PHYLOGENY AND DIVERGENCE TIME ESTIMATES FOR THE FERN GENUS *AZOLLA* (SALVINIACEAE)

Jordan S. Metzgar,* Harald Schneider,† and Kathleen M. Pryer[1],*

*Department of Biology, Duke University, Durham, North Carolina 27708, U.S.A.; and †Department of Botany, Natural History Museum, London SW7 5BD, United Kingdom

A phylogeny for all extant species of the heterosporous fern genus *Azolla* is presented here based on more than 5000 base pairs of DNA sequence data from six plastid loci (*rbcL*, *atpB*, *rps4*, *trnL-trnF*, *trnG-trnR*, and *rps4-trnS*). Our results are in agreement with other recent molecular phylogenetic hypotheses that support the monophyly of sections *Azolla* and *Rhizosperma* and the proposed relationships within section *Azolla*. Divergence times are estimated within *Azolla* using a penalized likelihood approach, integrating data from fossils and DNA sequences. Penalized likelihood analyses estimate a divergence time of 50.7 Ma (Eocene) for the split between sections *Azolla* and *Rhizosperma*, 32.5 Ma (Oligocene) for the divergence of *Azolla nilotica* from *A. pinnata* within section *Rhizosperma*, and 16.3 Ma (Miocene) for the divergence of the two lineages within section *Azolla* (the *A. filiculoides* + *A. rubra* lineage from the *A. caroliniana* + *A. microphylla* + *A. mexicana* complex).

Keywords: *Azolla*, divergence time estimates, ferns, freshwater plants, heterospory, Salviniaceae.

Introduction

Because of its great agricultural importance, *Azolla* Lam. (Salviniaceae) is "the most frequently studied genus of ferns in the world" (Lumpkin 1993, p. 339). *Azolla* has an obligate symbiotic association with the cyanobacterium *Anabaena azollae* Strasburger (Strasburger 1873; Peters and Meeks 1989; Plazinski et al. 1990; Baker et al. 2003). The nitrogen-fixing capabilities of this symbiosis have long been appreciated, promoting the use of *Azolla* as a "green manure" in rice paddies in Southeast Asia (Moore 1969; Peters and Meeks 1989; Wagner 1997).

Azolla is one of five fern genera to exhibit a heterosporous life cycle—a phenomenon observed in less than 1% of extant ferns (Pryer et al. 2004). Together with all other heterosporous ferns, *Azolla* is nested within leptosporangiate ferns and forms the sister lineage to a large clade comprised of tree ferns and polypods (Pryer et al. 2004; Schuettpelz et al. 2006; Schuettpelz and Pryer 2007). Molecular and morphological analyses have shown *Azolla* to be sister to *Salvinia*, together forming a clade that is sister to the three genera (*Marsilea*, *Regnellidium*, and *Pilularia*) that make up extant Marsileaceae (Rothwell and Stockey 1994; Hasebe et al. 1995; Pryer et al. 1995, 2004; Pryer 1999). Many authors (Tan et al. 1986; Peter and Meeks 1989; Saunders and Fowler 1993) have placed *Azolla* in its own monotypic family Azollaceae, but more recently, it has been recognized together with *Salvinia* in Salviniaceae (Smith et al. 2006).

The taxonomic history of *Azolla* has long been contentious and convoluted (for a thorough account, see Reid et al. 2006; but see also Dunham 1986; Dunham and Fowler 1987). *Azolla*

has traditionally been divided into two sections, *Azolla* and *Rhizosperma* (Mettenius 1847; Tan et al. 1986; Wagner 1997; Reid et al. 2006), based primarily on reproductive structures (e.g., section *Azolla* is characterized by three floats in the megaspore apparatus, whereas section *Rhizosperma* has nine).

Section *Rhizosperma* is consistently considered to comprise two species (*Azolla nilotica* and *A. pinnata*; Saunders and Fowler 1992); however, little consensus has been reached on the diversity in section *Azolla*, which has varied from two to five species. For example, the status of *A. rubra* as either a distinct species or a variety of *A. filiculoides* has long been debated (Svenson 1944), although most recently it has been treated as a distinct species (Zimmerman et al. 1989; Saunders and Fowler 1993; Reid et al. 2006). Likewise, the taxonomic status and rank of three sympatric New World taxa (*A. caroliniana*, *A. mexicana*, and *A. microphylla*) have also generated considerable debate. Numerous authors maintain them as distinct species (Svenson 1944; Moore 1969; Saunders and Fowler 1993). Others have contributed evidence or reviews suggesting that they be regarded as a single species (Zimmerman et al. 1989, 1991; Evrard and Van Hove 2004) or that the extent of the taxonomic conundrum is such that some of the names may have been misapplied and will require a detailed study of type material (Dunham 1986; Dunham and Fowler 1987). Reid et al. (2006) produced the first molecular phylogeny of *Azolla* and hypothesized that *A. caroliniana* was distinct whereas *A. microphylla* and *A. mexicana* should be considered the same species.

