

Rooting Phylogenetic Trees with Distant Outgroups: A Case Study from the Commelinoid Monocots

Sean W. Graham,* Richard G. Olmstead,† and Spencer C. H. Barrett‡

*Department of Biological Sciences, University of Alberta, Canada; †Department of Botany, University of Washington; and

‡Department of Botany, University of Toronto, Canada

Phylogenetic rooting experiments demonstrate that two chloroplast genes from commelinoid monocot taxa that represent the closest living relatives of the pickerelweed family, Pontederiaceae, retain measurable signals regarding the position of that family's root. The rooting preferences of the chloroplast sequences were compared with those for artificial sequences that correspond to outgroups so divergent that their signal has been lost completely. These random sequences prefer the three longest branches in the unrooted ingroup topology and do not preferentially root on the branches favored by real outgroup sequences. However, the rooting behavior of the artificial sequences is not a simple function of branch length. The random outgroups preferentially root on long terminal ingroup branches, but many ingroup branches comparable in length to those favored by random sequences attract no or few hits. Nonterminal ingroup branches are generally avoided, regardless of their length. Comparisons of the ease of forcing sequences onto suboptimal roots indicate that real outgroups require a substantially greater rooting penalty than random outgroups for around half of the least-parsimonious candidate roots. Although this supports the existence of nonrandomized signal in the real outgroups, it also indicates that there is little power to choose among the optimal and nearly optimal rooting possibilities. A likelihood-based test rejects the hypothesis that all rootings of the subtree using real outgroup sequences are equally good explanations of the data and also eliminates around half of the least optimal candidate roots. Adding genes or outgroups can improve the ability to discriminate among different root locations. Rooting discriminatory power is shown to be stronger, in general, for more closely related outgroups and is highly correlated among different real outgroups, genes, and optimality criteria.

Introduction

The root of a tree or clade represents its first and deepest split, and it therefore provides the arrow of time for polarizing the historical sequence of all subsequent evolutionary events. An incorrectly rooted tree can result in profoundly misleading inferences of taxonomic relationship and character evolution, and so determining the root location with accuracy is a critical component of any phylogenetic analysis. Unfortunately, opportunities for artifactual rootings may exist in many phylogenetic studies, including some of the most critical nodes on the Tree of Life (e.g., Philippe and Forterre 1999; Donoghue and Doyle 2000; Graham and Olmstead 2000), because of the relatively long branches that often connect ingroup and outgroup taxa. When sufficiently long, the outgroup branches can result in spurious rootings (e.g., Felsenstein 1978; Hendy and Penny 1989; Miyamoto and Boyle 1989; Wheeler 1990), and so it is important to choose outgroup taxa that are not too distantly related to the ingroup. The outgroup(s) chosen to root an ingroup need not be from the sister-group (Nixon and Carpenter 1993), but using taxa that are as closely related to the ingroup as possible should reduce any long-branch artifacts, by minimizing the distance between the root node and the first outgroup node (e.g., Wheeler 1990; Maddison, Ruvolo, and Swofford 1992; Smith 1994; Swofford et al. 1996). However, it is also not always obvious what the closest relatives to the

ingroup are, and even where this is known with some confidence, the ingroup may still be quite distantly related to its closest living relatives.

The amount of signal that outgroups provide regarding the root location, and whether this reflects history or spurious long-branch problems, are addressed relatively rarely (see Stiller and Hall 1999; Barkman et al. 2000; Qiu et al. 2001; Huelsenbeck, Bollback, and Levine 2002, for recent examples). One recent theoretical framework, the Relative Apparent Synapomorphy Analysis (RASA)-based methods of Lyons-Weiler, Hoelzer, and Tausch (1996, 1998) and Lyons-Weiler and Hoelzer (1997), provides a tree-independent approach for assessing the amount and quality of rooting signal. However, this general approach has recently attracted some criticism (see Simmons et al. 2002). We do not address RASA-based methods here but instead explore some tree-based methods for examining different rooting possibilities when the unrooted ingroup topology is known with some confidence, as it is with the pickerelweed family, Pontederiaceae. There is substantial ambiguity in the location of the root of Pontederiaceae (Eckenwalder and Barrett 1986; Kohn et al. 1996; Barrett and Graham 1997; Graham et al. 1998), but the monophyly of the family is very robustly supported (Graham and Barrett 1995), unrooted trees of this family are well corroborated by multiple, highly congruent lines of evidence from the chloroplast genome, and they find strong support from bootstrap analysis of the chloroplast data (Graham et al. 1998). Rooting ambiguity thus appears to derive solely from a weakness in the rooting signal provided by the available outgroup sequences.

Recent studies indicate that this family of aquatic monocots has its closest living relatives among the com-

Key words: phylogenetic signal, rooted trees, unrooted trees, outgroup, Pontederiaceae, monocotyledons, long-branch attraction.

Address for correspondence and reprints: Sean W. Graham, Department of Biological Sciences, CW405 Biological Sciences Centre, University of Alberta, Edmonton, Alberta, Canada T6G 2E9. E-mail: swgraham@ualberta.ca.

Mol. Biol. Evol. 19(10):1769–1781. 2002

© 2002 by the Society for Molecular Biology and Evolution. ISSN: 0737-4038

melinoid monocots, a very large and diverse array of taxa that includes the grasses, sedges, gingers, and palms (reviewed in Graham and Barrett 1995). Several subtle nonmolecular characters indicate a relatively close relationship of Pontederiaceae with Haemodora-ceae and Philydraceae (or both) (e.g., Simpson 1987; Steinecke and Hamann 1989; Tillich 1994, 1995; Givnish et al. 1999), whereas various molecular studies suggest that Commelinaceae may be the sister-group to Pontederiaceae and that Hanguanaceae is related to all four families (e.g., Chase et al. 1995, 2000; Givnish et al. 1999). The most recent classification scheme of the monocots (Chase et al. 2000) includes these five families in the order Commelinales, as the sister-group of Zingiberales (the gingers and relatives; see also Stevenson et al. 2000).

However, these higher-order relationships find only weak support from the available data. The five families of the Commelinales are all quite distinct from one another from morphological and molecular perspectives (e.g., Dahlgren, Clifford, and Yeo 1985, pp. 149–150, 323–344, 374–387; Duvall et al. 1993), and estimates of their age based on *rbcL* data suggest that they diverged from one another in the late Cretaceous (Bremer 2000). In the case of Pontederiaceae at least, this may predate substantially the diversification of the extant members of the family (see Barrett and Graham 1997). This evolutionary distinctness may explain why numerous permutations of relationships among the members of Commelinales and relatives have been observed in phylogenetic studies based on different sets of taxa and various molecular and morphological markers (e.g., Chase et al. 1993, 1995, 2000; Graham and Barrett 1995; Davis et al. 1998; Givnish et al. 1999; Stevenson et al. 2000; Neyland 2002). Most of these data sets provide very poor bootstrap support for any particular relationship in the order, and they neither indicate strong support for the sister-group status of any particular taxon to Pontederiaceae nor of any particular root position within the family (Kohn et al. 1996; Barrett and Graham 1997; Graham et al. 1998).

It is therefore of interest to ask how much signal the nearest outgroups provide for rooting Pontederiaceae, compared with those less closely related, and to investigate whether the optimal rootings determined using the nearest outgroups are a consequence of long-branch attraction. A related goal is to clarify the position of Pontederiaceae in monocot phylogeny. Several authors (e.g., Hillis 1996; Graybeal 1998; Soltis et al. 1998; Swofford and Poe 1999) have also noted that adding taxa and characters (or both) to a phylogenetic analysis can improve the accuracy of phylogenetic estimation, and so it would be valuable to address the extent to which adding data (genes or outgroups) improves our ability to assess where the root split of the family lies. We address these questions by sampling multiple outgroup taxa to Pontederiaceae for two chloroplast genes across a broad sample of monocots. Although our study focuses exclusively on this small family of commelinoid monocots, the insights gained from exploring rooting in this family are likely to be broadly applicable to any

phylogenetic study where ambiguity in tree rooting may be a function of distantly related outgroups.

