.2 Phylogeny construction

An initial phylogeny was generated in PAUP* 4 (version 4.0d81; Swofford, 2002) by a Heuristic search under the following conditions: optimality criterion = parsimony, unlimited maxtrees, collapse and multrees options in effect, TBR on. These conditions were retained for all subsequent analyses, the only difference being that a topological constraint tree was enforced. The constraint tree was constructed by hand in the PAUP input file, to make families and genera monophyletic, and match the topology of the majority rule consensus of all most parsimonious trees. The remaining polytomies at high taxonomic levels were resolved by reference to the APG (2003) phylogeny (but see analysis G, below). Polytomies within families were resolved by taking the topology from a most parsimonious tree chosen at random. The final constraint tree therefore contained no polytomies. This topology is shown in Fig. 1.

1.3 Molecular dating: rate-smoothing

Linder et al. (2005) found that, of three methods that accommodate rate variation for dating nodes within a phylogeny, PL was the least sensitive to undersampling effects. Therefore, PL was used in the initial analysis (A) and in subsequent analyses that tested various effects of taxon sampling (B–F) and phylogeny topology (G; see below). Of the other methods, maximum likelihood is much more time-consuming, which was prohibitive for running hundreds of parallel analyses, as this study required. For the same reason, Bayesian methods (e.g., Ronquist & Huelsenbeck, 2003; Drummond et al., 2006) could not be used in the current study. Conversely, rate-smoothing by PL and NPRS can be achieved very quickly by the program (Sanderson, 2002). Therefore, PL was used for all analyses in this study, except that certain analyses were also rate-smoothed using NPRS in order to compare the two methods. Dating was carried out by exporting trees from PAUP in NeXus format and moving each individual tree into an input file for the r8s program. Input files specified rate-smoothing either by PL (with TN algorithm) or NPRS (with Powell algorithm).

1.4 Fossil calibration: fixed age point (tricolpate pollen)

The r8s program requires one calibration point of fixed age, and for this, as in Anderson et al. (2005), the first appearance of tricolpate pollen at 124 mya was used. The most correct way to employ this calibration

Table 1 Accessions included in the analysis, their source, and in which analyses they were present

| Species | Genbank accession number | Analyses that each accession was present (+) or absent (-) for: | | | | | Reference for sequence |
|---|--------------------------|---|------|-----|---|---|----------------------------------|
| | | A,G,H,I | B, J | D,L | E | F | |
| Berberidaceae | | | | | | | |
| <i>Achlys triphylla</i> | L75868 | + | +/- | - | + | + | Kim & Jansen, 1996 |
| <i>Berberis thunbergii</i> | AF139878 | + | +/- | - | + | + | Wang et al., unpubl. data, 1999. |
| <i>Bongardia chrysogonum</i> | L75870 | + | +/- | - | + | + | Kim & Jansen, 1996 |
| <i>Caulophyllum robustum</i> 1 | D85695 | + | + | + | + | + | Kofuji et al., 1994 |
| <i>Caulophyllum robustum</i> 2 | AF190441 | + | + | + | + | + | Xiang et al., 2000 |
| <i>Caulophyllum thalictroides</i> 1 | L08760 | + | + | + | + | + | Chase et al., 1993 |
| <i>Caulophyllum thalictroides</i> 2 | AF190442 | + | + | + | + | + | Xiang et al., 2000 |
| <i>Diphylleia cymosa</i> | L75866 | + | +/- | - | + | + | Kim & Jansen, 1996 |
| <i>Dysosma versipellis</i> | AF079454 | + | +/- | - | + | + | Wang et al., 2005 |
| <i>Epimedium koreanum</i> | L75869 | + | +/- | - | + | + | Kim & Jansen, 1996 |
| <i>Jeffersonia diphylla</i> | L75867 | + | +/- | - | + | + | Kim & Jansen, 1996 |
| <i>Mahonia bealei</i> (= <i>Berberis bealei</i>) | L12657 | + | +/- | - | + | + | Qiu et al., 1993 |
| <i>Nandina domestica</i> | L37920 | + | +/- | - | + | + | Hoot et al., 1995 |
| <i>Podophyllum hexandrum</i> | AF079455 | + | + | + | + | + | Wang et al., 2005 |
| <i>Podophyllum peltatum</i> 1 | AF093716 | + | + | + | + | + | Hoot et al., 1999 |
| <i>Podophyllum peltatum</i> 2 | AF139877 | + | + | + | + | + | Wang et al., unpubl. data, 1999. |
| <i>Podophyllum peltatum</i> 3 | AF197591 | + | + | + | + | + | Qiu et al., 1999 |
| <i>Podophyllum peltatum</i> 4 | AF203488 | + | + | + | + | + | Xiang et al., 2000 |
| <i>Ranzania japonica</i> | L75853 | + | +/- | - | + | + | Kim & Jansen, 1996 |
| Buxaceae | | | | | | | |
| <i>Buxus sempervirens</i> 1 | AF093717 | + | +/- | - | + | + | Hoot et al., 1999 |
| <i>Buxus sempervirens</i> 2 | AF543712 | + | +/- | - | + | + | Hilu et al., unpubl. data, 2002 |
| <i>Pachysandra axillaris</i> | AF203485 | + | + | + | + | + | Xiang et al., 2000 |
| <i>Pachysandra procumbens</i> 1 | AF061993 | + | + | + | + | + | Qiu et al., 1998 |
| <i>Pachysandra procumbens</i> 2 | AF093718 | + | + | + | + | + | Hoot et al., 1999 |
| <i>Pachysandra procumbens</i> 3 | AF203486 | + | + | + | + | + | Xiang et al., 2000 |
| <i>Sarcococca confusa</i> | AF197588 | + | +/- | - | + | + | Qiu et al., 1999 |
| <i>Styloceras laurifolium</i> | AF093733 | + | +/- | - | + | + | Hoot et al., 1999 |
| Ceratophyllaceae | | | | | | | |
| <i>Ceratophyllum demersum</i> | DM77030 | + | + | + | + | + | Les et al., 1991 |
| <i>Ceratophyllum submersum</i> | AF197599 | + | + | + | + | + | Qiu et al., 1999 |
| Chloranthaceae | | | | | | | |
| <i>Chloranthus angustifolius</i> | AY236839 | + | + | + | + | + | Zhang & Renner, 2003 |
| Circaeasteraceae | | | | | | | |
| <i>Circaeaster agrestis</i> | AF093720 | + | + | + | + | + | Hoot et al., 1999 |
| Didymelaceae | | | | | | | |
| <i>Didymeles perrieri</i> | AF061994 | + | +/- | + | + | + | Qiu et al., 1998 |
| Eupteleaceae | | | | | | | |
| <i>Euptelea polyandra</i> | L12645 | + | + | + | + | + | Qiu et al., 1993 |
| Gunneraceae | | | | | | | |
| <i>Gunnera lobata</i> | AF307919 | + | + | + | + | + | Wardle et al., 2001 |
| <i>Gunnera monoica</i> | AF307918 | + | + | + | + | + | Wardle et al., 2001 |
| Kingdoniaceae | | | | | | | |
| <i>Kingdonia uniflora</i> | AF093719 | + | + | + | + | + | Hoot et al., 1999 |
| Lardizabalaceae | | | | | | | |
| <i>Decaisnea fargesii</i> | L37916 | + | + | + | + | + | Hoot et al., 1995 |
| <i>Lardizabala biternata</i> | L37919 | + | + | + | + | + | Hoot et al., 1995 |
| <i>Sargentodoxa cuneata</i> | AF093731 | + | + | + | + | + | Hoot et al., 1999 |
| <i>Sinofranchetia chinensis</i> | L37921 | + | + | + | + | + | Hoot et al., 1995 |
| Menispermaceae | | | | | | | |
| <i>Cissampelos pareira</i> | AF197590 | + | + | + | + | + | Qiu et al., 1999 |
| <i>Cocculus trilobus</i> | L12642 | + | + | + | + | + | Qiu et al., 1993 |
| <i>Menispermum canadense</i> | AF093726 | + | + | + | + | + | Hoot et al., 1999 |
| <i>Tinospora caffra</i> | L37923 | + | + | + | + | + | Hoot et al., 1995 |
| Myrothamnaceae | | | | | | | |
| <i>Myrothamnus flabellifolia</i> | AF060707 | + | +/- | - | + | + | Qiu et al., 1998 |

