Journal of Systematics and Evolution 46 (2): 130–141 (2008)
(formerly Acta Phytotaxonomica Sinica)

doi: 10.3724/SP.J.1002.2008.07165
http://www.plantsystematics.com

Inference of phylogenetic relationships among key angiosperm lineages using a compatibility method on a molecular data set

Yin-Long QIU* George F. ESTABROOK*

(*Department of Ecology & Evolutionary Biology, The University of Michigan,* Ann Arbor, MI 48109-1048, USA)

Abstract Phylogenetic relationships among the five key angiosperm lineages, *Ceratophyllum,* Chloranthaceae, eudicots, magnoliids, and monocots, have resisted resolution despite several large-scale analyses sampling taxa and characters extensively and using various analytical methods. Meanwhile, compatibility methods, which were explored together with parsimony and likelihood methods during the early development stage of phylogenetics, have been greatly under-appreciated and not been used to analyze the massive amount of sequence data to reconstruct the basal angiosperm phylogeny. In this study, we used a compatibility method on a data set of eight genes (mitochondrial *atp1, matR,* and *nad5,* plastid *atpB, matK, rbcL,* and *rpoC2,* and nuclear 18S rDNA) gathered in an earlier study. We selected two sets of characters that are compatible with more of the other characters than a random character would be with at probabilities of $p^{M<0.1}$ and $p^{M<0.5}$ respectively. The resulting data matrices were subjected to parsimony and likelihood bootstrap analyses. Our unrooted parsimony analyses showed that *Ceratophyllum* was immediately related to eudicots, this larger lineage was immediately related to magnoliids, and monocots were closely related to Chloranthaceae. All these relationships received 76%–96% bootstrap support. A likelihood analysis of the 8 gene $p^{M<0.5}$ compatible site matrix recovered the same topology but with low support. Likelihood analyses of other compatible site matrices produced different topologies that were all weakly supported. The topology reconstructed in the parsimony analyses agrees with the one recovered in the previous study using both parsimony and likelihood methods when no character was eliminated. Parts of this topology have also been recovered in several earlier studies. Hence, this topology plausibly reflects the true relationships among the five key angiosperm lineages.

Key words angiosperm, *Ceratophyllum,* character analysis, Chloranthaceae, compatibility, eudicots, magnoliids, monocots, phylogenetic method, phylogeny.

Relationships among five key angiosperm lineages (*Ceratophyllum,* Chloranthaceae, eudicots, magnoliids, and monocots) near the base of the angiosperm phylogeny remain unresolved despite several recent studies that sample a large number of taxa and characters and use various analytical methods (Chase et al., 1993; Qiu et al., 1999, 2005, 2006a; Doyle & Endress, 2000; Graham & Olmstead, 2000; Soltis et al., 2000; Hilu et al., 2003; Moore et al. 2007). A number of factors might be responsible for this phylogenetic conundrum: rapid radiation, extinction, evolutionary rate heterogeneity among different characters and different lineages, character state paucity in DNA sequence evolution that causes a disproportionately large number of back mutations, and lack of extensive fossil evidence. While it can certainly be hoped that with more genes sequenced this problem may be solved eventually, it is also worth exploring more analytical methods to untangle these difficult nodes in the angiosperm phylogeny. In this study, we use a compatibility-based method on a data set that was analyzed recently to attempt to resolve relationships among basal angiosperms (Qiu et al., 2006a). Our goals are to resolve relationships among these angiosperm lineages and to evaluate the usefulness of this compatibility-based method to this difficult phylogenetic problem.

Compatibility-based methods have not been used widely in recent phylogenetic studies. Hence we present a brief review of their history here. In the middle of the last century, a few systematic biologists began to include explicit phylogenetic concepts to compare characters, with an understanding that not all characters are equally useful for inferring evolutionary relationships among organisms. Wilson (1965) and Camin & Sokal (1965) each proposed related but distinct operational tests for the phylogenetic compatibility of a pair of characters based on the pattern of their character states within a group of related taxa. Hennig (1966) was among the first to advocate the use of compatibility to recognize characters that were phylogenetically in conflict so as to resolve them

Received: 13 December 2007 Accepted: 18 February 2008

* Authors for correspondence. Tel.: 1-734-764-8279 (YLQ), 1-734-764-6219 (GFE); Fax: 1-734-763-0544; E-mail: <ylqiu@umich.edu; gfe@umich. edu>.

explicitly. Le Quesne (1969) used the test of Wilson (1965) together with a heuristic algorithm to select characters estimated to be phylogenetically most reliable. Estabrook (1972a, b) reviewed these and other concepts of that time, and incorporated evolutionary history into the evaluation of characters. Through the 1970's and early 1980's many systematists applied compatibility concepts to evaluate characters and estimate phylogenetic relationships (e.g., Estabrook et al., 1977; Estabrook & Anderson, 1978; Meacham, 1980; Wiley, 1981). Wilson's (1965) concept of character compatibility was generalized, and the mathematical soundness of related concepts, with algorithms to implement them in practice, was established by Estabrook and his collaborators in several studies (Estabrook et al., 1975, 1976a, b; Estabrook & McMorris, 1977, 1980; Estabrook & Meacham, 1980; Meacham, 1983). Estabrook (1983), Meacham (1984), and Meacham & Estabrook (1985) reviewed the use of character compatibility analysis at a somewhat later time, and Estabrook (1997, 2008) gave a more recent review of compatibility-related concepts and questions. An early attempt to use character compatibility analysis with molecular data was made by Boulter et al. (1979). They used concepts, presented by Fitch (1975), Estabrook & Landrum (1975), and Sneath et al. (1975), to generalize compatibility concepts applicable to molecular data, and devised an algorithm to apply them to amino acid sequences to estimate relationships among 10 families of flowering plants. More recently, Pisani (2004) and Gupta & Sneath (2007) have used compatibility-based methods to investigate phylogenetic relationships in arthropods and bacteria, respectively.

1 Methods

To understand and evaluate the concepts and methods that we apply here, it is important to have an explicit definition of character compatibility. For a collection S of species or other evolutionary units (EUs), a qualitative character is a partition of S into character states of EUs that share a common property with respect to some basis for comparison. Two qualitative characters for S are defined to be compatible if there exists a tree with the EUs in S at all the branch tips (and perhaps some of the interior nodes) on which the states of both characters could evolve without homoplasy. Although we may not know which is the historically true phylogenetic tree for S, conceptually we define a true qualitative character to be one whose states can evolve on this true tree without homoplasy. Note that all true characters will be compatible with each other, and for any pair of incompatible characters at least one of them is false, i.e., suggesting a relationship that is phylogenetically false.

In our view, there are three categories of characters for any group of organisms: (1) those that accurately reflect relationships among lineages within the group and exhibit no homoplasy on the true phylogenetic tree for those lineages, (2) those that have experienced parallel, convergent or reversed evolution, and (3) those that contain human error, whether in the form of poor character definition or inaccurate coding of morphological characters, or in the form of poor alignment in molecular sequences. Characters in the first category are always compatible with each other, and they will make a group of mutually compatible characters. From such a group of characters, an accurate, if not completely resolved, representation of evolutionary relationships can be made. In the second category of characters, some could also be compatible with one another if they have experienced similar selection pressure resulting in parallel, convergent or reversed evolution. For example, the floral characters whose states reflect the wind pollination syndrome of species in the now defunct angiosperm subclass Hamamelidae are so often compatible with one another that they have misled botanists for nearly a century to incorrectly recognize that taxon (see Cronquist, 1981; Qiu et al., 1998). Similarly, genes in the mitochondrial genomes of the angiosperm genera *Plantago* and *Pelargonium* seem to have experienced accelerated evolution in comparison to those of other angiosperms, and in most phylogenetic analyses this phenomenon would lead to mis-placement of these two taxa because of random compatibility generated by the limited character states of DNA sequence evolution (Cho et al., 2004; Parkinson et al., 2005). However, we believe that true causes of the compatibility among characters of this kind are more likely to be brought to light if more characters of the entire organism are investigated carefully. Finally, characters in the third category are compatible among themselves in unpredictable ways.