Materials and Methods

Data Source and Matrix Construction

Partial coding sequences for the two chloroplast genes examined (*ndhF* and *rbcL*) were obtained by manual and automated DNA sequencing of PCR products, using methods and primers outlined in Olmstead and Sweere (1994), Graham et al. (1998) and Graham and Olmstead (2000). A total of 1343 bp of DNA sequence was obtained from *rbcL*, and approximately 490 bp of sequence was obtained from around the 3'-end of *ndhF*, the most variable part of this gene (see Graham et al. 1998). The region of *ndhF* sequenced corresponds to bp 1457–1946 of *Oryza sativa ndhF* (GenBank accession number X15901). The *ndhF* and *rbcL* sequences for 24 taxa in Pontederiaceae were obtained by Graham et al. (1998). Most of the *ndhF* sequences for outgroup taxa were obtained for the current study, and most of the other *rbcL* sequences were obtained directly from GenBank or were provided by other workers; collection details and a list of the sequence sources are provided in table 1. Manual sequence alignment was performed using criteria provided in Graham et al. (2000). Alignment gaps were not required for *rbcL*, but sixteen indels were inferred in *ndhF*. These were all parsimony uninformative, apart from a single indel shared by two varieties of *Pontederia cordata*, and another shared by *Acorus*, *Spathiphyllum*, *Sagittaria* and *Gymnostachys*. Alignment gaps were treated as missing data. In five cases (table 1) a composite “placeholder” taxon was represented by sequences from two different species in the same genus. The two genes were considered separately and in concert for the rooting experiments described below.

Inference of Monocot Phylogeny

The local position of Pontederiaceae among the sampled monocot taxa was examined in a maximum parsimony analysis using PAUP* 4, beta versions 4–10 (Swofford 1999). Heuristic searches were performed with all character and character-state changes equally weighted and with “MulTrees” and “Steepest Descent” options activated. To minimize the risk of finding only local optima, 100 random addition replicates were performed (Maddison 1991). Branch support among the outgroup taxa was assessed using bootstrap analysis (Felsenstein 1985), with 100 bootstrap replicates and one random addition sequence per bootstrap replicate.

Rooting Experiments

An unrooted, most-parsimonious topology of 24 taxa of Pontederiaceae derived from three chloroplast data sets (fig. 2 in Graham et al. 1998) was chosen to perform a series of rooting experiments for real and random outgroup sequences using PAUP*. This unrooted tree is one of the ten most-parsimonious trees found using a chloroplast restriction-site data set (Kohn et al.

1996) and is nearly identical in topology to trees inferred from a combined analysis of *ndhF* and *rbcL* (Graham et al. 1998). The same topology was also inferred from parsimony analysis of various combinations of the three chloroplast data sets that included the restriction-site data set, and most branches on it were robustly supported by parsimony bootstrap analyses of the three data sets combined (Graham et al. 1998).

Rooting experiments were performed using various outgroups (both individually and in combination) chosen to represent closely and distantly related taxa in the commelinoid monocots and elsewhere. Artificial outgroup sequences were also constructed to mimic real outgroups that have had their phylogenetic signal completely eroded by the passage of time. These random sequences were generated using MacClade version 3.07 (Maddison, W. P. and Maddison, D. R., 1992), with base frequencies determined from a broad sample of monocots (A, 28%; C, 18%; G, 21%; T, 32%). No significant heterogeneity in base frequencies was detected across pairwise comparisons of the taxa ($\chi^2 = 25.73$, $P = 0.999$, $df = 144$). A reduced two-gene matrix for Pontederiaceae comprising the characters variable in the family was appended to random sequences of the same length. These characters were considered because they all have the potential to be parsimony-informative in the context of data sets that include outgroup taxa.

One set of parsimony analyses was performed to explore the extent to which random outgroup sequences preferentially root on long branches. The unrooted topology of Pontederiaceae that we chose from Graham et al. (1998) was enforced as a backbone constraint in a series of parsimony-based Branch-and-Bound searches that permitted individual random outgroups to attach to one or more optimal root position(s). The frequency of favored rootings on each branch in the unrooted ingroup topology was noted, with fractional scores assigned when a random sequence hit multiple optimal branches.

Variation in the degree of suboptimality across the possible roots for the ingroup topology was also assessed for a variety of real and random outgroup combinations. Tree scores for different rootings were estimated for individual random outgroups (maximum parsimony only) and using a variety of real outgroups (maximum parsimony and likelihood). The 45 possible roots of the 24-taxon ingroup topology were generated using the "All rootings" option in MacClade. The resulting NEXUS text files were edited to include different outgroup possibilities. For the parsimony analyses, tree scores for the different possible root positions were determined with all characters equally weighted and unordered. Models for the likelihood-based rooting analyses were chosen using the hierarchical likelihood ratio tests described in Huelsenbeck and Crandall (1997). The hierarchy of models tested was the same as the example given in Huelsenbeck and Crandall (p. 453; fig. 4), except that (1) the "General time-reversible" (GTR) model was substituted for the HKY85 model used there; (2) a model with a clock was not assessed; and (3) the final paired hypotheses compared here considered models with and without invariable sites.

Model parameters were estimated from the data using the unrooted topology of Pontederiaceae for the combined two-gene data set and also separately for each individual gene. These models and parameter estimates were used in all subsequent likelihood analyses that included the (real) outgroups. The models chosen using the likelihood ratio tests were the "GTR + Γ + I" model for the *rbcL* data set and the combined *rbcL* and *ndhF* data set, and the "GTR + Γ " model for the *ndhF* data set ($P < 0.01$ for all significantly different comparisons with Bonferonni corrections performed; results not shown). The former model accommodates unequal base composition, the proportion of invariable characters (I), and nonuniform substitution rates within and between nucleotide characters (the GTR matrix was used to address uneven character-state transition rates, and the gamma [Γ] shape distribution parameter [α] was used to accommodate among-site rate variation); the latter model differs only by not directly accounting for invariable sites.

Evidence for the existence of historical signal in real outgroups concerning the position of the root of the ingroup tree has come from comparisons among real and random outgroup sequences of the decrease in parsimony observed between the optimal and next best rootings (Miyamoto and Boyle 1989). We extend this logic by examining all suboptimal rootings for random outgroup sequences. For each outgroup sequence, parsimony scores for all possible rootings were compared with that of the optimal root. The resulting rooting penalties required to place the root in suboptimal locations (the score of the suboptimal root location minus score of the optimal root location) were ranked, with tied ranks being broken arbitrarily. The mean and standard deviation in the rooting penalty across the random outgroups for each rank were compared with the penalty for the corresponding rank for the real outgroups. For each real outgroup or outgroup combination considered here, Shimodaira-Hasegawa tests with RELL (resampling estimated log-likelihood) estimates of the test distribution (Shimodaira and Hasegawa 1999 and see Goldman, Anderson, and Rodrigo 2000) were also performed to assess the null hypothesis that all 45 possible rootings of the Pontederiaceae subtree were equally good explanations of the data.

Results

Placement of Pontederiaceae in the Monocots

The phylogeny of the monocots inferred here using two chloroplast genes (fig. 1) indicates relationships that are broadly similar to those seen in other recent studies (e.g., Chase et al. 1995, 2000; Davis et al. 1998). As with those studies, it suffers from poor bootstrap support for the majority of the backbone of the inferred phylogeny. There is good support for most of the sampled monocot orders, as defined by APG (1998) and Chase et al. (2000), within the very coarse limits of our taxon sampling (table 1 and fig. 1). The two sampled members of Asparagales are depicted as the sister-group of the commelinoid monocots. In line with these studies, our

