

Familial concepts and relationships in the conifers based on *rbcL* and *matK* sequence comparisons

C. J. QUINN¹, R. A. PRICE² & P. A. GADEK³

Summary. Analysis of sequence data for the chloroplast encoded genes *rbcL* and *matK* for 40 taxa representing all conifer families and subfamilies, rooted on *Ginkgo biloba*, provides robust support for the family concepts of *Araucariaceae*, *Cupressaceae sensu lato*, *Pinaceae* and *Sciadopityaceae*. Whereas *Phyllocladus* and *Cephalotaxus* are robustly linked with, but separated from *Podocarpaceae sensu stricto* and *Taxaceae*, respectively, these distinctions decay at far fewer steps (+3.7 and +1.7 steps, respectively) than do any of the other families (>25 steps). It is argued that neither of these monogeneric families, *Phyllocladaceae* and *Cephalotaxaceae*, is warranted, and that recognition of a more broadly defined *Podocarpaceae* and *Taxaceae* is more informative of the relationships among conifers. The analysis provides strong support for *Wollemia* being sister to *Agathis*, rather than it being the first diverging lineage within the family. It also strongly supports sister relationships between *Araucariaceae* and *Podocarpaceae*, between *Cupressaceae sensu lato* and *Taxaceae sensu lato*, and between the latter clade and *Sciadopityaceae*. *Pinaceae* are strongly supported as the first lineage of extant conifers to diverge.

INTRODUCTION

The nature of the seed bearing structures has been central to all classifications of the gymnosperms, and attempts to provide a unifying interpretation of the seed cones of conifers, specifically the nature of the ovule supporting structure, have generated much interest and controversy. The controversy stemmed from attempts to establish the homologies of the highly reduced conditions found in some families, and especially the single ovule terminating a short lateral shoot in the taxads. While there is now abundant evidence available from morphology, anatomy and palaeobotany to support the interpretation of the ovuliferous scale as a highly reduced axillary fertile shoot (Florin 1954; Rothwell 1982), many other morphological characters are often highly cryptic on relationships. Attempts at phylogenetic reconstruction of conifers from morphological data (e.g. Hart 1987; Doyle 1996; Gadek *et al.* 2000; Nixon *et al.* 1994) highlight the considerable homoplasy, and consequent weakness of character support for the hypotheses generated from such analyses. Nixon *et al.* (1994: 517) have discussed this problem, pointing out that such high levels of homoplasy result from errors in interpretation of homologies in broad comparisons.

Accepted for publication February 2002.

¹ School of Biological Science, University of New South Wales, Sydney, NSW 2052, Australia. Author for correspondence: present address, Royal Botanic Gardens, Mrs Macquaries Rd, Sydney, NSW 2000, Australia.
email: chris.quinn@rbgsyd.nsw.gov.au

² P.O. Box 8082, Athens, GA 30603-8082, U.S.A.

³ School of Tropical Biology, James Cook University of North Queensland, PO Box 6811, Cairns, Qld 4870, Australia.

As a result of detailed investigation of early fossil conifers, Florin (1951) concluded that taxads represented a quite separate lineage from other conifers, being distinguished by their seeds terminating axillary fertile shoots. Despite the general weight of evidence from, for example, leaf and wood anatomy (Stewart & Rothwell 1993) and embryology (Doyle & Brennan 1971) suggesting strong affinities between taxads and other conifers (Harris 1976), this ordinal separation of *Taxaceae* from remaining conifers received considerable acceptance for some time (e.g., Sporne 1965). An alternative concept of a division between *Pinales*, with obvious cones, and a more broadly defined *Taxales* including those families in which the cone is either very reduced or absent (viz., *Taxaceae*, *Cephalotaxaceae* and *Podocarpaceae*), has also had its adherents (e.g., Li 1953; Keng 1975; Chen & Wang 1990; Ehrendorfer 1991). Miller (1999) has recently reported on a cladistic analysis of extant and basal fossil conifers using 11 characters of ovulate cones and one of pollen grains. He concludes that there were three groups of Late Paleozoic conifers, and that differences between modern conifer families stem largely from their development from different Paleozoic groups. *Taxaceae* were again considered the earliest lineage to separate, with *Pinaceae* being the second to diverge. A recent elegant reanalysis of the female reproductive structures in taxads, however, provides a unifying concept for the interpretation of ovule bearing structures throughout the conifers, and hence removes the basis for these ordinal separations (Stützel & Rößekamp 1999).

A range of unique features displayed by *Sciadopitys* and *Phyllocladus* has led to the widespread acceptance of each of these genera being referred to monogeneric families, *Sciadopityaceae* and *Phyllocladaceae* (Page 1990; Farjon 1998).

The increasing range of molecular data now becoming available for conifers and their relatives has significantly impacted on our understanding of relationships. Analyses are now published for *Araucariaceae* (Gilmore & Hill 1997; Setoguchi *et al.* 1998) and *Podocarpaceae* (Kelch 1998; Conran *et al.* 2000) using one chloroplast gene, *Cupressaceae sensu lato* (Gadek *et al.* 2000) and *Taxaceae* (Cheng *et al.* 2000; Price in press) using two regions, and *Pinaceae* (Wang *et al.* 2000) using a gene from each of the three genomes. Several studies have examined relationships between some of these families (Chase *et al.* 1993; Chaw *et al.* 1997; Stefanovic *et al.* 1998; Cheng *et al.* 2000). These broader analyses have been based on rather limited and sometimes uneven taxon sampling, and the authors are not unanimous in their conclusions on familial concepts. Both the family status of *Sciadopityaceae* (Chase *et al.* 1993; Cheng *et al.* 2000) and the submergence of *Taxodiaceae* in *Cupressaceae* (Brunsfeld *et al.* 1994; Gadek *et al.* 2000) are now well established. A close relationship between *Araucariaceae* and *Podocarpaceae* has been identified on the basis of 18S (Chaw *et al.* 1997) and 28S sequence data (Stefanovic *et al.* 1998), and the first of these authors interpreted their results as support for the recognition of *Phyllocladaceae*. An analysis of *rbcL* sequence data, however, has placed *Phyllocladus* within *Podocarpaceae* (Conran *et al.* 2000), although support for that topology was weak.

These molecular analyses have demonstrated clearly that *Taxaceae* are derived within conifers, being identified as close relatives of *Cephalotaxaceae* on both *matK* and *rbcL* data (Cheng *et al.* 2000; Price in press), but these authors differ on their estimates of support for the separate families. There is broad agreement, however,

on the placement of this clade close to *Cupressaceae sensu lato* (Cheng *et al.* 2000; Quinn & Price in press). The *Pinaceae* has been identified as the first diverging lineage of conifers in all analyses that have been rooted outside the order (Chase *et al.* 1993; Chaw *et al.* 1997; Quinn & Price in press).

Wollemia nobilis, the recently discovered extant representative of a separate and putatively ancient lineage within *Araucariaceae*, has been the subject of intense morphological and molecular investigation. Its affinities have been assessed differently on two analyses of *rbcL* data: sister to *Agathis* (Gilmore & Hill 1997) or sister to *Agathis+Araucaria* (Setoguchi *et al.* 1998).

Resolution of remaining contentious points on the phylogeny of conifers requires a less conserved gene than either 18S or 28S rDNA. We present sequence data from two chloroplast genes, *rbcL* and *matK*, for a balanced representation of genera across all putative families, from which we aim to obtain a more robust estimate of relationships.

METHODS

Taxa included in the study were chosen to represent the diversity of the currently recognized families (Table 1). Authorities for binomials, voucher specimens and GenBank numbers for all sequences are given in Table 1.

