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

ERIC SCHUETTPELZ1 and SARA B. HOOT2,3
1Department of Biology, Duke University, Durham, North Carolina 27708, U.S.A.
2Department of Biological Sciences, University of Wisconsin, Milwaukee, Wisconsin 53201, U.S.A.
3Author 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; Solis 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 (Madison 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 (Swoford 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 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 276 taxa in 16 major clades of Isoëtes (Appendix 1). Except for L. coromandelina (Coromandelina) and its closest (but still very distant) relatives. The clock assumption is violated (Huelsenbeck et al. 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 Lycopsid s (including Huperzia, Lycopodiella, Lycopsid, 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 this 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, TMMef + G, using the hierarchical likelihood ratio test approach as implemented in Modeltest v3.0b (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 = Isoetes; S = Selaginella; L = Lycopodium s.l.), with statistics and parameter estimates corresponding to the maximum parsimony and maximum likelihood analyses.

<table>
<thead>
<tr>
<th>Data</th>
<th>rbcL (I)</th>
<th>rbcL (I + S)</th>
<th>rbcL (I + L)</th>
<th>atpB-rbcL spacer (I)</th>
<th>ITS region (I)</th>
<th>Combined data (I)</th>
<th>Combined data (I + S)</th>
<th>Combined data (I + L)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alignment length</td>
<td>1300</td>
<td>1300</td>
<td>1300</td>
<td>730</td>
<td>660</td>
<td>2690</td>
<td>2690</td>
<td>2690</td>
</tr>
<tr>
<td>Variable characters</td>
<td>66</td>
<td>393</td>
<td>275</td>
<td>91</td>
<td>270</td>
<td>427</td>
<td>754</td>
<td>636</td>
</tr>
<tr>
<td>Missing data (%)</td>
<td>0.856</td>
<td>1.181</td>
<td>0.719</td>
<td>5.882</td>
<td>3.730</td>
<td>2.930</td>
<td>12.918</td>
<td>12.716</td>
</tr>
<tr>
<td>MP analyses</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Informative characters</td>
<td>39</td>
<td>263</td>
<td>178</td>
<td>69</td>
<td>199</td>
<td>307</td>
<td>531</td>
<td>446</td>
</tr>
<tr>
<td>Number of trees</td>
<td>1</td>
<td>1</td>
<td>5</td>
<td>8</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Tree length</td>
<td>82</td>
<td>565</td>
<td>383</td>
<td>113</td>
<td>486</td>
<td>684</td>
<td>1167</td>
<td>984</td>
</tr>
<tr>
<td>CI</td>
<td>0.829</td>
<td>0.807</td>
<td>0.770</td>
<td>0.876</td>
<td>0.739</td>
<td>0.769</td>
<td>0.783</td>
<td>0.765</td>
</tr>
<tr>
<td>RI</td>
<td>0.860</td>
<td>0.801</td>
<td>0.788</td>
<td>0.916</td>
<td>0.795</td>
<td>0.822</td>
<td>0.810</td>
<td>0.808</td>
</tr>
<tr>
<td>ML analyses</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Best-fit model</td>
<td>TrN + I + G</td>
<td>TIMef + G</td>
<td>TrN + G</td>
<td>K81uf + G</td>
<td>HKY + G</td>
<td>TIM + I + G</td>
<td>TIM + I + G</td>
<td>TrN + I + G</td>
</tr>
<tr>
<td>Frequency (A)</td>
<td>0.2727</td>
<td>0.2500</td>
<td>0.2745</td>
<td>0.3281</td>
<td>0.1723</td>
<td>0.2583</td>
<td>0.2526</td>
<td>0.2606</td>
</tr>
<tr>
<td>Frequency (C)</td>
<td>0.1992</td>
<td>0.2500</td>
<td>0.1984</td>
<td>0.1732</td>
<td>0.2748</td>
<td>0.2141</td>
<td>0.2206</td>
<td>0.2137</td>
</tr>
<tr>
<td>Frequency (G)</td>
<td>0.2388</td>
<td>0.2500</td>
<td>0.2342</td>
<td>0.1670</td>
<td>0.2852</td>
<td>0.2388</td>
<td>0.2453</td>
<td>0.2353</td>
</tr>
<tr>
<td>Frequency (T)</td>
<td>0.2893</td>
<td>0.2500</td>
<td>0.2929</td>
<td>0.3337</td>
<td>0.2677</td>
<td>0.2888</td>
<td>0.2815</td>
<td>0.2904</td>
</tr>
<tr>
<td>Rate (A-C)</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
<tr>
<td>Rate (A-G)</td>
<td>3.5776</td>
<td>2.5026</td>
<td>6.3140</td>
<td>2.5993</td>
<td>2.0437</td>
<td>2.3400</td>
<td>3.3925</td>
<td>3.3925</td>
</tr>
<tr>
<td>Rate (A-T)</td>
<td>1.0000</td>
<td>0.3095</td>
<td>1.0000</td>
<td>0.3144</td>
<td>0.6474</td>
<td>0.5561</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
<tr>
<td>Rate (C-G)</td>
<td>1.0000</td>
<td>0.3095</td>
<td>1.0000</td>
<td>0.3144</td>
<td>0.6474</td>
<td>0.5561</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
<tr>
<td>Rate (C-T)</td>
<td>7.4101</td>
<td>5.1885</td>
<td>13.0066</td>
<td>2.5993</td>
<td>3.2493</td>
<td>4.1333</td>
<td>5.5894</td>
<td>5.5894</td>
</tr>
<tr>
<td>Rate (G-T)</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
<tr>
<td>Ti-Tv ratio</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>1.4563</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Alpha (Shape)</td>
<td>0.9371</td>
<td>0.4026</td>
<td>0.2420</td>
<td>0.2906</td>
<td>0.4252</td>
<td>0.6515</td>
<td>0.6280</td>
<td>0.6507</td>
</tr>
<tr>
<td>Pinvar</td>
<td>0.7779</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.6132</td>
<td>0.3700</td>
<td>0.4730</td>
</tr>
<tr>
<td>Number of trees</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>ln L</td>
<td>-2382.82795</td>
<td>-4464.43866</td>
<td>-3764.81484</td>
<td>-1681.06660</td>
<td>-3243.11715</td>
<td>-7659.59792</td>
<td>-9808.83993</td>
<td>-9106.22731</td>
</tr>
<tr>
<td>ln L (clock enforced)</td>
<td>-2393.41843</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
</tbody>
</table>
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 topol-
ysis 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 ap-
proaches: equally weighted maximum parsimony (MP), maximum
likelihood (ML), and Bayesian inference (BI). To assess branch sup-
port with the MP and ML approaches, we conducted bootstrap
analyses (MPBS and MLBS). Because the individual data sets in-
dicated largely congruent results, the data were combined and
analyzed in unison.