Table 1
Sources of Monocot DNA Sequences Employed in the Current Study

Order (<i>sensu</i> Chase et al. 2000), Exemplar	Family	Voucher (or citation details)	GenBank Accession No.
Acorales			
<i>Acorus calamus</i>	Acoraceae	Denver Botanic Gard., CO, no voucher; RGO 97-149 DNA French 232, CHR8	<i>ndhF</i> (AY007647) <i>rbcL</i> (M91625 ⁹)
Alismatales			
<i>Gymnostachys anceps</i>	Araceae	WJ Kress 42-3417 Howard 4325, FTG	<i>ndhF</i> (AY007653) <i>rbcL</i> (M91629 ⁹)
<i>Sagittaria latifolia</i>	Alismataceae	SCH Barrett s.n., TRT DH Les s.n., CONN	<i>ndhF</i> (AY007657) <i>rbcL</i> (L08767 ⁹)
<i>Spathiphyllum wallisii</i>	Araceae	MW Chase 210, NCU	<i>ndhF</i> (AY007658); <i>rbcL</i> (AJ235807 ⁹)
Arecales			
<i>Caryota mitis</i>	Arecaceae	SW Graham 1020, TRT (Gaut et al. 1992)	<i>ndhF</i> (U79227) <i>rbcL</i> (M81811 ⁹)
Asparagales			
<i>Cyanostrum cordifolium</i>	Cyanostraceae	SW Graham & SCH Barrett 2, TRT	<i>ndhF</i> (U79228); <i>rbcL</i> (U41572 ⁹)
<i>Narcissus elegans</i>	Amaryllidaceae	SCH Barrett 1434, TRT MW Chase 617, K	<i>ndhF</i> (U79216); <i>rbcL</i> (AF116972 ⁹)
Commelinales			
24 taxa	Pontederiaceae	(Voucher information in Kohn et al. 1996)	<i>ndhF</i> (U41598-U41622 ⁹) <i>rbcL</i> (U41573-U41597 ⁹)
<i>Anigozanthos</i> sp.	Haemodoraceae	MG Simpson 6-VI-95A, SDSU	<i>ndhF</i> (U79226)
and <i>A. flavidus</i>		MW Chase 159, NCU	<i>rbcL</i> (AJ404843 ⁹)
<i>Hanguana malayana</i>	Hanguanaceae	Sirirugsa s.n.	<i>ndhF</i> (AY007654)
		PJ Rudall s.n., K	<i>rbcL</i> (AJ404842 ⁹)
<i>Philydrium lanuginosum</i>	Philydreae	SW Graham & SCH Barrett 1, TRT	<i>ndhF</i> (U41622 ⁹); <i>rbcL</i> (U41596 ⁹)
<i>Tradescantia zebrina</i>	Commelinaceae	SW Graham 1030, TRT J Cortesi 7324, GMUF	<i>ndhF</i> (U79229) <i>rbcL</i> (L05042 ⁹)
Dioscoreales			
<i>Dioscorea bulbifera</i>	Dioscoreaceae	EPO Biology, U. Colorado, Boulder, CO, no voucher; RGO 97-151 DNA (Kato et al. 1995)	<i>ndhF</i> (AY007652) <i>rbcL</i> (D28327 ⁹)
<i>Tacca chantrieri</i>	Taccaceae	MW Chase 175, NCU	<i>ndhF</i> (AY007659); <i>rbcL</i> (AJ235810 ⁹)
Pandanales			
<i>Cyclanthus bipartitus</i>	Cyclanthaceae	Hammel 15585	<i>ndhF</i> (AY007650); <i>rbcL</i> (AY007660)
<i>Pandanus utilis</i>	Pandanaceae	Hahn 6898	<i>ndhF</i> (AY007656)
and <i>P. veitchii</i>		Duvall 19920301, UCR	<i>rbcL</i> (M91632 ⁹)
Liliales			
<i>Lilium superbum</i>	Liliaceae	MW Chase 112, NCU	<i>ndhF</i> : (AY007655); <i>rbcL</i> (L12682 ⁹)
Poales			
<i>Ananas comosus</i>	Bromeliaceae	SW Graham 1000, TRT [Duvall et al. 1993]	<i>ndhF</i> (U79225) <i>rbcL</i> (L19977 ⁹)
<i>Carex echinochloe</i>	Cyperaceae	A Muasya 1051, K (A. M. Muasya et al., personal communication)	<i>ndhF</i> (AF191810 ⁹) <i>rbcL</i> (Y12997 ⁹)

Table 1
Continued

Order (<i>sensu</i> Chase et al. 2000), Exemplar	Family	Voucher (or citation details)	GenBank Accession No.
<i>Flagellaria indica</i>	Flagellariaceae	(LG Clark et al., personal communication) MW Chase 206, NCU	<i>ndhF</i> (U22008 ^a) <i>rbcL</i> (L12678 ^a)
<i>Oryza sativa</i>	Poaceae	(Hiratsuka et al. 1989)	Genes from complete chloroplast genome (X15901 ^a)
<i>Typha angustifolia</i>	Typhaceae	SW Graham 1040, TRT	<i>ndhF</i> (U79230)
and <i>T. latifolia</i>		Bradley 24974, GMUF	<i>rbcL</i> (M91634 ^a)
<i>Zea mays</i>	Poaceae	(Maier et al. 1995)	Genes from complete chloroplast genome (X86563 ^a)
Zingiberales			
<i>Canna sp.</i>	Cannaceae	SW Graham 1010, TRT	<i>ndhF</i> (AY007649)
and <i>C. indica</i>		RK Godfrey 57928	<i>rbcL</i> (L. Prince, unpublished sequence)
Unassigned commelinoids			
<i>Calcectasia intermedia</i>	Dasygogonaceae	MW Chase 456, K	<i>ndhF</i> (AY007648); <i>rbcL</i> (Chase et al. 1995, no number ^a)
<i>Dasygogon hookeri</i>	Dasygogonaceae	MW Chase 430	<i>ndhF</i> (AY007651)
and <i>D. bromelifolius</i>		PJ Rudall s.n., K	<i>rbcL</i> (Chase et al. 1995, no number ^a)

^a Previously published sequences.

data support the recent inclusion (APG 1998) of Dasygogonaceae and Hanguanaceae within the commelinoid monocots, with moderate bootstrap support (72%). The former family is poorly supported as the sister-group of Zingiberales, the latter as the sister-group of Commelinaceae in the Commelinales. As in other studies with broader taxon and gene sampling, we did not infer strongly supported relationships of taxa at the base of the commelinoid monocots, including those involving the palms, pineapples and cattails (Arecaceae, Bromeliaceae, and Typhaceae, respectively). The order Commelinales, which includes Pontederiaceae and four other families in the most recent ordinal treatment of Chase et al. (2000), finds further corroboration here, but with poor support (33%). Philydraceae, represented here by *Philydrum lanuginosum*, is resolved as the sister-group of Pontederiaceae, with only 38% bootstrap support.

Long Branches and the Root of Pontederiaceae

Pontederiaceae is strongly supported as monophyletic (fig. 1), as was found previously by Graham and Barrett (1995). Relationships within the family in the analysis of monocot phylogeny are essentially identical to those found by Graham et al. (1998) using the same genes and fewer outgroup taxa. Most branches within the family are well supported by bootstrap analysis when outgroups are excluded from analysis (see Graham et al. 1998, fig. 2). However, the optimal position of the root of the family in the parsimony search, between *Heteranthera-Hydrothrix* and the other taxa in the family, is poorly supported by bootstrap analysis (fig. 1; support values excluded for most branches above the root in Pontederiaceae for clarity—see Graham et al. 1998 for bootstrap values), as in previous studies of the family with fewer outgroup taxa.

More than 70% of the “hits” by random outgroups were on the three longest branches within the unrooted ingroup, which range in length from 16–20 steps under ACCTRAN optimization (these three branches attracted 16%, 20% and 35% of the hits, respectively; fig. 2). Wheeler (1990) predicted that the probability of a random sequence rooting on a given ingroup branch should be proportional to its length. This relationship holds for the terminal branches in figure 2 (correlation between length and number of hits: $r = 0.743$, $P < 0.001$, $df = [24 \text{ terminal branches} - 2] = 22$). However, the relationship breaks down completely for the nonterminal branches ($r = 0.331$, $P = 0.142$, $df = [21 - 2] = 19$). Most nonterminal branches that were of the same order of length as the longest three had few or no hits from the random sequences sampled here (fig. 2). These disfavored branches include the one preferred by the real outgroups under maximum parsimony (fig. 1) and the neighboring branch favored under maximum likelihood for most outgroup combinations (see below). These two branches have a parsimony-based length of 15 and 6 steps, respectively, on the unrooted tree under ACCTRAN optimization. A range of other branches were also favored as optimal rooting locations by the random outgroup sequences, but each of these accounted for less

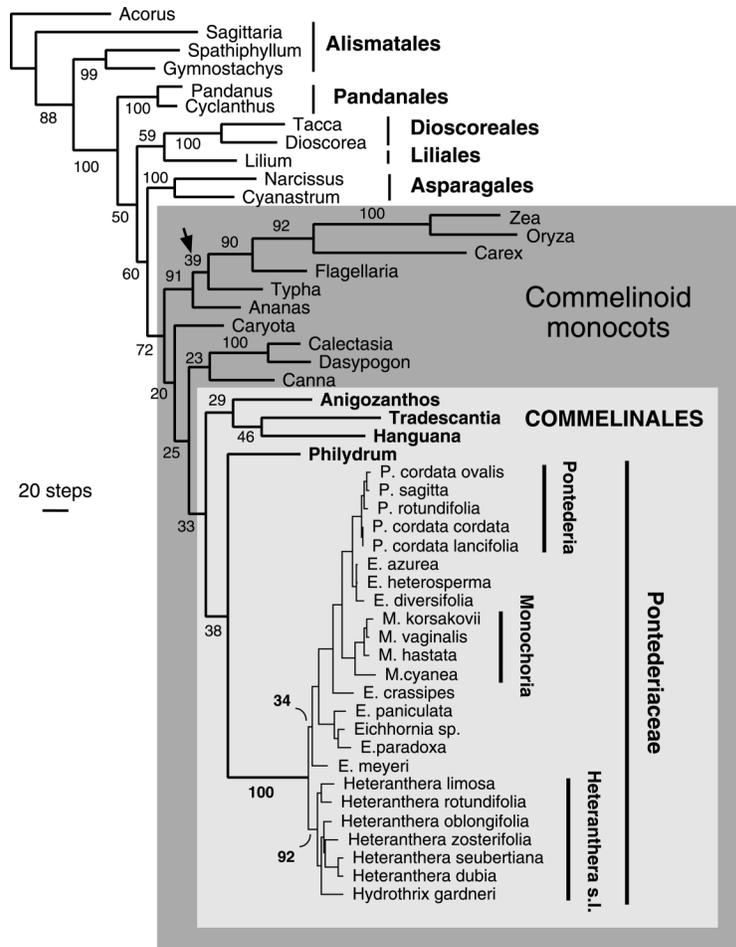


FIG. 1.—Phylogeny of the monocotyledons based on parsimony analysis of *ndhF* and *rbcL* sequence data combined. The tree (one of eight most-parsimonious) is represented as a phylogram (length = 2743 steps; consistency index = 0.457; retention index = 0.632) with branch lengths computed using ACCTRAN optimization. Two of three branches not found across all most-parsimonious trees involve short branches in Pontederiaceae (not shown); a third (indicated with an arrow) involves two different arrangements near the base of the commelinoid monocot clade. Bootstrap values are indicated beside branches but are excluded from all branches in Pontederiaceae, except for the two branches around the root. Taxon names follow Chase et al. (2000).