Table 1 Continued

| Species | Genbank accession number | Analyses that each accession was present (+) or absent (-) for: | | | | | Reference for sequence |
|--------------------------------------|--------------------------|---|------|-----|---|---|-------------------------------------|
| | | A,G,H,I | B, J | D,L | E | F | |
| Nelumbonaceae | | | | | | | |
| <i>Nelumbo lutea</i> | AF543715 | + | + | + | + | + | Hilu et al., unpubl. data, 2002 |
| <i>Nelumbo nucifera</i> | M77033 | + | + | + | + | + | Les et al., 1991 |
| Papaveraceae | | | | | | | |
| <i>Corydalis nobilis</i> | AF093722 | + | + | + | + | + | Hoot et al., 1999 |
| <i>Dicentra eximia</i> | L37917 | + | + | + | + | + | Hoot et al., 1995 |
| <i>Papaver orientale</i> | L08764 | + | + | + | + | + | Chase et al., 1993 |
| <i>Sanguinaria canadensis</i> | L01951 | + | + | + | + | + | Albert et al., 1992 |
| Platanaceae | | | | | | | |
| <i>Platanus orientalis</i> | AY858644 | + | + | + | + | + | Han et al., unpubl. data, 2004 |
| Proteaceae | | | | | | | |
| <i>Bellendena montana</i> | U79177 | + | + | + | + | + | Plunkett et al., unpubl. data, 1996 |
| <i>Carnarvonia araliifolia</i> | U79164 | + | + | + | + | + | Plunkett et al., unpubl. data, 1996 |
| <i>Dryandra drummondii</i> | U79167 | + | + | + | + | + | Plunkett et al., unpubl. data, 1996 |
| <i>Gevuina avellana</i> | U79169 | + | + | + | + | + | Plunkett et al., unpubl. data, 1996 |
| <i>Isopogon latifolius</i> | U79179 | + | + | + | + | + | Plunkett et al., unpubl. data, 1996 |
| <i>Lomatia silaifolia</i> | U79171 | + | + | + | + | + | Plunkett et al., unpubl. data, 1996 |
| <i>Macadamia ternifolia</i> | U79172 | + | + | + | + | + | Plunkett et al., unpubl. data, 1996 |
| <i>Persoonia lanceolata</i> | U79178 | + | + | + | + | + | Plunkett et al., unpubl. data, 1996 |
| <i>Petrophile biloba</i> | U79181 | + | + | + | + | + | Plunkett et al., unpubl. data, 1996 |
| <i>Placospermum coriaceum</i> | AF093729 | + | + | + | + | + | Hoot et al., 1999 |
| <i>Protea repens</i> | U79182 | + | + | + | + | + | Plunkett et al., unpubl. data, 1996 |
| <i>Roupala macrophylla</i> | AF093728 | + | + | + | + | + | Hoot et al., 1999 |
| <i>Teloepa speciosissima</i> | U79175 | + | + | + | + | + | Plunkett et al., unpubl. data, 1996 |
| <i>Adenanthos obovatus</i> | U79165 | - | - | - | + | + | Plunkett et al., unpubl. data, 1996 |
| <i>Brabejum stellatifolium</i> | U79166 | - | - | - | + | + | Plunkett et al., unpubl. data, 1996 |
| <i>Embothrium coccineum</i> | U79168 | - | - | - | + | + | Plunkett et al., unpubl. data, 1996 |
| <i>Grevillea robusta</i> 1 | AF197589 | - | - | - | + | + | Qiu et al., 1999 |
| <i>Grevillea robusta</i> 2 | AF193973 | - | - | - | + | + | Parkinson et al., 1999 |
| <i>Hakea myrtooides</i> | U79170 | - | - | - | + | + | Plunkett et al., unpubl. data, 1996 |
| <i>Lambertia inermis</i> | L11190 | - | - | - | + | + | Morgan & Soltis, 1993 |
| <i>Leucadendron lauroolum</i> | U79180 | - | - | - | + | + | Plunkett et al., unpubl. data, 1996 |
| <i>Orothamnus zeyheri</i> | U37132 | - | - | - | + | + | Barker et al., unpubl. data, 1995 |
| <i>Roupala macrophylla</i> | AF093728 | - | - | - | + | + | Hoot et al., 1999 |
| <i>Stenocarpus sinuatus</i> | U79174 | - | - | - | + | + | Plunkett et al., unpubl. data, 1996 |
| <i>Xylomelum pyriforme</i> | U79176 | - | - | - | + | + | Plunkett et al., unpubl. data, 1996 |
| Ranunculaceae | | | | | | | |
| <i>Aquilegia canadensis</i> | AY392755 | + | + | + | + | + | Duvall et al., 2006 |
| <i>Clematis</i> sp. | AF193972 | + | + | + | + | + | Parkinson et al., 1999 |
| <i>Coptis trifolia</i> | AF093730 | + | + | + | + | + | Hoot et al., 1999 |
| <i>Glaucidium palmatum</i> | AF093723 | + | + | + | + | + | Hoot et al., 1999 |
| <i>Hydrastis canadensis</i> | AF093725 | + | + | + | + | + | Hoot et al., 1999 |
| <i>Ranunculus macranthus</i> | DQ069502 | + | + | + | + | + | Leebens-Mack et al., 2005 |
| <i>Xanthorhiza simplicissima</i> | L12669 | + | + | + | + | + | Qiu et al., 1993 |
| <i>Anemone virginiana</i> | DQ006121 | - | - | - | - | + | Kress et al., 2005 |
| <i>Asteropyrum cavaleriei</i> | AF079453 | - | - | - | - | + | Wang et al., 2005 |
| <i>Beesia calthifolia</i> | AF079452 | - | - | - | - | + | Wang et al., 2005 |
| <i>Caltha appendiculata</i> | AF307908 | - | - | - | - | + | Wardle et al., 2001 |
| <i>Ficaria verna</i> | AF386100 | - | - | - | - | + | Ehrendorfer & Samuel, 2001 |
| <i>Hepatica nobilis</i> | AF386099 | - | - | - | - | + | Ehrendorfer & Samuel, 2001 |
| <i>Hepatica transsilvanica</i> | AF386089 | - | - | - | - | + | Ehrendorfer & Samuel, 2001 |
| <i>Psychrophila novae-zealandiae</i> | AF307907 | - | - | - | - | + | Wardle et al., 2001 |
| <i>Pulsatilla grandis</i> | AF386094 | - | - | - | - | + | Ehrendorfer & Samuel, 2001 |
| <i>Thalictrum dioicum</i> | DQ006122 | - | - | - | - | + | Kress et al., 2005 |
| Sabiaceae | | | | | | | |
| <i>Meliosma simplicifolia</i> | AF197587 | + | +/- | - | + | + | Qiu et al., 1999 |
| <i>Meliosma veitchiorum</i> | AF206793 | + | +/- | - | + | + | Soltis et al., 1999 |
| <i>Sabia</i> sp. | L12662 | + | +/- | - | + | + | Qiu et al., 1993 |

Table 1 Continued

| Species | Genbank accession number | Analyses that each accession was present (+) or absent (-) for: | | | | | Reference for sequence |
|---------------------------------|--------------------------|---|------|-----|---|---|------------------------|
| | | A,G,H,I | B, J | D,L | E | F | |
| Tetracentraceae | | | | | | | |
| <i>Tetracentron sinense</i> | L12668 | + | +/- | - | + | + | Qiu et al., 1993 |
| Trochodendraceae | | | | | | | |
| <i>Trochodendron aralioides</i> | L01958 | + | +/- | - | + | + | Albert et al., 1992 |

Analyses C and K combined data from analyses A and B, and I and J, respectively. +/-, accession was removed in some replicates for this analysis. References for sequences are as follows: **Albert** VA, Williams SE, Chase MW. 1992. Carnivorous plants: phylogeny and structural evolution. *Science* 257: 1491–1495. **Chase** MW, Soltis DE, Olmstead RG et al. 1993. Phylogenetics of seed plants: an analysis of nucleotide sequences from the plastid gene *rbcL*. *Annals of the Missouri Botanic Garden* 80: 528–580. **Duval** MR, Mathews S, Mohammad NA, Russell TJM. 2006. Placing the monocots: conflicting signal from trigenomic analyses. *Aliso* 22: 79–90. **Ehrendorfer** F, Samuel R. 2001. Contribution to a molecular phylogeny and systematics of *Anemone* and related genera (Ranunculaceae-Anemoninae). *Acta Phytotaxonomica Sinica* 39: 293–308. **Hoot** SB, Culham A, Crane PR. 1995. The utility of *atpB* gene sequences in resolving phylogenetic relationships: Comparison with *rbcL* and 18S ribosomal DNA sequences in the Lardizabalaceae. *Annals of the Missouri Botanic Garden* 82: 194–207. **Hoot** SB, Magallon S, Crane PR. 1999. Phylogeny of basal eudicots based on three molecular data sets: *atpB*, *rbcL*, and 18S nuclear ribosomal DNA sequences. *Annals of the Missouri Botanical Garden* 86: 1–32. **Kim** Y, Jansen RK. 1996. Phylogenetic implications of *rbcL* and ITS sequence variation in the Berberidaceae. *Systematic Botany* 21: 381–396. **Kofuji** R, Ueda K, Yamaguchi K et al. 1994. Molecular phylogeny in the Lardizabalaceae. *Journal of Plant Research* 107: 339–348. **Kress** WJ, Wurdack KJ, Zimmer EA et al. 2005. Use of DNA barcodes to identify flowering plants. *Proceedings of the National Academy of Sciences USA* 102: 8369–8374. **Leebens-Mack** J, Raubeson LA, Cui L et al. 2005. Identifying the basal angiosperm node in chloroplast genome phylogenies: Sampling one's way out of the Felsenstein zone. *Molecular Biology and Evolution* 22: 1948–1963. **Les** DH, Garvin DK, Wimpee CF. 1991. Molecular evolutionary history of ancient aquatic angiosperms. *Proceedings of the National Academy of Sciences USA* 88: 10119–10123. **Morgan** DR, Soltis DE. 1993. Phylogenetic relationships among members of Saxifragaceae sensu lato based on *rbcL* sequence data. *Annals of the Missouri Botanic Garden* 80: 631–660. **Parkinson** CL, Adams KL, Palmer JD. 1999. Multigene analyses identify the three earliest lineages of extant flowering plants. *Current Biology* 9: 1485–1488. **Qiu** Y-L, Chase MW, Hoot SB et al. 1998. Phylogenetics of the Hamamelidae and their allies: parsimony analyses of nucleotide sequences of the plastid gene *rbcL*. *International Journal of Plant Sciences* 159: 891–905. **Qiu** YL, Chase MW, Les DH et al. 1993. Molecular phylogenetics of the Magnoliidae: cladistic analysis of nucleotide sequences of the plastid gene *rbcL*. *Annals of the Missouri Botanic Garden* 80: 587–606. **Qiu** YL, Lee J, Bernasconi-Quadroni F et al. 1999. The earliest angiosperms: evidence from mitochondrial, plastid and nuclear genomes. *Nature* 402: 404–407. **Soltis** PS, Soltis DE, Chase MW. 1999. Angiosperm phylogeny inferred from multiple genes as a tool for comparative biology. *Nature* 402: 402–404. **Wang** W, Li RQ, Chen ZD. 2005. Systematic position of *Asteropyrum* (Ranunculaceae) inferred from chloroplast and nuclear sequences. *Plant Systematics and Evolution* 255: 41–54. **Wardle** P, Ezcurra C, Ramirez C et al. 2001. Floristic comparison of Southern Andean and New Zealand plant communities. *New Zealand Journal of Botany* 39: 69–108. **Xiang** QY, Soltis DE, Soltis PS, Manchester SR, Crawford DJ. 2000. Timing the eastern Asian–eastern North American floristic disjunction: molecular clock corroborates paleontological estimates. *Molecular Phylogenetics and Evolution* 15: 462–472. **Zhang** LB, Renner SS. 2003. The deepest splits in Chloranthaceae as resolved by chloroplast sequences. *International Journal of Plant Sciences* 164(S5): S383–S392.

point would be as a maximum age for divergence between eudicots and *Ceratophyllum* and a minimum age for crown radiation within eudicots, as tricolpate pollen must have evolved after the first event and before the second. However, as r8s requires a fixed calibration point, the age of 124 mya must be fixed to one of these nodes. If a minimum age were to be assigned, using the older point would be the correct approach because it is more conservative, but for fixed ages neither point is inherently more conservative. In this study, however, there are several other fossils that constrain minimum ages in the phylogeny (see below) but none that constrain maximum ages; therefore the eudicot calibration point ought to have a much stronger effect as a maximum than a minimum age constraint. If the maximum age 124 mya is applied to the crown node there is no constraint to the age of the stem node, whereas the reverse is not true. Hence it should be more conservative to fix the age of 124 mya to the later point of crown eudicot radiation. This also allows for tricolpate pollen to have arisen a little earlier than 124 mya, as indicated by the morphological diversity (Friis et al., 2006) and geographical range (Hughes & McDougall, 1990; Doyle, 1992) of

such pollen at this time. A further argument for this is that the actual sequence of events was: (1) *Ceratophyllum* diverged from eudicots; (2) tricolpate pollen first evolved; and (3) crown radiation of eudicots. The first fossilization of tricolpate pollen certainly occurred after event 2 (fossilization), and could have been before or after event 3 (radiation of crown eudicots). For this reason, on balance of probability the age of 124 mya is likely closer to event (3), that is, crown radiation of eudicots.