Since the database includes sequences derived from several separate laboratories, detailed methods used in each of these are to be found in Brunsfeld *et al.* (1994), Conran *et al.* (2000), Gadek *et al.* (2000) and Price (in press). In general, total DNA was extracted from fresh leaves or leaves dried in silica gel crystals. Double stranded templates were amplified using the PCR primers described in the above references, and sequenced on an automated sequencer. Problems with amplification of *matK* from *Podocarpaceae* using the standard primers necessitated the design of new primers for this family. Some combinations of primers yielded a paralogous locus from several taxa. The paralogues aligned well with sequences from other podocarps and had no internal stops, but formed a separate clade outside the family. Successful amplification within *Podocarpaceae* was achieved with the following primer combinations: forward primers, *trnK-TX2* (5'-TTCGTTCTTCCTCATGACGAA-3') or *trnK-POD1* (5'-CAAATGTGTAAGAGAAATAGT-3'; 200 bp before the start codon) and reverse primer *trnK-TX4R* (5'-TCTCTCCCCAAACCGTGCTTG-3'; 150 bp after the stop codon). These primers were designed on highly conserved regions of sequence identified by comparison of sequences of the gene and the surrounding *trnK* intron from *Taxus baccata* (Price, unpublished data) and *Picea mariana* (GenBank # AF059343).

Sequences were viewed in MacClade Version 4.0 (Maddison & Maddison 2000) to assist with the positioning of segments affected by insertion/deletion mutations (indels), and translated to check for stop codons. Deleted segments were treated as missing data in the analyses, and potentially informative indels scored as additional characters (present/absent) that were added to the sequence database.

Trees were rooted using *Ginkgo biloba* as the outgroup, this taxon being related to but unequivocally outside *Coniferales* in recent analyses of seed plants (Chase *et*

TABLE 1. Details of included taxa and sequence accession numbers. Binomials according to Farjon (1998).

	GenBank accessions	
	<i>matK</i>	<i>rbcL</i>
CONIFERALES		
ARAUCARIACEAE		
<i>Agathis vitiensis</i> (Seem.) Benth. & Hook.	AF456372	U87756
<i>A. robusta</i> (C. Moore ex F. Muell.) F. Muell.	AF456371	AF249665
<i>Araucaria araucana</i> (Molina) K. Koch	AF456373	AF248664
<i>A. heterophylla</i> (Salisb.) Franco	AF456374	U96462
<i>A. hunsteinii</i> K. Schum.	AF456375	U87749
<i>Wollemia nobilis</i> W. G. Jones <i>et al.</i>	AF456377	U87757
CEPHALOTAXACEAE		
<i>Cephalotaxus oliveri</i> Mast.	AF457108	AF456387
CUPRESSACEAE sensu lato		
<i>Athrotaxis laxifolia</i> Hook.	AF152176	L25754
<i>Callitris rhomboidea</i> R. Br. ex Rich.	AF152180	L12537
<i>Cryptomeria japonica</i> D. Don	AF152184	L25751
<i>Cunninghamia lanceolata</i> (Lamb.) Hook.	AF152185	L25757
<i>Cupressus sempervirens</i> L.	AF152187	L12571
<i>Diselma archeri</i> Hook. f.	AF152193	L12572
<i>Juniperus conferta</i> Parl.	AF152197	L12573
<i>Libocedrus plumosa</i> (D. Don) Sarg.	AF152200	L12574
<i>Sequoia sempervirens</i> (D. Don) Endl.	AF152209	L25755
<i>Thuja occidentalis</i> L.	AF152214	L12578
<i>Taiwania cryptomerioides</i> Hayata	AF152211	L25756
PINACEAE		
<i>Abies bracteata</i> Hook. & Arn.	AF456365	AF456380
<i>Cedrus deodara</i> (Roxb. ex D. Don) G. Don	AF456366	AF456381
<i>Picea pungens</i> Engelm.	AF456367	AF456382
<i>Pinus longaeva</i> D. K. Bailey	AF456368	AF456383
<i>Tsuga mertensiana</i> (Bong.) Carrière	AF456369	AF456384
PODOCARPACEAE sensu lato		
<i>Afrocarpus falcatus</i> (Thunb.) C. N. Page	AF457111	AF249589
<i>Dacrydium cupressinum</i> Sol. ex G. Forst.	AF457112	AF249634
<i>Halocarpus kirkii</i> (F. Muell. ex Parl.) Quinn	AF457117	AF249640
<i>Lepidothamnus laxifolius</i> (Hook. f.) Quinn	AF457114	AF249643
<i>Phyllocladus trichomanoides</i> D. Don	AF456376	AF249654
<i>Podocarpus elatus</i> R. Br. ex Endl.	AF457113	AF249606
<i>Prumnopitys ferruginea</i> (G. Benn. ex D. Don) de Laub.	AF457115	AF249656
<i>Saxegothaea conspicua</i> Lindl.	AF457116	AF249662

SCIADOPITYACEAE

<i>Sciadopitys verticillata</i> (Thunb.) Sieb. & Zucc.	AF457107	L25753
--	----------	--------

TAXACEAE

<i>Amentotaxus argotaenia</i> (Hance) Pilg.	AF152219	L12580
<i>Austrotaxus spicata</i> Compton	AF456378	AF456385
<i>Pseudotaxus chienii</i> (W. C. Cheng) W. C. Cheng	AF456379	AF456386
<i>Taxus baccata</i> L.	AF457109	AF456388
<i>T. chinensis</i> var. <i>mairei</i> (Lemée & Lév.) S. Y. Hu ex T. S. Liu	AB024001	AB027316
<i>Torreya nucifera</i> (L.) Sieb. & Zucc.	AB024003	AB027317
<i>T. taxifolia</i> Arn.	AF457110	AF456389

OUTGROUPS**CYCADACEAE**

<i>Cycas circinalis</i> L.	AF410164	L12674
----------------------------	----------	--------

GINKGOACEAE

<i>Ginkgo biloba</i> L.	AF456370	AJ235804
-------------------------	----------	----------

NYMPHAEACEAE

<i>Nymphaea odorata</i> Aiton	AF092988	M77034
-------------------------------	----------	--------

MAGNOLIACEAE

<i>Magnolia denudata</i> Desr.	AF123465	–
<i>M. stellata</i> Maxim.	–	AF238057

al. 1993; Chaw *et al.* 1997). Alternative roots on *Cycadales* (*Cycas* spp.) and on basal angiosperms (*Magnolia* spp. and *Nymphaea odorata*) were also employed to observe the effect of more distant roots. Heuristic searches were performed in PAUP Version 4.0b8 (Swofford 2001) using tree bisection reconnection branch-swapping and the MULPARS option. Heuristic analyses involving 100 replicates of random taxon addition were employed to search for multiple islands of trees. Initial searches were unweighted, but in order to downweight highly homoplasious characters, successive approximations weighting (SW; Farris 1969) was then performed according to the rescaled consistency index (RC) of each character. These analyses were repeated until tree length stabilised. MacClade was used to calculate the number of steps per site, and to plot these by codon position. Branch lengths for trees were calculated using the ACCTRAN (accelerated transformation optimisation) option in PAUP. Relative support for the clades identified by parsimony analysis was estimated by bootstrap (Felsenstein 1985) in PAUP using 500 replicates, and decay analyses (Donoghue *et al.* 1992) using PAUP and AutoDecay version 4.0.2' (Eriksson 2000) with ten replicates of random taxon addition on each constraint tree. Constraint analyses were conducted to test alternative topologies against the data using constraint trees constructed in MacClade and imported into PAUP. Analyses were then performed in PAUP using 100 replicates of random taxon addition.

RESULTS

Because of incompleteness of sequences and the position of the sequencing primer within the 5' end of the gene, the first 24 bp of the *rbcL* gene were excluded from analyses. Of the remaining 1404 sites, 455 (32.4%) were variable and 306 (21.8%) potentially informative. Uncorrected pairwise sequence divergences within the ingroup for *rbcL* ranged from 0.15% (between *Agathis robusta* and *A. macrophylla*) to 11.5% (between *Lepidothamnus* and *Callitris*).