than 5% of the total hits. In total, less than 10% of the branches favored as root locations by random outgroups occurred on nonterminal branches on the unrooted topology, despite their representing more than 40% of the total tree length and comprising half of the top ten branch lengths on the unrooted topology under ACCTRAN optimization.

Rooting Experiments

Using the two-gene data set, parsimony- and likelihood-based methods largely concurred on the extent of signal in the outgroup taxa for rooting the ingroup when the four closest outgroup taxa were used to root the ingroup tree (figs. 3 and 4; note that the outgroup subtree of Commelinales taxa used is that implied in fig. 1). Rooting penalties were highly correlated between the parsimony and likelihood optimality criteria, at least for the case considered here (the four Commelinales taxa and both genes combined; table 2). Most of the outgroups considered in table 2 had the same optimal root location under maximum likelihood (indicated in figs. 2

and 5), but for several outgroups where this was not the optimal root under maximum likelihood it was only marginally suboptimal (results not shown). This root was also only marginally suboptimal for the maximum parsimony case considered in table 2 (three steps longer, the third best root under parsimony).

A stronger parsimony penalty for real outgroups compared with random outgroups was used by Miyamoto and Boyle (1989) as evidence of rooting signal, although they focused only on the optimal and first suboptimal root locations. Here, the parsimony penalties required to force the root to the twenty-second ranked (and worse) rootings for the random outgroup sequences were on average more than two standard deviations closer to optimal length than those observed using the four real outgroups together (fig. 3). However, *Acorus* was not well distinguished from random sequences (fig. 3). The other real outgroups considered here (table 2) largely performed between these two extremes. One intermediate example, a composite outgroup involving Commelinaceae and Hanguanaceae, is shown in figure 3.

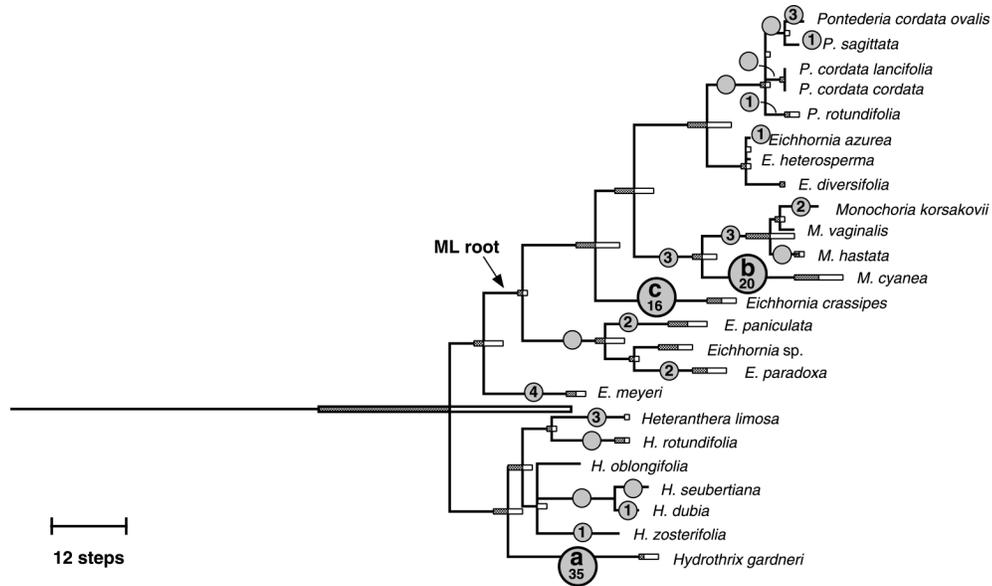


FIG. 2.—Propensity for random outgroups to root on different branches within Pontederiaceae. Branch lengths for Pontederiaceae were calculated from the combined *ndhF* and *rbcL* data; the bars at each node show the minimum, average, and maximum branch lengths across the most-parsimonious reconstructions. Circled values indicate the percentage of optimal rootings by 100 random outgroups on the unrooted topology of Pontederiaceae. Fractional values were applied to random outgroups that rooted on multiple branches. Empty circles indicate that some hits were found, but with a frequency <0.5%. The three labeled values (a, b, and c) are the top hits by the random outgroups. The rooting is the optimal one under parsimony; the optimal root in various maximum likelihood analyses (see text) is also indicated. Outgroup taxa were excluded from the figure for clarity; the branch subtending the ingroup is from the root of Pontederiaceae to the first outgroup node only.

The same general patterns held in the likelihood analyses. Of those outgroup combinations examined, the experiment involving the composite Commelinales outgroup (Commelinaceae, Philydraceae, Haemodoraceae and Hanguanaceae) rejected a substantial fraction of suboptimal root locations (20 of 44 suboptimal roots; figs. 4 and 5; table 2). In general, the roots rejected by the different outgroup permutations considered here

were also subsets of those rejected using the composite Commelinales outgroups and both genes combined (bracketed values in table 2). Some other outgroups performed as well as this or better (summarized in figs. 4 and 5; table 2), including one of the four Commelinales taxa when used individually as an outgroup family (*Anigozanthos*, representing Haemodoraceae). For the other three families of Commelinales (Commelinaceae, Han-

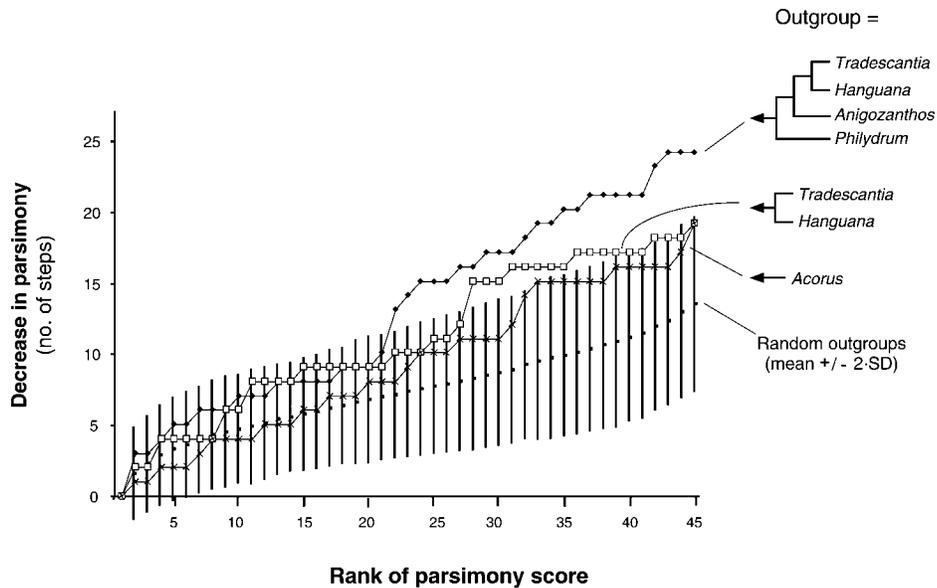


FIG. 3.—Ranked decrease in parsimony when different outgroups (real and random) are forced to suboptimal root locations in Pontederiaceae. The real outgroups considered are ([*Philydrum*, [*Anigozanthos*, [*Hanguana*, *Tradescantia*]]]), ([*Tradescantia*, *Hanguana*]), and (*Acorus*). All cases are ranked independently. For the random outgroups, the mean penalty (plus or minus two standard deviations) was calculated across tree scores at each rank.