In this study, we use a data set of six plastid loci (*rbcL*, *atpB*, *rps4*, *trnL-trnF*, *trnG-trnR*, and *rps4-trnS*) to examine the phylogeny of extant *Azolla* taxa; only one of the six loci (*trnL-trnF*) duplicates data from Reid et al. (2006), although we sequenced this locus independently. This allows us to

[1] Author for correspondence; e-mail pryer@duke.edu.

critically evaluate the traditional sectional classification and the conclusions of Evrard and Van Hove (2004) and Reid et al. (2006) regarding the *A. caroliniana-microphylla-mexicana* (CAR-MIC-MEX) complex. We also integrate data from *Azolla*'s rich fossil record to estimate divergence times for all well-supported nodes in our molecular phylogeny.

Material and Methods

Taxonomic Sampling

Twelve accessions of *Azolla* were studied here (table 1): three from section *Rhizosperma* and nine from section *Azolla*. *Salvinia minima* and *Marsilea botryocarpa* were used as outgroup taxa to root the topology.

DNA Isolation, Amplification, and Sequencing

Genomic DNA was extracted from silica-dried leaf material using the DNeasy kit (Qiagen, Valencia, CA) according to the manufacturer's instructions. Six plastid loci (*rbcL*, *atpB*, *rps4*, *trnL-trnF*, *trnG-trnR*, and *rps4-trnS*) were amplified separately for each accession using the polymerase chain reaction (PCR), according to established protocols (Pryer et al. 2004). PCR products were cleaned using Montage columns (Millipore, Bedford, MA) according to the manufacturer's protocol. Sequencing reactions were performed for both strands of all purified PCR products using Big Dye Terminator Cycle Sequencing reagents (Applied Biosystems, Foster City, CA). All sequencing reactions were completed on either an ABI 3700 or an ABI 3730XL automated sequencer (Applied Biosystems). Primer information for amplification and sequencing reactions is shown in table 2. All sequence reads were examined for contamination using the National Center for Biotechnology Information nucleotide-nucleotide BLAST (blastn) function (Altschul et al. 1997). DNA sequence chromatograms were corrected and assembled into contiguous alignments using Sequencher (Gene Codes, Ann Arbor, MI). Seventy-five sequences used in this study were generated specifically for this project and are available in GenBank (table 1).

Sequence Alignments

Sequence alignments were performed by eye using MacClade, version 4.06 (Maddison and Maddison 2003). The protein-coding *rbcL*, *atpB*, and *rps4* data sets were easily aligned (without gaps) across all *Azolla* accessions and outgroup taxa. The noncoding *trnG-trnR*, *trnL-trnF*, and *rps4-trnS* data were also easily aligned but only across all *Azolla* accessions. Because numerous ambiguously aligned regions resulted when the outgroup sequences were added, they were not included in the noncoding locus data sets. The noncoding alignments for the *Azolla* taxa alone did have some insertions and deletions, but none was ambiguous; therefore, no regions were excluded, nor was any gap coding method employed.

Data Set Combinability

Using PAUP*, version 4.0b10 (Swofford 2002), separate maximum parsimony (MP) and maximum likelihood (ML) bootstrap analyses were run for each data set. The MP analy-

ses were run for 1000 replicates with 10 random addition sequence replicates each and tree-bisection-reconnection (TBR) branch swapping. The ML analyses were run for 500 replicates, each with 10 random addition sequence replicates. In addition, four independent Bayesian Markov chain Monte Carlo (B/MCMC) analyses, each using four chains, were run for each data set, using MrBayes, version 3.1.1 (Ronquist and Huelsenbeck 2003). These analyses were run for 10 million generations, with trees being sampled every 1000 generations; the first 2.5 million generations were discarded as the burn-in phase. The 18 resultant topologies (six from MP, six from ML, and six from B/MCMC) were inspected for topological conflicts using a threshold of 70% bootstrap proportion or higher for the MP and ML analyses and a threshold of 0.95 posterior probability or higher for the B/MCMC analyses. No topological conflict was visually detected among data sets or analytical methods, and hence all six data sets were combined into a single data set.