1.5 Fossil calibration: minimum age points

For fossil calibration, the minimum node ages used by Anderson et al. (2005; q.v. for details of fossils) were employed in all analyses, that is, divergence of Menispermaceae from Ranunculaceae/ Berberidaceae; ≥ 91 mya); divergence of Platanaceae from Proteaceae (≥ 110 mya); divergence of Didymelaceae from Buxaceae (≥ 95 mya); divergence of Buxales from branch leading to core eudicots (≥ 110 mya). A calibration point of ≥ 98 mya for the divergence of Sabiaceae was not used because this node was always older due to older dates enforced on nodes directly descended from this

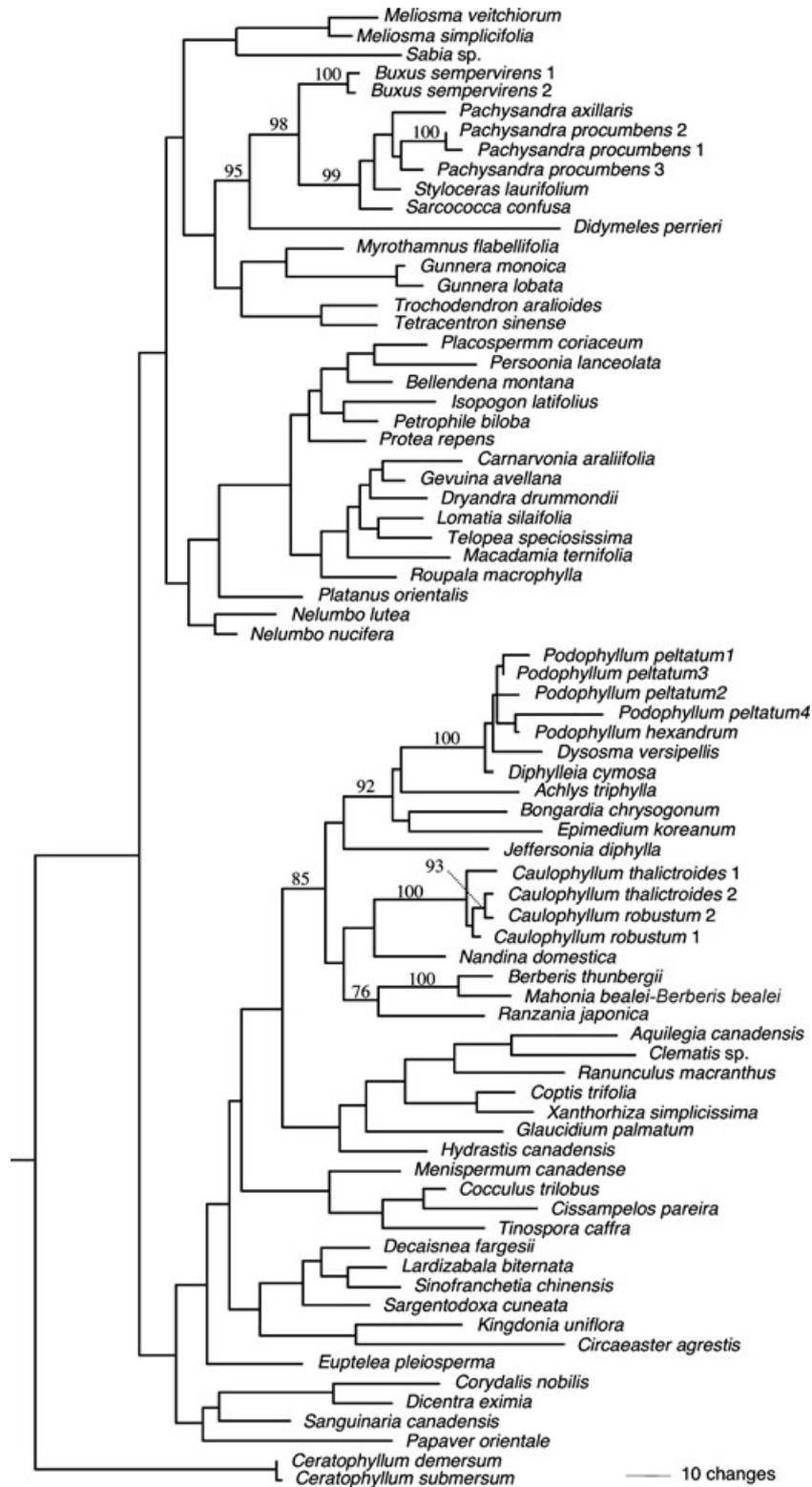


Fig. 1. Phylogram for taxa included in analysis A, comprising the original dataset with bootstrapping. Bootstrap support values over 70% shown for nodes within Buxales and Berberidaceae only.

one. A minimum age of 112 mya for divergence of Ranunculales from all other eudicots (i.e. crown radiation of eudicots) was included only when the fixed 124 mya calibration point was moved from crown to stem node (analysis H, see below). The ≥ 95 mya calibration point could not be used when *Didymeles* Thou. was not present in the dataset, as this genus was the only representative of Didymelaceae available. Hence this fossil was excluded in analyses that determined the effects of deleting individual taxa, (analyses B, C, J and K; see below). However, the effect of including or excluding this fossil was compared using the complete dataset and found to be negligible, except on *Pachysandra* in analyses where sampling of related taxa was substantially reduced (analyses D and L; see below). From this, the effect of removing or including the ≥ 95 mya calibration point on analyses B, C, J and K can be assumed to be likewise very small.

1.6 Nodes to be dated

Within each of the three target genera a single node, henceforth referred to as the “target node” was identified whose age would be calculated in each analysis. In *Pachysandra*, the basal node within the genus also marked divergence of the Asian *Pach. axillaris* from the American *Pach. procumbens*; however, this was not the case in *Caulophyllum* or *Podophyllum*. In *Podophyllum*, the American *Pod. peltatum* was apparently non-monophyletic, whereas in *Caulophyllum* apparently neither species was monophyletic, at least for *rbcL* haplotypes. This is probably due to Genbank sequences having the wrong species name on one accession from each genus. This is plausible because of the two examples of Genbank accessions identified as the wrong family noted above, and the relative similarity of the two species in each genus. Otherwise, unlikely scenarios of haplotype or lineage sorting leaving signatures ~ 5 million years later have to be invoked. For this study, therefore, the basal node in each genus is used as the target node, with the caveat that for *Podophyllum* and *Caulophyllum* there is a small chance that this node in fact precedes speciation.

1.7 Sources of node age variation

1.7.1 Analysis A: original dataset with bootstrapping Initial age estimates were obtained for the three target nodes based on the original dataset and constraint tree. A confidence interval for the age of each node was generated by bootstrapping. Following a protocol supplied by Dr. V. Savolainen (Royal Botanic Garden, Kew, UK; pers. comm., 2005) 100 bootstrap replicate trees were generated using PAUP*, allowing unlimited maxtrees but with tree topology constrained as above.

Hence, branch lengths could vary between replicates but topology did not. The 100 replicate trees obtained were exported from PAUP in NeXus format and each was copied individually into an input file for r8s (Sanderson, 2002), which was used to calculate node ages for each replicate using PL (analyses A, C–G) or NPRS (analyses I, K and L). From these 100 replicate dates, mean, standard deviation (SD) and 95% confidence ranges ($\text{mean} \pm 2.26 * \text{SD}$) were generated for the age of each target node. This method is not perfect for estimating confidence ranges (Prof. Jeffrey L. Thorne, North Carolina State University, Raleigh, NC, USA; pers. comm., 2008), but is used here because no better method for putting ranges on nodes ages calculated using PL or NPRS is known to the author, and because it is directly comparable with the methods used for examining taxon sampling variation. Bootstrapping was used in all analyses except B and J.

1.7.2 Analysis B: 112 variant trees generated by taxon sampling variation This analysis was designed to test the extent to which target node ages might be altered by the removal (or by inference, the addition) of up to three taxon samples from the dataset. This effect is henceforth referred to as “taxon sampling variation” to distinguish it from “taxon sampling density”, that is, the effects of adding or removing larger sets of taxa from specific clades, which is tested for in other analyses (see below).

For each target genus, 10 taxon samples were selected that occupied close phylogenetic positions. For *Caulophyllum* and *Podophyllum* this set was the same, that is, the single representatives of the 10 other genera of Berberidaceae that were included in the original dataset: *Achlys* (L75866), *Bongardia* (L75870), *Diphylleia* (L75866), *Dysosma* (AF079454), *Epimedium* (L75869), *Jeffersonia* (L75867), *Berberis* (AF139878), *Mahonia* (L12657), *Ranzania* (L75853), and *Nandina* (L37920). These were collectively referred to as sampling set B (Fig. 2). For *Pachysandra*, only four taxon samples were available within Buxaceae, so the remainder were chosen from adjacent clades. The set used, referred to as sampling set A, therefore comprised the following: *Buxus sempervirens* (AF093717), *Buxus sempervirens*, (AF543712), *Styloceras* (AF093733), *Sarcococca* (AF197588), *Sabia* (L12662), *Meliosma simplicifolia* (AF197587), *Meliosma veitchiorum* (AF206793), *Trochodendron* (L01958), *Tetracentron* (L12668), and *Myrothamnus* (AF060707) (Fig. 2).