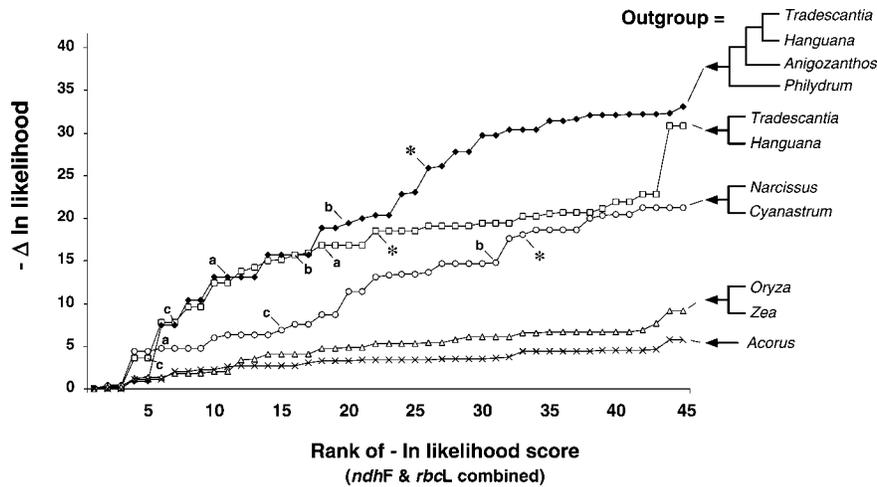


FIG. 4.—Ranked decrease in likelihood when various outgroup combinations are forced to suboptimal root locations in Pontederiaceae. All cases are ranked independently. Asterisks indicate the $-\ln L$ scores above which roots are significantly suboptimal for that outgroup set, as inferred using a series of Shimodaira-Hasegawa tests (table 2 and fig. 5). The top three branches favored by the random outgroups under the parsimony criterion (labels a, b, and c; fig. 2) also are indicated for three of the outgroup sets.

guanaceae and Philydraceae, represented by *Tradescantia*, *Hanguana* and *Philydrum*, respectively), only a handful of candidate roots in Pontederiaceae could be rejected using the exemplar genera individually as outgroups (table 2). However, in combination, the exemplars from these families can reject as many or more branches as the four outgroups combined: one example, the combination of Commelinaceae and Hanguanaceae, is shown in figures 4 and 5 and table 2.

Two of the most distantly related outgroup cases examined (*Acorus* by itself and the two representatives of Poaceae considered together) were not able to reject any root candidates (table 2 and fig. 4). The two Poaceae

taxa are on relatively long branches within the commelinoid monocots (fig. 1), and *Acorus* is presumed to be the sister-group of the rest of the monocots here. However, one noncommelinoid outgroup case considered here (the two representatives of Asparagales taken together) was able to reject a substantial fraction of the root locations (13 of 20) rejected by the four outgroups used together (table 2 and fig. 4).

When the rooting experiment with the four Commelinales taxa was repeated using individual genes, a smaller subset of these candidate roots could be rejected than was possible with the two genes considered together (table 2). The penalties observed using each gene

Table 2
Score Correlation^a and Ability to Discriminate Suboptimal Roots of Pontederiaceae for Various Outgroups, Data Sets and Optimality Criteria

Outgroup ^b	Optimality Criterion	Data Set	r^c	No. of Significantly Suboptimal Roots ^d
<i>Acorus</i> (Acoraceae)	ML	<i>ndhF</i> + <i>rbcL</i>	0.866	0
Poaceae ^e	ML	<i>ndhF</i> + <i>rbcL</i>	0.843	0
Asparagales ^e	ML	<i>ndhF</i> + <i>rbcL</i>	0.922	13 (13)
<i>Anigozanthos</i>	ML	<i>ndhF</i> + <i>rbcL</i>	0.935	20 (18)
<i>Philydrum</i>	ML	<i>ndhF</i> + <i>rbcL</i>	0.898	5 (3)
<i>Hanguana</i>	ML	<i>ndhF</i> + <i>rbcL</i>	0.782	4 (4)
<i>Tradescantia</i>	ML	<i>ndhF</i> + <i>rbcL</i>	0.721	6 (6)
(<i>Han</i> , <i>Tra</i>)	ML	<i>ndhF</i> + <i>rbcL</i>	0.797	24 (18)
(<i>Phi</i> , [<i>Ani</i> , [<i>Han</i> , <i>Tra</i>]])	MP (Fitch)	<i>ndhF</i> + <i>rbcL</i>	0.942	n/a
(<i>Phi</i> , [<i>Ani</i> , [<i>Han</i> , <i>Tra</i>]])	ML	<i>rbcL</i>	0.867	2 (2)
(<i>Phi</i> , [<i>Ani</i> , [<i>Han</i> , <i>Tra</i>]])	ML	<i>ndhF</i>	0.947	14 (14)

^a Correlation (across root positions) of scores with those seen using the two-gene data set, the outgroup combination (*Phi*, [*Ani*, [*Han*, *Tra*]]), and the GTR + Γ + I likelihood model. Twenty root positions were worse than the optimal root for the latter case.

^b Abbreviations: *Ani* = *Anigozanthos* (Haemodoraceae); *Han* = *Hanguana* (Hanguanaceae); *Phi* = *Philydrum* (Philydraceae); *Tra* = *Tradescantia* (Commelinaceae); ML = Maximum likelihood; and MP = Maximum parsimony.

^c $P < 0.001$. $df = (45 \text{ ingroup branches} - 2) = 43$. Bonferonni corrections were performed for these correlation tests.

^d Assessed using Shimodaira-Hasegawa tests (see text). Numbers in parentheses indicate root locations that are rejected by that case and also by the case with the two-gene data set, the outgroup combination (*Phi*, [*Ani*, [*Han*, *Tra*]]), and the GTR + Γ + I likelihood model.

^e Poaceae = (*Oryza*, *Zea*); Asparagales = (*Cyanastrum*, *Narcissus*).

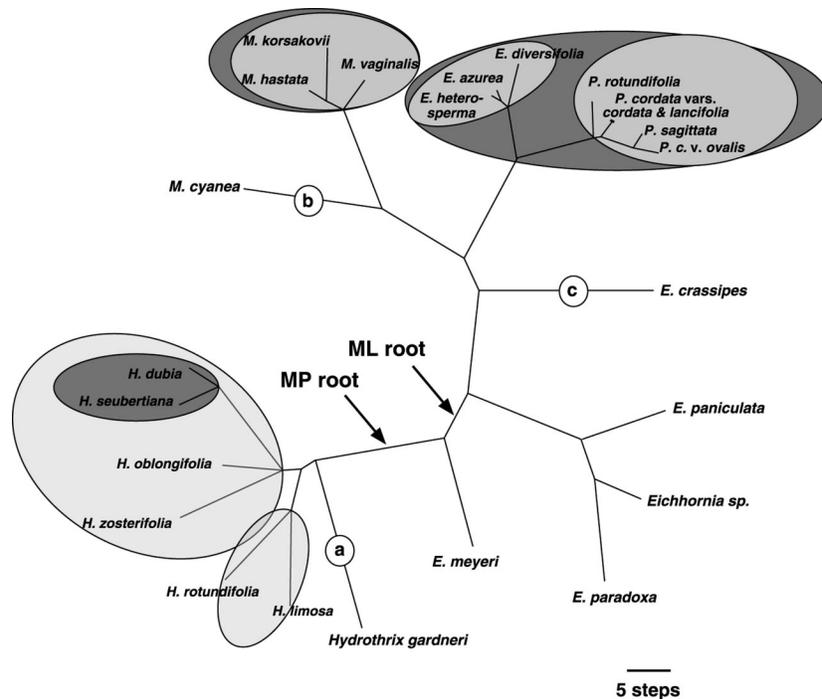


FIG. 5.—Branches that are significantly suboptimal in a likelihood analysis (Shimodaira-Hasegawa tests; $P < 0.05$; table 2 and fig. 4) of the combined DNA sequence data when outgroups are forced to different root locations in Pontederiaceae. The three dark circles indicate locations rejected when the composite outgroup (*Philydrum*, [*Anigozanthos*, [*Tradescantia*, *Hanguana*]]) was employed; the five lighter circles indicate those rejected when the composite outgroup (*Tradescantia*, *Hanguana*) was employed; branches in the intersection of circles were rejected in both instances. Branch lengths are parsimony-based. The optimal parsimony- and likelihood-based roots for the case with the four outgroup taxa, and the three branches favored by the majority of random outgroups with maximum parsimony analysis (labels a, b, and c; fig. 2) also are indicated.

individually were nonetheless highly correlated to those observed using both genes together (table 2). A similarly strong correlation to this case was also seen with regard to rooting penalties for all permutations of outgroup taxa, genes, and optimality criteria considered. Thus, this correlation was apparent even where no or few roots could be rejected using likelihood (e.g., for the four Commelinales considered together but using *rbcL* alone, or for *Acorus* or Poaceae using both genes; table 2). In comparison to other outgroups, the slopes of the likelihood-penalty surface were flattest for the latter two cases (Poaceae and *Acorus*; fig. 4).