Problems with primer specificity meant that some *matK* sequences were incomplete: *Callitris*, *Diselma* and *Juniperus* had between 73 and 140 bp undetermined at the 5' end; *Afrocarpus*, *Agathis heterophylla* and *A. robusta* had 205, 89 and 123 undetermined positions, respectively, near the middle of the gene; only 712 bp at the 5' end were determined for *Agathis macrophylla*. Alignment required a total of 25 indels in *matK*, ranging from one to three complete codons, and another two near the stop codon that consisted of 4 and 10 bp, respectively. Thirteen of these, all involving multiples of 3 bp, were potentially informative and were scored and added to the database (Table 2). Of the 1566 sites in *matK*, 989 were variable (63.2%) and 712 (45.4%) potentially informative. Pairwise divergences within *matK* ranged from 1.1% (between *Wollemia* and *Agathis robusta*) to 25.1% (between *Callitris* and either *Tsuga* or *Cedrus*).

Unweighted heuristic search of the *rbcL* data yielded a single island of 190 trees of 1050 steps (RC = 0.42). Three rounds of SW heuristic search gave a single tree of 443.6 steps (RC = 0.72, retention index (RI) = 0.90; Fig. 1). There is strong support (100% bootstrap) for the monophyly of *Araucariaceae*, *Cupressaceae sensu lato*, *Pinaceae*, and *Podocarpaceae* inclusive of *Phyllocladus*. *Taxaceae*, on the other hand,

TABLE 2. Positions of potentially informative insertion/deletion mutations (indels) in the *matK* database, and number of origins of each required on the SW trees (Figs 3 & 4).

Indel	positions affected	# origins
a	160 – 162	1
b	220 – 225	1
c	259 – 261	1
d	262 – 267	1
e	283 – 285	1
f	358 – 360	1
g	561 – 566	2
h	573 – 578	1
i	597 – 599	1
j	625 – 627	1
k	784 – 786	1
l	1450 – 1452	1
m	1456 – 1461	1

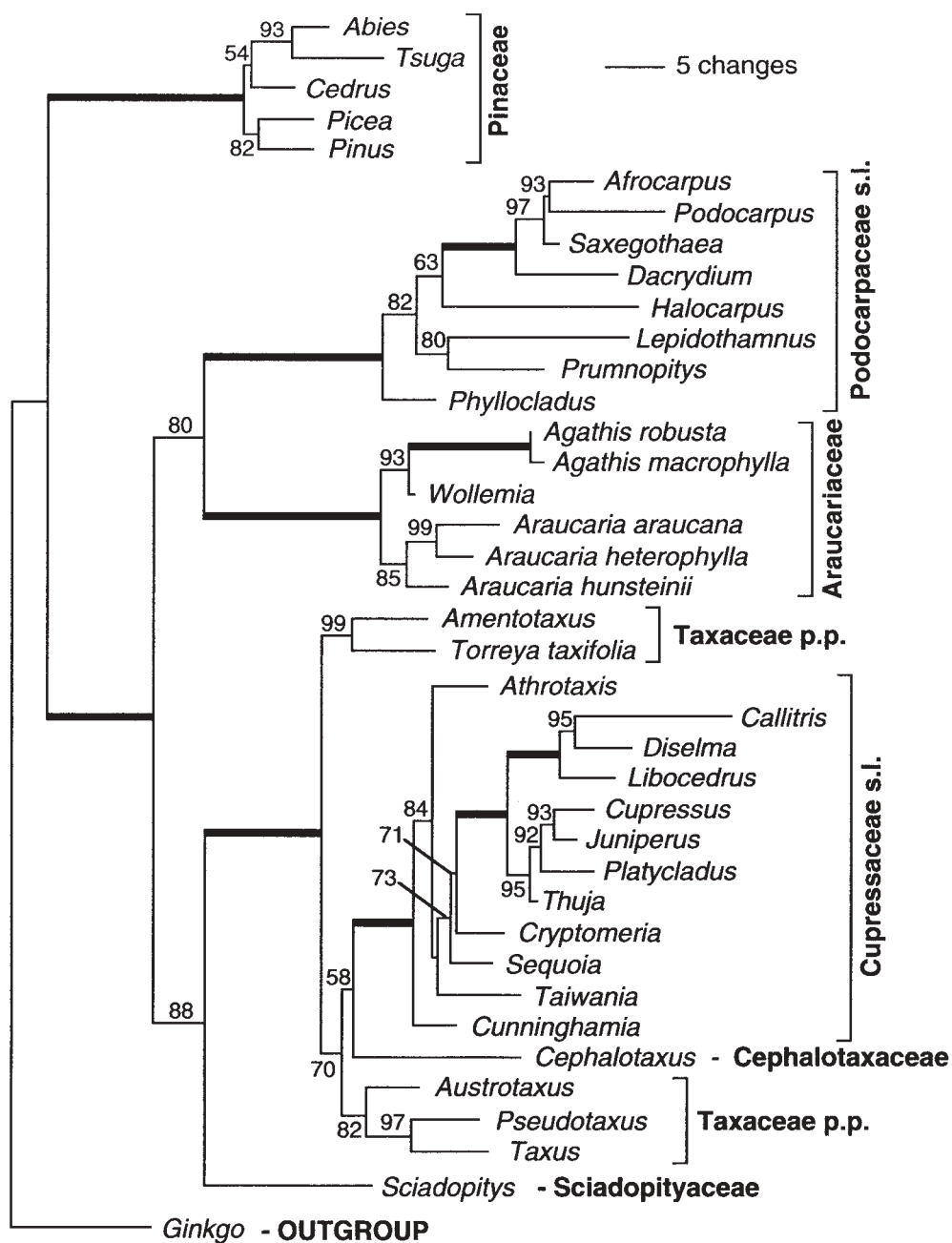


FIG. 1. Tree obtained from heuristic search with successive approximations weighting (SW) of the *rbcl* sequence data. Tree length 443.6 steps; CI = 0.68 excluding uninformative characters, RI = 0.90, RC = 0.72. Thick branches received 100% SW bootstrap support; values between 50% and 100% are shown above branches.