MP analyses of the four data sets (three markers plus the com-
bined data set) were performed using PAUP* version 4.0b10 (Swol-
ford 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 v3.06 (Posada and Crandall 1998). Heuristic ML searches
were conducted in PAUP* using the appropriate model of evolu-
tion 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 gener-
ations (1,000 trees) were excluded as the burn-in phase, adequate
for the analyses to reach stationarity. A consensus tree, with av-
erage branch lengths and Bayesian posterior probabilities (BPP),
was computed using the ‘sumtree’ 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 strat-
egy: 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 wheth-
er 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 se-
quence evolution and parameter estimates as identified by Mo-
deltest and the topology resulting from the ML search. We then
estimated the likelihood of observing the data given the model
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 Iso-
ëtes using the outgroup approach, we incorporated in parallel theour 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 (Shi-
modaira 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
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. se-
quencies, 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 Selagi-
ella sequences yielded significant (P ≤ 0.05) results;
33 of 64 comparisons between Isoëtes and Lycopodium
s.l. were significant. Each of the significant compari-
sions indicated a higher rate of evolution in the out-
group. Eighteen of 28 comparisons among the out-
group species were also significant (most of these were
between Selaginella and Lycopodium s.l.), but only six
of the 120 comparisons between ingroup sequences re-
vealed 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 comparis-
sions between ingroup species pairs (Fig. 1). However,
when comparisons between ingroup and outgroup
species were also considered, observed distances be-
gan to approach an asymptote while estimated dis-
tances 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 re-
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
received significant support from all three measures: MPBS \(\geq 70\); MLBS \(\geq 70\); and BPP \(\geq 95\). 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. occitii, 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. malinerniana, 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).
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 infragenic 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
two resolved partitions receiving significant support from all three support measures (MLBS $\geq 70$; MPBS $\geq 70$; and BPP $\geq 0.95$; 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-
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...
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