Discussion

Monocot Phylogeny and the Placement of Pontederiaceae

Our data corroborate the broad pattern of phylogenetic relationships indicated in recent studies (e.g., Givnish et al. 1999; Chase et al. 2000; Stevenson et al. 2000). Pontederiaceae belongs in the commelinoid monocots, near Commelinaceae, Haemodoraceae, Hanguanaceae and Philydraceae. As was found in these earlier studies, the precise interrelationships of these five families to one another was poorly supported.

Signal and Bias in the Closest Relatives of Pontederiaceae for Rooting the Family

Poor support for the root of Pontederiaceae is not a function of uncertain relationships within the family

because the two ingroup branches created at the root split using the real outgroups (fig. 1) correspond to a single very strongly supported branch in the unrooted version of the tree (Graham et al. 1998). This lack of robust rooting support is similar to that observed with random outgroups. To demonstrate this, a subset of random outgroup sequences were employed as outgroups in parsimony-based bootstrap analyses. For seven out of 10 random sequences examined, no single ingroup branch was favored by $\geq 50\%$ of bootstrap replicates (results not shown). Support for the remaining three cases was in the 60%–85% range, representing only moderately robust rootings by random outgroups. Thus, although long-branch attraction may influence rooting decisions in Pontederiaceae (fig. 2), it is unlikely to result in a robustly supported wrong answer, at least at the current level of nucleotide sampling.

The comparable bootstrap support for rootings seen with real and random outgroups begs the question of whether any signal remains in the real outgroups for rooting Pontederiaceae. Comparisons of the degree of suboptimality of correspondingly ranked alternative rootings for real and random outgroup sequences demonstrate that the real outgroups do possess significant historical signal (fig. 3). The idea of investigating the various rooting possibilities of outgroups on unrooted trees goes back to Lundberg (1972). This approach was recommended by Nixon and Carpenter (1993) and Swofford et al. (1996) when it is suspected that long

branches connecting the ingroup and outgroups may distort the estimation of phylogenetic relationships within the ingroup, above and beyond uncertainty over the point of root attachment. Miyamoto and Boyle (1989) were the first to explore the rooting potential of random outgroups (using “Lundberg rooting experiments”) as a baseline for comparing the behavior of divergent real sequences. They cite Rohlf and Fisher (1968) in this regard, but the latter authors used random sequences for a quite different purpose. Recent studies that examine the rooting potential of random versus real sequences include Sullivan and Swofford (1997) and Stiller and Hall (1999). The parsimony-based test used here to assess rooting signal is based on the one suggested by Miyamoto and Boyle (1989), but we had to examine rootings substantially less optimal than the first suboptimal root to demonstrate phylogenetic signal among the real outgroups for rooting Pontederiaceae (fig. 3).

Likelihood-based explorations of the possible rootings (table 2 and figs. 4 and 5) further confirm that many candidate root locations in the family are significantly poorer explanations of the data than the optimal cases (figs. 4 and 5). Both parsimony- and likelihood-based methods using the four closest outgroup families in tandem suggest that roughly half of all possible root locations within the family can be rejected with confidence by these outgroups.

Several lines of reasoning lend credence to the optimal roots favored by the real outgroups, despite their lack of robust support by bootstrap analysis or the Shimodaira-Hasegawa tests. First, the optimal roots favored by the real outgroups under maximum parsimony and likelihood (figs. 2 and 5) are preferred by none of the random outgroups considered here (fig. 2). Second, although real outgroups do not reject the three roots favored by most of the random outgroups, at an alpha level of 0.05, these rooting possibilities are generally not in the top 20% of rooting candidates favored by real outgroups (fig. 4). Third, a substantial fraction of the homoplasy accumulated independently in different outgroups (that fraction of homoplasy not shared through common ancestry) should erode signal concerning the root of Pontederiaceae but in different ways in each terminal outgroup lineage. Although the signal that we detected was weak (sometimes to the point of not being able to reject any alternative rooting position), the pattern of rooting preference was still highly consistent among real outgroups (see the correlations listed in table 2), despite these substantial and presumably independent opportunities in each lineage for the accumulation of misinformative characters (e.g., Swofford et al. 1996), and for the loss of informative characters due to multiple hits. Because there has been little detectable erosion in the rooting correlations due to substitutions on the terminal Commelinales branches, it seems reasonable to assume that the rooting signal in toto among these outgroups has not been substantially biased by long-branch effects in the sense of Felsenstein (1978) and Hendy and Penny (1989).

One previously published root of Pontederiaceae based on a restriction-site data set (Kohn et al. 1996)

was on a branch favored by many of the random outgroups (labeled “c” in fig. 2). The possibility that this was an artifactual rooting was noted by Kohn et al. It has been suggested that restriction-site data generally have minimal utility at deeper levels of phylogenetic analysis (Olmstead and Palmer 1994; Soltis and Soltis 1998), and this may be the case with regard to outgroup-based rooting of Pontederiaceae. For the purposes of the evolutionary reconstructions that Kohn et al. were examining, their overall results did not differ markedly from those performed using the root indicated with the DNA sequence data (Kohn et al. 1996; Barrett and Graham 1997). However, even if root estimation has been nudged only slightly off course, this may reduce the utility of the phylogenetic inference in other reconstructions of character evolution, or for other purposes, such as the generation of classifications that better reflect phylogeny. Pursuing a more confident inference of root placement for Pontederiaceae is thus a crucial goal for future study.

Future Prospects

Using the four outgroup representatives in Commelinales as a composite outgroup was more useful for rejecting suboptimal roots than most of these outgroup families used individually (table 2). For example, when exemplars representing Commelinaceae or Hanguanaceae were used individually, we were able to reject only a few suboptimal root positions. In combination however, the discriminatory power of these two families was among the very best of those examined (table 2 and figs. 4 and 5). Breaking long branches by further addition of taxa should therefore improve our ability to infer the root of Pontederiaceae. This general result has been noted by many workers (e.g., Graybeal 1998; Hillis 1998). However, Smith (1994; and see Hendy and Penny 1989) pointed out that it is better to have a denser representation of outgroups in the sister-group than to sample heavily in less closely related taxa. Although we agree in principle that multiple taxa should be sampled in the sister-group (where practical or possible), sampling somewhat more distantly related taxa is important too because this should help break up the branch between the ingroup root and the sister-group.

Sampling outgroups beyond the sister-group serves a further purpose—it helps test the idea that the sister-group really is the sister-group. It may not always be clear in advance what the sister-group is, as was the case here. The sequences examined here for the exemplar of Commelinaceae, the family suggested by previous molecular studies to be the sister-group of Pontederiaceae (e.g., Chase et al. 1995) retained little robust signal when used individually to root the family, although they performed very well in combination with another poorly performing outgroup family (Hanguanaceae). Before we can add a denser sampling of taxa within the sister-group and other close relatives of Pontederiaceae, we need to more confidently identify the sister-group of the family. An ongoing study of commelinoid monocot phylogeny based on multiple genes is aimed at addressing

this question. Preliminary results (S. W. Graham, unpublished data) provide extremely strong support for Haemodoraceae as the sister-group of Pontederiaceae (and intriguingly, the single representative of this family examined in the current study, *Anigozanthos*, rejected as many root positions as the four Commelinales outgroups combined; table 2). Adding further genes should further improve our ability to reject suboptimal roots, as the combined discriminatory power of two genes was substantially greater than for either gene individually, in terms of the number of ingroup roots that could be rejected with confidence (table 2).

Future taxon sampling within Pontederiaceae should focus on sampling those branches that could not be rejected with confidence. Currently unsampled members of Pontederiaceae are likely to place in sectors of the tree that have been rejected as rooting candidates here (see Barrett and Graham 1997), and so these unsampled lineages are unlikely to include the point of attachment of the root. It is noteworthy that the candidate branches that cannot be rejected with confidence as points of attachment for the root of Pontederiaceae are largely restricted to the interior backbone of the unrooted tree. This backbone is generally strongly disfavored by random outgroup roots, suggesting that if the true root of a clade is to be found on its nonterminal branches (as may be the case here), long-branch attraction is unlikely to generate an artifactual rooting that suggests otherwise. The tendency observed here for random long-branches to disfavor interior parts of the unrooted tree, and to have no rooting preference on the interior backbone based on branch length, has not been noted previously. It would be valuable to explore this property in other phylogenies to determine if it is a widespread phenomenon.