In a real phylogenetic study, we do not know which characters belong to which category. However, because characters in the first category are compatible among themselves and should form a core group of compatible characters, they may appear among a surprisingly large number of characters with which a given character may be compatible. We here introduce a character concept that has been previously proposed by Meacham (1994), the COSLAC, which is a

*C*haracter that is c*O*mpatible *w*ith a *S*urprisingly *LA*rge number of other *C*haracters. The first category of characters should be COSLACs. Fewer second category characters should be among COSLACs, and very few third category characters should belong to COSLACs. Thus we select characters that qualify as COSLACs for further analysis, because among such characters should be relatively more first category characters and relatively fewer second and third category characters.

To discover which characters are COSLACs, we use the criterion of Estabrook and Landrum (1975), Fitch (1975), and Sneath et al. (1975) to test the compatibility of a pair of characters, as shown in Fig. 1. Notice that this algorithm does not require the construction of any phylogenetic tree. For each character (or position in the aligned sequences), we compare it with every other character, counting the *N*umber of other *C*haracters with which it is *C*ompatible (NCC hereafter).

One might naively think that the characters compatible with the most other characters would be more likely to be COSLACs. However, Meacham (1981) showed that, depending on the number of character states and the distribution of taxa through those states, some characters are more likely than others to be compatible at random with other characters. For this reason it is important to know whether a given character is compatible with many other characters as much as we would expect of a random character. To address this issue, we need to calculate the probability that a random character would be compatible with at least as many other characters as was a given character. It would, however, be an impossibly complicated problem to calculate such a probability using the closed procedures of Meacham (1981) for each character in a large data set, such as the one we analyze here. To avoid this problem, Meacham (1994) estimated very close approximations to these probabilities using simulation. We will use his approach here.

To estimate this probability for a given character, we replace it with one chosen at random equiprobably from all possible characters with the same number of states and the same number of EUs in each state (distribution of the states among the EUs will be almost always different in the random character than in the given character). We then compare this random character to each of the other observed characters in the data set, and count the number of them with which it is compatible. We repeat this process 10000 times. The probability that a random character would be compatible with at least as many other characters as was the given character (NCC) can now be estimated as: the number of simulated characters that were compatible with *NCC* or more other characters divided by the number of simulated characters, in our case 10000. Note that the other characters that are compatible with the random character may or may not be the same as those compatible with the given character, and only the number may be equal or larger. This probability, termed p^{MANY} here, can be construed as the realized significance of *NCC* for the given character. An *NCC* equal to the expected number of other characters with which a random character would be compatible would have a significance of $p = 0.5$. A character with a realized significance of *NCC* substantially less than $p^{\mathrm{MANY}} = 0.5$ would be compatible with surprisingly many other characters, i.e., less likely to be a random character and thus qualified as a COSLAC. This could be grounds for including such a character in the data set used subsequently for further phylogenetic analysis.

A computer program called MEACHAM (available at www-Personal.umich.edu/~gfe/) was developed based on the fast algorithm of Estabrook and McMorris (1977) and was used here to identify COSLACs in the 8-gene matrix used in an earlier study (Qiu et al., 2006a). The eight genes used in that study were: mitochondrial *atp1, matR,* and *nad5,* plastid *atpB, matK, rbcL,* and *rpoC2,* and nuclear 18S rRNA gene. Because highly divergent taxa in the data set present problems for proper identification of COSLACs, the gymnosperms, *Amborella*, Nymphaeales, and Austrobaileyales used in the original data set were excluded from the analyses here. As a result, 144 taxa representing *Ceratophyllum*, Chloranthaceae, eudicots, magnoliids, and monocots were used in this study. Removal of *Amborella*, Nymphaeales, Austrobaileyales and gymnosperms prevents them from influencing resolution of relationships among the five key angiosperm lineages, and previous studies have demonstrated that these five lineages make a strongly supported monophyletic group (Qiu et al., 1999, 2005, 2006a; Hilu et al., 2003).

Each of the eight genes was analyzed individually using the program MEACHAM to identify COSLACs. To illustrate the output file from analyses, we present a sample from the nuclear 18S rRNA gene in Table 1, which shows a list of selected sites, the number of other compatible characters with a given site (NCC), and realized significance of *NCC* for the site. We provide the following explanation to help interpret this output file. For example, site 34 is compatible with 297 of the other informative sites of

A, matrix

	character		
evolutionary unit	I	II	III
a	A	C	A
b	C	A	A
c	A	G	T
d	G	T	T
e	C	C	C
f	G	C	G
g	C	C	G

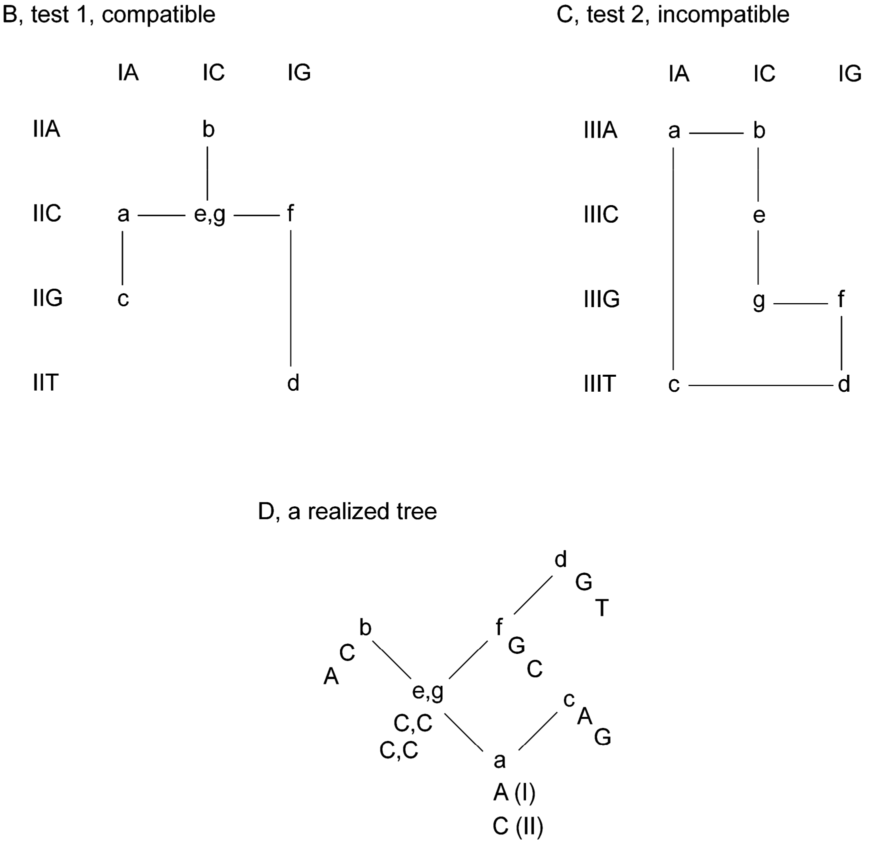


Fig. 1. An example to demonstrate tests of potential compatibility for three qualitative characters (I, II, and III) in a study of seven evolutionary units (a – g). **A** shows a matrix of character state distribution of three characters in seven evolutionary units. **B** and **C** illustrate two tests of character compatibility. In each test, states of one character label the row and those of the other label the column; each evolutionary unit is placed in the cell whose row and column labels indicate the states that it manifests. Moving only from one occupied cell to another in a straight line horizontally or vertically but never retracing a path already taken, if you can return to an occupied cell you have already visited then the two characters are incompatible, as for I and III in test 2. Otherwise, the two characters are compatible, as for I and II in test 1. **D** presents a realized tree from two compatible characters, I and II.

