Systematic Botany (2006), 31(2): pp. 258–270

Inferring the Root of *Isoëtes*: Exploring Alternatives in the Absence of an Acceptable Outgroup

ERIC SCHUETTPELZ[1] and SARA B. HOOT[2,3]

[1]Department of Biology, Duke University, Durham, North Carolina 27708, U.S.A.
[2]Department of Biological Sciences, University of Wisconsin, Milwaukee, Wisconsin 53201, U.S.A.
[3]Author for correspondence (hoot@uwm.edu)

Communicating Editor: Thomas A. Ranker

ABSTRACT. The ancient divergence between the lycopsid genus *Isoëtes* and its closest living relative (*Selaginella*) has resulted in considerable morphological and genetic disparity, yet within *Isoëtes* there is remarkable morphological and genetic uniformity. This has made it difficult to identify the phylogenetic root of the genus. In this study, we addressed this problem and characterized the early branching patterns within *Isoëtes* using an expanded set of taxa and three molecular markers. We assessed the saturation in the molecular data sets, tested for differences in evolutionary rate, determined the stability of the ingroup topology, and evaluated the applicability of the molecular clock. We then explored three alternative rooting approaches: outgroup, midpoint, and maximum likelihood under the assumption of a molecular clock. Attempts to infer the root of *Isoëtes* using the outgroup approach were severely hindered by the effects of saturation, but the results from midpoint rooting and the enforcement of the molecular clock were highly consistent among the data sets. We identify the root of *Isoëtes* to be located among three major, highly supported clades.

KEYWORDS: Isoëtaceae, *Isoëtes*, lycophyte, phylogenetic root, midpoint rooting, relative rate comparisons.

Isoëtes, comprising approximately 200 extant species, is an ancient genus of primarily aquatic lycopsids that is phylogenetically isolated from all other living plants. Lycopsid fossils assigned to the *Isoëtes* lineage have been reported from the Late Devonian (Pigg 1992, 2001). These early fossils were often tree-like in habit and not very similar in appearance to extant *Isoëtes*, but nonetheless mark the divergence of the *Isoëtes* lineage from its nearest extant relative, the lycopsid genus *Selaginella*. The ancient divergence between *Isoëtes* and *Selaginella* has led to considerable morphological and genetic differentiation (Manhart 1994; Wikström and Kenrick 1997, 2001; Duff and Nickrent 1999; Soltis et al. 1999; Rydin and Wikström 2002).

Within *Isoëtes*, there is remarkable morphological and genetic uniformity (Taylor and Hickey 1992; Hoot and Taylor 2001; Rydin and Wikström 2002). Fossils that are morphologically similar to extant *Isoëtes* appear by the Jurassic (Skog and Hill 1992; Retallack 1997; Pigg 2001), and limited morphological change has apparently occurred since this time. The simple and conserved morphology of *Isoëtes* provides few variable characters, presenting a daunting challenge to those working on the systematics of the genus. Conserved molecular markers such as the plastid *rbcL* gene can be phylogenetically informative globally within *Isoëtes* (Rydin and Wikström 2002). However, because of limited genetic differentiation, it is necessary to utilize more variable DNA regions to yield species-level phylogenies with reasonable resolution and branch support (Hoot and Taylor 2001; Hoot et al. 2004).

The isolation and infrageneric uniformity of *Isoëtes* have together made it difficult to definitively identify the phylogenetic root of the genus. Morphological studies of *Isoëtes* initially led to its subdivision into a small, putatively relictual grade (subgenus *Euphyllum*) and a diverse, derived clade (subgenus *Isoëtes*; Hickey 1986, 1990). Later work formally recognized two sections within the latter subgenus: section *Coromandelina*, restricted to the Indian subcontinent, and the cosmopolitan section *Isoëtes* (Taylor and Hickey 1992). An initial phylogenetic study of subgenus *Isoëtes* (Hoot and Taylor 2001)—employing multiple molecular markers—was rooted in accordance with the morphological classification, placing the root between the two recognized sections. However, a subsequent molecular study (Rydin and Wikström 2002)—based on a single plastid gene but including several outgroup taxa as well as an exemplar from subgenus *Euphyllum*—found a rooting apparently inconsistent with the original morphological hypotheses. Further sequencing suggested that this conflict may not be so great; the *rbcL* sequence newly generated from Indian material of *I. coromandelina* (section *Coromandelina*; Appendix 1) is very unlike the sequence attributed to this species in the study of Rydin and Wikström (2002), which shows strong affinities to the North American species complex (Hoot and Taylor 2001). More comprehensive sequencing within *Isoëtes* also revealed that some major lineages were either unrepresented or underrepresented in the earlier works (e.g., Hoot and Taylor 2001; Rydin and Wikström 2002). Questions concerning the nature of the morphological and molecular disagreement, combined with the sampling problems, necessitated further analyses to infer the root of *Isoëtes* and characterize the early branching patterns within the genus.

Attempting to identify the root of *Isoëtes* using the

outgroup approach is a complicated endeavor. Outgroup rooting, although by far the most commonly used rooting approach, can be challenging when the nearest sister group is distantly related and when there is relatively little variation within the ingroup (Maddison et al. 1984; Wheeler 1990; Huelsenbeck et al. 2002). *Isoëtes* seems to exemplify these challenges, especially from a molecular standpoint. Non-coding molecular markers are simply unalignable between *Isoëtes* and other lycopsids. Coding regions, while alignable, provide only a limited number of characters that segregate among *Isoëtes* species. These characters, which are potentially informative with regard to both resolving and rooting the *Isoëtes* topology, are likely saturated between *Isoëtes* and its closest (but still very distant) relatives. The paucity of polarizing characters and the lack of consensus among them due to saturation, can compromise analyses.

Unfortunately, few alternatives exist for rooting a phylogenetic tree; none of which are commonly utilized in the literature. The choices—asymmetrical step-matrices, nonreversible models of DNA substitution, midpoint rooting, and the enforcement of a molecular clock—all require certain assumptions (not unlike the outgroup approach, which requires the assumption that the outgroup falls outside the group of interest). Asymmetrical step-matrices and nonreversible models of DNA substitution both assume that rates of substitution differ depending on the direction of the change. This may be a reasonable assumption, but such approaches have been shown to do a poor job of discriminating among possible rootings and are therefore of limited use (Yang 1994; Huelsenbeck et al. 2002). The two remaining approaches, midpoint rooting and the enforcement of a molecular clock, both assume at least some degree of rate constancy and perform well when rates of substitution are constant (Huelsenbeck et al. 2002). Although many studies have suggested that rate constancy is an exception rather than the norm (e.g., Li and Wu 1985; Britten 1986; Li 1997; Muse 2000), it has been shown that even the more severe approach of rooting through the enforcement of the molecular clock may be able to correctly identify the root when the clock assumption is violated (Huelsenbeck et al. 2002). Midpoint rooting further relaxes the clock assumption and should therefore be even more robust to ubiquitous rate heterogeneity (Swofford et al. 1996).

Here, our objective is to infer the root of *Isoëtes* using an expanded set of taxa and three molecular markers: the plastid protein coding *rbcL* gene (*rbcL*), the plastid non-coding *atpB-rbcL* intergenic spacer (*atpB-rbcL* spacer), and the internal transcribed spacers of nuclear ribosomal DNA (ITS region). We assess the saturation in these data and test for differences in evolutionary rate where possible. To determine the stability of the ingroup topology, we analyze the ingroup data using maximum parsimony, maximum likelihood, and Bayesian inference. We then evaluate the applicability of the molecular clock to the various data sets and explore three alternative rooting approaches: outgroup, midpoint, and maximum likelihood under the assumption of a molecular clock.