Phylogenetic Analyses

As noted, an excess of ambiguously aligned regions resulted in the alignments of the noncoding data sets (*trnG-trnR*, *trnL-trnF*, and *rps4-trnS*) when outgroups were included. Rather than excluding much of the data for these three loci, we performed two sets of phylogenetic analyses to maximize the information available (similar to the approach in Nagalingum et al. 2007). Analysis 1 included all *Azolla* accessions, plus the outgroups, for only the protein-coding loci (*atpB*, *rbcL*, and *rps4*). Because this analysis unequivocally established section *Rhizosperma* as the monophyletic sister to the rest of the genus (see "Results"), it was then used as an outgroup in analysis 2, which focused on exploring relationships within section *Azolla*. Analysis 2 included only the *Azolla* accessions but incorporated all six loci (*atpB*, *rbcL*, *rps4*, *rps4-trnS*, *trnG-trnR*, and *trnL-trnF*). No ambiguous regions were present in these alignments, and no sequence data were excluded.

For both analyses 1 and 2, MP and ML analyses were run using PAUP*, version 4.0b10 (Swofford 2002), and B/MCMC analyses were performed using MrBayes, version 3.1.1 (Ronquist and Huelsenbeck 2003). All MP heuristic searches were for 1000 replicates with TBR branch swapping, and all MP bootstrap analyses were performed using 1000 replicates, each with 10 random addition sequence replicates and TBR branch swapping. The ML heuristic searches were run with TBR branch swapping, and the ML bootstrap analyses all included 500 replicates with 10 random addition sequence replicates each and TBR branch swapping. Each B/MCMC analysis was performed using four independent tree searches. Each of these searches used four chains, running for 10 million generations apiece. The B/MCMC analyses were performed with data partitioned by locus, flat priors, and trees being sampled every 1000 generations. The likelihood versus generation scores were plotted for all four searches, and stationarity was determined to have occurred after 2,500,000 generations in each analysis. The prestationarity trees were excluded as the burn-in period. For each B/MCMC analysis, post-burn-in trees from all four runs were pooled, and a majority-rule consensus tree with average branch lengths and posterior probabilities computed from 30,000 trees, using the "sumt" command in MrBayes.

Table 1

Specimens Sampled in This Study, Voucher Information, and Database Accession Numbers

Taxon	Voucher information	GenBank accession nos.						Fern DNA database number[a]
		rbcL	*atpB*	*rps4*	*rps4-trnS*	*trnL-trnF*	*trnG-trnR*	
Section *Azolla:*								
Azolla caroliniana Willd.	USA: Ohio, W. Tuleke s.n. (DUKE); CA3001 in IRRI[b]	EF520919	EF520869	EF520906	EF520932	EF520882	EF520894	2105
A. filiculoides Lam.	Brazil: Paraná, Rolândia, R. Loeb Caldenhof s.n. (DUKE)	EF520926	EF520876	EF520913	EF520939	EF520889	EF520901	2112
A. mexicana Schlecht. & Cham. 1	USA: California, Sutter County, Sutter Basin of Sacramento Valley, D. W. Rains s.n. (DUKE)	EF520922	EF520872	EF520909	EF520935	EF520885	EF520897	2108
A. mexicana 2	USA: California, Butte County, Graylodge State Waterfowl Area, S. Ela s.n. (DUKE); ME2001 in IRRI[b]	EF520923	EF520873	EF520910	EF520936	EF520886	EF520898	2109
A. mexicana 3	Columbia: Leticia, Zimmerman s.n. (DUKE); ME2016 in IRRI[b]	EF520924	EF520874	EF520911	EF520937	EF520887	EF520899	2110
A. microphylla Kaulf. 1	Ecuador: Galapagos, Santa Cruz Island, T. Lumpkin s.n. (DUKE)	EF520929	EF520879	EF520916	EF520942	EF520892	EF520904	2116
A. microphylla 2	Paraguay: Chaco, C. Shock s.n. (DUKE)	EF520921	EF520871	EF520908	EF520934	EF520884	EF520896	2107
A. microphylla 3	Paraguay: Nuevo Italia, Logo Cañada, Watanabe s.n. (DUKE); MI4075 in IRRI[b]	EF520928	EF520878	EF520915	EF520941	EF520891	EF520903	2114
A. rubra R. Br.	Australia: Victoria, Melbourne, N. Nagalingum s.n. (DUKE)	EF520930	EF520880	EF520917	EF520943	EF520893	EF520905	3428
Section *Rhizosperma:*								
A. nilotica Dcne. ex Mett.	Sudan: Kosti, T. Lumpkin s.n. (DUKE); NI5001 in IRRI[b]	EF520925	EF520875	EF520912	EF520938	EF520888	EF520900	2111
A. pinnata R. Br. 1	China: Shandong province, Tancheng District, Bai Kai-Ze s.n. (DUKE); PI0022 in IRRI[b]	EF520920	EF520870	EF520907	EF520933	EF520883	EF520895	2106
A. pinnata 2	Malaysia: Penang, Bumbong Lima, Butterworth s.n. (DUKE); PI0002 in IRRI[b]	EF520927	EF520877	EF520914	EF520940	EF520890	EF520902	2113
Outgroup taxa:								
Marsilea botryocarpa Ballard	Kenya: Tsavo East National Park, R. Faden s.n. (UC)	DQ643294	DQ643259	DQ536326	NA	NA	NA	462
Salvinia minima Baker	USA: Louisiana, S. Fredericq s.n. (DUKE)	EF520931	EF520881	EF520918	NA	NA	NA	2026