Table 1 A sample of output file of compatible analysis of the nuclear 18S rRNA gene*

Site	*NCC*	p^{MANY}	p^{FEW}	Site	*NCC*	p^{MANY}	p^{FEW}
25	289	0.705	0.324	144	9	0.572	0.463
34	297	0.452	0.575	146	276	0.045	0.962
36	111	0.081	0.928	150	168	0.972	0.034
39	28	0.888	0.123	154	252	0.872	0.142
42	201	0.171	0.846	160	147	0.072	0.938
96	57	0.016	0.988	162	302	0.331	0.691
101	18	0.909	0.097	169	201	0.988	0.014
102	99	0.023	0.981	170	127	0.291	0.735

* *NCC* indicates number of other compatible characters for a selected site; p^{MANY} and p^{FEW} represent respectively the probabilities that at least as MANY and at most as FEW other informative sites would be potentially compatible with a random site with the same frequency of EU's among its states as observed, estimated by 10000 simulations per site.

the gene. A random site was chosen 10000 times equiprobably from all possible sites with the same number of EUs exhibiting each nucleotide as for site 34. For 4520 of these random sites (p^{MANY} = 0.452), the number of other informative sites of 18S with which these random sites were compatible was greater than or equal to 297; for 5750 of these random sites (p^{FEW} = 0.575), the number of other informative sites of 18S with which these random sites were compatible was less than or equal to 297. Thus site 34 is compatible with about as many other sites as would be expected of a random site.

Site 96 is compatible with 57 of the other informative sites of 18S. Of 10000 random sites, only 160 were compatible with 57 or more of the other informative sites of 18S (p^{MANY} = 0.016), and 9880 (nearly all) were compatible with 57 or fewer other informative sites of 18S (p^{FEW} = 0.988). Site 96 has too many other sites (even though *NCC* = 57) compatible with it to seem like a random site, because only very few random sites were compatible with 57 or more other sites. Thus, site 96 is significantly non-random (p^{MANY} = 0.016) and qualifies as a COSLAC.

When *NCC* has been calculated for each site of a gene and the probability *p* that a random character will be compatible with at least *NCC* other sites has been estimated by simulation, two subsets of sites are chosen for further phylogenetic analysis: sites with $p^{\mathrm{MANY}} \leqslant 0.1$, which are compatible with surprisingly many other sites; and sites with $p^{\mathrm{MANY}} \leqslant 0.5$, which are compatible with at least as many other sites as would be expected of a random character. Basically, two categories of COSLACs are identified according to different levels of realized significance (i.e., probabilities that the characters selected for further analysis are better than random characters).

The resulting matrices were analyzed using both parsimony (Swofford, 2003) and maximum likelihood (Posada & Crandall, 1998; Guindon & Gascuel, 2003) bootstrap (Felsenstein, 1985) methods to investigate phylogenetic relationships among *Ceratophyllum,* Chloranthaceae, eudicots, magnoliids, and monocots. The search details are available upon request.

2 Results and Discussion

2.1 Character compatibility in the eight genes of the five angiosperm lineages

The numbers of sites in each of the eight genes with various levels of realized significance are presented in Table 2. The sites with realized significance $p^{\mathrm{MANY}} \leqslant 0.1$ and $p^{\mathrm{MANY}} \leqslant 0.5$ of at least as MANY other compatible characters as *NCC* represents really high and high quality COSLACs, respectively; their numbers are listed under $No^{\mathrm{M<0.1}}$ and $No^{\mathrm{M<0.5}}$ in Table 2. These sites were used in the parsimony and maximum likelihood bootstrap analyses to reconstruct phylogenetic relationships among the five key angiosperm lineages.

The levels of compatibility shown in Table 2 are strikingly low. Day et al. (1998) used the compatibility criterion described here to measure the phylogenetic randomness of 102 published data sets, of which only 7 had comparably low levels of compatibility. Only about half of the informative sites are compatible with more other sites than would be expected of a random site (see *NR* in Table 2); this is what we would expect if all the sites were random. On the other hand, the data are clearly non-random, which is shown by the observed number of sites with realized significance $p^{\mathrm{MANY}} \leqslant 0.1$ of at least as MANY other compatible characters as *NCC* being far greater than the expected number of such sites for the random data (see $No^{\mathrm{M<0.1}}$ in Table 2).

Less than half of the sites in plastid *atpB* and

Table 2 The results of compatible analyses of the eight genes*

Gene	No^{T}	No^{I}	$No^{M<0.1}$	$No^{M<0.5}$	$No^{F>0.2}$	$No^{F>0.4}$	*NR*
Nuclear							
18S	1755	350	95 (35)	167 (175)	248	153	0.48
Mitochondrial							
atp1	1330	373	97 (37)	144 (187)	235	215	0.39
matR	2153	709	202 (71)	366 (355)	539	292	0.52
nad5	1248	218	65 (22)	114 (109)	169	81	0.52
Total	4731	1300	364 (130)	624 (651)	943	588	0.48
Plastid							
atpB	1506	568	95 (57)	228 (284)	367	296	0.40
matK	1851	1222	395 (122)	733 (611)	995	494	0.60
rbcL	1043	561	180 (56)	339 (281)	447	188	0.60
rpoC2	3173	1864	534 (186)	1004 (932)	1363	749	0.54
Total	7573	4215	1204 (421)	2304 (2108)	3174	1727	0.54
Grand total	14059	5865	1663 (586)	3095 (2933)	4365	2468	0.52