Materials and Methods

Sampling. Sampling of *Isoëtes* was determined using a placeholder strategy, selecting taxa to represent all of the well-supported major lineages found either in previously published phylogenies (Hoot and Taylor 2001; Rydin and Wikström 2002) or in preliminary analyses for this study. Our total ingroup sampling consisted of 16 *Isoëtes* taxa (Appendix 1). Except for *I. coromandelina* (section *Coromandelina*), all of these would have traditionally been placed in section *Isoëtes*. Overall, nine species are shared between this study and that of Hoot and Taylor (2001), and four species are shared between this study and that of Rydin and Wikström (2002); only two species sampled here were common to both earlier studies. Five previously unsampled taxa were incorporated, including two different populations of *I. australis* (because of substantial sequence differences, perhaps indicative of cryptic speciation). We also used a placeholder approach in selecting species to serve as outgroups in later analyses. Based on previous phylogenetic studies of lycopsids (Wikström and Kenrick 2001; Korall and Kenrick 2002), we selected four species from among the major clades of *Selaginella* (the sister group to *Isoëtes*) as well as four species from among the major clades of *Lycopodium* s.l. (including *Huperzia*, *Lycopodiella*, *Lycopodium*, and *Phylloglossum;* phylogenetically more distant but genetically less divergent than *Selaginella*).

DNA Sequencing and Alignment. DNA extraction, amplification, PCR product purification, sequencing, and alignment of the three markers used in the study (*rbcL, atpB-rbcL* spacer, and ITS region) were as described in Hoot and Taylor (2001), except that PCR products were purified using either Wizard (Promega) or QIAquick (Qiagen) columns and sequencing reactions were run on either an ABI 373-Stretch (Applied Biosystems) or a CEQ 2000 (Beckman Coulter) automated sequencer. Gaps were treated as missing data and not scored. Statistics relevant to the various data sets, including the percentage of missing data, can be found in Table 1. The alignments used in this study are available in TreeBASE (study accession number S1409).

Evaluating Saturation and Relative Rate Differences. To assess the level of nucleotide saturation at the *rbcL* locus between *Isoëtes* and the outgroups, we conducted pairwise comparisons among the included species. We plotted uncorrected ''p'' distances (observed) against maximum likelihood distances (expected), under the assumption that observed differences will increase linearly with increasing expected differences in the absence of saturation but will approach an asymptote with increasing expected differences in the presence of saturation (Philippe et al. 1994). All pairwise distances were calculated using PAUP* v4.0b10 (Swofford 2002) with maximum likelihood distances estimated using two models: (1) the simplest model of sequence evolution (Jukes and Cantor 1969); and (2) the best-fitting model of sequence evolution identified for this 24 taxon *rbcL* data set, TIMef+G, using the hierarchical likelihood ratio test approach as implemented in Modeltest v3.06 (Posada and Crandall 1998).

To determine whether differences in evolutionary rate were present between *Isoëtes* and the outgroups or within these three genera, we conducted a series of pair-wise relative rate comparisons. For each pair-wise comparison, a three-taxon tree was constructed, using *Ginkgo* (an exemplar from the euphyllophyte lineage sister to the lycophytes) as the outgroup. Two models were compared—one with (null) and one without (alternative) the constraint of equal rates between the two focal species. The likelihoods corresponding to each of these models were evaluated using the likelihood ratio test statistic (Goldman 1993). All 276 pair-wise comparisons were made in an automated fashion using the program

TABLE 1. Summary of molecular data sets phylogenetically analyzed in this study (I = *Isoëtes*; S = *Selaginella*; L = *Lycopodium* s.l.), with statistics and parameter estimates corresponding to the maximum parsimony and maximum likelihood analyses.

	rbcL (I)	*rbcL* (I + S)	*rbcL* (I + L)	*atpB-rbcL* spacer (I)	ITS region (I)	Combined data (I)	Combined data (I + S)	Combined data (I + L)
Data								
Alignment length	1300	1300	1300	730	660	2690	2690	2690
Variable characters	66	393	275	91	270	427	754	636
Missing data (%)	0.856	1.181	0.719	5.882	3.750	2.930	12.918	12.716
MP analyses								
Informative characters	39	263	178	69	199	307	531	446
Number of trees	1	1	5	8	3	1	1	1
Tree length	82	565	383	113	486	684	1167	984
CI	0.829	0.807	0.770	0.876	0.739	0.769	0.783	0.765
RI	0.860	0.801	0.788	0.916	0.795	0.822	0.810	0.808
ML analyses								
Best-fit model	TrN + I + G	TIMef + G	TrN + G	K81uf + G	HKY + G	TIM + I + G	TIM + I + G	TrN + I + G
Frequency (A)	0.2727	0.2500	0.2745	0.3281	0.1723	0.2583	0.2526	0.2606
Frequency (C)	0.1992	0.2500	0.1984	0.1732	0.2748	0.2141	0.2206	0.2137
Frequency (G)	0.2388	0.2500	0.2342	0.1650	0.2852	0.2388	0.2453	0.2353
Frequency (T)	0.2893	0.2500	0.2929	0.3337	0.2677	0.2888	0.2815	0.2904
Rate (A-C)	1.0000	1.0000	1.0000	1.0000	—	1.0000	1.0000	1.0000
Rate (A-G)	3.5776	2.5026	6.3140	2.5993	—	2.0437	2.3400	3.3925
Rate (A-T)	1.0000	0.3095	1.0000	0.3144	—	0.6474	0.5561	1.0000
Rate (C-G)	1.0000	0.3095	1.0000	0.3144	—	0.6474	0.5561	1.0000
Rate (C-T)	7.4101	5.1885	13.0066	2.5993	—	3.2493	4.1333	5.5894
Rate (G-T)	1.0000	1.0000	1.0000	1.0000	—	1.0000	1.0000	1.0000
Ti-Tv ratio	—	—	—	—	1.4563	—	—	—
Alpha (Shape)	0.9371	0.4026	0.2420	0.2906	0.4252	0.6515	0.6280	0.6507
Pinvar	0.7779	0.0000	0.0000	0.0000	0.0000	0.6132	0.3700	0.4730
Number of trees	1	1	2	1	1	1	1	2
ln L	−2382.82795	−4464.43866	−3764.81484	−1681.06660	−3243.11715	−7659.59792	−9808.83993	−9106.22731
ln L (clock enforced)	−2393.41843	—	—	−1694.23111	−3270.28887	−7684.22422	—	—

HyPhy 0.99 (Kosakovsky Pond et al. 2005), with the GTR+I+G model of sequence evolution (as identified using Modeltest for this 25 taxon *rbcL* data set) and globally estimated parameters.

Ingroup Analyses. To assess the stability of the ingroup topology and to evaluate the combinability of the three data sets (*rbcL*, ITS region, and *atpB-rbcL* spacer), we first analyzed sequences from the 16 *Isoëtes* samples only. We utilized three analytical approaches: equally weighted maximum parsimony (MP), maximum likelihood (ML), and Bayesian inference (BI). To assess branch support with the MP and ML approaches, we conducted bootstrap analyses (MPBS and MLBS). Because the individual data sets indicated largely congruent results, the data were combined and analyzed in unison.

MP analyses of the four data sets (three markers plus the combined data set) were performed using PAUP* version 4.0b10 (Swofford 2002) employing the branch and bound search option, saving all shortest trees. For the ML analyses, the best-fitting model of nucleotide substitution was identified for each data set using the hierarchical likelihood ratio test approach, as implemented in Modeltest v.3.06 (Posada and Crandall 1998). Heuristic ML searches were conducted in PAUP* using the appropriate model of evolution with the associated parameter estimates, 100 random addition sequence replicates, and TBR branch swapping with the MULTrees option in effect.

Bayesian inference (BI) was conducted using MrBayes version 3.1 (Huelsenbeck and Ronquist 2001; Ronquist and Huelsenbeck 2003). Using the model as identified by Modeltest and flat priors, four chains (three heated) were run for 10 million generations, sampling trees every 1,000 generations. The first 1 million generations (1,000 trees) were excluded as the burn-in phase, adequate for the analyses to reach stationarity. A consensus tree, with average branch lengths and Bayesian posterior probabilities (BPP), was computed using the 'sumt' command.

MPBS analyses with 1,000 pseudoreplicates were conducted in PAUP* using a branch and bound search strategy. MLBS analyses were also completed in PAUP* using a full heuristic search strategy: 100 pseudoreplicates, each with 10 random addition sequence replicates, TBR branch swapping, and MULTrees on.

Testing for the Presence of a Molecular Clock. To assess whether or not a global molecular clock (i.e., the same rate of sequence evolution across the entire phylogeny) could be applied to any of the individual ingroup data sets or the combined ingroup data set, we compared likelihood estimates using the likelihood ratio test (Goldman 1993). For each data set, we utilized the model of sequence evolution and parameter estimates as identified by Modeltest and the topology resulting from the ML search. We then estimated the likelihood of observing the data given the model and topology provided, without a clock enforced and with a clock enforced (rooting the tree at every possible internode) in PAUP*, and compared these estimates using the likelihood ratio test.