Analytical Simplifications and Consequences

To facilitate tree-score estimation in a reasonable time frame, parameters for the likelihood model were estimated using the Pontederiaceae subtree only, and these values were used in all subsequent likelihood analyses. To investigate whether it would make a substantial difference to estimate model parameters directly, the Shimodaira-Hasegawa analyses were repeated for one of the outgroup cases (*Philydrum*, with both genes considered simultaneously), but with the "GTR + Γ + I" model parameters estimated separately for each root position. Estimated model parameters (not shown) were neither found to differ substantially across rootings nor to those estimated using the Pontederiaceae subtree alone. The number and identity of significantly suboptimal roots according to the Shimodaira-Hasegawa tests were also found to be very similar to those obtained when parameters were derived from the Pontederiaceae subtree alone (results not shown). The use of this analytical shortcut therefore likely has little or no effect on our overall conclusions.

Appropriate corrections to significance levels are made in Shimodaira-Hasegawa's test to account for the multiple comparisons being performed (Goldman, An-

derson, and Rodrigo 2000), in our case across the 45 possible root positions. However, multiple sets of Shimodaira-Hasegawa tests were performed using different outgroup combinations, and it would also be valuable to correct the significance levels to take account of this level of hypothesis testing. However, because tests using related outgroups are likely to be strongly correlated (table 2), adjusting the alpha-level to account for multiple tests is probably not appropriate, and it is not clear what the appropriate correction would be.

Goldman, Anderson, and Rodrigo (2000) emphasize the importance of an honest choice of a priori hypothesis topologies to the conclusions generated using the Shimodaira-Hasegawa test. We erred on the side of being conservative here and considered all possible rooted versions of the unrooted topology of Pontederiaceae. However, the test assumes that the topologies considered include the true one (Goldman, Anderson, and Rodrigo 2000), and our results should be viewed with the caveat that although the true, unrooted chloroplast phylogeny of Pontederiaceae is likely very similar in shape to the unrooted topology considered here, it is not guaranteed to be identical to it.

Conclusions

Incorrectly rooted trees may result in profoundly misleading evolutionary and taxonomic inferences, and this may be a relatively widespread phenomenon in phylogenetic studies. The approaches presented here may be useful in any study where distantly related outgroups may lead to artifactual or ambiguous rootings of the ingroup subtree. For the case study examined here, the available data do not yet permit a conclusive rooting of Pontederiaceae, but the general pattern of rooting preferences for the outgroups in the commelinoid monocots (and beyond) has apparently not been degraded, even in the face of the substantial erosion of phylogenetic signal on long outgroup branches. Our results highlight some general areas for future research, including the different rooting behavior of random outgroup sequences on terminal versus nonterminal ingroup branches. The variety of approaches employed here concur on the nature and strength of the signal in real, distant outgroups for inferring the position of the root of Pontederiaceae, and the ability to discriminate against suboptimal root locations is shown to be substantially improved by adding outgroup taxa and characters.

Acknowledgments

We thank Michael Donoghue for bringing Miyamoto and Boyle's work on tree rooting to our attention, Mark Chase, William Hahn, John Kress, Thomas Lemieux, Linda Prince, Michael Simpson, Alan Yen, the man from Del Monte and others for access to plant material or unpublished sequences, and two reviewers for helpful suggestions on the manuscript. This work was funded in part by NSF grant DEB 9727025 to R.G.O. and by research grants to S.C.H.B. and S.W.G. from NSERC.

LITERATURE CITED

- APG (ANGIOSPERM PHYLOGENY GROUP). 1998. An ordinal classification for the families of flowering plants. *Ann. Mo. Bot. Gard.* **85**:531–553.
- BARKMAN, T. J., G. CHENERY, J. R. MCNEAL, J. LYONS-WEILER, W. J. ELLISENS, G. MOORE, A. D. WOLFE, and C. W. DEPAMPHILIS. 2000. Independent and combined analyses of sequences from all three genomic compartments converge on the root of flowering plant phylogeny. *Proc. Natl. Acad. Sci. USA* **97**:13166–13171.
- BARRETT, S. C. H., and S. W. GRAHAM. 1997. Adaptive radiation in the aquatic plant family Pontederiaceae: insights from phylogenetic analysis. Pp. 225–258 in T. J. GIVNISH and K. J. SYTSMAS, eds. *Molecular evolution and adaptive radiation*. Cambridge University Press, Cambridge, U.K.
- BREMER, K. 2000. Early Cretaceous lineages of monocot flowering plants. *Proc. Natl. Acad. Sci. USA* **97**:4707–4711.
- CHASE, M. W., M. R. DUVAL, H. G. HILLS et al. (11 co-authors). 1995. Molecular phylogenetics of Liliaceae. Pp. 109–137 in P. J. RUDALL, P. J. CRIBB, D. F. CUTLER, and C. J. HUMPHRIES, eds. *Monocotyledons: systematics and evolution*. Royal Botanic Gardens, Kew, U.K.
- CHASE, M. W., D. E. SOLTIS, R. G. OLMSTEAD et al. (43 co-authors). 1993. Phylogenetics of seed plants: an analysis of nucleotide sequences from the plastid gene *rbcl*. *Ann. Mo. Bot. Gard.* **80**:528–580.
- CHASE, M. W., D. E. SOLTIS, P. S. SOLTIS et al. (13 co-authors). 2000. Higher-level systematics of the monocotyledons: an assessment of current knowledge and a new classification. Pp. 1–24 in K. L. WILSON and D. A. MORRISON, eds. *Monocots: systematics and evolution*. CSIRO, Melbourne.
- DAHLGREN, R. M. T., H. T. CLIFFORD, and P. F. YEO. 1985. *The families of the monocotyledons: structure, evolution and taxonomy*. Springer-Verlag, Berlin.
- DAVIS, J. I., M. P. SIMMONS, D. W. STEVENSON, and J. F. WENDEL. 1998. Data decisiveness, data quality, and incongruence in phylogenetic analysis: an example from the monocotyledons using mitochondrial *atpA* sequences. *Syst. Biol.* **47**:282–310.
- DONOGHUE, M. J., and J. A. DOYLE. 2000. Seed plant phylogeny: demise of the anthophyte hypothesis? *Curr. Biol.* **10**:R106–R109.
- DUVAL, M. R., M. T. CLEGG, M. W. CHASE et al. (11 co-authors). 1993. Phylogenetic hypotheses for the monocotyledons constructed from *rbcl* sequencing data. *Ann. Mo. Bot. Gard.* **80**:607–619.
- ECKENWALDER, J. E., and S. C. H. BARRETT. 1986. Phylogenetic systematics of Pontederiaceae. *Syst. Bot.* **11**:373–391.
- FELSENSTEIN, J. 1978. Cases in which parsimony and compatibility methods will be positively misleading. *Syst. Zool.* **27**:401–410.
- . 1985. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* **39**:783–791.
- GAUT, B. S., S. V. MUSE, W. D. CLARK, and M. T. CLEGG. 1992. Relative rates of nucleotide substitution at the *rbcl* locus of monocotyledonous plants. *J. Mol. Evol.* **35**:292–303.
- GIVNISH, T. J., T. M. EVANS, J. C. PIRES, and K. J. SYTSMAS. 1999. Polyphyly and convergent morphological evolution in Commelinales and Commelinidae: evidence from *rbcl* sequence data. *Mol. Phyl. Evol.* **12**:360–385.
- GOLDMAN, N., J. P. ANDERSON, and A. G. RODRIGO. 2000. Likelihood-based tests of topologies in phylogenetics. *Syst. Biol.* **49**:652–670.
- GRAHAM, S. W., and S. C. H. BARRETT. 1995. Phylogenetic systematics of Pontederiales: implications for breeding-system evolution. Pp. 415–441 in P. J. RUDALL, P. J. CRIBB, D. F. CUTLER, and C. J. HUMPHRIES, eds. *Monocotyledons: systematics and evolution*. Royal Botanic Gardens, Kew, U.K.
- GRAHAM, S. W., J. R. KOHN, B. R. MORTON, J. E. ECKENWALDER, and S. C. H. BARRETT. 1998. Phylogenetic congruence and discordance among one morphological and three molecular data sets from Pontederiaceae. *Syst. Biol.* **47**:545–567.
- GRAHAM, S. W., and R. G. OLMSTEAD. 2000. Utility of 17 chloroplast genes for inferring the phylogeny of the basal angiosperms. *Am. J. Bot.* **87**:1712–1730.
- GRAHAM, S. W., P. A. REEVES, A. C. E. BURNS, and R. G. OLMSTEAD. 2000. Microstructural changes in noncoding chloroplast DNA: interpretation, evolution, and utility of indels and inversions in basal angiosperm phylogenetic inference. *Int. J. Plant Sci.* **161**:S83–S96.
- GRAYBEAL, A. 1998. Is it better to add taxa or characters to a difficult phylogenetic problem? *Syst. Biol.* **47**:9–17.
- HENDY, M. D., and D. PENNY. 1989. A framework for the quantitative study of evolutionary trees. *Syst. Zool.* **38**:297–309.
- HILLS, D. M. 1996. Inferring complex phylogenies. *Nature* **383**:130.
- . 1998. Taxonomic sampling, phylogenetic accuracy, and investigator bias. *Syst. Biol.* **47**:3–8.
- HIRATSUKA, J., H. SHIMADA, R. WHITTIER et al. (16 co-authors). 1989. The complete sequence of the rice (*Oryza sativa*) chloroplast genome: intermolecular recombination between distinct tRNA genes accounts for a major plastid DNA inversion during the evolution of the cereals. *Mol. Gen. Genet.* **217**:185–194.
- HUELSENBECK, J. P., J. P. BOLLBACK, and A. M. LEVINE. 2002. Inferring the root of a phylogenetic tree. *Syst. Biol.* **51**:32–43.
- HUELSENBECK, J. P., and K. A. CRANDALL. 1997. Phylogeny estimation and hypothesis testing using maximum likelihood. *Ann. Rev. Ecol. Syst.* **28**:473–466.
- KATO, H., R. TERAUCHI, F. H. UTECH, and S. KAWANO. 1995. Molecular systematics of the Trilliacae *sensu lato* as inferred from *rbcl* sequence data. *Mol. Phyl. Evol.* **4**:184–193.
- KOHN, J. R., S. W. GRAHAM, B. R. MORTON, J. J. DOYLE, and S. C. H. BARRETT. 1996. Reconstruction of the evolution of reproductive characters in Pontederiaceae using phylogenetic evidence from chloroplast DNA restriction-site variation. *Evolution* **50**:1454–1469.
- LUNDBERG, J. G. 1972. Wagner networks and ancestors. *Syst. Zool.* **21**:398–413.
- LYONS-WEILER, J., and G. A. HOELZER. 1997. Escaping from the Felsenstein Zone by detecting long branches in phylogenetic analysis. *Mol. Phyl. Evol.* **8**:375–384.
- LYONS-WEILER, J., G. A. HOELZER, and R. J. TAUSCH. 1996. Relative Apparent Synapomorphy Analysis (RASA) I: the statistical measurement of phylogenetic signal. *Mol. Biol. Evol.* **12**:749–757.
- . 1998. Optimal outgroup analysis. *Biol. J. Linn. Soc.* **64**:493–511.
- MADDISON, D. R. 1991. The discovery and importance of multiple islands of most-parsimonious trees. *Syst. Zool.* **40**:315–328.
- MADDISON, D. R., M. RUVOLO, and D. L. SWOFFORD. 1992. Geographic origins of human mitochondrial DNA: phylogenetic evidence from control region sequences. *Syst. Biol.* **41**:111–124.
- MADDISON, W. P., and D. R. MADDISON. 1992. MacClade: analysis of phylogeny and character evolution. Version 3.0.