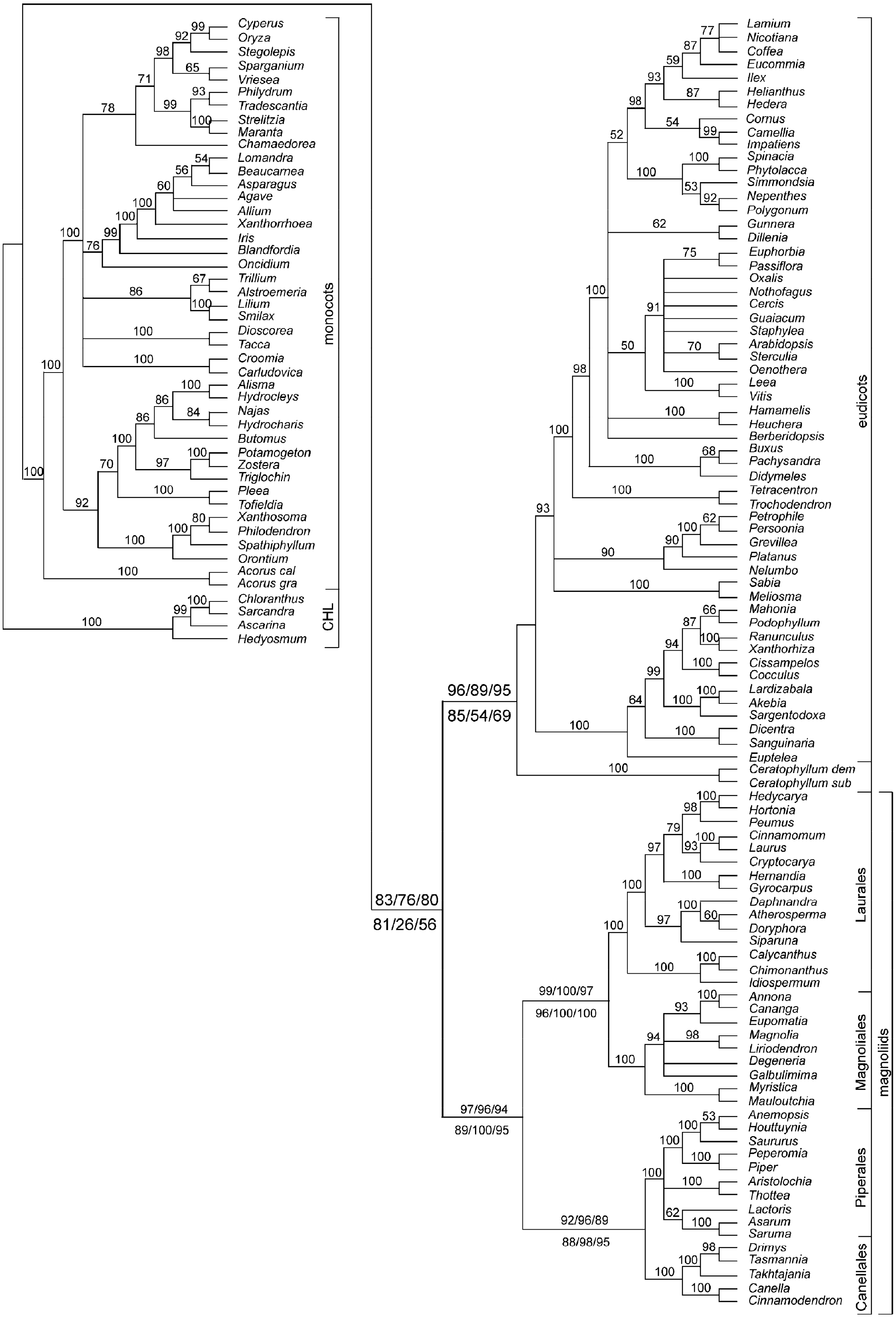
* No^{T} = number of total sites; No^{I} = number of informative sites; $No^{M<0.1}$ and $No^{M<0.5}$ represent respectively numbers of sites with realized significance $p^{MANY} \leqslant 0.1$ and $p^{MANY} \leqslant 0.5$ of at least as MANY other compatible characters as *NCC* (the numbers in parentheses represent the expected number of sites if the data were random); $No^{F>0.2}$ and $N^{F>0.4}$ represent respectively numbers of sites with realized significance $p \geqslant 0.2$ and $p \geqslant 0.4$ of at most as FEW other compatible characters as *NCC*; $NR = N^{M<0.5}/No^{I}$, i.e., the fraction of informative sites more compatible than expected of a random site.

mitochondrial *atp1* are compatible with more other sites than would be expected of a random character. Of the remaining genes, the chloroplast genes have slightly more sites that are compatible with one another, although levels are still low. These low levels of compatibility might have been caused by several factors mentioned at the beginning of the paper: rapid radiation, extinction, evolutionary rate heterogeneity among different characters and different lineages, and character state paucity in DNA sequence evolution that causes a disproportionately large number of back mutations. They are consistent with the difficulty that has been experienced in several earlier studies attempting to elucidate relationships among these angiosperm lineages using molecular data (Chase et al., 1993; Qiu et al., 1999, 2005, 2006a; Graham & Olmstead, 2000; Soltis et al., 2000; Hilu et al., 2003). Previous studies have also detected high levels of homoplasy in morphological characters among key basal angiosperm lineages (Donoghue & Doyle, 1989; Doyle & Endress, 2000). Hence, these observations highlight the need of conducting phylogenetic analysis using refined character sets to resolve relationships among the key angiosperm lineages.

2.2 Phylogenetic relationships among the key angiosperm lineages inferred from COSLACs in the eight genes

Figure 2 shows the bootstrap consensus tree from an unrooted parsimony analysis of the 8 gene matrix composed of $p^{M<0.1}$ sites (COSLACs at the $p^{MANY} \leqslant 0.1$ level of significance). In this tree, eudicots are immediately related to *Ceratophyllum* with 96% bootstrap support; this lineage is in turn immediately related to magnoliids with 83% bootstrap support. The latter value can also be interpreted as support for a close relationship between monocots and Chloranthaceae as the tree is an unrooted network. In all trees shown here, Chloranthaceae are placed at the bottom being sister to all other taxa because some phylogenetic analyses have indicated that they may represent the lineage splitting from other angiosperms right after Austrobaileyales (Doyle & Endress, 2000; Qiu et al., 2006a), and because this family also has the oldest fossil record among all angiosperms (Friis et al., 1986, 1999; Eklund et al., 2004). We use this topology merely for the convenience of presentation.

In parsimony bootstrap analyses of three other matrices, one made of the 8 gene $p^{M<0.5}$ sites and two consisted of the 4 plastid gene $p^{M<0.1}$ and $p^{M<0.5}$ sites respectively, topologies of the bootstrap consensus trees are all identical to the one shown in Fig. 2 in terms of relationships among the key lineages, and topologies within the key lineages are all similar to those shown in Fig. 2. Thus, we will not present those trees here and instead provide only bootstrap values for the important nodes in Fig. 2. In all three analyses, the relationships among the five key angiosperm lineages receive moderate (75%–90%) to strong (>90%) bootstrap support. No parsimony bootstrap analysis was performed on the matrices composed of