Rooting. We employed three strategies to identify the position of the root within *Isoëtes*: (1) the outgroup rooting approach, (2) midpoint rooting of the ingroup topologies, and (3) maximum likelihood searches with the molecular clock enforced. To root *Isoëtes* using the outgroup approach, we incorporated in parallel the four species selected from *Selaginella* (the sister group to *Isoëtes*) and the four species selected from *Lycopodium* s.l. (phylogenetically more distant but genetically less divergent than *Selaginella*). Only *rbcL* sequences were alignable between the ingroup and outgroup taxa, so analyses using the outgroup approach were limited to the *rbcL* data and the combined data (with missing data in place of ITS region and *atpB-rbcL* spacer sequences for the outgroups). These data were analyzed as described above for the ingroup-only analyses (MP, ML, BI, MPBS, and MLBS), and the resulting trees rooted with the four included outgroup species. To further assess the stability of the recovered (i.e., most likely) rootings, all possible bifurcating rootings were evaluated by employing the SH test (Shimodaira and Hasegawa 1999) in PAUP*.

Midpoint rooting for each of the data sets was conducted in PAUP* using the results (both topological and branch length) of the three analytical approaches (MP, ML, and BI). To identify the root using the molecular clock approach (and support for the resulting relationships), we conducted additional ML and MLBS searches. These ingroup-only analyses for each of the four data sets were conducted in an identical fashion as the ML and MLBS analyses described above, but with the molecular clock enforced.

Results

Data. Statistics relevant to the alignments of the three individual molecular data sets and the combined data set are provided in Table 1. Considerable differences in the number of ingroup-variable characters are present among the data sets, with the *atpB-rbcL* spacer yielding more variable characters (91) than the *rbcL* gene (66); and the nuclear ITS region yielding more variable characters (270) than the two plastid regions combined, despite its smaller size.

As previously noted, *atpB-rbcL* spacer and ITS region sequences are not alignable between *Isoëtes* and the outgroups. Plastid *rbcL* gene sequences are readily alignable and when outgroup sequences were added, considerably more variation was introduced into the data set. With the addition of the *Selaginella* sequences, the number of variable *rbcL* characters rose from 66 to 393; and with the addition of the *Lycopodium* s.l. sequences, the number rose to 275. Such discrepancies indicate that the divergences between *Isoëtes* and the outgroups are particularly deep or that the outgroup sequences are evolving considerably faster than the ingroup sequences, or both. Significant differences in rate were affirmed through the relative rate comparisons. All 64 comparisons between *Isoëtes* and *Selaginella* sequences yielded significant ($P \leq 0.05$) results; 33 of 64 comparisons between *Isoëtes* and *Lycopodium* s.l. were significant. Each of the significant comparisons indicated a higher rate of evolution in the outgroup. Eighteen of 28 comparisons among the outgroup species were also significant (most of these were between *Selaginella* and *Lycopodium* s.l.), but only six of the 120 comparisons between ingroup sequences revealed significant differences in the rate of molecular evolution.

Plots of observed (uncorrected ''p'') versus expected (maximum likelihood) distances among taxa showed an initial, linear increase corresponding to comparisons between ingroup species pairs (Fig. 1). However, when comparisons between ingroup and outgroup species were also considered, observed distances began to approach an asymptote while estimated distances continued to increase, indicating the probable onset of saturation (Fig. 1).

Unrooted Ingroup Trees. Phylogenetic analyses of the three data sets (*rbcL*, *atpB-rbcL* spacer, and ITS region), restricted to the ingroup taxa, yielded highly congruent unrooted trees, regardless of the analytical approach (ML, MP, or BI) employed (Fig. 2; see Table 1 for statistics corresponding to the various ingroup analyses). In the *rbcL* analyses, six of 13 bipartitions

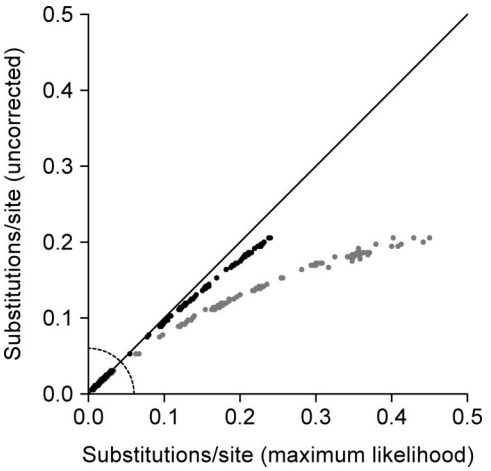

FIG. 1. Plots of observed (uncorrected "p") versus expected (maximum likelihood) distances among *Isoëtes* and outgroup *rbcL* sequences. Black points indicate comparisons of uncorrected "p" and JC (simplest model) maximum likelihood distances; gray points indicate comparisons of uncorrected "p" and TIMef+G (best-fitting model) maximum likelihood distances. Points between origin and dashed arc indicate *Isoëtes*-*Isoëtes* comparisons; points beyond dashed arc indicate *Isoëtes*-outgroup and outgroup-outgroup comparisons.

received significant support from all three measures: MPBS ≥ 70; MLBS ≥ 70; and BPP ≥ 95. Nine bipartitions in the *atpB-rbcL* spacer analyses and ten bipartitions in the ITS region analyses received significant support by all three measures. None of the well-supported partitions was in conflict among data sets (Fig. 2), and therefore the data were combined. Analyses of the combined data set yielded significant support for 11 bipartitions; only two bipartitions did not receive significant support from all three measures.

Evaluating Rate Constancy. Although pair-wise relative rate comparisons among *rbcL* sequences did not reveal significant differences in rate among the ingroup species (with a few exceptions), likelihood ratio tests used to assess the applicability of a global molecular clock suggested that the rate of molecular evolution was not constant within the ingroup. The enforcement of a global molecular clock (i.e., the same rate of sequence evolution across the entire phylogeny) resulted in a significantly worse likelihood ($p \leq 0.05$) for every possible rooting, for the individual and combined data sets, with only three exceptions. The global clock could not be rejected for three possible rootings of the *rbcL* tree (Fig. 2A; positions indicated by arrows).

Ingroup Rooting. The various attempts to infer the root of *Isoëtes* revealed different rootings, depending on the data set, analytical method, outgroups, and the rooting criterion employed. However, the differences observed were generally not well supported, and the results were largely consistent with one another.

When outgroups from *Selaginella* were included, maximum likelihood analyses of both the *rbcL* data set and the combined data set (with missing *atpB-rbcL* spacer and ITS region data for the outgroup taxa) yielded trees with *I. nuttallii* sister to the remainder of *Isoëtes* (Figs. 3A, 3B). The rooting resulting from a Bayesian analysis of the *rbcL* data was consistent with that from the maximum likelihood analyses (Fig. 3A); but Bayesian analysis of the combined data set yielded a tree in which *I. australis*, *I. capensis*, *I. coromandelina*, and *I. panamensis* were sister to all remaining species (tree not shown, but rooting consistent with Fig. 5D). Maximum parsimony analyses of the *rbcL* and combined data sets both yielded a rooting with *I. coromandelina* and *I. panamensis* resolved as sister to the remaining species (trees not shown, but rooting consistent with Fig. 6A).

When outgroups from *Lycopodium* s.l. were included, maximum likelihood analysis of the *rbcL* data set yielded a rooting with *I. melanopoda* sister to the remainder of *Isoëtes* (Fig. 4A). Maximum likelihood analysis of the combined data set resulted in two altogether different rootings: one with *I. nuttallii*, *I. orcuttii*, *I. abyssinica*, and *I. velata* sister to the remainder of *Isoëtes* (tree not shown); and another with a basal trichotomy (Fig. 4B). Bayesian analysis of the *rbcL* data set resolved *I. melanopoda* and *I. histrix* as sister to all other species of *Isoëtes*; and Bayesian analysis of the combined data set yielded a tree with *I. malinverniana*, *I. drummondii*, *I. taiwanensis*, *I. kirkii*, *I. melanopoda*, *I. setacea*, and *I. histrix* sister to the remaining included species (trees not shown). Maximum parsimony analysis of the *rbcL* data set yielded two different rootings: one with *I. melanopoda* sister to the remainder of *Isoëtes* (tree not shown, but rooting consistent with Fig. 4A), and the other with *I. drummondii*, *I. taiwanensis*, *I. kirkii*, *I. melanopoda*, *I. setacea*, and *I. histrix* sister to the other *Isoëtes* species (tree not shown; parsimony analysis of the combined data set yielded an identical rooting). Not one of the various outgroup rootings, however, received significant support from any measure. Regardless of whether *Selaginella* or *Lycopodium* s.l. outgroups were used, significant support was only present for more derived relationships within *Isoëtes* (Figs. 3, 4). Furthermore, the SH test could not reject any possible rooting with the *Selaginella* outgroups and only five possible rootings could be rejected with the *Lycopodium* s.l. outgroups (none of which were ever actually recovered).