To quantify the effects of taxon sampling variation, 10 variant datasets were generated each with a single taxon from sampling set B deleted, and then a further set of 45 datasets were generated with two taxa from

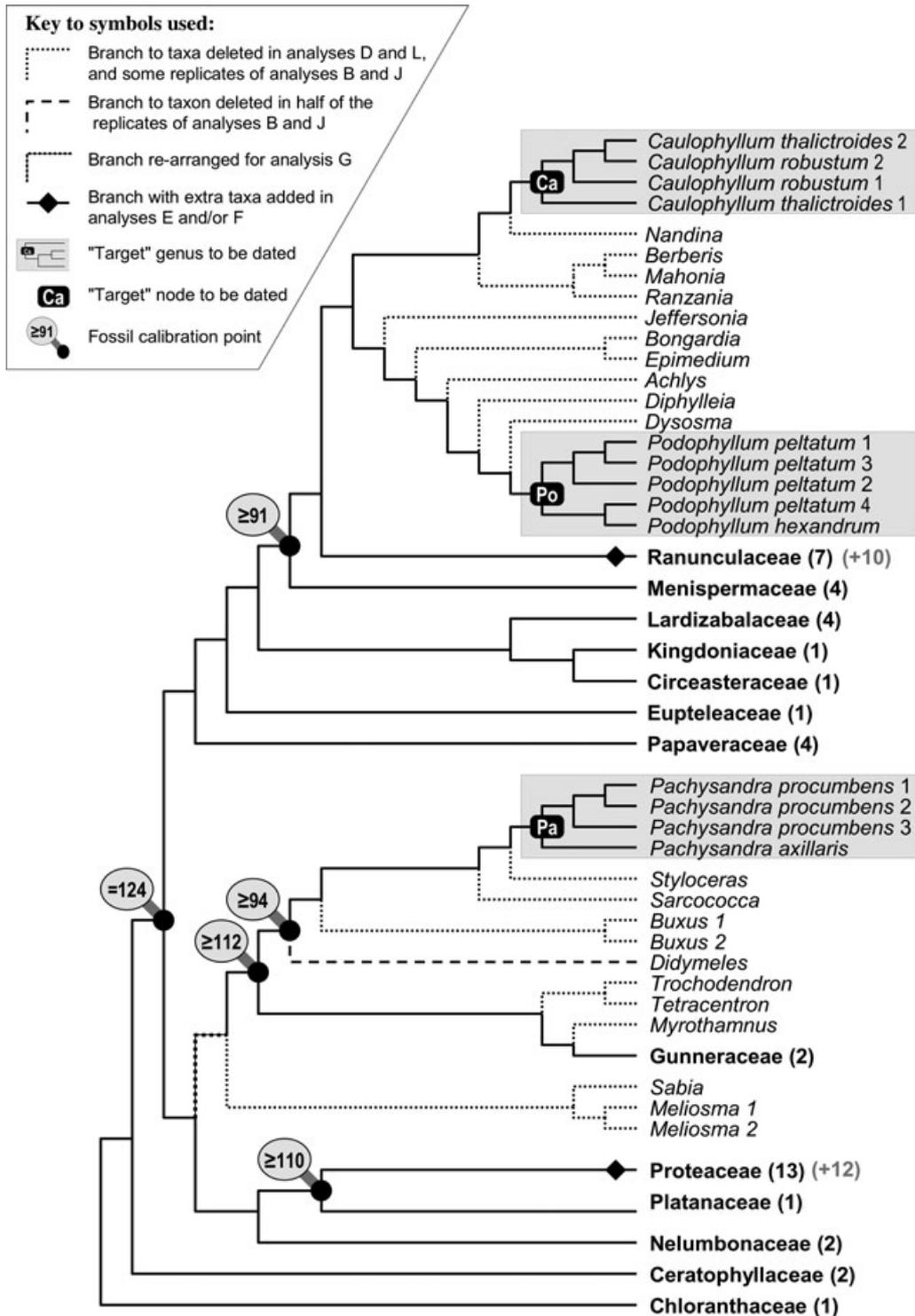


Fig. 2. Constraint tree used in all analyses, indicating calibration points, nodes to be dated ("target nodes"), taxa that were deleted for some analyses, and other alterations that were made.

sampling set B deleted. Including the original dataset with no taxa deleted, this made a total of 56 variant datasets. For *Caulophyllum*, a further 56 datasets were then created by removing all accessions of *Podophyllum* and also either no accessions, one accession or two accessions from sampling set B. Likewise, for *Podophyllum*, a further 56 datasets were then created by removing all accessions of *Caulophyllum* and also either no accessions, one accession or two accessions from sampling set B.

In the case of *Pachysandra*, 55 variant datasets in addition to the original one were generated by removing either one accession or two from sampling set A, and a further 56 were generated by doing the same but also removing *Didymeles* (AF061994) as well. Hence, for each of *Caulophyllum*, *Podophyllum* and *Pachysandra*, SD and confidence ranges for the initial divergence event within the genus were thus calculated from the 112 age estimates produced by deleting between 0 and 3 taxa from the dataset. In this way, an estimate of error due solely to minor variations in taxon sampling was calculated.

1.7.3 Analysis C: combined effects of bootstrapping and taxon sampling variation The age variation detected by bootstrapping the original dataset (analysis A), was combined with that due to effects of taxon sampling variation (analysis B) using the following procedure. For each target node, each of the 112 replicate ages generated by taxon sampling variation was converted into a proportion of the mean age for all replicates. For each node, each of these 112 proportions was multiplied by each of the 100 bootstrap replicate node ages from analysis A, producing a dataset of 11200 age estimates. This provided an approximation of the dataset that would have been obtained had each of the 112 variant trees from analysis B been bootstrapped, but took a hundredth of the time. A new mean, SD and confidence range was then calculated for each node from these 11200 age estimates.

1.7.4 Analysis D: taxon sampling density I. Effects of deleting 10 taxon samples from same clade as target node In this analysis 10 of the most closely related taxa to each target node were deleted, and the node age was then calculated as in analysis A. For *Pachysandra*, the 10 taxon samples that comprised sampling set A were deleted. Likewise, for *Caulophyllum* and *Podophyllum*, the 10 taxa that comprised sampling set B were deleted. Target nodes were then dated as in analysis A.

1.7.5 Analysis E: taxon sampling density II. Effects of increasing taxon sampling in a clade not close to target nodes The original analysis (A) included 13 representatives of Proteaceae; for analysis E this was

increased to 25. The constraint tree was altered to fix the positions of all these taxa to the topology in the majority rule consensus of most parsimonious trees. Otherwise dates were calculated as in analysis A.

1.7.6 Analysis F: taxon sampling density III. Effects of increasing taxon sampling in a clade sister to that containing target nodes The original analysis (A) included seven representatives of Ranunculaceae, the sister group to Berberidaceae that contains *Podophyllum* and *Caulophyllum*. For analysis F this was increased to 17. The added taxa from Proteaceae (analysis E) were also retained. The constraint tree was altered once again to fix the positions of the new taxa to the topology in the majority rule consensus of most parsimonious trees, and dates were otherwise calculated as in analysis A.

1.7.7 Analysis G: switching positions of Sabiales and Proteales in constraint tree In addition to taxon sampling density, choice of tree topology might also have an effect on the ages calculated for target nodes. However, in large phylogenies there are often areas where the topology has not been agreed by researchers. Among basal eudicots, the relative positions of Sabiales and Proteales differ between published phylogenies. In the APG II (2003) phylogeny, Proteales branch off first, followed by Sabiales then Buxales, on the branch leading to core eudicots. This topology is followed in all analyses except this one (G). However, in many other studies the relative positions of Proteales and Sabiales are reversed, that is, the latter branches first (see figure and refs. in Anderson et al., 2005). In analysis G, therefore, the constraint tree was altered so that Sabiales branched off before Proteales; in other respects the dates were calculated as for analysis A.

1.7.8 Analysis H: fixing tricolpate pollen age (124 mya) to stem node instead of crown node for eudicots In all other analyses, the fixed age point of 124 mya is assigned to the crown divergence node for eudicots, for the reasons given above. However, it is valuable to know the extent to which this choice might affect the results, especially as Anderson et al. (2005) preferred to fix this age to the stem node, that is, divergence of *Ceratophyllum* from eudicots. Hence, for analysis H, the age of 124 mya is fixed to the stem node, and the age constraint of ≥ 112 mya for the crown node was applied. Otherwise this analysis was the same as analysis A.

1.7.9 Analyses I, J, K and L: using NPRS instead of PL Using NPRS to smooth branch lengths instead of PL has been shown to increase sensitivity of node ages to a reduction in taxon sampling density (Linder et al., 2005). Here, the effects of using NPRS instead of PL are tested on a much deeper and more unevenly sampled

phylogeny than in Linder et al.'s (2005) work. Analyses I, J, K and L follow the same methods as analyses A, B, C and D, respectively, except that branch lengths are smoothed using NPRS instead of PL.

2 Results

2.1 Analysis A: original dataset

For *Caulophyllum*, the initial node age estimate (analysis A) was 9.8 mya, with a bootstrap confidence range of 3.1–19.0 mya. For *Podophyllum*, the initial node age estimate was 12.9 mya, with a confidence range of 6.8–32.3 mya. For *Pachysandra*, the initial node age estimate was 19.7 mya, with a confidence range of 11.2–31.9 mya (Table 2).

2.2 Analysis B: 112 variant trees generated by taxon sampling variation

The initial age estimates in *Podophyllum* and *Pachysandra* were sensitive to the effects of removing even a single taxon from the dataset. Removing *Meliosma simplicifolia* increased the age of the *Pachysandra* node from 19.7 to 23.5 mya (19%). Likewise removing *Diphylleia* increased the age of the *Podophyllum* node from 12.9 to 17.4 (35%). Removing both *Styloceras* A. Juss. and *Sarcococca* Lindl. (the two closest sister genera to *Pachysandra*; Fig. 1) increased the target node age from 19.7 to 26.4 (34%). For *Podophyllum*, removing both *Achlys* DC. and *Diphylleia* Michx. increased the age of the target node from 12.9 to 24.0, an increase of 85%, whereas removing *Achlys* and *Caulophyllum* reduced the *Podophyllum* node age to 10.2 (a 19% decrease). Conversely, the highest and lowest initial values for *Caulophyllum* when two taxa were deleted were 8.6 and 11.1, representing a 12% decrease and 13% increase from the initial value of 9.8 mya, respectively.