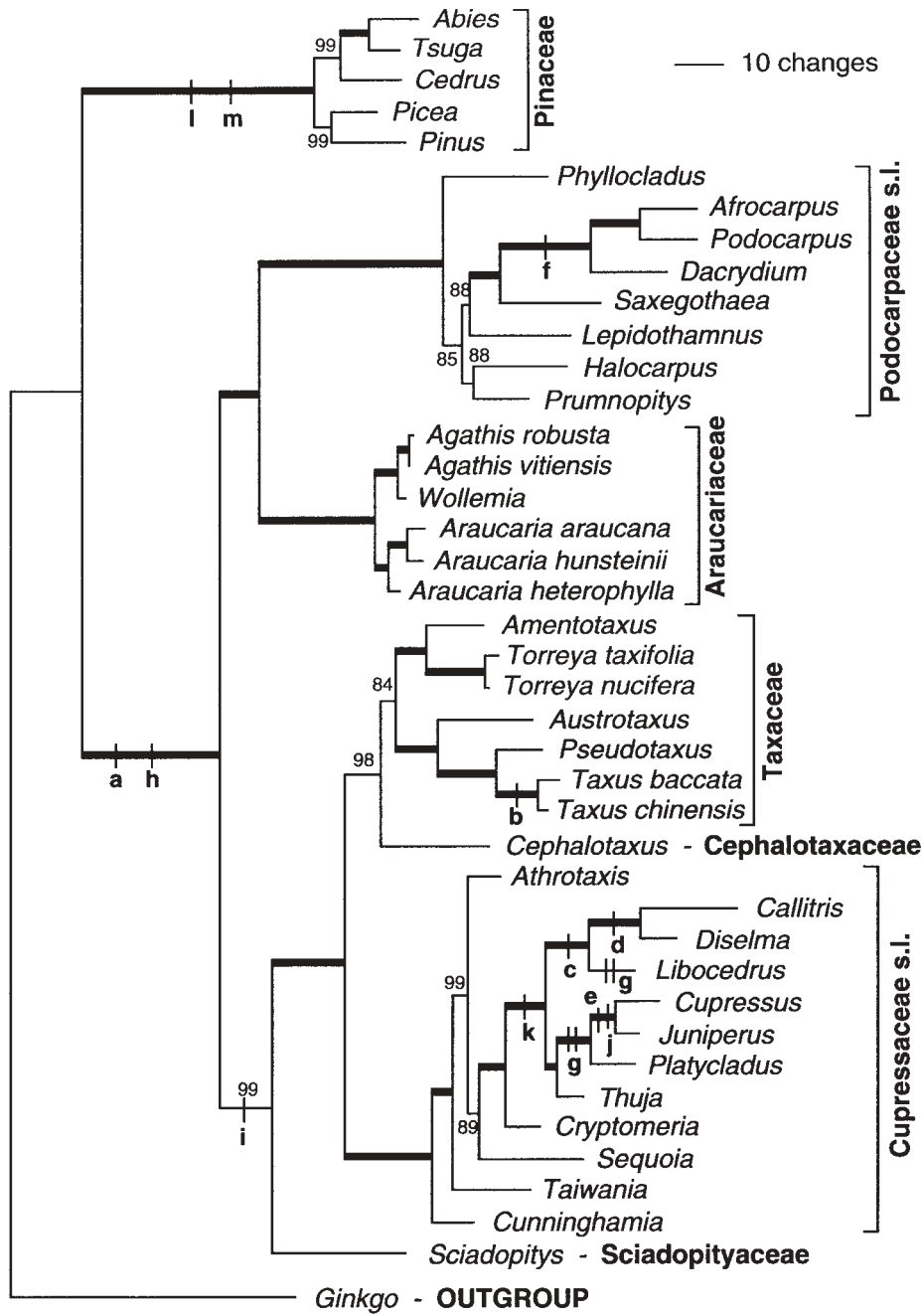


FIG. 2. Tree obtained from heuristic search with successive approximations weighting (SW) of the *matK* sequence data. Tree length 1124.6 steps; CI = 0.68 excluding uninformative characters; RI = 0.89; RC = 0.69. Thick branches received 100% SW bootstrap support; values between 50% and 100% are shown on the branches. Distributions of informative indels *a-m* (Table 2) are mapped on the tree: single bar indicates unique origin; double bar indicates parallelism.

comprise two separate lineages which, together with *Sciadopitys* and *Cephalotaxus*, are grouped with *Cupressaceae sensu lato*. There is strong support (100%) for *Sciadopitys* being the first of these lineages to diverge, but there is little support for the order of divergence of the other three lineages. Apart from the early separation of *Pinaceae* from remaining conifers, relationships between the families are for the most part not strongly supported.

Heuristic search of the *matK* database yielded a single island of only two equally parsimonious trees of 2577 steps (RC = 0.46). Three rounds of SW heuristic search gave a single tree of 1124.6 steps (RC = 0.69, RI = 0.89; Fig. 2). Again, there is robust support (100%) for the monophyly of *Araucariaceae*, *Cupressaceae sensu lato*, *Pinaceae* and *Podocarpaceae* inclusive of *Phyllocladus*. *Taxaceae* appear monophyletic, receiving moderate support (84%), with *Cephalotaxus* the strongly supported (98%) sister taxon. Relationships between families are strongly supported (98 – 100%).

Given the congruence between the topologies in Figs 1 and 2 in all the well-supported aspects, with differences being restricted to relationships between *Cephalotaxus* and the two clades of *Taxaceae* that are poorly supported in the *rbcL* analysis, we felt justified in conducting an analysis of the combined data. Initial unweighted heuristic search yielded a single island containing two equally parsimonious trees of 3639 steps (RI = 0.78, RC = 0.44), which differed only in the order of association of *Cephalotaxus* with the two clades of *Taxaceae*. The strict consensus of these trees is shown in Fig. 3. There is now robust support for all the families (100% bootstrap, decay >50 steps) except *Taxaceae* and the monotypic *Sciadopityaceae*. *Taxaceae* appear paraphyletic, consisting of two well supported clades, the relationships of which to *Cephalotaxus* are unresolved. Relationships between the remaining families are strongly resolved, the weakest feature (87%, +7) being the placement of *Sciadopitys* as sister to the very strong clade (100%, +34) comprising *Cupressaceae*, *Cephalotaxaceae* and *Taxaceae*. When either *Cycas* or the two basal angiosperms were included and used as roots, the topology within the conifers remained unaltered.

Three rounds of SW heuristic search yielded a single tree of 1564.6 steps (RC = 0.70, RI = 0.89; Fig. 4) in which all currently recognized families are monophyletic. *Araucariaceae*, *Cupressaceae sensu lato* and *Pinaceae* are each recognized as extremely robust clades (100% bootstrap, >25 steps decay). *Phyllocladus* is included within a highly robust assemblage of *Podocarpaceae sensu lato* (100%, +52.5 steps), there being less support (96%, +3.5) for its placement as the first diverging lineage within the clade. There is a well supported sister relationship (100%, +7.3) between *Araucariaceae* and *Podocarpaceae sensu lato*. *Cephalotaxus* is the well supported (98%, +4.3) sister to *Taxaceae*. The latter again comprises two very robust subclades (100%, +9.1 for *Amentotaxus* + *Torreya*; 100%, +7.9 for *Austrotaxus* + *Pseudotaxus* + *Taxus*), but support for the monophyly of *Taxaceae* is distinctly weaker than for any other family (88%, +1.7 steps). The combined *Taxaceae* + *Cephalotaxaceae* clade is robustly grouped (100%, +21.8) with *Cupressaceae sensu lato*, and *Sciadopitys* is the well supported sister to this combined clade (100%, +7.9). There is extremely strong support (100%, +30.5) for *Pinaceae* being the first lineage to diverge.