Midpoint rooting of the *rbcL* and *atpB-rbcL* spacer trees identified the position of the root as dividing *Isoëtes* into a clade consisting of *I. australis*, *I. coromandelina*, and *I. panamensis* and a clade containing all oth-

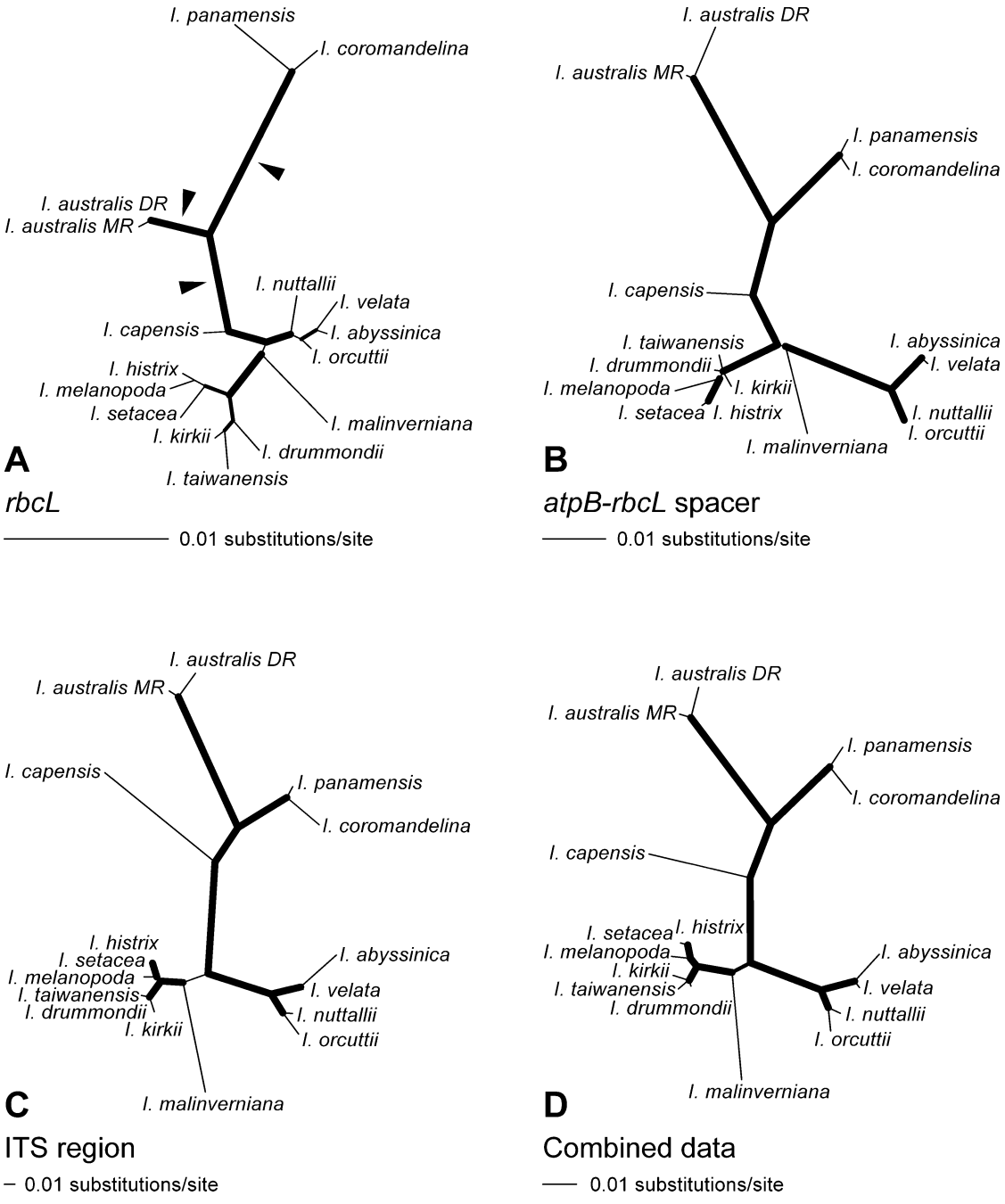

FIG. 2. Unrooted *Isoëtes* trees resulting from maximum likelihood analyses. Heavy bold indicates significant support from all three measures (MPBS ≥ 70; MLBS ≥ 70; BPP ≥ 0.95). Lighter bold indicates significant support from one or two measures. Arrows (in A) indicate the three possible rootings that yield trees for which the molecular clock cannot be rejected.

er *Isoëtes* species (rooted ML trees shown in Figs. 5A, 5B; MP and BI rootings identical to ML, not shown). The earliest bifurcation within this larger clade, in turn, separated *I. capensis* from the remaining taxa. Midpoint rooting of the ITS region and combined data trees revealed a somewhat different topology, placing *I. capensis* as sister to the *I. australis*, *I. coromandelina*, and *I. panamensis* clade (rooted ML trees shown in Figs. 5C, 5D; MP and BI rootings identical to ML, not shown).