The SD values obtained from all 112 trees were always smaller than those derived by bootstrapping (analysis A), but varied greatly between genera. The SD for *Caulophyllum* was very small, at 0.47, leading to a narrow confidence range of 8.4–10.6 mya. However, the SD value for *Podophyllum*, 2.73, was large enough to generate a very broad confidence range of 7.6–20.0 mya, wider than that derived previously for divergence within this genus (Xiang et al., 2000). The range for *Pachysandra* was 17.6–25.0 mya (Table 2).

2.3 Analysis C: combined effects of bootstrapping and taxon sampling variation

Combining the effects of taxon sampling variation and bootstrapping led to only slight expansions of

the confidence ranges, relative to bootstrapping alone, in each case. The change for *Caulophyllum* was tiny (~0.1 mya in each direction), but for *Podophyllum* the upper age limit increased by 1.7, from 32.3 to 34.0, and the range for *Pachysandra* expanded by ~0.6 mya in each direction, relative to analysis A (Table 2).

2.4 Analysis D: effects of deleting 10 taxon samples close to target node

Reducing sampling of Berberidaceae from 12 genera to 2 had almost no effect on the initial estimate and its confidence range within *Caulophyllum*. (Table 2; Fig. 3). However, deleting the same 10 taxa expanded the confidence range for *Podophyllum*, increasing the upper limit by 2.6 from 32.3 to 34.9, although the initial date estimate was almost unchanged. In *Pachysandra*, the effect of removing 10 adjacent taxa was dramatic. The initial estimate increased by more than 10 my, from 19.7 to 31.3, the bootstrap mean by a similar amount, and the SD from 4.56 to 6.01. Because of this, the confidence range shifted dramatically back in time, from 11.2–31.9 mya before deletion to 19.5–46.7 mya afterwards. This and analysis L were the only analyses in which the removal of the minimum age constraint of 95 mya for divergence of *Didymeles* from Buxaceae had any effect, making the mean and confidence range limits for *Pachysandra* approximately 2.5 my older.

2.5 Analysis E: increased taxon sampling in Proteaceae

Increasing taxon sampling from 13 to 25 taxa in Proteaceae, a family not phylogenetically close to any of the target nodes, had negligible effects on dates and ranges in *Caulophyllum* and *Pachysandra*. In *Podophyllum*, however, it caused slight increases in bootstrap mean and SD, leading to an increase in upper confidence limit from 32.3 to 35.8.

2.6 Analysis F: increased taxon sampling in Ranunculaceae (and Proteaceae)

Increasing taxon sampling from 7 to 17 taxa in Ranunculaceae, the sister family to Berberidaceae, once again had a negligible effect on dates in *Caulophyllum*. Conversely, for *Podophyllum* this increased the initial estimate by 9.4 my, and its upper confidence limit by a similar amount (9.5 my), expanding the possible age range for this genus considerably (Table 2; Fig. 3). In *Pachysandra*, which is not closely related to Ranunculaceae, there was no effect on the initial estimate, but a slight increase (1.3 my) in its upper confidence limit relative to analysis E.

Table 2 Divergence times calculated for the basal ("target") node in each of three genera, in 12 different analyses

| <i>Caulophyllum</i> | | Initial value | Mean value | SD of mean | Confidence range (95%) |
|---------------------------|---|---------------|------------|------------|------------------------|
| Analysis name and summary | | | | | |
| A | Original dataset, bootstrapped | 9.8 | 11.1 | 3.51 | 3.1–19.0 |
| B | Taxon sampling variation | 9.8 | 9.5 | 0.47 | 8.4–10.6 |
| C | Combined variation from A & B | 9.8 | 11.0 | 3.55 | 3.0–19.1 |
| D | Ten taxa removed near target node | 9.8 | 10.9 | 3.47 | 3.1–18.8 |
| E | Taxon sampling increased (Proteaceae) | 9.8 | 10.8 | 3.15 | 3.6–17.9 |
| F | Taxon sampling increased (Ranunculaceae) | 10.8 | 11.3 | 3.42 | 3.5–19.0 |
| G | Topology altered (Sabiales and Proteales) | 10.1 | 10.9 | 3.15 | 3.7–18.0 |
| H | Eudicot fossil on stem not crown | 9.7 | 10.6 | 3.28 | 3.2–18.0 |
| I | As A, but using NPRS | 12.7 | 15.6 | 3.99 | 6.6–24.6 |
| J | As B, but using NPRS | 12.7 | 11.6 | 1.10 | 9.1–14.1 |
| K | As C, but using NPRS | 12.7 | 15.6 | 4.25 | 6.0–25.2 |
| L | As D, but using NPRS | 14.1 | 18.8 | 5.25 | 7.0–30.7 |
| <i>Podophyllum</i> | | Initial value | Mean value | SD of mean | Confidence range (95%) |
| Analysis name and summary | | | | | |
| A | Original dataset, bootstrapped | 12.9 | 16.2 | 7.12 | 0.1–32.3 |
| B | Taxon sampling variation | 12.9 | 13.8 | 2.73 | 7.6–20.0 |
| C | Combined variation from A & B | 12.9 | 16.2 | 7.89 | 0–34.0 |
| D | Ten taxa removed near target node | 12.9 | 17.0 | 7.91 | 0–34.9 |
| E | Taxon sampling increased (Proteaceae) | 12.9 | 17.5 | 8.09 | 0–35.8 |
| F | Taxon sampling increased (Ranunculaceae) | 22.4 | 21.3 | 9.06 | 0.8–41.7 |
| G | Topology altered (Sabiales and Proteales) | 13.7 | 18.4 | 7.96 | 0.4–36.3 |
| H | Eudicot fossil on stem not crown | 12.6 | 14.8 | 6.45 | 0.2–29.4 |
| I | As A, but using NPRS | 22.7 | 22.1 | 3.14 | 15.0–29.2 |
| J | As B, but using NPRS | 22.7 | 22.8 | 1.10 | 20.3–25.2 |
| K | As C, but using NPRS | 22.7 | 22.1 | 3.31 | 14.6–29.6 |
| L | As D, but using NPRS | 25.6 | 25.6 | 4.44 | 15.6–35.7 |
| <i>Pachysandra</i> | | Initial value | Mean value | SD of mean | Confidence range (95%) |
| Analysis name and summary | | | | | |
| A | Original dataset, bootstrapped | 19.7 | 21.5 | 4.56 | 11.2–31.9 |
| B | Taxon sampling variation | 19.7 | 21.3 | 1.63 | 17.6–25.0 |
| C | Combined variation from A & B | 19.7 | 21.4 | 4.79 | 10.5–32.2 |
| D | Ten taxa removed near target node | 31.3 | 33.1 | 6.01 | 19.5–46.7 |
| E | Taxon sampling increased (Proteaceae) | 19.7 | 21.3 | 4.42 | 11.3–31.3 |
| F | Taxon sampling increased (Ranunculaceae) | 19.7 | 22.2 | 4.71 | 11.6–32.9 |
| G | Topology altered (Sabiales and Proteales) | 19.7 | 21.5 | 4.07 | 12.3–30.7 |
| H | Eudicot fossil on stem not crown | 19.8 | 21.3 | 4.51 | 11.1–31.5 |
| I | As A, but using NPRS | 34.8 | 35.0 | 5.07 | 23.5–46.5 |
| J | As B, but using NPRS | 34.8 | 39.0 | 3.08 | 32.0–46.0 |
| K | As C, but using NPRS | 34.8 | 35.0 | 5.76 | 22.0–48.0 |
| L | As D, but using NPRS | 46.1 | 47.3 | 6.54 | 32.5–62.1 |

NPRS, non-parametric rate smoothing; SD, standard deviation.

2.7 Analysis G: switching positions of Sabiales and Proteales in constraint tree

As with other analyses, this switch of clade positions had little effect on *Caulophyllum*. The initial age estimate rose slightly from 9.8 to 10.1, but a reduction in SD from 3.52 to 3.15 caused a slight contraction of the confidence interval at both ends. In *Podophyllum* the initial age estimate rose by a small amount (0.8 my), whereas the upper confidence limit increased by 4 my, relative to analysis A. In *Pachysandra*, the only genus of the three directly descended from the nodes at which Proteales and Sabiales branch off, there was no change in the initial estimate, and a reduction in SD from 4.51 to

4.07, leading to a contraction of the confidence interval by ~1my at each end.

2.8 Analysis H: fixing tricolpate pollen age (124 mya) to stem node instead of crown node for eudicots

In *Caulophyllum*, this change (relative to analysis A) reduced initial age, mean age and SD by 2%, 4% and 6%, respectively, resulting in a contraction of the upper age limit by 1 my from 19 to 18 mya. In *Podophyllum* initial age, mean and SD reduced by 3%, 8% and 9%, respectively, and the upper age limit dropped by nearly 3 my (Table 2; Fig. 3). In *Pachysandra* all the differences