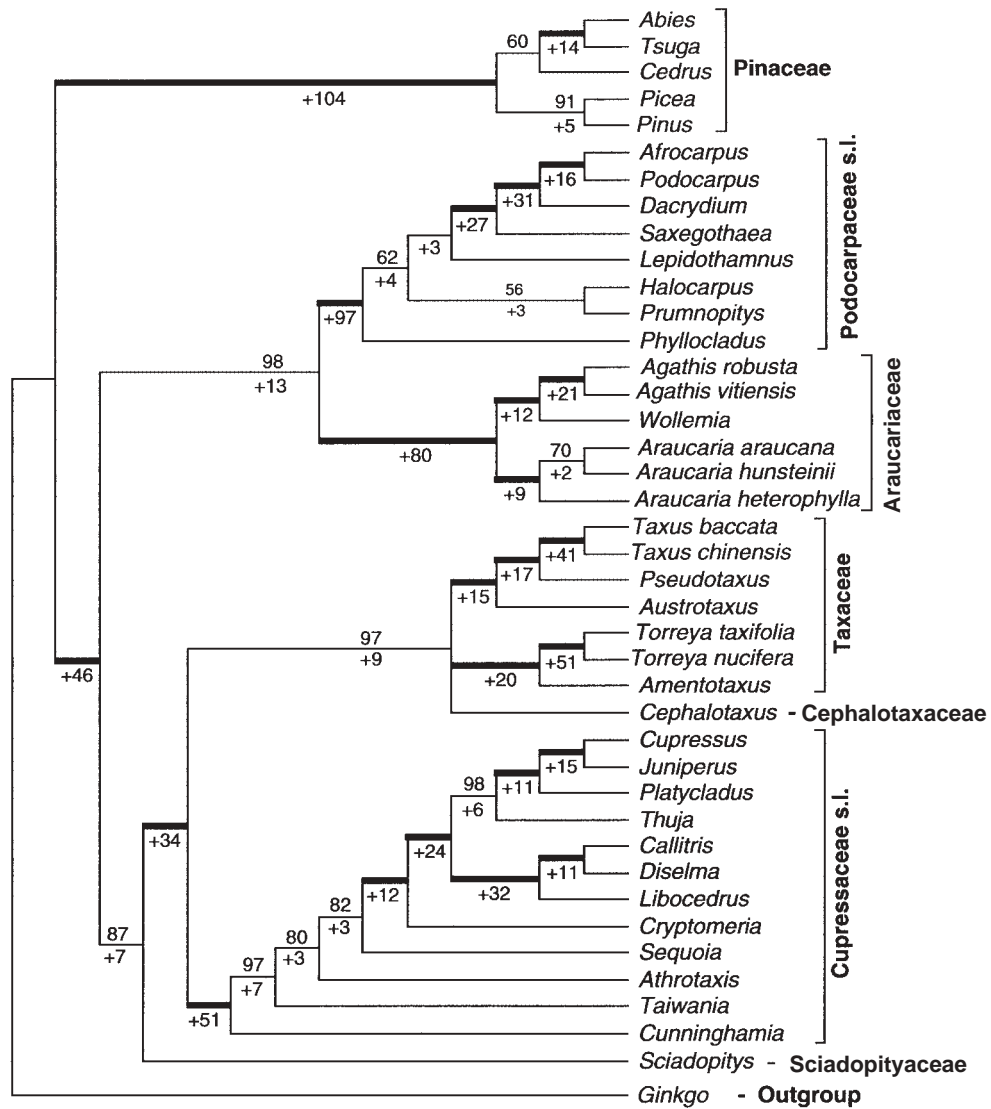


FIG. 3. Strict consensus of two equally parsimonious trees of 3639 steps found from unweighted heuristic search of the combined *rbcL* and *matK* data; CI = 0.50 excluding uninformative characters; RI = 0.78; RC = 0.44. Thick branches received 100% bootstrap support; other values >50% shown above branches. Decay values >1 shown below the branches.

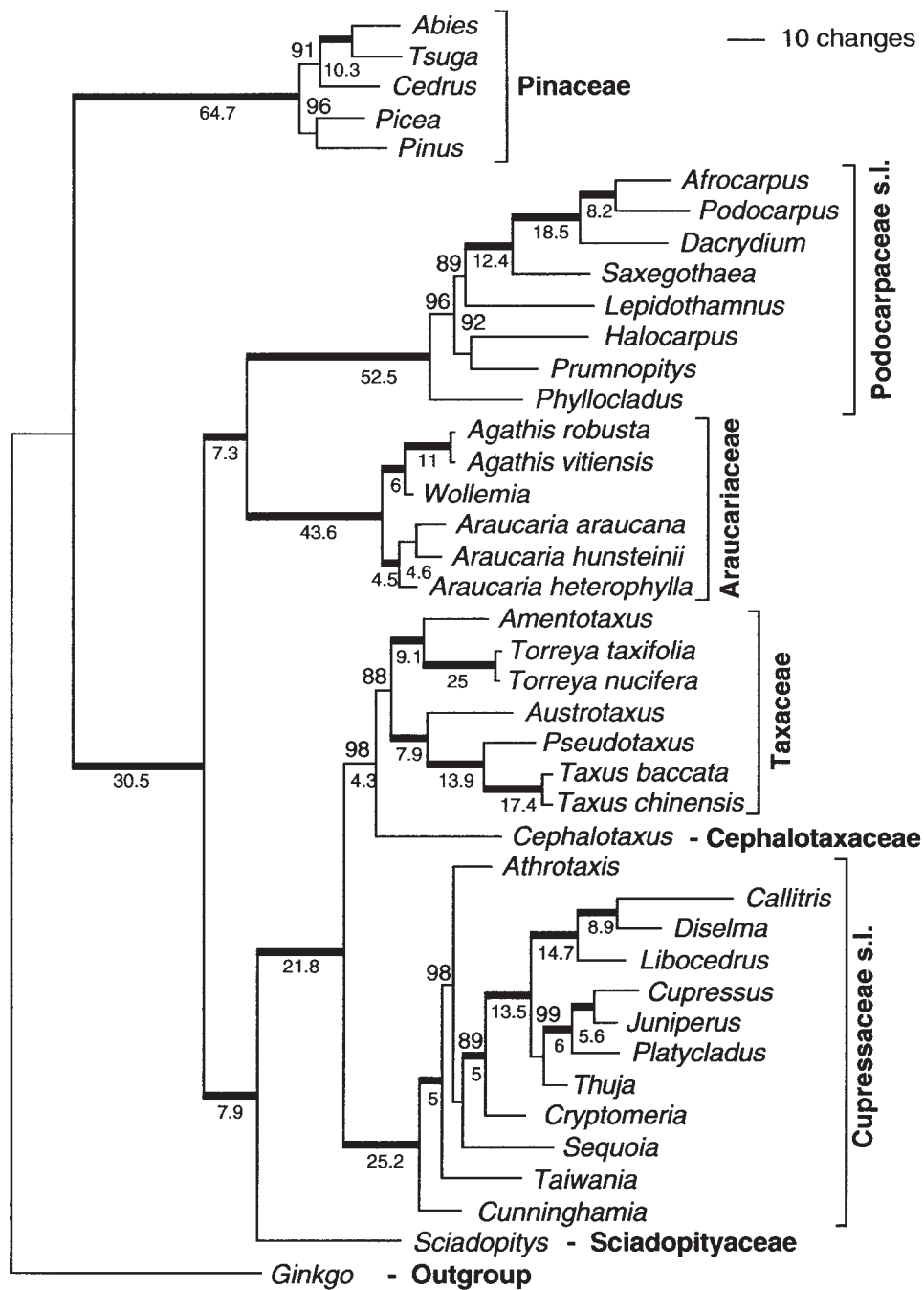


FIG. 4. Tree obtained from heuristic search with successive approximations weighting (SW) of the combined *matK* and *rbcL* data. Tree length 1564.6 steps; CI = 0.68 excluding uninformative characters; RI = 0.89; RC = 0.70. Thick branches received 100% SW bootstrap support; values between 50% and 100% shown above branches; weighted decay values >4 shown below branches.

DISCUSSION

Variable positions and informative sites within *matK* (63.2% and 45.4%, respectively) are twice as frequent as in *rbcL* (32.4% and 21.8%, respectively). This, along with the much reduced bias toward the third codon position (1.1:1:1.46 for *matK* cf. 1.7:1:5.2 for *rbcL*) and the frequent indels in the former locus, reflects the much less constrained nature of the *matK* locus which is now well documented (e.g., Liang & Hilu 1996; Gadek *et al.* 2000). The higher frequency of variable sites in the *matK* locus is therefore to a large extent the result of this lower level of functional constraint which allows more sites to tolerate change; it is not a direct reflection of an increased tendency of the locus to undergo change.

This becomes clear when the distribution of hypervariable sites is compared between the two loci (Table 3). Both the loci are AT rich, although this is more pronounced in the *matK* locus (65.1% AT) than in the *rbcL* locus (55.7% AT). An analysis of the number of steps on the combined SW tree against sites in the three codon positions is given in Table 3. The mean number of steps per variable site is highest at the third position for both loci (2.52 and 2.77, respectively for *rbcL* and *matK*), reflecting the lower levels of constraint at this position. However, individual sites at all three positions in both loci may be highly homoplasious: for *rbcL*, there are up to 10, 15 and nine steps per site in the first, second and third codon positions, respectively; for *matK* the maximum number of steps per site is eight, eight and nine, respectively, for the first, second and third positions. There are 138 sites that change five or more times when the *matK* data are plotted on the SW tree, but only 67 (48.6%) are at the third codon position. In contrast, there are a total of 43 such sites in *rbcL*, and 33 of them (76.7%) are at the third position. Hence, excluding the third codon position would remove much of the homoplasy in the *rbcL* data, but would be largely ineffective for the *matK* data. Cheng *et al.* (2000) claimed that the *matK* locus was saturated at the third position in their study of *Taxaceae* and *Cephalotaxaceae*, and translated their sequences before analysis. We consider that the existence of highly homoplasious sites at all three codon positions makes the successive approximations weighting method a more appropriate strategy for these data.