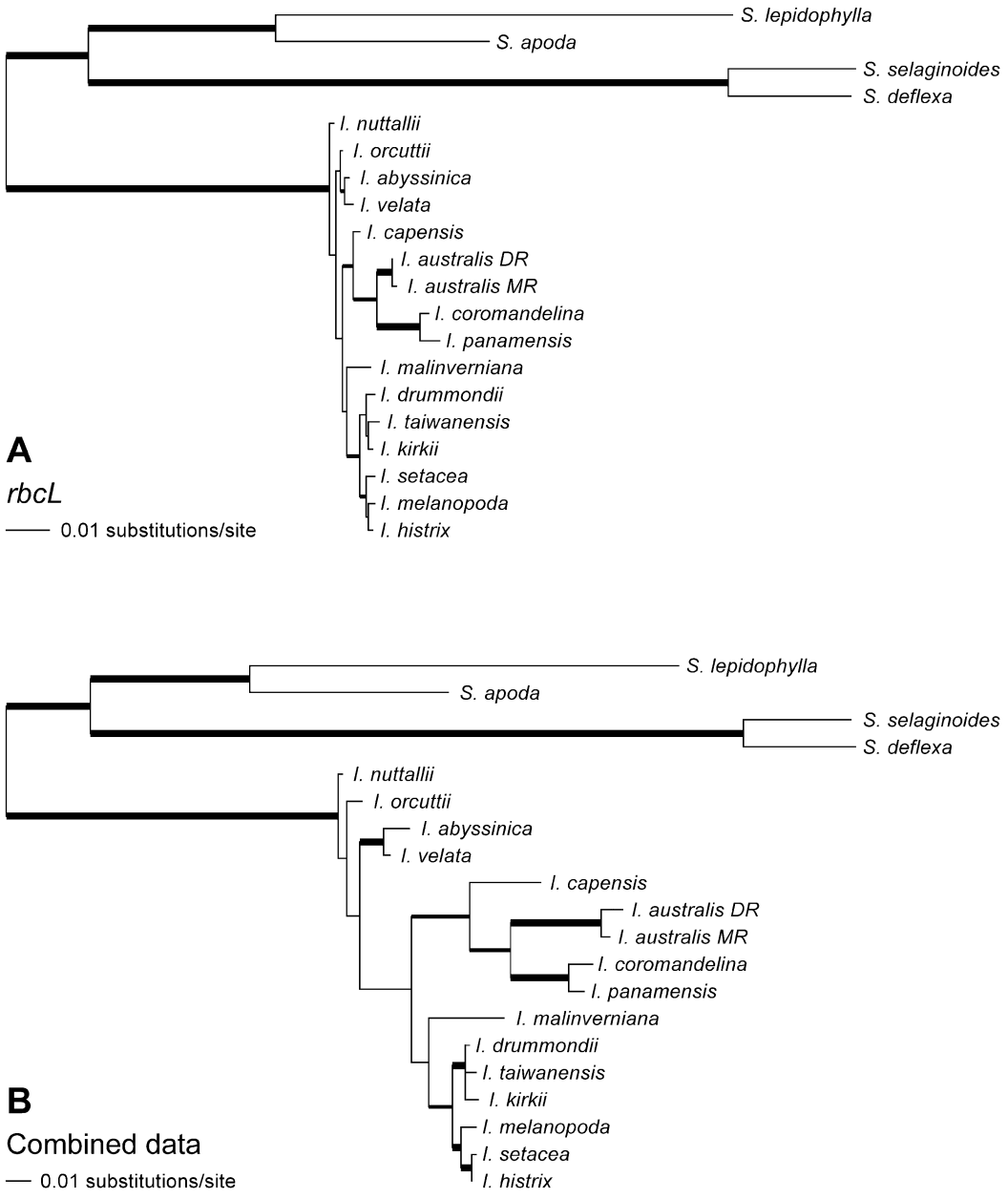

FIG. 3. Outgroup rooted *Isoëtes* trees resulting from maximum likelihood analyses including *Selaginella* outgroups. Heavy bold and lighter bold are as in Fig. 2.

Analyses utilizing maximum likelihood searches with the molecular clock enforced revealed rootings identical to those of midpoint rooting for each data set (Figs. 6B, 6C, 6D) except for the *rbcL* analysis, which resolved *I. coromandelina* and *I. panamensis* as sister to the rest of *Isoëtes* (Fig. 6A). This conflicting rooting, however, was not significantly supported by the maximum likelihood bootstrap analysis with the clock enforced.

DISCUSSION

Stability of the Unrooted Ingroup Topology. Based on our ingroup analyses of *Isoëtes* species, the infrageneric relationships appear both well resolved and well supported. Analyses of the individual data sets yielded very consistent results (Figs. 2A–C), regardless of the optimality criterion employed. Combined analysis of the data resulted in a fully resolved topology with all but

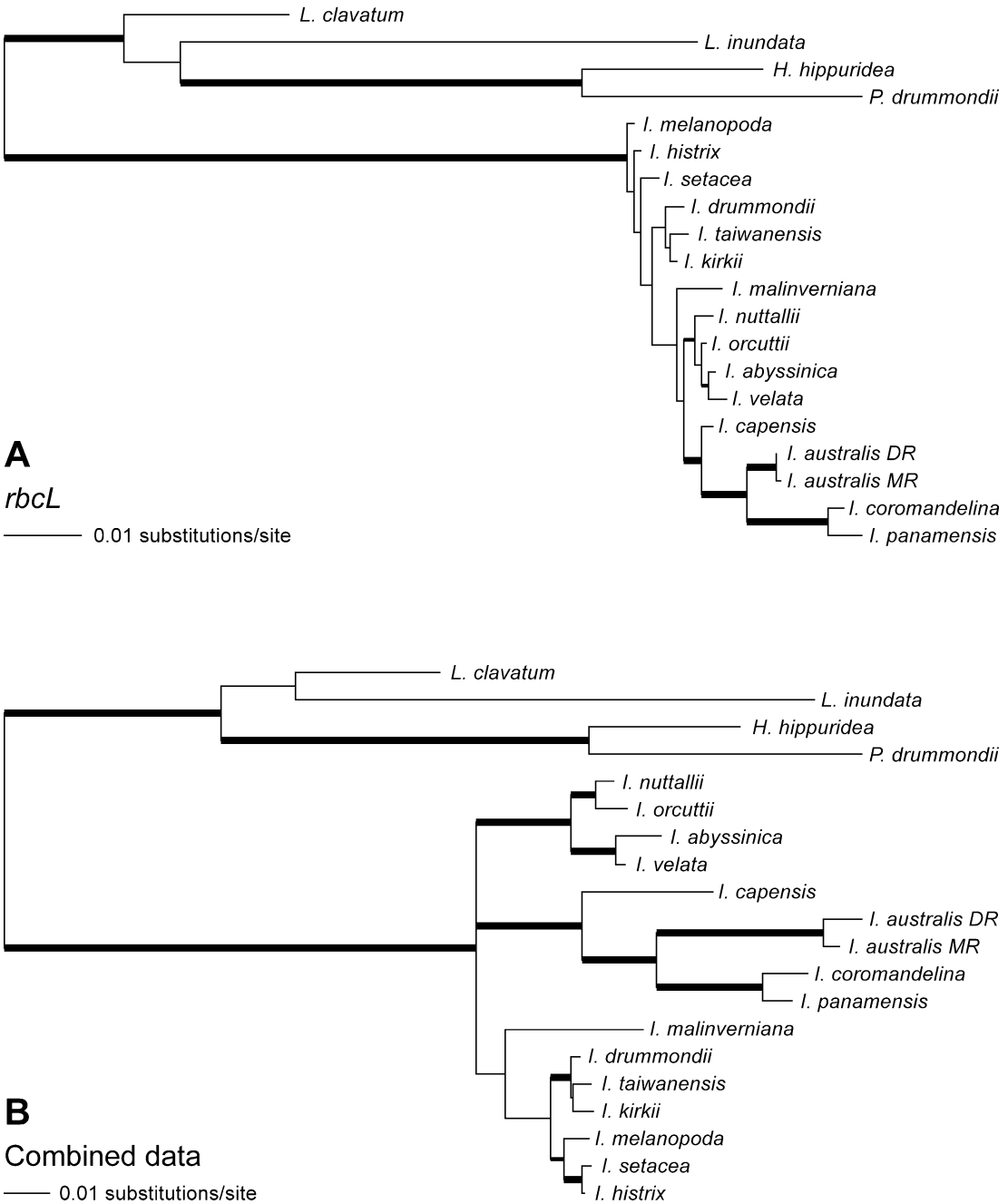

FIG. 4. Outgroup rooted *Isoëtes* trees resulting from maximum likelihood analyses including *Lycopodium* s.l. outgroups. Heavy bold and lighter bold are as in Fig. 2.

two resolved partitions receiving significant support from all three support measures (MLBS ≥ 70; MPBS ≥ 70; and BPP ≥ 0.95; Fig. 2D). However, the unrooted ingroup topology, by definition, does not identify the position of the root within the genus. Without this knowledge, one can say little about phylogenetic relationships, character evolution, or biogeography.

The Inferred Root of Isoëtes. Most pair-wise relative rate comparisons between *Isoëtes* and the outgroups revealed significant differences, and saturation plots suggested that substitutional saturation has occurred (Fig. 1). Extreme molecular evolutionary rate differences and saturation can present problems for typical outgroup rooting, as revealed in the results of our analyses incorporating outgroup species (Figs. 3, 4). Instead of simply polarizing the well-supported in-