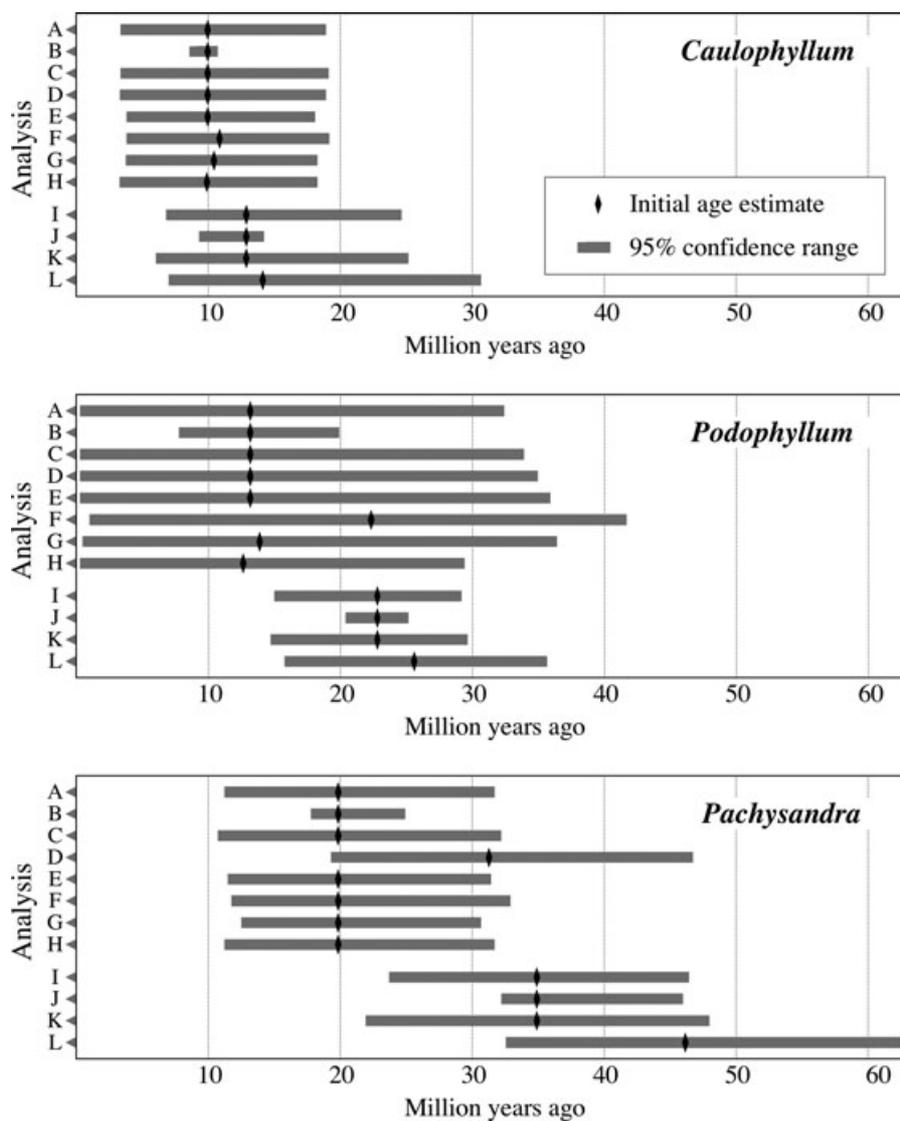


Fig. 3. Divergence times calculated by each of the 12 different analyses for the three target genera *Caulophyllum*, *Podophyllum* and *Pachysandra*. For each analysis the initial estimate (before confidence range calculation) and the 95% confidence range calculated by bootstrapping (analyses A, D–I and L), varying taxon sampling (analyses B and J) or combining the two (analyses C and K).

were negligible (<0.5%). Hence the confidence ranges in each case were either contracted or unaffected by switching the fixed calibration point from crown to stem node for eudicots.

2.9 Analysis I: as analysis A but using NPRS in place of PL

Using NPRS instead of PL caused an increase in all initial ages and bootstrap means. In *Caulophyllum* the initial estimate and confidence limits shifted backwards in time by between 2.5 and 5 my, and the width of the confidence interval increased slightly. In *Podophyllum* the initial age estimate increased dramatically by almost

10 my, from 12.9 to 22.7. However, the SD was more than halved, with the result that the confidence interval contracted dramatically relative to that of analysis A or any bootstrapped PL analysis. For *Pachysandra*, the initial age estimate increased by just over 15 my, and the confidence range widened slightly, from 11.2–31.9 to 23.5–46.5 (Table 2; Fig. 3).

2.10 Analyses J and K: as analyses B and C but using NPRS in place of PL

The confidence ranges generated by taxon sampling variation were narrower than those generated by bootstrapping, but this trend was less pronounced under

NPRS than under PL. In particular, the upper limit for *Pachysandra* (46.0) was only 0.5 mya younger than that generated by bootstrapping (46.5). In fact, if the taxon samples *Styloceras*, *Sarcococca* and *Didymeles* were removed from the dataset the age of the target node (48.3) fell outside the bootstrap confidence range.

When bootstrap and taxon sampling variation were combined, the expansions at either end of the confidence ranges for *Caulophyllum* and *Podophyllum* were small (<0.6 in each case) relative to analysis I, however the range for *Pachysandra* expanded by 1.5 my at each end.

2.11 Analysis L: as analysis D but using NPRS in place of PL

Deleting 10 related taxa increased the initial estimate for *Caulophyllum* by 1.4 my, and the upper confidence limit by 5.1 my to 30.7. This strongly contrasts to when PL was used (analysis D), when these deletions had almost no effect. In *Podophyllum* the upper age limit increased by a similar amount, and the initial estimate by 2.9. In *Pachysandra* the effect was greater still, with the initial estimate increased by 11.3 to 46.1 mya, the lower age limit by 10.5, and the upper by 15.6 my (Table 2; Fig. 3).

3 Discussion

Taxon sampling variation, the removal of up to 3 out of 11 accessions close to the target node in different combinations, created an age range 7.4 my wide in *Pachysandra* and 12.4 my wide in *Podophyllum*, but only 2.2 my wide in *Caulophyllum*. Crucially, however, the ranges due to taxon sampling variation were always much narrower than those derived by bootstrapping branch lengths (Fig. 3). As in other studies, NPRS gave older node ages than PL, and was more sensitive to taxon undersampling. Otherwise, there was a clear increase in the *Pachysandra* node age when taxon sampling was reduced by 10, but a similar reduction in taxon sampling density did not affect node ages in *Caulophyllum* or *Podophyllum*, and the other changes to the analysis generally only had small effects.

The initial analysis (A) calculated ages for target nodes within three genera, and generated confidence ranges by bootstrapping branch lengths, using PL for rate smoothing. The confidence ranges generated by bootstrapping alone were very wide, presumably a consequence of the distance between the target nodes and the closest calibration points. The initial age estimates for the two Berberidaceous genera, *Caulophyllum* and *Podophyllum*, were approximately 3 mya apart, 9.8 and 12.9, respectively, but the SD for *Podophyllum* was

twice that of *Caulophyllum*, so that the confidence range for the age of the former stretched from 0 to 32.3 mya. In contrast, the range for *Caulophyllum* was much narrower, from 3 to 19 mya, which was still much broader than a previous estimate (Xiang et al., 2000). Divergence in *Pachysandra* occurred 11.1–31.5 mya according to this analysis, and appears to be older than divergence in at least *Caulophyllum*, a result also indicated by Xiang et al. (2000).

3.1 Effects of taxon sampling variation

The confidence ranges generated by deleting between one and three taxa in various combinations (analysis B) were always much narrower than those generated by bootstrapping branch lengths, but nonetheless the confidence ranges for *Podophyllum* and *Pachysandra* spanned ~12.5 and 7.5 my, respectively, with upper and lower limits that could lead to very different biogeographic conclusions. Even the exclusion of a single taxon in a close phylogenetic position to the target node was found to alter its age by up to 4.5 mya or 35%, though in most cases the effect was smaller. For *Pachysandra*, deletion of the two *Buxus* accessions increases divergence time within the genus from 19.7 to 26.4 and the initial estimate of 12.9 mya for *Podophyllum* is almost doubled to 24.0 by removing *Achlys* and *Diphylleia*.

By contrast, the range due to taxon sampling variation for *Caulophyllum* was very small, indicating that dates within this genus are not sensitive to small scale taxon sampling effects. All dates in the range 8.4–10.6 are most compatible with vicariance due to climate cooling in Beringia (Milne & Abbott, 2002) so in this genus, unlike the others, choice of taxa sampled alone could not lead to different biogeographic conclusions. Overall, these results show that if bootstrapping is not carried out, the effects of small-scale taxon sampling variation could be significant, and the decision to include or exclude a few taxa outside the target group could possibly alter the conclusions reached. Depending on the taxon deleted, the age of a target node can go up or down, and it follows that the effects of adding a new taxon within Buxales or Berberidaceae might be similar to those of deleting a taxon.

The variation due to taxon sampling variation was, however, much smaller than that detected by bootstrapping branch lengths, and when the two effects were combined, the resultant range was only slightly greater than that revealed by bootstrapping alone. Relative to bootstrapping alone, the combined analysis increased the respective SDs of the target nodes in *Caulophyllum*, *Podophyllum* and *Pachysandra* by 1%, 11% and 6%, respectively, with the only age range limit that changed

by more than 1 my being the upper limit of *Podophyllum*, from 32.3 to 34.0. Therefore, even where taxon sampling variation alone caused significant variation in the initial node estimate, its effect when bootstrapping branch lengths is taken into consideration was sufficiently small that it can safely be ignored. Based on this, the use of bootstrapping can be tentatively assumed to control for small-scale taxon sampling effects close to the node to be dated. However, in cases where bootstrap confidence intervals are narrower, for example, because multiple genes have been used, the relative contribution of taxon sampling variation to node age variation might become more significant.

3.2 Effects of taxon sampling density in clades containing target nodes

In the original analysis (A), 12 genera of Berberidaceae were represented, *Caulophyllum*, *Podophyllum* and 10 others. Remarkably, removing all of the latter 10 genera from the dataset (analysis D) had almost no effect on the initial age estimate for *Caulophyllum* or its confidence range, with no value changing by more than 0.2 mya. In the case of *Podophyllum* the initial estimate was once again barely affected by the reduction in sampling; however, all of the mean, SD and upper confidence limits increased slightly, although this effect was smaller than that of altering sampling density (analyses E and F), tree topology (analysis G), the position of the fixed calibration point (analysis H) or sampling within *Podophyllum* (analysis I) (Table 2; Fig. 3). Overall, therefore, the effect of drastically reducing sampling density within Berberidaceae on nodes within *Caulophyllum* and *Podophyllum* were remarkably small.

In contrast, the effect on dates in *Pachysandra* of reducing sampling density by removing 10 accessions in phylogenetically close positions was large: the initial age estimate for the target node and its upper and lower confidence limits were increased by 11.6, 8.3 and 14.8 my, respectively (Table 2; Fig. 3). Hence the clade containing *Pachysandra* appears much more sensitive to taxon undersampling effects than that containing Buxaceae. This could be due to much lower taxon density in this clade before the 10 taxa were deleted, relative to that containing the Berberidaceae – the 10 closest taxa to *Pachysandra* included five that were outside the order Buxales (Fig. 2). When the 10 local taxa are deleted, *Pachysandra* is separated from the basal eudicot node by only three nodes, whereas *Caulophyllum* and *Podophyllum* are each separated by six (Fig. 2). When all taxa are present, the number of nodes separating each target node from the basal eudicot node is seven for *Pachysandra*, eight for *Caulophyllum* and 11 for *Podophyllum* (Fig. 2). Hence the drop in number of nodes is propor-

tionally greatest in *Pachysandra*, which could also account for the greater effect observed. Overall, this result indicates that undersampling effects could be more severe in clades that are already relatively poor in species. Hence it is not clear whether the date calculated for *Pachysandra* with the complete taxon set may still be affected by undersampling, or whether indeed there are enough extant lineages in or near the Buxales to counter this effect.