Note. NA = data not available for this taxon.
[a] Permanent record numbers in Fern DNA Database at http://www.pryerlab.net/DNA_database.shtml.
[b] International Rice Research Institute (IRRI), Manila.

Table 2

Primers Used for Amplifying and Sequencing DNA Sequences in Salviniales

Primer	Sequence (5′ to 3′)	Reference
trnL-trnF:		
TRNLC	TACGACGATCTYTCTAAACAAGC	Taberlet et al. 1991
TRNLD	GTCGATAAGCYTGAGCTTGTTTAG	Taberlet et al. 1991
TRNLE	GGYAAGATTGCTCAAATACCAG	Taberlet et al. 1991
TRNFF	CARGTTCGACAGCAAGTYTCTCG	Taberlet et al. 1991
rbcL:		
ESRBCL1F	ATGTCACCACAAACGGAGACTAAAGC	Korall et al. 2006
ES645F	AGAYCGTTTCYTATTYGTAGCAGAAGC	Korall et al. 2006
ES663R	TACRAATARGAAACGRTCTCTCCAACG	Korall et al. 2006
ESRBCL1361R	TCAGGACTCCACTTACTAGCTTCACG	Korall et al. 2006
atpB:		
ATPB672F	TTGATACGGGAGCYCCTCTWAGTGT	Wolf 1997
ATPB1163F	ATGGCAGAATRTTTCCGAGATRTYA	Wolf 1997
ATPB1419F	CRACATTTGCACATYTRGATGCTAC	Wolf 1997
ATPB1592R	TGTAACGYTGYAAAGTTTGCTTAA	Wolf 1997
ATPB609R	TCRTTDCCTTCRCGTGTACGTTC	Pryer et al. 2004
ATPE384R	GAATTCCAAACTATTCGATTAGG	Pryer et al. 2004
rps4:		
RPS5F	ATGTCCCGTTATCGAGGACCT	Nadot et al. 1995
ESRPS4IF	CGAGAATCTATSGAYTTGAATATAAA	Korall et al. 2006
ESRPS4IR	TTTATATTCAARTCSATAGATTCTCG	Korall et al. 2006
ESRPS4LF	GTCTCGKTATCGMGGACCTCG	Korall et al. 2006
TRNSR	TACCGAGGGTTCGAATC	Smith and Cranfill 2002
trnG-trnR:		
TRNG1F	GCGGGTATAGTTTAGTGGTAA	Nagalingum et al. 2007
TRNG63R	GCGGGAATCGAACCCGCATCA	Nagalingum et al. 2007
TRNG353R	TTGCTTMTAYGACTCGGTG	Korall et al. 2007
TRNR22R	CTATCCATTAGACGATGGACG	Nagalingum et al. 2007
atpB-rbcL:		
ESRBCL26R	GCTTTAGTCTCCGTTTGTGGTGACAT	Korall et al. 2007
ATPBSPACER703R	CCAATGATCTGAGTAATSTATCC	Korall et al. 2007
ATPB609R	TCRTTDCCTTCRCGTGTACGTTC	Pryer et al. 2004

Note.   Primers in italics were used both for amplifying and sequencing.