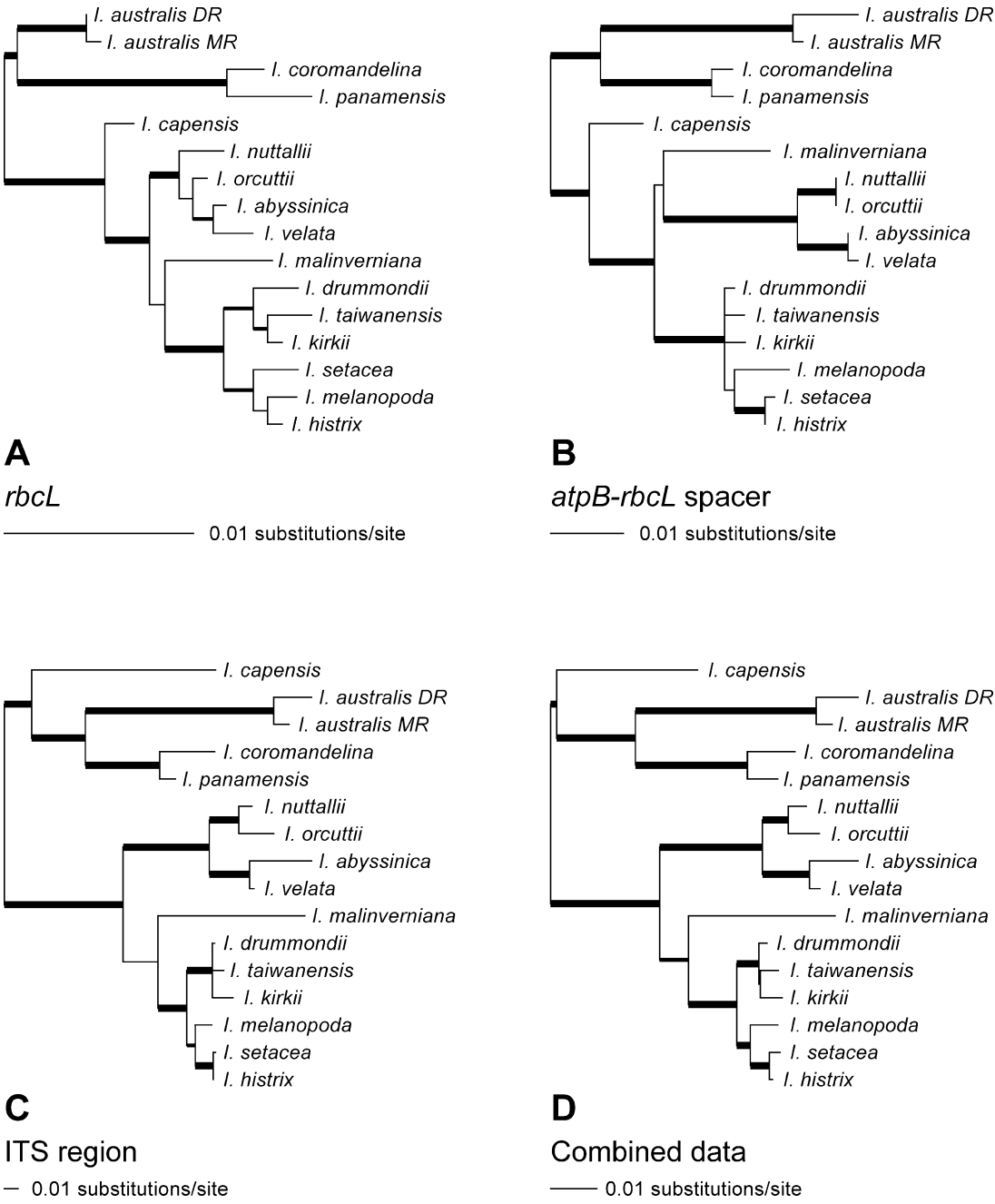

FIG. 5. Midpoint rooted *Isoëtes* trees resulting from maximum likelihood analyses. Heavy bold and lighter bold are as in Fig. 2.

group topology, these analyses resulted in questionable trees with little support throughout. The likelihood scores of the various possible rootings were all very similar, and the SH test was unable to discriminate among most of the possibilities. Many different rootings were explored in the bootstrap and Bayesian analyses, and depending on the replicate or generation, the outgroup taxa attached to the ingroup tree in very disparate locations. It should be noted, however, that the ingroup topology itself was not affected, as is sometimes the case when using distant outgroups (Swofford et al. 1996; Milinkovitch and Lyons-Weiler 1998; Holland et al. 2003). The problems with outgroup rooting of *Isoëtes* are apparently due to two underlying causes: the relative lack of signal useful for polarizing the ingroup topology (only 66 characters were variable

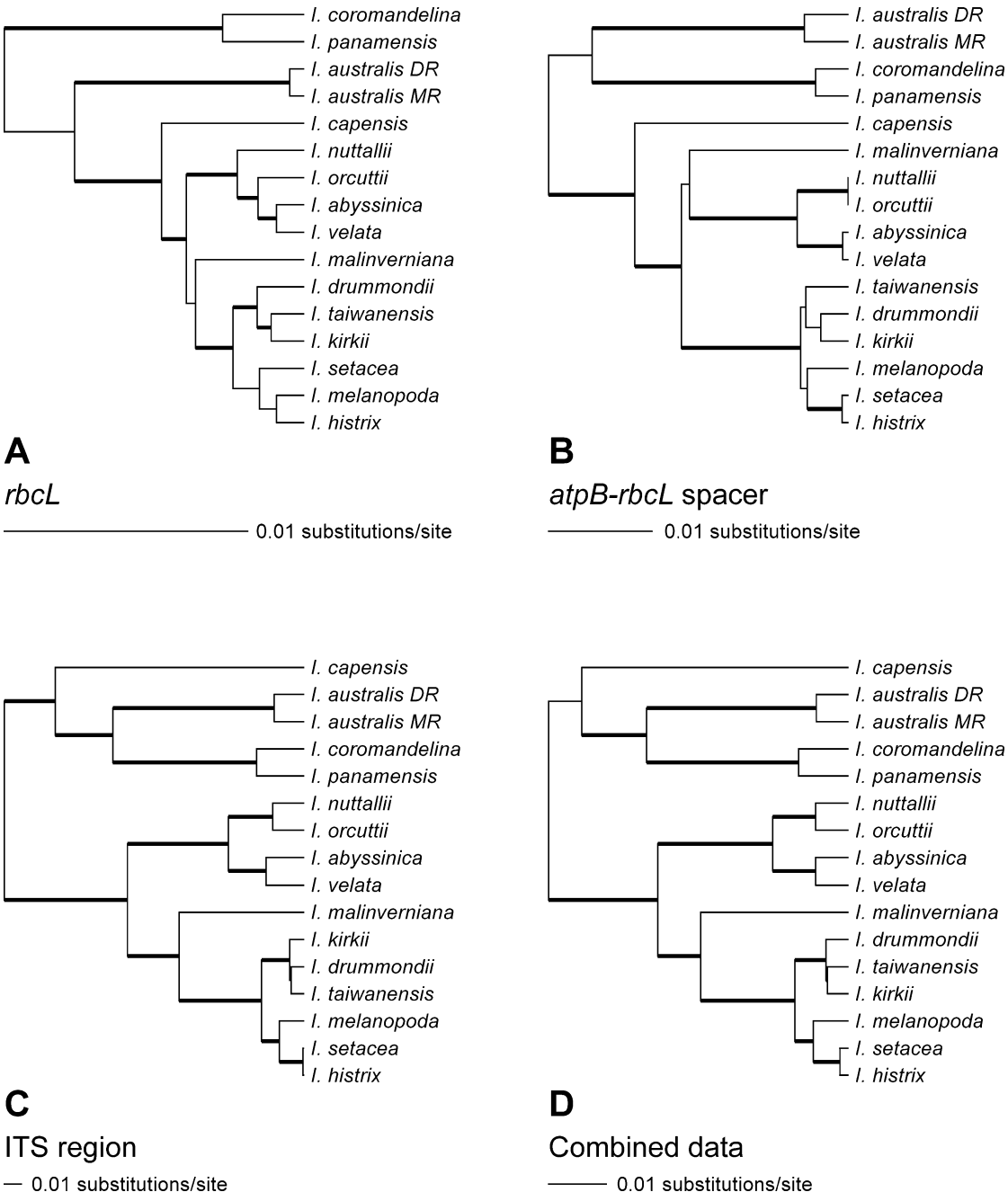

FIG. 6. Clock rooted *Isoëtes* trees resulting from maximum likelihood analyses with the molecular clock enforced. Lighter bold indicates significant support (MLBS ≥ 70) from maximum likelihood bootstrap analysis with the molecular clock enforced.

among the *Isoëtes* species at the *rbcL* locus) and, more importantly, the lack of consistency within the polarizing signal due to saturation. These problems necessitated the pursuit of alternative rooting approaches.