In Berberidaceae, this study indicates that sampling individuals from 12 genera within the family may be ample for a reasonably stable node age estimate. This is in direct contrast to results from the Restio clade (Linder et al., 2005), wherein taxon sampling effects began to appear when the total number of taxa sampled dropped below 150, which is still more than in the entire dataset of the current study. From this, sensitivity to taxon sampling certainly varies between clades, but more work would be needed to determine what causes such variation.

3.3 Effects of taxon sampling density in clades not containing target nodes

When taxon sampling in Ranunculaceae was increased from 7 to 17, the initial estimate for *Podophyllum* rose by 10 mya, the bootstrap mean by 5 mya and its SD by nearly 2 my, which extended the upper confidence limit by 9.5 my, from 32.3 to 41.7. However, otherwise increasing taxon sampling in Ranunculaceae and Proteaceae had small or negligible effects. Most notably *Caulophyllum*, a similar phylogenetic distance from Ranunculaceae, was virtually unaffected by increased sampling in Ranunculaceae: the initial age estimate increased by ~1 my but the confidence range was scarcely different from that in analysis A. Increased taxon sampling density in Proteaceae decreased the SD value for *Caulophyllum* slightly, but increased both mean and SD for *Podophyllum*, causing the upper confidence limit for the *Podophyllum* target node to increase by 3.5 mya. In 98 of 100 bootstrap replicates from analysis E, the age estimated by r8s for the divergence of Proteaceae from Platanaceae was 110, that is, the minimum age by which this node was constrained; only in two cases was that age exceeded. Hence the constraint on this node probably limited the effect that branch lengths in Proteaceae could have on dates for nodes elsewhere in the tree.

Unsurprisingly, branch lengths in *Pachysandra* were barely affected by changes in these phylogenetically distant clades. From this, sensitivity to sampling density in neighboring clades unsurprisingly decreases with taxonomic distance, but the effect also varies dramatically between genera.

3.4 Effects of relative positions of major clades

There is not yet a consensus on whether Proteales or Sabiales diverged first on the branch leading to Buxales and core eudicots. The APG II (2003) phylogeny indicates that Proteales diverged first, whereas some studies on basal eudicots indicate that Sabiales diverged first (Anderson et al., 2005, and references therein). Given the proximity of these nodes to the fixed calibration point of crown eudicot radiation, the choice of which topology to use might impact on the date estimated for any node in the phylogeny.

The effect of switching from a Proteales-first topology to a Sabiales-first topology (analysis G) on *Pachysandra* and *Caulophyllum* are minimal, the main changes in each case being a slight decrease in SD. Only in *Podophyllum* is there a marked change, with the bootstrap mean and SD increasing by 2.2 and 0.84 my, respectively, stretching the upper confidence limit back by an additional 4 my, to 36.3 my. This result is surprising given that *Podophyllum* is not directly descended from the nodes where Proteales and Sabiales diverge, whereas *Pachysandra* is. This result probably reflects the sensitivity of the target node in *Podophyllum* to various effects (Table 2, Fig. 3), whereas *Pachysandra* might have been shielded from the effects of the change by intervening fossil calibration points. Hence the effects of changing major clade positions were overall quite small, but could not be predicted by taxonomic distance from the area where the change is made. Hence, where analyses disagree in the branching order of major clades, it might be wise to run at least initial dating estimates using several competing topologies to determine whether these will provide differing ages for target nodes.

3.5 Tricolpate pollen fixed calibration point: stem or crown node?

Moving the eudicot pollen fossil from crown to stem node (analysis H) reduced both the mean and SD of the target node ages within *Caulophyllum* (mean from 11.1 to 10.6; SD from 3.51 to 3.28), and *Podophyllum* (mean from 16.2 to 14.8; SD from 7.12 to 6.45). There was no effect on *Pachysandra*, again presumably because of the effects of intervening calibration points. Because eudicot pollen provides the only fixed age constraint alongside many minimum age constraints, its effect on other node ages should be far stronger as a maximum age constraint than as a minimum age constraint. Intuitively, placing a maximum age constraint on a crown node should be more conservative as it allows the stem node to be any age above 124 mya, whereas placing the fixed constraint on the stem node confines the crown node to a narrow range between 124 and 112 mya (see Anderson et al., 2005). Moving the fixed

calibration point from crown node to stem node should therefore have the effects of making nodes with eudicots younger, and decreasing their range of variation. These effects are seen in both *Caulophyllum* and *Podophyllum*, contracting their confidence ranges at both ends, although mainly the upper end. Therefore, on this evidence, provided that other minimum age calibration points are also present, applying a fixed age constraint to crown rather than stem node would appear to be more conservative, and is recommended.

3.6 Effect of using NPRS in place of PL

Linder et al. (2005) found that NPRS overestimated the ages of more recent nodes in a phylogeny, and that reducing taxon sampling density exacerbated this effect. The current study shows that these effects are also seen in phylogenies containing multiple angiosperm orders and deep calibration points. In all three genera, using NPRS (analysis I) makes the initial age for the target node older, however the SD increases using NPRS in *Caulophyllum* and *Pachysandra* but decreases in *Podophyllum*. The same was true of the SD calculated using taxon sampling variation (analysis J). As with PL, the age ranges revealed by taxon sampling variation were far narrower than those revealed by bootstrapping, and combining the two (analysis K) had little effect relative to bootstrapping alone.

NPRS was more sensitive to undersampling than was PL. When 10 taxa from Berberidaceae to the target node were removed (analysis L), dates calculated for *Caulophyllum* increased by approximately 3 my, whereas when PL was used there was almost no change due to this deletion (Table 2, Fig. 3). In the case of *Podophyllum*, when PL was used the deletion had no effect on initial date estimate and changed the bootstrap mean value by less than 1 my and the upper age limit by 2.6 my; using NPRS, these three values increased by 2.9, 3.5 and 6.5 my, respectively. In the case of *Pachysandra*, the deletion increases age estimates by approximately 10 my whichever smoothing method is used (Table 2; Fig. 3). Overall, these results concur with those of Linder et al., (2005) in indicating that PL is less sensitive to undersampling effects than is NPRS. This suggests that PL should be preferred to NPRS when working with study groups with relatively few species.

3.7 Biogeographic inferences

For *Caulophyllum*, the current dataset indicates a time of divergence between 3 and 25 my, or 3–19 mya if PL is accepted to be more accurate than NPRS. Clearly this is a much broader range than that in Xiang et al.'s (2000) paper, but it does include a meaningful upper age limit and hence appears to exclude disjunction due

to North Atlantic land bridge vicariance followed by extinction in western Eurasia. It cannot, however, separate the hypotheses of divergence due to BLB severance and divergence due to prior climate cooling, as it is compatible with both.

Few biogeographic conclusions can be made about *Podophyllum*, as the age range in this genus is so wide. The range is thus unsurprisingly compatible with dates for divergence within *Podophyllum* calculated by Xiang et al. (2000) (3.7–10.1 mya) and by Liu et al. (2002) (4.6–8.4 mya). However, both of these estimates had to get around the absence of direct fossil calibration. Xiang et al. (2000) used rates from a different genus for calibration, whereas Liu et al. (2002) assumed a universal mutation rate and clock-like evolution for the internal transcribed spacer region. Neither can be regarded as a confident estimate. The very wide range for *Podophyllum* in the current study is certainly due to the very variable branch lengths within the genus. The number of substitutions between the target node and the five accessions included were: AF197591, one; AF079455, five; AF139877, five; AF093716, seven; AF203488, 24 (Fig. 1). If AF197591 and AF203488 are removed the bootstrap range contracts to 3.3–15.1, similar to that calculated for *Caulophyllum*, and still compatible with those from other studies (Xiang et al., 2000; Liu et al., 2002). Overall, the results from *Podophyllum* indicate that a high level of rate heterogeneity, such as seen within this genus, can not only lead to wide confidence ranges, but might also increase sensitivity to a wide range of other effects. The problem of atypical accessions could be solved by sampling a larger number of *Podophyllum* accessions and hence determining whether the atypical accessions are natural outliers or if they might contain sequencing errors.

The age range in *Pachysandra* is consistently between 11 and 33 my when PL is used. The exception, analysis D, can clearly be discounted in terms of the real age of this genus as it is a clear illustration of the effect of undersampling. Hence this study, like that of Xiang et al. (2000), indicates divergence too early to be due to BLB severance at 5.5 mya, but it cannot reject the hypothesis of North Atlantic land bridge vicariance, and is also compatible with vicariance caused by the loss of a connection across the BLB at an earlier point than 5.5 mya, for example, due to climate cooling.

3.8 Conclusions and future directions

This study marks a step towards the goal of a universal method of molecular dating, which includes meaningful upper ages as well as lower ages, and can be applied even where there are no useful fossil calibration points within the family of interest. It has shown that

a distant but reliable maximum age calibration point can be used, in conjunction with more recent minimum age fossils, provided undersampling effects are not too great and PL is used. Crucially, the variation revealed by bootstrapping branch lengths effectively absorbed variation due to taxon sampling effects, except where undersampling effects were extreme.

The wide confidence ranges clearly have limitations for biogeographic interpretation, but there are three ways in which the ranges could be narrowed. First, if multiple DNA regions are used, as in Wikström et al. (2001), the SD of all node ages should drop and age ranges will contract. The difficulty here is a practical one: the method requires sampling across a wide enough taxonomic range to incorporate the basal eudicot node in the phylogeny, and sequences for each region will need to be available for each taxon included. Fortunately DNA banks like that at Kew are beginning to make it practical to obtain multiple sequences from a wide taxonomic range relatively easily.