TABLE 3. Variability of loci when data plotted onto the tree obtained from successive approximations weighting analysis of the combined *rbcL* and *matK* data.

	codon position	variable positions	total steps	steps/variable site		# hypervariable sites	AT content
				mean	max		
<i>rbcL</i>	all	455	1062			43	55.7%
	1	92	176	1.19	10	7	
	2	54	106	1.96	15	3	
	3	309	780	2.52	9	33	
<i>matK</i>	all	989	2564			138	65.1 %
	1	311	683	2.19	8	41	
	2	276	659	2.39	8	30	
	3	402	1114	2.77	9	67	

There is a high degree of congruence between the topologies in Figs 1 and 2 which resulted from analyses of the data derived from the separate loci. In each case there is 100% bootstrap support for *Araucariaceae*, *Cupressaceae sensu lato*, *Podocarpaceae sensu lato*, and *Pinaceae*. The sole difference involves *Taxaceae*, which are resolved as a single clade on *matK* data, but as two separate clades on *rbcL*. Relationships between the family clades are also strongly congruent, with *Pinaceae* strongly supported as the first diverging lineage among extant conifers, *Podocarpaceae* sister to *Araucariaceae*, and *Sciadopityaceae* sister to a clade comprising the remaining families. The distribution of informative indels in *matK* (Fig. 2) is highly congruent with several features of the topology, including the positions of *Pinaceae* (indels *a* and *h*) and *Sciadopityaceae* (indel *i*). In Fig. 1 the relationships between the groups within the *Cupressaceae* + *Taxaceae* + *Cephalotaxaceae* clade are very weak: e.g., there is only 58% bootstrap support for the sister relationship between *Cephalotaxus* and *Cupressaceae sensu lato*. In Fig. 2, however, there is strong support (98%) for the placement of the former with the *Taxaceae* clade. These differences reflect the more conservative nature of the *rbcL* locus. Under these circumstances, the combining of the two databases in a single analysis is well justified.

Downweighting of the more homoplasious sites within all three codon positions of both loci has yielded a well supported estimate of the phylogeny of extant conifers. *Pinaceae*, *Cupressaceae sensu lato*, and *Araucariaceae* are each clearly resolved in all analyses, and each receives 100% bootstrap support and has a high decay value, as does the separation of the monotypic *Sciadopityaceae* from all other representatives. It is interesting to note that, although the unweighted heuristic search of the *rbcL* and *matK* data, both separately (Figs 1 & 2) and together (Fig. 3), failed to resolve the relationships between the two clades of *Taxaceae* and *Cephalotaxus*, the combined weighted analysis (Fig. 4) did provide support for monophyly of *Taxaceae* in line with the conclusions of Cheng *et al.* (2000). Nevertheless, the relatively modest level of bootstrap support for this family (88%), and in particular the low decay value for the clade (1.7 SW steps), indicates some need for caution on the order of divergence of these three lineages. There is clearly a much closer affinity between all three clades than there is between any of the other currently accepted conifer families, so that their assignment to a single family as proposed by Price (in press) has merit.

Embryological characters have been used to support a close link between *Taxaceae* and *Cephalotaxaceae* (Chen & Wang 1990). Both families were considered to have their closest relationship with *Podocarpaceae*, and all three families to belong to a separate lineage of conifers that merits recognition at the ordinal level. There is certainly no support for these latter views in the analysis presented here. A detailed reinvestigation of the female reproductive structures of *Taxaceae* has resulted in a unifying concept of ovule position in all members of the family (Stützel & Röwekamp 1999). These authors have further argued that the entire shoot bearing the axillary reproductive shoots is homologous with the compound coniferous cone of Florin (1954): ie., that the reproductive shoots in *Taxaceae* are homologous with the axillary fertile shoot (ovuliferous scale) in the model proposed by Florin (1954). This interpretation removes the basis for the distinction made between *Taxaceae* and other coniferous families, and is congruent with our molecular estimate of the phylogeny of the group.

The placement of *Phyllocladus* within a robust clade of podocarps clearly indicates that this genus, despite its unique vegetative morphology and other distinctive features (Tomlinson *et al.* 1989a, 1989b, 1997), has very close relationships with the representatives of the *Podocarpaceae* included in this analysis. The low decay values (+3.5 and +1.9 SW steps) for the separation of the *Prumnopitys* + *Halocarpus* subclade from *Phyllocladus* and *Lepidothamnus* indicates the need for some reservations on the order of these branches. Once again, they are of a quite different order to those separating the other families of conifers in Fig. 4, with the sole exception of *Taxaceae* and *Cephalotaxaceae* referred to above. The analyses of Conran *et al.* (2000) and Kelch (1998, 2002), which were based on much more intensive taxon sampling, also place *Phyllocladus* with *Podocarpaceae sensu stricto*, although the conservative nature of the *rbcL* and 18S sequence data these authors used yielded very low support values for the placement of the genus with respect to other podocarps. The apparent strength of the separation between *Phyllocladus* and the other three podocarps included in the analysis of Chaw *et al.* (1997) appears to have been an artifact of the very low taxon density; all three genera are relatively derived in Conran's and Kelch's analyses.

Constraint analyses performed on our combined SW data revealed that fixing *Lepidothamnus* or the *Prumnopitys* + *Halocarpus* subclade as the first diverging podocarp lineage required 5.4 and 3.7 additional steps, respectively. This is further demonstration of the weakness of support for the separation between *Phyllocladus* and the remaining podocarps in this database. Constraining *Phyllocladus* in the subclade with *Halocarpus* and *Prumnopitys*, in line with the topology obtained by Conran *et al.* (2000), required only 4 additional steps on our combined SW data. However, constraining *Phyllocladus*, *Prumnopitys* and *Lepidothamnus* in a clade sister to the remaining podocarps, in line with the topology obtained by Kelch (2002), required an additional 9.2 steps. Hence, our data tend to support the former pattern of relationship.

Despite the highly specialised vegetative morphology of *Phyllocladus* (Keng 1973, 1974; Quinn 1987; Tomlinson *et al.* 1989b) and the differences in its reproductive morphology (Tomlinson *et al.* 1989a) and pollination mechanism (Tomlinson *et al.* 1997), the family has been shown to be united by the embryological synapomorphy of binucleate E cells (Quinn 1986), and also by the tendency of its karyotype to undergo dysploid evolution, largely by a process that has been termed 'centric fusion' (Hair & Beuzenberg 1958; Hair 1963). The *rbcL* analysis of Conran *et al.* (2000) placed the genus as sister to *Halocarpus*, although they obtained very little support for this association. We have previously noted (Quinn & Price in press) that both genera are distinguished from other members of the family by having less than 20 major chromosome arms in their haploid complement, and this may be interpreted as independent evidence to support a sister relationship. The placement of *Saxegothaea* and *Lepidothamnus* as early diverging lineages in the *rbcL* analysis is not supported here, although the low taxon density in our analysis may have influenced the result.

Even so, the strength of support for the association of *Saxegothaea* with *Afrocarpus*, *Dacrydium* and *Podocarpus* (100%, +12.4) appears to contradict the placement in the analysis of Conran *et al.* (2000), although the strong association between *Podocarpus*

and *Afrocarpus* (100%, +8.2) is congruent with the results of analyses of 18S (Kelch 1998, 2002) and *rbcL* (Conran *et al.* 2000) sequence data, and of the morphological data (Kelch 1997). There is very little character support for the placement of the other podocarp representatives in Fig. 4, or in any of the other analyses; the inclusion of further taxa could well alter the pattern of relationships markedly. Nevertheless, further analyses with more intensive sampling of taxa, and perhaps even additional loci, are needed to obtain an unequivocal estimate of the phylogeny within this family. Considering the much improved support obtained from the combined database for the topology within *Podocarpaceae* (*cf.* Figs 1 and 4), it would seem that an expansion of the *matK* data to complement a selected subset of the existing *rbcL* database may be sufficient to provide a robust resolution of relationships between genera in the family.