Models of sequence evolution for all ML and B/MCMC analyses were determined using Modeltest, version 3.6 (Posada and Crandall 1998). For analysis 1, the TrN + G model was used for *atpB* and *rbcL*, HKY + G was used for *rps4*, and TrN + G was used for the combined data set. For analysis 2, the HKY model was used for *atpB*, *rps4*, and *rps4-trnS*. The HKY + G model was used for *trnL-trnF*, K81uf + G was used for *trnG-trnR*, and TrNef + G was used for *rbcL*.

Divergence Time Estimation

The topology obtained from the Bayesian inference search of the combined data set in analysis 2 was used for the penalized likelihood (PL) divergence date analysis. Because PL divergence time estimation programs can experience difficulty with topologies that have extremely short branch lengths, we followed the placeholder approach (as recommended by Sanderson [2003]) by deleting several redundant accessions. We removed *Azolla mexicana* 2, *A. mexicana* 3, *A. microphylla* 1, *A. microphylla* 3, and *A. pinnata* 2. This left us with at least one accession representing each commonly recognized species in *Azolla*. Because of the well-supported (Pryer et al. 2004) intergeneric relationships in the heterosporous fern clade ((*Salvinia*, *Azolla*) (*Marsilea* (*Regnellidium*,

Pilularia))), *Salvinia minima* was added to the tree as nearest sister genus to the *Azolla* clade, and *Marsilea botryocarpa* was added to the tree as nearest sister genus to the *Azolla-Salvinia* clade. Branch lengths for this topology were then re-estimated in PAUP*, version 4.0b10, using the combined *rbcL*, *atpB*, and *rps4* data sets and the ML parameters for the combined three-locus data set (TrN + G, as estimated by Modeltest). *Marsilea botryocarpa* was then pruned from the tree, leaving the *Salvinia-Azolla* node as the root node.

Using r8s (ver. 1.60; Sanderson 2003), we implemented two fossil constraints (table 3) in conjunction with this tree topology. The *Salvinia-Azolla* node was fixed as a calibration point at 89 Ma, using the earliest reports of *Azolla* megaspores (Collinson 1991). The section *Azolla* node was assigned a minimum age constraint of 13.65 Ma using a megaspore fossil of *A. filiculoides* (Mai 2001). This fossil indicates that the first divergence in section *Azolla* had occurred no later than this date. The fossil was not assigned to the node representing the most recent common ancestor of *A. filiculoides* and *A. rubra* because the megaspore morphology of these two species is essentially identical. Instead, the fossil was assigned to the next-lowest node (section *Azolla* node) because it is at least assignable to the *A. filiculoides/A. rubra* lineage.

Table 3

Node Numbers from Figure 2, Their Molecular Age Estimates, Mean Molecular Ages with Standard Deviations (SDs), Fossil Age Constraints, and Relevant Comments

Node	Molecular age estimate (Ma)	Mean molecular age ± SD (Ma)	Fossil age constraint (Ma)	Comments
1	89.0	89.0 ± .0	89.0	Fixed calibration point; earliest fossil *Azolla* megaspores (Santonian; Collinson 1991)
2	50.7	47.9 ± 3.5	NA	
3	32.5	31.0 ± 3.2	NA	
4	16.3	16.6 ± 2.6	13.65	Minimum age constraint; *A. filiculoides* fossil megaspore (Mai 2001)
5	3.1	3.4 ± 1.4	NA	
6	3.9	4.0 ± 1.7	NA	
7	2.3	1.9 ± 1.2	NA	

Note. NA = not available.

A r8s analysis using the three-locus data set was performed with a smoothing parameter of 1 as selected by a cross-validation procedure (Sanderson 2002). Using the procedure described in the r8s manual (Sanderson 2002), we ran 1000 bootstrap replicates. First, 1000 bootstrap data matrices were generated using PAUP*, version 4.0b10. Branch lengths for the tree used in the initial r8s analysis were then reestimated for each of the bootstrap data sets. Finally, divergence times for all the resulting bootstrap trees were calculated by r8s. Out of the 1000 replicates, 984 were successfully completed.