Second, where multiple deep calibration points are being used, the accuracy of molecular dating could be improved by increased sampling within the target genus, hence reducing the effect of atypical sequences, which the *Podophyllum* data shows can be large. Although the genera concerned might have only two or three species, if these species are 5 my old or more then ample sequence variation may be found at the infraspecific level. However, when multiple DNA regions are used (unless they are all from the plastid) caution is required because the within-species phylogeny resolved might differ between DNA regions. Finally, as this study indicates, increased taxon sampling, particularly in under-represented clades, will likely improve the accuracy of molecular dating.

This study has shown that maximum age ranges can be calculated even for genera that lack fossil records, but equally illustrates that the data currently available limits the precision and accuracy of such an approach, particularly as the need for the broadest possible taxon sampling meant that it was only possible to use one, slow-evolving gene (*rbcL*). A study that uses multiple sequences would certainly be valuable, although for practical reasons this might need to involve a smaller taxonomic group, as in Linder et al. (2005). Taxon sampling effects do not represent an insurmountable obstacle to the use of large phylogenies with deep calibration points to date more recent nodes, and the challenge now is to expand the method to genera outside of basal eudicots, and to narrow the ranges calculated. For this, new rate smoothing methods that are able to tackle very large phylogenies (e.g. Britton et al., 2007) will be valuable, provided the sensitivity of such

methods to taxon sampling density is understood. Moreover, Bayesian methods are rapidly developing, and already are capable of combining phylogeny inference with rate smoothing (Drummond et al., 2006). Here, sensitivity to taxon sampling needs to be determined, but as computing time decreases it should become possible to incorporate taxon sampling tests into such methods, and hence ultimately to control for taxon sampling variation automatically during the rate smoothing process.

Acknowledgements I thank Casja Anderson, James Richardson and Jenny Xiang for advice concerning this work; two anonymous referees for constructive comments on the MS; Jevgenia Viktorova and the England cricket team for keeping me sane during long weeks of data analysis; and NERC (fellowship NE/B500658/1) for the financial support that made this work possible.

References

- Anderson CL, Bremer K, Friis EM. 2005. Dating phylogenetically basal Eudicots using *rbcL* sequences and multiple fossil reference points. *American Journal of Botany* 92: 1737–1748.
- APG II. 2003. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG II. *Botanical Journal of the Linnean Society* 141: 399–436.
- Bell CD, Donoghue MJ. 2005. Dating the Dipsacales: comparing models, genes, and evolutionary implications. *American Journal of Botany* 92: 284–296.
- Britton T, Anderson CL, Jacquet D, Lundqvist S, Bremer K. 2007. Estimating divergence times in large phylogenetic trees. *Systematic Biology* 56: 741–752.
- Crane PR. 1989. Paleobotanical evidence on the early radiation of nonmagnoliid dicotyledons. *Plant Systematics and Evolution* 162: 165–191.
- de Queiroz A. 2005. The resurrection of oceanic dispersal in historical biogeography. *Trends in Ecology & Evolution* 20: 68–73.
- Drummond AJ, Ho SYW, Phillips MJ, Rambaut A. 2006. Relaxed phylogenetics and dating with confidence. *PLoS Biology* 4: 699–710.
- Donoghue MJ, Bell CD, Li J. 2001. Phylogenetic patterns in Northern Hemisphere plant geography. *International Journal of Plant Sciences* 162: S41–S52.
- Doyle JA. 1992. Revised palynological correlations of the lower Potomac Group (USA) and the Cocobeach sequence of Gabon (Barremian-Aptian). *Cretaceous Research* 13: 337–349.
- Feng Y, Oh S-H, Manos PS. 2005. Phylogeny and historical biogeography of the genus *Platanus* as inferred from nuclear and chloroplast DNA. *Systematic Botany* 30: 786–799.
- Friis EM, Pedersen KR, Crane PR. 2006. Cretaceous angiosperm flowers: innovation and evolution in plant reproduction. *Palaeogeography Palaeoclimatology Palaeoecology* 232: 251–293.
- Gaut BS. 1998. Molecular clocks and nucleotide substitution rates in higher plants. *Evolutionary Biology* 30: 93–120.
- Givnish TJ, Renner SS. 2004. Tropical intercontinental disjunctions: Gondwana breakup, immigration from the boreotropics, and transoceanic dispersal. *International Journal of Plant Sciences* 165: S1–S6.
- Gladenkov AY, Oleinik AE, Marincovich L, Barinov KB. 2002. A refined age for the earliest opening of Bering Strait. *Palaeogeography Palaeoclimatology Palaeoecology* 183: 321–328.
- Hughes NF, McDougall AB. 1990. Barremian-Aptian angiosperm pollen records from southern England. *Review of Palaeobotany and Palynology* 65: 145–151.
- Linder HP, Hardy CR, Rutschmann F. 2005. Taxon sampling effects in molecular clock dating: an example from the African Restionaceae. *Molecular Phylogenetics and Evolution* 35: 569–582.
- Liu J-Q, Chen Z-D, Lu A-M. 2002. Molecular evidence for the sister relationship of the eastern Asia–North American intercontinental species pair in the *Podophyllum* group (Berberidaceae). *Botanical Bulletin of Academia Sinica* 43: 147–154.
- Kay KM, Whittall JB, Hodges SA. 2006. A survey of nuclear ribosomal internal transcribed spacer substitution rates across angiosperms: an approximate molecular clock with life history effects. *BMC Evolutionary Biology* 6: 36–45.
- Magallon SA, Sanderson MJ. 2005. Angiosperm divergence times: the effect of genes, codon positions, and time constraints. *Evolution* 59: 1653–1670.
- Marshall CR. 2008. A simple method for bracketing absolute divergence times on molecular phylogenies using multiple fossil calibration points. *American Naturalist* 171: 726–742.
- McGlone MS. 2005. Goodbye Gondwana. *Journal of Biogeography* 32: 739–740.
- Milne RI. 2006. Northern hemisphere plant disjunctions: a window on Tertiary land bridges and climate change? *Annals of Botany* 98: 465–472.
- Milne RI, Abbott RJ. 2002. The origin and evolution of Tertiary relict floras. *Advances in Botanical Research* 38: 281–314.
- Moore MJ, Bell CD, Soltis PS, Soltis DE. 2007. Using plastid genome-scale data to resolve enigmatic relationships among basal angiosperms. *Proceedings of the National Academy of Sciences USA* 104: 19363–19368.
- Pennington RT, Cronk QCB, Richardson JA. 2004. Introduction and synthesis: plant phylogeny and the origin of major biomes. *Philosophical Transactions of the Royal Society of London series B-Biological Sciences* 359: 1455–1464.
- Renner SS. 2004. Plant dispersal across the tropical Atlantic by wind and sea currents. *International Journal of Plant Sciences* 165: S23–S33.
- Renner SS. 2005. Relaxed molecular clocks for dating historical plant dispersal events. *Trends in Plant Science* 10: 550–558.
- Renner SS, Clausing G, Meyer K. 2001. Historical biogeography of Melastomataceae: the roles of Tertiary migration and long-distance dispersal. *American Journal of Botany* 88: 1290–1300.
- Ronquist F, Huelsenbeck JP. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19: 1572–1574.
- Sanderson MJ. 2002. Estimating absolute rates of molecular evolution and divergence times: a penalized likelihood approach. *Molecular Biology and Evolution* 19: 101–113.

- Sanderson MJ, Doyle JA. 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.
- Sanderson MJ, Thorne JL, Wikström N, Bremer K. 2004. Molecular evidence on plant divergence times. *American Journal of Botany* 91: 1656–1665.
- Sanmartin I, Ronquist F. 2004. Southern hemisphere biogeography inferred by event-based models: plant versus animal patterns. *Systematic Biology* 53: 216–243.
- Shaul S, Graur D. 2002. Playing chicken (*Gallus gallus*): methodological inconsistencies of molecular divergence date estimates due to secondary calibration points. *Gene* 300: 59–61.
- Soltis PS, Soltis DE, Savolainen V, Crane PR, Barraclough TG. 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.
- Swofford DL. 2002. PAUP* phylogenetic analysis using parsimony (*and other methods). Version 4. Sunderland, MA: Sinauer Associates.
- Tavaré S, Marshall CR, Will O, Soligo C, Martin RD. 2002. Using the fossil record to estimate the age of the last common ancestor of extant primates. *Nature* 416: 726–729.
- Tiffney BH. 1985a. Perspectives on the origin of the floristic similarity between eastern Asia and eastern North America. *Journal of the Arnold Arboretum* 66: 73–94.
- Tiffney BH. 1985b. The Eocene North Atlantic land bridge: its importance in Tertiary and modern phylogeography of the Northern Hemisphere. *Journal of the Arnold Arboretum* 66: 243–273.
- Tiffney BH, Manchester SR. 2001. The use of geological and paleontological evidence in evaluating plant phylogeographic hypotheses in the northern hemisphere tertiary. *International Journal of Plant Sciences* 162: S3–S17.
- Wen J. 1999. Evolution of eastern Asian and eastern North American disjunct distributions in flowering plants. *Annual Review of Ecology and Systematics* 30: 421–455.
- White JM, Ager, TA, Adam, DP, Leopold EB, Liu G, Jette H, Schweger CE. 1997. An 18 million year record of vegetation and climate change in north-western Canada and Alaska: tectonic and global climatic correlates. *Palaeogeography Palaeoclimatology Palaeoecology* 130: 293–306.
- Wikström N, Savolainen V, Chase MW. 2001. Evolution of the Angiosperms: calibrating the family tree. *Proceedings of the Royal Society of London* 268: 2211–2220.
- Wolfe JA. 1994. An analysis of Neogene climates in Beringia. *Palaeogeography Palaeoclimatology Palaeoecology* 108: 207–216.
- Xiang QY, Manchester SR, Thomas DT, Zhang WH, Fan CZ. 2005. Phylogeny, biogeography, and molecular dating of Cornelian cherries (*Cornus*, Cornaceae): tracking Tertiary plant migration. *Evolution* 59: 1685–1700.
- Xiang QY, Soltis DE, Soltis PS, Manchester SR, Crawford DJ. 2000. Timing the eastern Asian–eastern North American floristic disjunction: molecular clock corroborates paleontological estimates. *Molecular Phylogenetics and Evolution* 15: 462–472.
- Yang Z, Yoder AD. 2003. Comparison of likelihood and Bayesian methods for estimating divergence times using multiple gene loci and calibration points with application to a radiation of cute-looking mouse lemur species. *Systematic Biology* 52: 705–716.