There is strong support (100%, +6) for the position of *Wollemia* as sister to *Agathis*, and the closeness of this relationship is emphasized by the low divergence between the sequences of *A. robusta* and *Wollemia*: 1.1% in *matK* and 1.7% in *rbcL*. These results agree with the placement determined on *rbcL* alone by both Gilmore & Hill (1997) and Conran *et al.* (2000), but the support for the topology is much higher in our analysis. Our results do not support the conclusions of Setoguchi *et al.* (1998) based on an *rbcL* database with much higher taxon density. Those authors obtained 79% bootstrap support for *Wollemia* having diverged prior to the separation of *Agathis* and *Araucaria*, and argued that this was congruent with fossil evidence of Jurassic age. Recent re-evaluation of the fossil evidence has led to the conclusion that the earliest unequivocal record of *Wollemia* is from Mid Cretaceous deposits in New South Wales (Chambers *et al.* 1998). Currently the earliest record for *Agathis* is Late Cretaceous (R. S. Hill, Adelaide University, pers. comm.). Both appear in the fossil record much more recently than *Araucaria*, for which there are widespread records throughout the Jurassic and Cretaceous of both hemispheres (Hill & Brobribb 1999). This is more consistent with *Wollemia* being sister to *Agathis* than with it being the first lineage to diverge within the family. It can be seen that whereas the *rbcL* data are equivocal on the relationships of *Wollemia*, with taxon representation affecting the topology markedly, there is very strong support for a closer relationship with *Agathis* in the *matK* data (Fig. 2). This is a good example of the imperative of sufficient sequence divergence as a prerequisite for robust resolution of relationships; further sampling of taxa is no substitute.

Finally, the highly robust support for the early divergence of *Pinaceae* is interesting in light of the more recent appearance of identifiable members of this family in the fossil record. Though *Araucariaceae* and *Podocarpaceae* appear in the Triassic (Bock 1954; Townrow 1967; Miller 1999), *Pinaceae* are not known for certain before the Early Cretaceous (Miller 1999). Yet the cladistic analysis based on a set of reproductive characters conducted on representatives of extant and basal fossil conifers (Miller 1999), and also that on extant conifers based on a wider range of characters (Hart 1987), have each placed the divergence of *Pinaceae* before the other two families. Constraining *Araucariaceae* and *Podocarpaceae* into a clade that is sister to all the other families requires an extra 31.3 steps on the combined SW tree (+46 steps on the unweighted tree),

indicating considerable weight in the molecular evidence against this arrangement. This conflict between the results of cladistic analyses and the known fossil record of the families needs to be addressed. Further support for the early separation of *Pinaceae* from other conifers comes from a broad survey for the presence/absence of the second intron of the mitochondrial gene, *nad1* (Gugerli *et al.* 2001), which provides evidence of the monophyly of the non-*Pinaceae* conifers. The strength of the molecular data on this point suggests that it is not going to be the molecular evidence that accommodates the fossil record. It appears more likely that the lineage leading to *Pinaceae* had diverged from other conifers long before the morphological synapomorphies by which the modern family is recognised had become established.

This early separation of *Pinaceae* conflicts with past taxonomies that have emphasized a distinction between either the *Taxaceae* or a more broadly defined *Taxales* as “conifers without cones” on the one hand, and the rest of the conifers on the other (e.g. Florin 1948; Chen & Wang 1990; Ehrendorfer 1991). Nor does it support the suggestion that a more proliferating axillary fertile shoot may be plesiomorphic in the order (Schweitzer 1996; Stützel & Røwekamp 1999). This would require the reduction to the typical conifer condition to have occurred at least four times, in *Pinaceae*, in *Sciadopityaceae*, in *Cupressaceae* and in the *Podocarpaceae-Araucariaceae* clade. A more parsimonious solution on the molecular estimate of the phylogeny is that the more highly proliferated state seen on the fertile long shoots of *Torreya californica* (Stützel & Røwekamp 1999) represents a derived condition.

Miller (1999) has already noted the conflict between the placement of *Taxaceae* in his analyses of mainly seed-cone characters and other estimates of relationships between living conifers. His analyses identified *Taxus* as having diverged from within *Utrechtiaceae*, whereas all other living families were linked with *Majonicaceae*. He saw this as providing a phylogenetic basis for the distinction of *Taxaceae* from all other conifer families. Both Hart (1987) and Price *et al.* (1993), however, found a sister relationship between *Taxus* and *Cephalotaxus*, and a grouping of both these taxa with members of *Cupressaceae sensu lato*, *Sciadopityaceae* and *Araucariaceae*, a pattern that is consistent with our own results. As noted by Miller (1999), this divergence of results may indicate a need to reconsider his interpretation of the ovulate fructification in *Taxaceae*. Constraining *Taxaceae* in a sister relationship with all other living conifers in our combined analysis required an additional 76.6 steps on the SW tree (+126 steps on the unweighted tree). Hence, there is very strong evidence against this relationship in these molecular data.

The results of our analyses lead us to conclude that neither of the monogeneric families, *Cephalotaxaceae* and *Phyllocladaceae*, is warranted. *Taxaceae* should be redefined to include the former, whereas the latter should be resubmerged in the *Podocarpaceae*. We believe our results should also finally put to rest any argument for a division of the conifers at the ordinal level.

ACKNOWLEDGEMENTS

This work was supported by Australian Research Council grants AD9031851 and D18415783. The following are acknowledged for their assistance in obtaining materials: Dr Chris Page of Truro, Cornwall; the Directors of the Royal Botanic Gardens at Kew, Melbourne, and Sydney; Mr Ross Ingram of Mt Tomah Botanic Garden; Ms Kathy Vohs, University of Melbourne. We acknowledge Mr Ken Hill, Royal Botanic Gardens, Sydney, for DNA and *rbcL* sequences of *Araucariaceae*; Dr John Conran, University of Adelaide, for DNA and *rbcL* sequences of *Podocarpaceae*; Ms Margaret Heslewood for technical assistance and comments on the manuscript.

REFERENCES

- Bock, W. (1954). *Primaraucaria*, a new araucarian genus from the Virginia Triassic. *J. Paleontol.* 28: 32 – 42.
- Brunsfeld, S. J., Soltis, P. S., Soltis, D. E., Gadek, P. A., Quinn, C. J., Strenge, D. D. & Ranker, T. A. (1994). Phylogenetic relationships among the genera of *Taxodiaceae* and *Cupressaceae*: evidence from *rbcL* sequences. *Syst. Bot.* 19: 253 – 262.
- Chambers, T. C., Drinnan, A. N. & McLoughlin, S. (1998). Some morphological features of Wollemi Pine (*Wollemia nobilis*: *Araucariaceae*) and their comparison to Cretaceous plant fossils. *Int. J. Plant Sci.* 159: 160 – 171.
- Chase, M. W., *et al.* (1993). Phylogenetics of seed plants: an analysis of nucleotide sequences from the plastid gene *rbcL*. *Ann. Missouri Bot. Gard.* 80: 528 – 580.
- Chaw, S-M., Zharkikh, A., Sung, H. S., Lau, T-C. & Li, W-H. (1997). Molecular phylogeny of extant gymnosperms and seed plant evolution: analysis of nuclear 18S rRNA sequences. *Molec. Biol. Evol.* 14: 56 – 68.
- Chen, Z-K. & Wang, F-H. (1990). On the embryology and relationship of *Cephalotaxaceae* and *Taxaceae*. *Cathaya* 2: 41 – 52.
- Cheng, Y., Nicolson, R. G., Tripp, K. & Chaw, S-M. (2000). Phylogeny of *Taxaceae* and *Cephalotaxaceae* genera inferred from chloroplast *matK* gene and nuclear rDNA ITS region. *Molec. Phylogenet. Evol.* 14: 353 – 365.
- Conran, J. G., Wood, G. A., Martin, P. G., Dowd, J. M., Quinn, C. J., Gadek, P. A. & Price, R. A. (2000). Generic relationships within and between the gymnosperm families *Podocarpaceae* and *Phyllocladaceae* based on an analysis of the cp gene *rbcL*. *Austral. J. Bot.* 48: 715 – 724.
- Donoghue, M. J., Olmstead, R. G., Smith, J. F. & Palmer, J. D. (1992). Phylogenetic relationships of *Dipsacales* based on *rbcL* sequences. *Ann. Missouri Bot. Gard.* 79: 333 – 345.
- Doyle, J. A. (1996). Seed plant phylogeny and the relationships of *Gnetales*. *Int. J. Pl. Sci.* 157 Supplement: 3 – 39.
- & Brennan, M. (1971). Cleavage polyembryony in conifers and taxads — a survey. Podocarps, taxads and taxodioids. *Sci. Proc. Roy. Dublin Soc., Ser. A* 4: 57 – 88.
- Ehrendorfer, F. (1991). Gymnosperma, Nacktsamer. In P. Sitte, H. Ziegler, F. Ehrendorfer & A. Bresinsky (eds.), *Lehrbuch der Botanik für Hochschulen*. 712 – 730. Fischer, Stuttgart.