Cyperus
Oryza
Stegolepis
Sparganium
Vriesea
Philydrum
Tradescantia
Strelitzia
Maranta
Chamaedorea
Lomandra
Beaucarnea
Asparagus
Agave
Allium
Xanthorrhoea
Iris
Blandfordia
Oncidium
Trillium
Alstroemeria
Lilium
Smilax
Dioscorea
Tacca
Croomia
Carludovica
Alisma
Hydrocleys
Najas
Hydrocharis
Butomus
Potamogeton
Zostera
Triglochin
Pleea
Tofieldia
Xanthosoma
Philodendron
Spathiphyllum
Orontium
Acorus cal
Acorus gra
Chloranthus
Sarcandra
Ascarina
Hedyosmum
monocots
CHL
Lamium
Nicotiana
Coffea
Eucommia
Ilex
Helianthus
Hedera
Cornus
Camellia
Impatiens
Spinacia
Phytolacca
Simmondsia
Nepenthes
Polygonum
Gunnera
Dillenia
Euphorbia
Passiflora
Oxalis
Nothofagus
Cercis
Guaiacum
Staphylea
Arabidopsis
Sterculia
Oenothera
Leea
Vitis
Hamamelis
Heuchera
Berberidopsis
Buxus
Pachysandra
Didymeles
Tetracentron
Trochodendron
Petrophile
Persoonia
Grevillea
Platanus
Nelumbo
Sabia
Meliosma
Mahonia
Podophyllum
Ranunculus
Xanthorhiza
Cissampelos
Cocculus
Lardizabala
Akebia
Sargentodoxa
Dicentra
Sanguinaria
Euptelea
Ceratophyllum dem
Ceratophyllum sub
eudicots
Hedycarya
Hortonia
Peumus
Cinnamomum
Laurus
Cryptocarya
Hernandia
Gyrocarpus
Daphnandra
Atherosperma
Doryphora
Siparuna
Calycanthus
Chimonanthus
Idiospermum
Annona
Cananga
Eupomatia
Magnolia
Liriodendron
Degeneria
Galbulimima
Myristica
Mauloutchia
Anemopsis
Houttuynia
Saururus
Peperomia
Piper
Aristolochia
Thottea
Lactoris
Asarum
Saruma
Drimys
Tasmannia
Takhtajania
Canella
Cinnamodendron
Laurales
Magnoliales
Piperales
Canellales
magnoliids
83/76/80
81/26/56
96/89/95
85/54/69
97/96/94
89/100/95
99/100/97
96/100/100
92/96/89
88/98/95

either mitochondrial or nuclear gene COSLAC sites because some search replicates found a huge number of equally parsimonious trees and the analyses could not be finished within a reasonable amount of time.

In a maximum likelihood bootstrap analysis of the 8 gene $p^{M<0.5}$ site matrix, we obtained a consensus tree with a topology virtually identical to that shown in Fig. 2, but with only 54% and 26% bootstrap values for the close relationships between eudicots and *Ceratophyllum* and between this larger lineage and magnoliids, respectively. These results are again shown in Fig. 2 to save space. Our maximum likelihood bootstrap analyses of five other matrices produced four topologies that differed from the one shown in Fig. 2 in terms of relationships among these key angiosperm lineages. In consideration of space limitation and presentation conciseness, we provide only schematic diagrams of these trees that depict relationships among the lineages with bootstrap values indicated on the important nodes. These matrices and the resulting trees are: (1) the 8 gene $p^{M<0.1}$ site matrix and Fig. 3A, (2) the 3 mitochondrial gene $p^{M<0.1}$ site matrix and Fig. 3B, (3) the 3 mitochondrial gene $p^{M<0.5}$ site matrix and Fig. 3C, and (4) the 4 plastid gene $p^{M<0.1}$ and $p^{M<0.5}$ site matrices and Fig. 3D. In contrast to the parsimony bootstrap analyses, likelihood bootstrap analyses of all six matrices recovered very low bootstrap values for relationships among these lineages, whether the topologies were identical to or different from the one shown in Fig. 2.

The moderate to strong bootstrap support for relationships among *Ceratophyllum*, Chloranthaceae, eudicots, magnoliids, and monocots, recovered in the parsimony analyses shown above gives some indication that these relationships may be resolved soon. Though complete resolution of a difficult phylogenetic problem should receive consistent internal support within the data of a study and also have external corroboration from evidence of other studies, a high bootstrap value is an indication of strong internal support and can usually be taken as an early sign that the problem may be near resolution (Nei et al., 1998; Qiu et al., 2006a). In this case, the moderately to strongly supported relationships among the five key angiosperm lineages are also recovered in a maximum likelihood bootstrap analysis of one matrix (Fig. 2), albeit with low support. Moreover, in maximum likelihood bootstrap analyses of the 3 mitochondrial gene $p^{M<0.1}$ and $p^{M<0.5}$ site matrices (Fig. 3: B, C), the overall topology of both trees would be identical to that in Fig. 2 if Chloranthaceae were attached to monocots. The mitochondrial genes exhibited accelerated evolution in *Acorus* and alismatids and were excluded from the data set (Qiu et al., 2006a). This data removal might have been responsible for the different placement of monocots in these two analyses. Hence, it is likely that the topology in Fig. 2 reflects the true underlying evolutionary relationships among these five angiosperm lineages. Still, we would caution that this result should serve only as a hypothesis for further test in future studies.

The results of maximum likelihood analyses are puzzling in several aspects: (1) they differ by data sets, (2) the bootstrap values are uniformly low, and (3) they are different from those obtained by the parsimony analyses. When such sharply different results are obtained from likelihood and parsimony analyses, one may be inclined to think that the parsimony analyses have probably suffered from the systematic errors present in the data due to the particular character and taxon distribution shaped by extinction, evolutionary rate heterogeneity among different characters and different lineages, and sampling error in the study design (Felsenstein, 1978). While this possibility cannot be excluded, it is worth pointing out that the phylogenetic pattern recovered by the parsimony analyses is also present in some of the likelihood analysis results (Fig. 2; Fig. 3: B, C). Perhaps in this case likelihood analyses have failed to detect the true historical pattern due to over-parameterization. It should also be realized that likelihood analysis computer programs are at an early stage of development. Further, maximum likelihood methods can become strongly biased and statistically inconsistent when sequence evolutionary rates change non-identically over time (Kolaczkowski & Thornton, 2004). Therefore, we are not particularly concerned by the poor results of the likelihood analyses in this study even though they certainly caution us to be careful in interpreting the results generated by the compatibility

←

Fig. 2. Results of bootstrap analyses of various compatible site matrices and one matrix with all sites. The tree shown is the parsimony bootstrap consensus tree from the 8 gene $p^{M<0.1}$ site matrix. For other matrices, only bootstrap values (all from parsimony analyses except indicated) for the nodes under investigation in this study are provided and they are shown in boldface and large font, in the following order: above the branch, 8 gene $p^{M<0.1}$ site matrix / 8 gene $p^{M<0.5}$ site matrix / 4 plastid gene $p^{M<0.1}$ site matrix; below the branch, 4 plastid gene $p^{M<0.5}$ matrix / 8 gene $p^{M<0.5}$ site matrix with maximum likelihood analysis / 8 gene all site matrix (without editing sites) with parsimony analysis. Bootstrap values from analyses of all these matrices are also provided for monophyly of, and relationships within, magnoliids. Abbreviations: *Acorus cal, Acorus calamus; Acorus gra, Acorus gramineus; Ceratophyllum dem, Ceratophyllum demersum; Ceratophyllum sub, Ceratophyllum submersum;* CHL, Chloranthaceae.