![Diagram of Isoëtes trees resulting from maximum likelihood analyses with the molecular clock enforced. Lighter bold indicates significant support (MLBS $\geq 70$) from maximum likelihood bootstrap analysis with the molecular clock enforced.](image)
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 Isoetes 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 Isoetes 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 Isoetes into two subgenera (species-poor Euphyllum and species-rich Isoetes; Hickey 1986, 1990) and a subsequent subdivision of subgenus Isoetes into two sections (species-poor Coromandelina and species-rich Isoetes; 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 Isoetes 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 Isoetes 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 Isoetes relationships.

**Acknowledgements.** The authors thank W.-L. Chiou, C. Cuten, G. Gastony, S. Leonard, L. Musselman, C. Prada, J. Schneller, W.C. Taylor, H.-E. Wannberg, 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. Prys for helpful comments and suggestions. This work was supported in part by an NSF grant (DEB991460) to S.B. Hoot.

**Literature Cited**


**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.

<table>
<thead>
<tr>
<th>Species Name</th>
<th>Authors and Accession Numbers</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Isoetes drummondii</em> A. Braun; Hoot s.n. (UWM)</td>
<td>DQ294243, DQ280355, DQ284993</td>
</tr>
<tr>
<td><em>Isoetes histrix</em> Bory &amp; Durieu; Rydin and Wikström (2002)</td>
<td>AF404497; Wannborg NR5350 (MIL): DQ280356, DQ284994</td>
</tr>
<tr>
<td><em>Isoetes kirkii</em> A. Braun; Woodland &amp; Cutten s.n. (MIL)</td>
<td>DQ294244, DQ280357, Taylor et al. (2004): AY641100</td>
</tr>
<tr>
<td><em>Isoetes malinverniana</em> Ces. &amp; De Not.; Zurich Botanic Garden accession 15.9.02</td>
<td>DQ294245, DQ280358, DQ284995</td>
</tr>
<tr>
<td><em>Isoetes melanopoda</em> J. Gay &amp; Durieu; Manhart (1994): L11054</td>
<td>Leonard s.n. (MIL): DQ280359, DQ284996</td>
</tr>
<tr>
<td><em>Isoetes nuttallii</em> A. Braun; Taylor s.n. (MIL)</td>
<td>DQ294246, DQ280360, DQ284997</td>
</tr>
<tr>
<td><em>Isoetes orcuttii</em> Eaton; Taylor s.n. (MIL)</td>
<td>DQ294247, DQ280361, DQ284998</td>
</tr>
<tr>
<td><em>Isoetes panamensis</em> Maxon &amp; C.V. Morton; Taylor 6087 (MIL)</td>
<td>DQ294248, DQ280362, DQ284999</td>
</tr>
<tr>
<td><em>Isoetes setacea</em> Lam.; Prada s.n. (MIL)</td>
<td>DQ294249, DQ280363, DQ285000</td>
</tr>
<tr>
<td><em>Isoetes taiwanensis</em> De Vol; Chiu s.n. (MIL): DQ294250, DQ280364, Taylor et al. (2004): AY641101</td>
<td></td>
</tr>
<tr>
<td><em>Phylloglossum drummondii</em> Kunze; Wikström and Kenrick (1997): Y07939</td>
<td></td>
</tr>
<tr>
<td><em>Selaginella apoda</em> (L.) Spring; Korall et al. (1999): AJ010854</td>
<td></td>
</tr>
<tr>
<td><em>Selaginella de¯exa</em> Brack.; Korall et al. (1999): AF093253</td>
<td></td>
</tr>
<tr>
<td><em>Selaginella lepidophylla</em> (Hook. &amp; Grev.) Spring; Korall et al. (1999): AF093254</td>
<td></td>
</tr>
<tr>
<td><em>Selaginella selaginoides</em> (L.) P. Beauv.; Korall et al. (1999): Y07940</td>
<td></td>
</tr>
</tbody>
</table>