- Eriksson, T. (2000). AutoDecay version 4.0.2'. (Program distributed by author) Department of Botany, Stockholm University, Stockholm.
- Farjon, A. (1998). World checklist and bibliography of conifers. Royal Botanic Gardens, Kew.
- Farris, J. S. (1969). A successive approximations approach to character weighting. *Syst. Zool.* 18: 374 – 385.
- Felsenstein, J. (1985). Confidence limits on phylogenies: an approach using bootstrap. *Evolution* 39: 783 – 791.
- Florin, R. (1948). On the morphology and relationships of the *Taxaceae*. *Bot. Gaz.* (Crawfordsville) 110: 31 – 39.
- (1951). Evolution of cordaites and conifers. *Acta Horti Berg.* 15: 285 – 388.
- (1954). The female reproductive organs of conifers and taxads. *Biol. Rev. Cambridge Philos. Soc.* 29: 367 – 389.
- Gadek, P. A., Alpers, D. L., Heslewood, M. M. & Quinn, C. J. (2000). Relationships within *Cupressaceae sensu lato*: a combined morphological and molecular approach. *Amer. J. Bot.* 87: 1044 – 1057.
- Gilmore, S. & Hill, K. D. (1997). Relationships of the Wollemi Pine (*Wollemia nobilis*) and a molecular phylogeny of the *Araucariaceae*. *Telopea* 7: 177 – 302.
- Gugerli, F., Sperisen, C., Büchler, U., Brunner, I., Brodbeck, S., Palmer, J. D. & Qiu, Y.L. (2001). The evolutionary split of *Pinaceae* from other conifers: evidence from an intron loss and multigene phylogeny. *Molec. Phylogenet. Evol.* 21: 167 – 175.
- Hair, J. B. (1963). Cytological relationships of the southern podocarps. Pacific basin biogeography, pp. 401 – 414. Bishop Museum, Hawaii.
- & Beuzenberg, E. J. (1958). Chromosomal evolution in the *Podocarpaceae*. *Nature* 181: 1584 – 1586.
- Harris, T. M. (1976). The Mesozoic gymnosperms. *Rev. Palaeobot. Palynol.* 21: 119 – 134.
- Hart, J. A. (1987). A cladistic analysis of conifers: preliminary results. *J. Arnold Arbor.* 68: 269 – 307.
- Hill, R. S. & Brobribb, T. J. (1999). Southern conifers in time and space. *Austral. J. Bot.* 47: 639 – 696.
- Kelch, D. G. (1997). The phylogeny of the *Podocarpaceae* based on morphological evidence. *Syst. Bot.* 22: 113 – 131.
- (1998). Phylogeny of *Podocarpaceae* — comparison of evidence from morphology and 18S rDNA. *Amer. J. Bot.* 85: 986 – 996.
- (2002). Phylogenetic assessment of the monotypic genera *Sundacarpus* and *Manoao* (*Coniferales: Podocarpaceae*) utilizing evidence from 18S rDNA sequences. *Austral. Syst. Bot.* 15: 29 – 35.
- Keng, H. (1973). On the family *Phyllocladaceae*. *Taiwania* 18: 142 – 145.
- (1974). The phylloclade of *Phyllocladus* and its possible bearing on the branch systems of progymnosperms. *Ann. Bot. (London)* 38: 757 – 764.
- (1975). A new scheme of classification of conifers and taxads. *Taxon* 24: 289 – 292.
- Li, H-L. (1953). Present distribution and habitats of the conifers and taxads. *Evolution* 7: 245 – 261.

- Liang, H. & Hilu, K. W. (1996). Application of the *matK* gene sequences to grass systematics. *Canad. J. Bot.* 74: 125 – 134.
- Maddison, W. P. & Maddison, D. R. (2000). *MacClade 4.0*. Sinauer, Sunderland, MA.
- Miller, C. N. (1999). Implications of fossil conifers for the phylogenetic relationships of living families. *Bot. Rev. (Lancaster)* 65: 239 – 277.
- Nixon, K. C., Crepet, W. L., Stevenson, D. & Frits, E. M. (1994). A reevaluation of seed plant phylogeny. *Ann. Missouri Bot. Gard.* 81: 484 – 533.
- Page, C. N. (1990). *Coniferophytina*. In: K. U. Kramer & P. S. Green (eds.), *The families and genera of vascular plants*. Vol. 1. Springer-Verlag, Berlin.
- Price, R. A., Thomas, J., Strauss, S. H., Gadek, P. A., Quinn, C. J. & Palmer, J. D. (1993). Familial relationships of conifers from *rbcL* sequence data. *Amer. J. Bot. (Suppl.)* 80: 172.
- (in press). Generic and familial relationships of the *Taxaceae* from *rbcL* and *matK* sequence comparisons. *Acta Hort.*
- Quinn, C. J. (1986). Embryogeny in *Phyllocladus*. *New Zealand J. Bot.* 24: 575 – 579.
- (1987). The *Phyllocladaceae* H. Keng — a critique. *Taxon* 36: 559 – 565.
- & Price, R. (in press). Relationships of southern conifers. *Acta Hort.*
- Rothwell, G. W. (1982). New interpretation of the earliest conifers. *Rev. Palaeobot. Palynol.* 37: 7 – 28.
- Schweitzer, H. (1996). *Voltzia hexagona* (Bischoff) GEINITZ aus dem mittleren Perm Westdeutschlands. *Palaeontographica, Abt. B, Paläophytol.* 239: 1 – 22.
- Setoguchi, H., Osawa, T. A., Pintaud, J.-C., Jaffre, T. & Veillon, J.-M. (1998). Phylogenetic relationships within *Araucariaceae* based on *rbcL* gene sequences. *Amer. J. Bot.* 85: 1507 – 1516.
- Sporne, K. R. (1965). *The morphology of gymnosperms: the structure and evolution of primitive seed plants*. Hutchinson, London.
- Stefanovic, S., Jager, M., Deutsch, J., Broutin, J. & Masselot, M. (1998). Phylogenetic relationships of conifers inferred from partial 28S rRNA gene sequences. *Amer. J. Bot.* 85: 688 – 697.
- Stewart, W. N. & Rothwell, G. N. (1993) *Paleobotany and the evolution of plants* 2nd. edn. Cambridge University Press, Cambridge.
- Stützel, T. & Röwekamp, I. (1999). Female reproductive structures in *Taxales*. *Flora* 194: 145 – 157.
- Swofford, D. L. (2001). *PAUP**. Phylogenetic analysis using parsimony (*and other methods). Version 4.0b8. Sinauer, Sunderland, MA.
- Tomlinson, P. B., Takaso, T. & Rattenbury, J. A. (1989a). Cone and ovule ontogeny in *Phyllocladus* (*Podocarpaceae*). *Bot. J. Linn. Soc.* 99: 209 – 221.
- , ——— & ——— (1989b). Developmental shoot morphology in *Phyllocladus* (*Podocarpaceae*). *Bot. J. Linn. Soc.* 99: 223 – 248.
- , Braggins, J. E. & Rattenbury, J. A. (1997). Contrasted pollen capture mechanisms in *Phyllocladaceae* and certain *Podocarpaceae* (*Coniferales*). *Amer. J. Bot.* 84: 214 – 223.
- Townrow, J. A. (1967). On *Rissikia* and *Mataia*, podocarpaceous conifers from the Lower Mesozoic of southern lands. *Pap. Proc. Roy. Soc. Tasmania* 101: 103 – 136.
- Wang, X.-Q., Tank, D. C. & Sang, T. (2000). Phylogeny and divergence times in *Pinaceae*: evidence from three genomes. *Molec. Biol. Evol.* 17: 773 – 781.