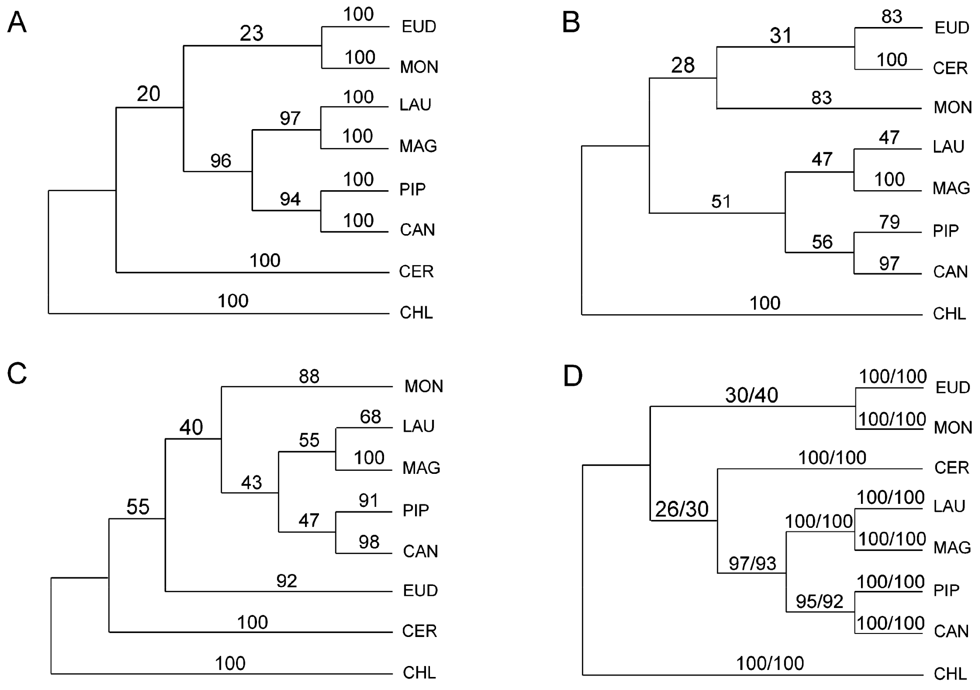


Fig. 3. Schematic presentation of maximum likelihood bootstrap consensus trees of various compatible site matrices. **A,** 8 gene $p^{M<0.1}$ site matrix. **B,** 3 mitochondrial gene $p^{M<0.1}$ site matrix. **C,** 3 mitochondrial gene $p^{M<0.5}$ site matrix. **D,** 4 plastid gene $p^{M<0.1}$ / $p^{M<0.5}$ site matrices. Bootstrap values from analyses of all these matrices are also provided for monophyly of, and relationships within, magnoliids.

Abbreviations: CAN, Canellales; CER, *Ceratophyllum;* CHL, Chloranthaceae; EUD, eudicots; LAU, Laurales; MAG, Magnoliales; MON, monocots; PIP, Piperales.

method we used here.

Because compatibility and parsimony methods are more closely related to each other than either is to likelihood methods (Felsenstein, 2004), it is fair to ask whether uniform increase of bootstrap values on relationships among the key angiosperm lineages in the parsimony analyses, but not the likelihood analyses we performed on the refined data sets, is caused by this factor. On the other hand, parsimony methods have been shown to be robust in analysis of most real world as well as simulated data sets (Hillis et al., 1994; Kolaczkowski & Thornton, 2004). Hence, we will leave it for future studies to determine whether the increase of bootstrap values in these analyses is due to the superior capability conferred by the combined use of the compatibility and parsimony methods to detect the true phylogenetic signal, or the long branch attraction problem of the parsimony method worsened by its closely related cousin.

Some of the relationships reconstructed here have also been obtained earlier by other studies. The close relationship between *Ceratophyllum* and eudicots was seen in three analyses with large data sets (Soltis et al., 2000; Hilu et al., 2003; Qiu et al., 2005), but all with low bootstrap support. Magnoliids were placed as sister to eudicots (not including *Ceratophyllum*) in another study with a large data set, but again with low bootstrap support (Zanis et al., 2002). Chloranthaceae were shown to be sister to monocots with 74%–81% jackknife values (Hilu et al., 2003). Finally, all the relationships reconstructed among the five angiosperm lineages in this study were recovered by both parsimony and likelihood analyses of the 8 gene matrix and most of its various partitions when all characters were included in an earlier study, but all with <50% bootstrap support (Qiu et al., 2006a). It is difficult to assess at present whether agreement of these results from the earlier studies with the ones obtained here can serve as evidence to support a conclusion that the true phylogenetic relationships among the five key angiosperm lineages are correctly reconstructed.

Recently, Moore et al. (2007) reported moderately supported relationships among *Ceratophyllum*, Chloranthaceae, eudicots, magnoliids, and monocots, with a maximum likelihood analysis of 61 plastid genes from 45 seed plants. Eudicots were shown to be sister to *Ceratophyllum*, and this larger lineage was then sister to monocots. The clade of these three lineages was then sister to a clade composed of *Chloranthus* and magnoliids. The parsimony analyses performed in that study consistently failed to recover these relationships. Two factors should be kept in mind when we examine these results. One is that phylogenomic analyses are extremely sensitive to taxon sampling (Stefanovic et al., 2004; Leebens-Mack et al., 2005; Wolf et al., 2005; Qiu et al., 2006b; Lemieux et al., 2007). The other is that all of the genes used in Moore et al. (2007) were from a single organellar genome. It remains to be seen whether these two factors are responsible for the different results obtained in that study and ours here.

2.3 Usefulness of the compatibility method

Did eliminating less compatible characters help reconstruct phylogenetic relationships among the key angiosperm lineages? The answer is largely a positive one in our opinion, as there is a consistent phylogenetic pattern emerging from all performed parsimony analyses and some likelihood analyses (Figs. 2; Fig. 3: B, C), which not only agrees with the one recovered before when no character was eliminated (Qiu et al., 2006a) but also is more strongly supported. Nevertheless, the likelihood analysis results of some matrices are not congruent with this pattern, and the underlying causes of these differences remain to be determined.

The compatibility method we used here adopts a very strict criterion in detecting historical signals for phylogenetic reconstruction. Even though it is related to parsimony methods (Felsenstein, 2004), it is sufficiently different that it deserves to be explored for its usefulness for solving difficult phylogenetic problems, especially when it can help maximize phylogenetic signal retrieval from existing data. Today, molecular systematic studies do have the luxury of gathering a large amount of data because of rapid progress in sequencing technology. However, building large data sets without careful evaluation of the quality of data unnecessarily lowers the efficiency of research, and thus delays resolution of difficult phylogenetic problems. Many large data sets gathered for difficult phylogenetic problems have high levels of homoplasy (e.g., Chase et al., 1993; Qiu et al., 1999, 2005, 2006a, b; Doyle & Endress, 2000; Graham & Olmstead, 2000; Soltis et al., 2000; Hilu et al., 2003; Stefanovic et al., 2004; Leebens-Mack et al., 2005; Wolf et al., 2005; Lemieux et al., 2007). This seems to be where compatibility methods can make a contribution to the solution of difficult phylogenetic problems.

In this study, we specifically examined increase of bootstrap values for the two key nodes, i.e., the close relationships between *Ceratophyllum* and eudicots, and between this larger lineage and magnoliids. These relationships were reconstructed before, when all characters of the 8 genes were analyzed by both parsimony and likelihood methods, but the bootstrap support was low: 49% and 31% respectively (both from a parsimony analysis) (Qiu et al., 2006a). Because the analysis in Qiu et al. (2006a) differed from the ones conducted here in having gymnosperms, *Amborella*, Nymphaeales and Austrobaileyales in the data set, we removed these taxa to generate a data set with identical taxon sampling to the 8 gene $p^{M<0.1}$ and $p^{M<0.5}$ site matrices so that the contribution of both taxon and character removal to the increase of bootstrap values could be partitioned. A parsimony bootstrap analysis of the resulting data set increased bootstrap values from 49% and 31% to 69% and 56% for these two nodes (Fig. 2). Hence, removal of the distantly related taxa did increase bootstrap values, but not to the extent as observed in the compatibility analyses performed in this study. Comparisons of bootstrap values at these two nodes from analyses of the matrix with all characters and the matrices with only COSLACs show that elimination of less compatible characters increases bootstrap values at least by16%–20%, and often more. Therefore, these analyses demonstrate that exclusion of distantly related taxa and elimination of less compatible characters can help increase confidence levels on resolution of difficult phylogenetic problems.