- Computer program and documentation distributed by Sinauer, Sunderland, Mass.
- MAIER, R. M., K. NECKERMAN, G. L. IGLOI, and H. KÖSSEL. 1995. Complete sequence of the maize chloroplast genome: gene content, hotspots of divergence and fine tuning of genetic information by transcript editing. *J. Mol. Biol.* **251**: 614–628.
- MIYAMOTO, M. M., and S. M. BOYLE. 1989. The potential importance of mitochondrial DNA sequence data to eutherian mammal phylogeny. Pp. 437–450 in B. FERNHOLM, K. BREMER, and H. JÖRNVALL, eds. *The hierarchy of life*. Elsevier Press, Amsterdam.
- NEYLAND, R. 2002. A phylogeny inferred from large-subunit (26S) ribosomal DNA sequences suggests that Burmanniales are polyphyletic. *Aust. Syst. Bot.* **15**:19–28.
- NIXON, K. C., and J. M. CARPENTER. 1993. On outgroups. *Cladistics* **9**:413–426.
- OLMSTEAD, R. G., and J. D. PALMER. 1994. Chloroplast DNA systematics: a review of methods and data analysis. *Am. J. Bot.* **81**:1205–1224.
- OLMSTEAD, R. G., and J. A. SWEERE. 1994. Combining data in phylogenetic systematics: an empirical approach using three molecular data sets in the Solanaceae. *Syst. Biol.* **43**: 467–481.
- PHILIPPE, H., and P. FORTERRE. 1999. The rooting of the universal tree of life is not reliable. *J. Mol. Evol.* **49**:509–523.
- QIU, Y.-L., J. LEE, B. A. WHITLOCK, F. BERNASCONI-QUADRONI, and O. DOMBROVSKA. 2001. Was the ANITA rooting of the angiosperm phylogeny affected by long-branch attraction? *Mol. Biol. Evol.* **18**:1745–1753.
- ROHLF, F. J., and D. R. FISHER. 1968. Tests for hierarchical structure in random data sets. *Syst. Zool.* **17**:407–412.
- SHIMODAIRA, H., and M. HASEGAWA. 1999. Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Mol. Biol. Evol.* **16**:1114–1116.
- SIMMONS, M. P., C. P. RANDLE, J. V. FREUDENSTEIN, and J. W. WENZEL. 2002. Limitations of Relative Apparent Synapomorphy Analysis (RASA) for measuring phylogenetic signal. *Mol. Biol. Evol.* **19**:14–23.
- SIMPSON, M. G. 1987. Pollen ultrastructure of the Pontederiaceae: evidence for exine homology with the Haemodora-ceae. *Grana* **26**:113–126.
- SMITH, A. B. 1994. Rooting molecular trees: problems and strategies. *Biol. J. Linn. Soc.* **51**:279–292.
- SOLTIS, D. E., and P. S. SOLTIS. 1998. Choosing an approach and an appropriate gene for phylogenetic analysis. Pp. 1–42 in P. S. SOLTIS, D. E. SOLTIS, and J. J. DOYLE, eds. *Molecular systematics of plants II: DNA sequencing*. Kluwer, Boston, Mass.
- SOLTIS, D. E., P. S. SOLTIS, M. E. MORT, M. W. CHASE, V. SAVOLAINEN, S. B. HOOT, and C. M. MORTON. 1998. Inferring complex phylogenies using parsimony: an empirical approach using three large DNA data sets for angiosperms. *Syst. Biol.* **47**:32–42.
- STEINECKE, H., and U. HAMANN. 1989. Embryologisch-systematische Untersuchungen an Haemodoraceen. *Bot. Jahrb. Syst.* **111**:247–262.
- STEVENSON, D. W., J. I. DAVIS, J. V. FREUDENSTEIN, C. R. HARDY, M. P. SIMMONS, and C. D. SPECHT. 2000. A phylogenetic analysis of the monocotyledons based on morphological and molecular character sets, with comments on the placement of *Acorus* and Hydatellaceae. Pp. 17–24 in K. L. WILSON and D. A. MORRISON, eds. *Monocots: systematics and evolution*. CSIRO, Melbourne.
- STILLER, J. W., and B. D. HALL. 1999. Long-branch attraction and the rDNA model of early eukaryotic evolution. *Mol. Biol. Evol.* **16**:1270–1279.
- SULLIVAN, J., and D. L. SWOFFORD. 1997. Are guinea pigs rodents? The importance of adequate models in molecular phylogenetics. *J. Mammal. Evol.* **4**:77–86.
- SWOFFORD, D. L. 1999. PAUP* 4.0 (beta version). Computer program and documentation distributed by Sinauer, Sunderland, Mass.
- SWOFFORD, D. L., G. J. OLSEN, P. J. WADDELL, and D. M. HILLIS. 1996. Phylogenetic inference. Pp. 407–514 in D. M. HILLIS, C. MORITZ, and B. K. MABLE, eds. *Molecular systematics*. 2nd edition. Sinauer, Sunderland, Mass.
- SWOFFORD, D. L., and S. POE. 1999. Taxon sampling revisited. *Nature* **398**:299–300.
- TILlich, H.-J. 1994. Untersuchungen zum Bau der Keimpflanzen der Philydraceae und Pontederiaceae (Monocotyledoneae). *Sendtnera* **2**:171–186.
- TILlich, H.-J. 1995. Seedlings and systematics in monocotyledons. Pp. 303–352 in P. J. RUDALL, P. J. CRIBB, D. F. CUTLER, and C. J. HUMPHRIES, eds. *Monocotyledons: systematics and evolution*. Royal Botanic Gardens, Kew, U.K.
- WHEELER, W. C. 1990. Nucleic acid sequence phylogeny and random outgroups. *Cladistics* **6**:363–367.
- ELIZABETH KELLOGG, reviewing editor

Accepted June 3, 2002