In most instances, enforcing the molecular clock across *Isoëtes* data sets resulted in a significantly worse likelihood score (Table 1). However, for the *rbcL* data there were three rootings consistent with a constant rate of molecular evolution (i.e., the clock could not be rejected when the tree was rooted in these locations; Fig. 2A). Furthermore, although relative rates tests can sometimes lack power (Bromham et al. 2000), such pair-wise comparisons for the *rbcL* sequences (the only data alignable to the outgroup sequences) revealed

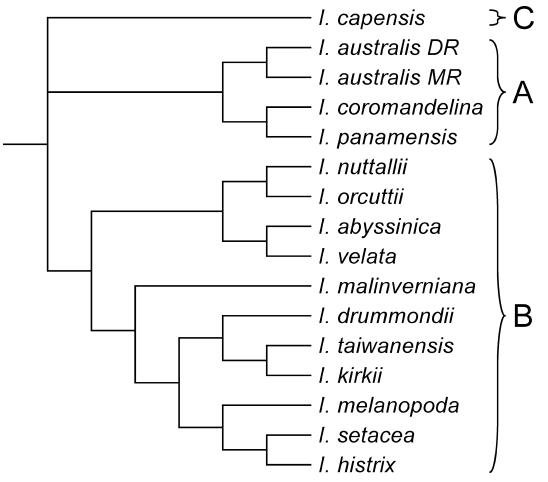

FIG. 7. Current best estimate for the position of the *Isoëtes* root, yielding a basal trichotomy. Clades A and B correspond approximately to the two major clades resolved by Rydin and Wikström (2002); clade C is newly recognized here.

only a few significant rate differences among the ingroup species. These suggestions of relative rate constancy, although by no means conclusive, validated our only choice—exploring rooting approaches that require at least some indication of a constant rate.

Our results utilizing midpoint rooting and the enforcement of the molecular clock were highly consistent among the data sets. Midpoint rooting of the *rbcL* and *atpB-rbcL* spacer trees yielded identical results, placing *I. australis*, *I. coromandelina*, and *I. panamensis* sister to the remaining *Isoëtes* species (Figs. 5A, 5B). The ITS region and the combined data moved the root by one node to place *I. capensis* as sister to *I. australis*, *I. coromandelina*, and *I. panamensis* (Figs. 5C, 5D). For the *atpB-rbcL* spacer, ITS region, and combined data, results when enforcing the clock were identical to those from midpoint rooting (Figs. 6B, 6C, 6D) but clock-enforced maximum likelihood bootstrap support for the basal split was only significant with the ITS data. When the clock was enforced with the *rbcL* data, *I. coromandelina* and *I. panamensis* (without *I. australis*) were resolved as sister to the remaining species but clock-enforced maximum likelihood bootstrap support was again lacking (Fig. 6A). These generally consistent results suggest that relative rates (among taxa) were similar across the loci. Considering that the three data sets we used were of disparate origins—plastid protein coding, plastid non-coding, and nuclear non-coding—there is reason to believe that the approximate position of the root, as identified here, is correct.

Accounting for the limited disagreement between the plastid and nuclear data sets, we tentatively identify the root of *Isoëtes* to be located among three major, highly supported clades (Fig. 7). Two of these, A and B, are comparable in composition to the two major clades resolved in an earlier molecular study of the genus (Rydin and Wikström 2002), although the *I. nuttallii/I. orcuttii/I. abyssinica/I. velata* subclade (of clade B) was unsampled in that work. The third major clade, designated here as C, was also unsampled in that previous study.

Our results, like those of Rydin and Wikström (2002), are not consistent with earlier morphology-based hypotheses, which recognized a basal division in *Isoëtes* into two subgenera (species-poor *Euphyllum* and species-rich *Isoëtes*; Hickey 1986, 1990) and a subsequent subdivision of subgenus *Isoëtes* into two sections (species-poor *Coromandelina* and species-rich *Isoëtes*; Taylor and Hickey 1992). Rydin and Wikström (2002) found an exemplar of the *Euphyllum* group to be embedded within clade A; our results support the inclusion of a *Coromandelina* group exemplar in this clade as well. The disagreement between molecules and morphology, however, is not a rooting issue. It is impossible to root a better-sampled ingroup tree based on sequence data in such a way that it will be consistent with the morphology-based hypothesis (the *Euphyllum* and *Coromandelina* exemplars do not share the same position within clade A).

Recommendations. The results of our efforts to infer the root of *Isoëtes* generally converged on a basal trichotomy hypothesis (Fig. 7). Although some of the approaches we explored are not ideal, there are essentially no alternatives given the existing data. The usefulness of outgroups in rooting is well understood, and it is our hope that enough appropriate data—variable within the ingroup but not plagued by the effects of saturation relative to the outgroup—will eventually be available to infer the root of *Isoëtes* with such an approach. There is also considerable potential in using gene duplications to locate the phylogenetic root, without the need for the inclusion of outgroup taxa (Iwabe et al. 1989; Mathews and Donoghue 1999). Until the necessary data are acquired, and additional analyses are possible, we suggest that the rooting supported here be utilized in future studies focusing on infrageneric *Isoëtes* relationships.

ACKNOWLEDGEMENTS. The authors thank W.-L. Chiou, F. Cutten, G. Gastony, S. Leonard, L. Musselman, C. Prada, J. Schneller, W. C. Taylor, H.-E. Wanntorp, D. Woodland, for providing leaf material and/or DNA; N. Napier for help with some of the *rbcL* sequencing; and D. J. Hearn, J. S. Metzgar, N. S. Nagalingum, and especially K. M. Pryer for helpful comments and suggestions. This work was supported in part by an NSF grant (DEB9981460) to S. B. Hoot.

LITERATURE CITED

BRITTEN, R. J. 1986. Rates of DNA sequence evolution differ between taxonomic groups. *Science* 231: 1393–1398.

BROMHAM, L., D. PENNY, A. RAMBAUT, and M. D. HENDY. 2000. The power of relative rates tests depends on the data. *Journal of Molecular Evolution* 50: 296–301.

DUFF, R. J. and D. L. NICKRENT. 1999. Phylogenetic relationships of land plants using mitochondrial small-subunit rDNA sequences. *American Journal of Botany* 86: 372–386.

GOLDMAN, N. 1993. Statistical tests of models of DNA substitution. *Journal of Molecular Evolution* 36: 182–198.

HASEBE, M., R. KOFUJI, M. ITO, M. KATO, K. IWATSUKI, and K. UEDA. 1992. Phylogeny of gymnosperms inferred from *rbcL* gene sequences. *Botanical Magazine, Tokyo* 105: 673–679.

HICKEY, R. J. 1986. The early evolutionary and morphological diversity of *Isoëtes*, with descriptions of two new neotropical species. *Systematic Botany* 11: 309–321.

———. 1990. Studies of neotropical *Isoëtes* L. I. *Euphyllum*, a new subgenus. *Annals of the Missouri Botanical Garden* 77: 239–245.

HOLLAND, B. R., D. PENNY, and M. D. HENDY. 2003. Outgroup misplacement and phylogenetic inaccuracy under a molecular clock—a simulation study. *Systematic Biology* 52: 229–238.

HOOT, S. B. and W. C. TAYLOR. 2001. The utility of nuclear ITS, a leafy homolog intron, and chloroplast *atpB-rbcL* spacer region data in phylogenetic analyses and species delimitation in *Isoëtes*. *American Fern Journal* 91: 166–177.

———, N. S. NAPIER, and W. C. TAYLOR. 2004. Revealing unknown or extinct lineages within *Isoëtes* (Isoëtaceae) using DNA sequences from hybrids. *American Journal of Botany* 91: 899–904.

HUELSENBECK, J. P. and F. RONQUIST. 2001. MrBayes: Bayesian inference of phylogeny. *Bioinformatics* 17: 754–755.

———, J. P. BOLLBACK, and A. M. LEVINE. 2002. Inferring the root of a phylogenetic tree. *Systematic Biology* 51: 32–43.

IWABE, N., K.-I. KUMA, M. HASEGAWA, S. OSAWA, and T. MIYATA. 1989. Evolutionary relationship of archaebacteria, eubacteria, and eukaryotes inferred from phylogenetic trees of duplicated genes. *Proceedings of the National Academy of Sciences USA* 86: 9355–9359.

JUKES, T. H. and C. R. CANTOR. 1969. Evolution of protein molecules. Pp. 21–132 in *Mammalian protein metabolism*, ed. H. N. Munro. New York: Academic Press.

KORALL, P. and P. KENRICK. 2002. Phylogenetic relationships in Selaginellaceae based on *rbcL* gene sequences. *American Journal of Botany* 89: 506–517.