Recently, several other authors have also experimented with identifying and eliminating problematic characters to optimize performance of phylogenetic methods on difficult problems (Brinkmann & Philippe, 1999; Philippe et al., 2000; Burleigh & Mathews, 2004; Pisani, 2004; Gupta & Sneath, 2007). While it may be too early to generalize the usefulness of compatibility methods to help solve difficult phylogenetic problems, the results from this study are certainly encouraging. Hence, we suggest that they should be explored and added to the toolbox of phylogeneticists in the effort to reconstruct the tree of life.

Acknowledgements We thank Bin WANG for help with some analyses and James A. DOYLE for suggestions. YLQ was supported by an Early Career Award (DEB 0332298) and an ATOL grant (DEB 0431239) from the National Science Foundation, USA.

References

Boulter D, Peacock D, Guise A, Gleaves JT, Estabrook G. 1979. Relationships between the partial amino-acid sequences of plastocyanin from members of ten families of flowering plants. Phytochemistry 18: 603–608.

Brinkmann H, Philippe H. 1999. Archaea sister group of bacteria? Indications from tree reconstruction artifacts in ancient phylogenies. Molecular Biology and Evolution 16: 817–825.

Burleigh J, Mathews S. 2004. Phylogenetic signal in nucleotide data from seed plants: implications for resolving the seed plant tree of life. American Journal of Botany 91: 1599–1613.

Camin JH, Sokal RR. 1965. A method for deducing branching sequences in phylogeny. Evolution 19: 311–326.

Chase MW, Soltis DE, Olmstead RG, Morgan D, Les DH, Mishler BD, Duvall MR, Price RA, Hills HG, Qiu Y-L, Kron KA, Rettig JH, Conti E, Palmer JD, Manhart JR, Sytsma KJ, Michaels HJ, Kress WJ, Karol KG, Clark WD, Hedren M, Gaut BS, Jansen RK, Kim K-J, Wimpee CF, Smith JF, Furnier GR, Strauss SH, Xiang Q-Y, Plunkett GM, Soltis PS, Swensen S, Williams SE, Gadek PA, Quinn CJ, Eguiarte LE, Golenberg E, Learn GH Jr, Graham SW, Barrett SCH, Dayanandan S, Albert VA. 1993. Phylogenetics of seed plants: an analysis of nucleotide sequences from the plastid gene *rbc*L. Annals of the Missouri Botanical Garden 80: 528–580.

Cho Y, Mower JP, Qiu Y-L, Palmer JD. 2004. Mitochondrial substitution rates are extraordinarily elevated and variable in a genus of flowering plants. Proceedings of the National Academy of Sciences USA 101: 17741–17746.

Cronquist A. 1981. An integrated system of classification of flowering plants. New York: Columbia University Press.

Day WHE, Estabrook GF, McMorris FR. 1998. Measuring the phylogenetic randomness of biological data sets. Systematic Biology 47: 604–616.

Donoghue MJ, Doyle JA. 1989. Phylogenetic analysis of angiosperms and the relationships of Hamamelidae. In: Crane PR, Blackmore S eds. Evolution, systematics, and fossil history of the Hamamelidae. Vol. 1. Oxford: Clarendon Press. 17–45.

Doyle JA, Endress PK. 2000. Morphological phylogenetic analysis of basal angiosperms: Comparison and combination with molecular data. International Journal of Plant Sciences 161: S121–S153.

Eklund H, Doyle JA, Herendeen PS. 2004. Morphological phylogenetic analysis of living and fossil Chloranthaceae. International Journal of Plant Sciences 165: 107–151.

Estabrook GF. 1972a. Cladistic methodology: a discussion of the theoretical basis for the induction of evolutionary history. Annual Review of Ecology and Systematics 3: 427–456.

Estabrook GF. 1972b. Theoretical concepts in systematic and evolutionary studies. Progress in Theoretical Biology 2: 23–86.

Estabrook GF. 1983. The causes of incompatibility. In: Felsenstein J ed. Numerical taxonomy. NATO ASI Series G. #1. Berlin: Springer-Verlag. 279–295.

Estabrook GF. 1997. Ancestor-descendant relations and incompatible data: motivation for research in discrete mathematics. In: Mirkin B, McMorris FR, Roberts FS, Rzhetsky A eds. Mathematical hierarchies and biology. Providence, Rhode Island: American Mathematical Society. 1–28.

Estabrook GF. 2008. Fifty years of character compatibility concepts at work. Journal of Systematics and Evolution 46: 109–129.

Estabrook GF, Anderson WR. 1978. An estimate of phylogenetic relationships within the genus *Crusea* (Rubiaceae) using character compatibility analysis. Systematic Botany 3: 179–196.

Estabrook GF, Johnson CS, McMorris FR. 1975. An idealized concept of the true cladistic character. Mathematical Biosciences 23: 263–272.

Estabrook GF, Johnson CS, McMorris FR. 1976a. An algebraic analysis of cladistic characters. Discrete Mathematics 16: 141–147.

Estabrook GF, Johnson CS, McMorris FR. 1976b. A mathematical foundation for analysis of cladistic character compatibility. Mathematical Biosciences 29: 181–187.

Estabrook GF, Landrum L. 1975. A simple test for the possible simultaneous divergence of two amino acid positions. Taxon 24: 609–613.

Estabrook GF, McMorris FR. 1977. When are two qualitative taxonomic characters compatible? Journal of Mathematical Biology 4: 195–200.

Estabrook GF, McMorris FR. 1980. When is one estimate of evolutionary relationships a refinement of another? Journal of Mathematical Biology 10: 367–373.

Estabrook GF, Meacham CA. 1980. How to determine the compatibility of undirected character state trees. Mathematical Biosciences 46: 251–256.

Estabrook GF, Strauch JG Jr, Fiala KL. 1977. An application of compatibility analysis to the Blackiths' data on Orthopteroid insects. Systematic Zoology 26: 269–276.

Felsenstein J. 1978. Cases in which parsimony or compatibility methods will be positively misleading. Systematic Zoology 27: 401–410.

Felsenstein J. 1985. Confidence limits on phylogenies—an approach using the bootstrap. Evolution 39: 783–791.

Felsenstein J. 2004. Inferring phylogenies. Sunderland, Massachusetts: Sinauer.

Fitch WM. 1975. Toward finding the tree of maximum parsimony. In: Estabrook GF ed. Proceedings of the Eighth International Conference on Numerical Taxonomy. San Francisco: W. H. Freeman. 189–230.

Friis EM, Crane PR, Pedersen KR. 1986. Floral evidence for Cretaceous chloranthoid angiosperms. Nature 320: 163–164.

Friis EM, Pedersen KR, Crane PR. 1999. Early angiosperm diversification: The diversity of pollen associated with angiosperm reproductive structures in Early Cretaceous floras from Portugal. Annals of the Missouri Botanical Garden 86: 259–296.

Graham SW, Olmstead RG. 2000. Utility of 17 plastid genes for inferring the phylogeny of the basal angiosperms. American Journal of Botany 87: 1712–1730.

Guindon S, Gascuel O. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Systematic Biology 52: 696–704.

Gupta RS, Sneath PHA. 2007. Application of the character compatibility approach to generalized molecular sequence data: Branching order of the proteobacterial subdivisions. Journal of Molecular Evolution 64: 90–100.

Hennig W. 1966. Phylogenetic systematics. Chicago: University of Illinois Press.

Hillis DM, Huelsenbeck JP, Cunningham CW. 1994. Application and accuracy of molecular phylogenies. Science 264: 671–677.

Hilu KW, Borsch T, Müller K, Soltis DE, Soltis PS, Savolainen

V, Chase MW, Powell MP, Alice LA, Evans R, Sauquet H, Neinhuis C, Slotta TAB, Rohwer JG, Campbell CS, Chatrou LW. 2003. Angiosperm phylogeny based on *mat*K sequence information. American Journal of Botany 90: 1758–1776.

Kolaczkowski B, Thornton JW. 2004. Performance of maximum parsimony and likelihood phylogenetics when evolution is heterogeneous. Nature 431: 980–984.

Le Quesne WJ. 1969. A method of selection of characters in numerical taxonomy. Systematic Zoology 18: 201–205.

Leebens-Mack J, Raubeson LA, Cui LY, Kuehl JV, Fourcade MH, Chumley TW, Boore JL, Jansen RK, dePamphilis CW. 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.

Lemieux C, Otis C, Turmel M. 2007. A clade uniting the green algae *Mesostigma viride* and *Chlorokybus atmophyticus* represents the deepest branch of the Streptophyta in chloroplast genome-based phylogenies. BMC Biology 5: 2.

Meacham CA. 1980. Phylogeny of the Berberidaceae with an evaluation of classifications. Systematic Botany 5: 149–172.

Meacham CA. 1981. A probability measure for character compatibility. Mathematical Biosciences 57: 1–18.

Meacham CA. 1983. Theoretical and computational considerations of the compatibility of qualitative taxonomic characters. In: Felsenstein J ed. Numerical taxonomy. NATO ASI Series G. #1. Berlin: Springer-Verlag. 304–314.

Meacham CA. 1984. Evaluating characters by character compatibility analysis. In: Duncan TO, Stuessy TF eds. Cladistics: Perspectives on the reconstruction of evolutionary history. New York: Columbia University Press. 152–165.

Meacham CA. 1994. Phylogenetic relationships at the basal radiation of angiosperms: further study by probability of character compatibility. Systematic Botany 19: 506–522.

Meacham CA, Estabrook GF. 1985. Compatibility methods in systematics. Annual Review of Ecology and Systematics 16: 431–446.

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.

Nei M, Kumar S, Takahashi K. 1998. The optimization principle in phylogenetic analysis tends to give incorrect topologies when the number of nucleotides or amino acids used is small. Proceedings of the National Academy of Sciences USA 95: 12390–12397.

Parkinson CL, Mower JP, Qiu Y-L, Shirk AJ, Song KM, Young ND, dePamphilis CW, Palmer JD. 2005. Multiple major increases and decreases in mitochondrial substitution rates in the plant family Geraniaceae. BMC Evolutionary Biology 5: 73.

Philippe H, Lopez P, Brinkmann H, Budin K, Germot A, Laurent J, Moreira D, Muller M, Le Guyader H. 2000. Early-branching or fast-evolving eukaryotes? An answer based on slowly evolving positions. Proceedings of the Royal Society of London Series B-Biological Sciences 267: 1213–1221.

Pisani D. 2004. Identifying and removing fast-evolving sites using compatibility analysis: An example from the Arthropoda. Systematic Biology 53: 978–989.

Posada D, Crandall KA. 1998. MODELTEST: testing the model of DNA substitution. Bioinformatics 14: 817–818.

Qiu Y-L, Chase MW, Hoot SB, Conti E, Crane PR, Sytsma KJ, Parks CR. 1998. Phylogenetics of the Hamamelidae and their allies: Parsimony analyses of nucleotide sequences of the plastid gene *rbc*L. International Journal of Plant Sciences 159: 891–905.

Qiu Y-L, Li L, Hendry TA, Li R, Taylor DW, Issa MJ, Ronen AJ, Vekaria ML, White AM. 2006a. Reconstructing the basal angiosperm phylogeny: evaluating information content of the mitochondrial genes. Taxon 55: 837–856.

Qiu Y-L, Lee J, Bernasconi-Quadroni F, Soltis DE, Soltis PS, Zanis M, Zimmer EA, Chen Z, Savolainen V, Chase MW. 1999. The earliest angiosperms: evidence from mitochondrial, plastid and nuclear genomes. Nature 402: 404–407.

Qiu Y-L, Dombrovska O, Lee J, Li L, Whitlock BA, Bernasconi-Quadroni F, Rest JS, Davis CC, Borsch T, Hilu KW, Renner SS, Soltis DE, Soltis PS, Zanis MJ, Cannone JJ, Gutell RR, Powell M, Savolainen V, Chatrou LW, Chase MW. 2005. Phylogenetic analysis of basal angiosperms based on nine plastid, mitochondrial, and nuclear genes. International Journal of Plant Sciences 166: 815–842.

Qiu Y-L, Li L, Wang B, Chen Z, Knoop V, Groth-Malonek M, Dombrovska O, Lee J, Kent L, Rest J, Estabrook GF, Hendry TA, Taylor DW, Testa CM, Ambros M, Crandall-Stotler B, Duff RJ, Stech M, Frey W, Quandt D, Davis CC. 2006b. The deepest divergences in land plants inferred from phylogenomic evidence. Proceedings of the National Academy of Sciences USA 103: 15511–15516.

Sneath PHA, Sackin MJ, Ambler RP. 1975. Detecting evolutionary incompatibilities from protein sequences. Systematic Zoology 24: 311–332.

Soltis DE, Soltis PS, Chase MW, Mort ME, Albach DC, Zanis M, Savolainen V, Hahn WH, Hoot SB, Fay MF, Axtell M, Swensen SM, Prince LM, Kress WJ, Nixon KC, Farris JS. 2000. Angiosperm phylogeny inferred from 18S rDNA, *rbc*L, and *atp*B sequences. Botanical Journal of the Linnean Society 133: 381–461.

Stefanovic S, Rice DW, Palmer JD. 2004. Long branch attraction, taxon sampling, and the earliest angiosperms: *Amborella* or monocots? BMC Evolutionary Biology 4: 35.

Swofford DL. 2003. PAUP*4.0b10: Phylogenetic analysis using parsimony. Sunderland, Massachusetts: Sinauer.

Wiley EO. 1981. Phylogenetics: The theory and practice of phylogenetic systematics. New York: John Wiley & Sons.

Wilson EO. 1965. A consistency test for phylogenies based on contemporaneous species. Systematic Zoology 14: 214–220.

Wolf PG, Karol KG, Mandoli DF, Kuehl J, Arumuganathan K, Ellis MW, Mishler BD, Kelch DG, Olmstead RG, Boore JL. 2005. The first complete chloroplast genome sequence of a lycophyte, *Huperzia lucidula* (Lycopodiaceae). Gene 350: 117–128.

Zanis MJ, Soltis DE, Soltis PS, Mathews S, Donoghue MJ. 2002. The root of the angiosperms revisited. Proceedings of the National Academy of Sciences USA 99: 6848–6853.