———, ———, and J. P. THERRIEN. 1999. Phylogeny of Selaginellaceae: evaluation of generic/subgeneric relationships based on *rbcL* gene sequences. *International Journal of Plant Sciences* 160: 585–594.

KOSAKOVSKY POND, S. L., S. D. W. FROST, and S. V. MUSE. 2005. HyPhy: hypothesis testing using phylogenies. *Bioinformatics* 21: 676–679.

LI, W.-H. 1997. *Molecular evolution*. Sunderland: Sinauer Associates.

——— and C.-I. WU. 1985. Rates of nucleotide substitution are evidently higher in rodents than in man. *Molecular Biology and Evolution* 4: 74–77.

MADDISON, W. P., M. J. DONOGHUE, and D. R. MADDISON. 1984. Outgroup analysis and parsimony. *Systematic Zoology* 33: 83–103.

MANHART, J. R. 1994. Phylogenetic analysis of green plant *rbcL* sequences. *Molecular Phylogenetics and Evolution* 3: 114–127.

MATHEWS, S. and M. J. DONOGHUE. 1999. The root of angiosperm phylogeny inferred from duplicate phytochrome genes. *Science* 286: 947–950.

MILINKOVITCH, M. C. and J. LYONS-WEILER. 1998. Finding optimal ingroup topologies and convexities when the choice of outgroups is not obvious. *Molecular Phylogenetics and Evolution* 9: 348–357.

MUSE, S. V. 2000. Examining rates and patterns of nucleotide substitution in plants. *Plant Molecular Biology* 42: 25–43.

PHILIPPE, H., U. SORHANNUS, A. BARION, R. PERASSO, F. GASSE, and A. ADOUTTE. 1994. Comparison of molecular and paleontological data in diatoms suggests a major gap in the fossil record. *Journal of Evolutionary Biology* 7: 247–265.

PIGG, K. B. 1992. Evolution of isoetalean lycopsids. *Annals of the Missouri Botanical Garden* 79: 589–612.

———. 2001. Isoëtalean lycopsid evolution: from the Devonian to the present. *American Fern Journal* 91: 99–114.

POSADA, D. and K. A. CRANDALL. 1998. Modeltest: testing the model of DNA substitution. *Bioinformatics* 14: 817–818.

RETALLACK, G. J. 1997. Earliest Triassic origin of *Isoëtes* and quillwort evolutionary radiation. *Journal of Paleontology* 71: 500–521.

RONQUIST, F. and J. P. HUELSENBECK. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19: 1572–1574.

RYDIN, C. and N. WIKSTRÖM. 2002. Phylogeny of *Isoëtes* (Lycopsida): resolving basal relationships using *rbcL* sequences. *Taxon* 51: 83–89.

SHIMODAIRA, H. and M. HASEGAWA. 1999. Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Molecular Biology and Evolution* 16: 1114–1116.

SKOG, J. E. and C. R. HILL. 1992. The Mesozoic herbaceous lycopsids. *Annals of the Missouri Botanical Garden* 79: 648–675.

SOLTIS, P. S., D. E. SOLTIS, P. G. WOLF, D. L. NICKRENT, S.-M. CHAW, and R. L. CHAPMAN. 1999. The phylogeny of land plants inferred from 18S rDNA sequences: pushing the limits of rDNA signal? *Molecular Biology and Evolution* 16: 1774–1784.

SWOFFORD, D. L. 2002. *PAUP*: Phylogenetic analysis using parsimony (*and other methods)*, version 4.0b10. Sunderland: Sinauer Associates.

———, G. J. OLSEN, P. J. WADDELL, and D. M. HILLIS. 1996. Phylogenetic inference. Pp. 407–514 in *Molecular systematics*, second edition, ed. D. M. Hillis, C. Moritz, and B. K. Mable. Sunderland: Sinauer Associates.

TAYLOR, W. C. and R. J. HICKEY. 1992. Habitat, evolution, and speciation in *Isoëtes*. *Annals of the Missouri Botanical Garden* 79: 613–622.

———, A. R. LEKSCHAS, Q. F. WANG, X. LIU, N. S. NAPIER, and S. B. HOOT. 2004. Phylogenetic relationships of *Isoëtes* (Isoëtaceae) in China as revealed by nucleotide sequences of the nuclear ribosomal ITS region and the second intron of a *LEAFY* homolog. *American Fern Journal* 94: 196–205.

WHEELER, W. C. 1990. Nucleic acid sequence phylogeny and random outgroups. *Cladistics* 6: 363–368.

WIKSTRÖM, N. and P. KENRICK. 1997. Phylogeny of Lycopodiaceae (Lycopsida) and the relationships of *Phylloglossum drummondii* Kunze based on *rbcL* sequences. *International Journal of Plant Sciences* 158: 862–871.

——— and ———. 2001. Evolution of Lycopodiaceae (Lycopsida): estimating divergence times from *rbcL* gene sequences by use of nonparametric rate smoothing. *Molecular Phylogenetics and Evolution* 19: 177–186.

YANG, Z. 1994. Estimating the pattern of nucleotide substitution. *Journal of Molecular Evolution* 39: 105–111.

APPENDIX 1

Species sampled, with voucher information and GenBank accession numbers. For newly reported sequences, collection and deposition information is provided; for previously published sequences, a literature citation is provided. GenBank accession numbers are listed in the following order: *rbcL*, *atpB-rbcL* spacer, ITS region.

Ginkgo biloba L.; Hasebe et al. (1992): D10733. *Huperzia hippuridea* (Christ) Holub; Wikström and Kenrick (1997): Y07931. *Isoëtes abyssinica* Chiov.; *Gastony 97–101* (MIL): DQ294238, DQ280350, DQ284988. *Isoëtes australis* S. Williams (*DR*); *Taylor 6376* (MIL): DQ294239, DQ280351, DQ284989. *Isoëtes australis* S. Williams (*MR*); *Seigler & Maslin 14701a* (MIL): DQ294240, DQ280352, DQ284990. *Isoëtes capensis* Duthie; *Musselman 99201* (MIL): DQ294241, DQ280353, DQ284991. *Isoëtes coromandelina* L. f.; *Srivastava s.n. Aug 97* (MIL): DQ294242, DQ280354, DQ284992. *Isoëtes*

drummondii A. Braun; *Hoot s.n.* (UWM): DQ294243, DQ280355, DQ284993. *Isoëtes histrix* Bory & Durieu.; Rydin and Wikström (2002): AF404497; *Wanntorp NR5350* (MIL): DQ280356, DQ284994. *Isoëtes kirkii* A. Braun; *Woodland & Cutten s.n.* (MIL): DQ294244, DQ280357, Taylor et al. (2004): AY641100. *Isoëtes malinverniana* Ces. & De Not.; *Zurich Botanic Garden accession 15.9.02*: DQ294245, DQ280358, DQ284995. *Isoëtes melanopoda* J. Gay & Durieu; Manhart (1994): L11054; *Leonard s.n.* (MIL): DQ280359, DQ284996. *Isoëtes nuttallii* A. Braun; *Taylor s.n.* (MIL): DQ294246, DQ280360, DQ284997. *Isoëtes orcuttii* Eaton; *Taylor s.n.* (MIL): DQ294247, DQ280361, DQ284998. *Isoëtes panamensis* Maxon & C.V. Morton; *Taylor 6087* (MIL): DQ294248, DQ280362, DQ284999. *Isoëtes setacea* Lam.; *Prada s.n.* (MIL): DQ294249, DQ280363, DQ285000. *Isoëtes taiwanensis* De Vol; *Chiou s.n.* (MIL): DQ294250, DQ280364, Taylor et al. (2004): AY641101. *Isoëtes velata* A. Braun; *Prada s.n.* (MIL): DQ294251, DQ280365, DQ285001. *Lycopodiella inundata* (L.) Holub; Wikström and Kenrick (1997): Y07938. *Lycopodium clavatum* L.; Wikström and Kenrick (1997): Y07936. *Phylloglossum drummondii* Kunze; Wikström and Kenrick (1997): Y07939. *Selaginella apoda* (L.) Spring; Korall et al. (1999): AJ010854. *Selaginella deflexa* Brack.; Korall et al. (1999): AF093253. *Selaginella lepidophylla* (Hook. & Grev.) Spring; Korall et al. (1999): AF093254. *Selaginella selaginoides* (L.) P. Beauv.; Korall et al. (1999): Y07940.