Results

Phylogenetic Analyses

Analysis 1. The MP analysis of the combined three-locus data set (*atpB*, *rbcL*, and *rps4*) produced two equally most parsimonious trees (916 steps, consistency index [CI] = 0.882, retention index [RI] = 0.816), and the strict consensus tree was well resolved (not shown). The ML analysis resulted in one most likely tree ($-\ln L = 8530.8154$; tree not shown). The B/MCMC analysis returned a majority-rule consensus tree with a well-resolved backbone (fig. 1A). Analysis 1 allowed us to robustly confirm the monophyly of *Azolla* from all three measures of support—Bayesian posterior probability (PP = 1.00) and ML and MP bootstrap percentages ($BP_{ML} = 99$ and $BP_{MP} = 100$). We also established the reciprocal monophyly of sections *Azolla* ($BP_{MP} = 100$; $BP_{ML} = 100$; PP = 1.00) and *Rhizosperma* ($BP_{MP} = 94$; $BP_{ML} = 100$; PP = 1.00). The well-supported position of section *Rhizosperma* as sister to the rest of the genus allowed us to use it as the outgroup in analysis 2. No well-supported conflicts were observed among the three analytical methods.

Analysis 2. The MP analysis of the six-locus combined data set resulted in six equally most parsimonious trees (960 steps, CI = 0.969, RI = 0.971), which yielded a well-resolved strict consensus tree (tree not shown). The ML tree search returned one most likely tree ($-\ln L = 13,042.541$; tree not shown). The Bayesian analysis resulted in a majority-rule consensus tree with a well-resolved backbone (fig. 1B). All three analytical methods returned highly congruent topologies without conflicts. *Azolla nilotica* and the accessions of

A. pinnata are sister taxa in a well-supported clade ($BP_{MP} = 100$; $BP_{ML} = 100$; PP = 1.00). This group is sister to the rest of the genus. In this larger clade, *A. filiculoides* and *A. rubra* are sister taxa ($BP_{MP} = 100$; $BP_{ML} = 100$; PP = 1.00) and form a sister group to the *A. caroliniana* clade (CAR-MIC-MEX) ($BP_{MP} = 100$; $BP_{ML} = 100$; PP = 1.00). There is little resolution within this group, but our single accession of *A. caroliniana* is consistently sister to all of the *A. microphylla* and *A. mexicana* accessions ($BP_{MP} = 100$; $BP_{ML} = 100$; PP = 1.00); the latter two species are both not monophyletic. The *A. mexicana* 1 accession was sister to the remaining *A. microphylla* and *A. mexicana* accessions with strong support ($BP_{MP} = 88$; $BP_{ML} = 86$; PP = 1.00); in addition, *A. mexicana* 2 and *A. microphylla* 2 were always sister taxa ($BP_{MP} = 99$; $BP_{ML} = 100$; PP = 1.00), as were *A. mexicana* 3 and *A. microphylla* 3.

There was no conflict between the topologies recovered in analyses 1 and 2 (fig. 1A, 1B). However, we were able to obtain a more resolved topology by conducting analysis 2, which resulted in three well-supported nodes within the CAR-MIC-MEX clade, compared with one in analysis 1 (well supported: $BP_{MP} \geq 70$; $BP_{ML} \geq 70$; PP ≥ 0.95).

Divergence Time Estimates

The average branch lengths of the topology used for the divergence time estimation were based on the *atpB*, *rbcL*, and *rps4* data and are shown in figure 1C. Age estimates from the PL analysis, as well as age estimates and standard deviations from the PL bootstrap analysis for all well-supported nodes in the Bayesian consensus tree, are presented in table 3. The age estimate data are graphically depicted in figure 2 as a chronogram plotted against the geologic timescale.

Our results from the PL r8s analysis estimate the divergence of sections *Azolla* and *Rhizosperma* at 50.7 Ma (Eocene; fig. 2, node 2). *Azolla nilotica* and *A. pinnata* (section *Rhizosperma*) are estimated to have diverged at 32.5 Ma (Oligocene; fig. 2, node 3). The *A. filiculoides-A. rubra* (FIL-RUB) clade and the CAR-MIC-MEX clade (section *Azolla*) diverged at 16.3 Ma (Miocene; fig. 2, node 4), and *A. filiculoides* and *A. rubra* diverged at 3.1 Ma (Pliocene; fig. 2, node 5). *Azolla caroliniana* (CAR-MIC-MEX species complex) diverged from *A. microphylla* and *A. mexicana* at 3.9 Ma (Pliocene; fig. 2,
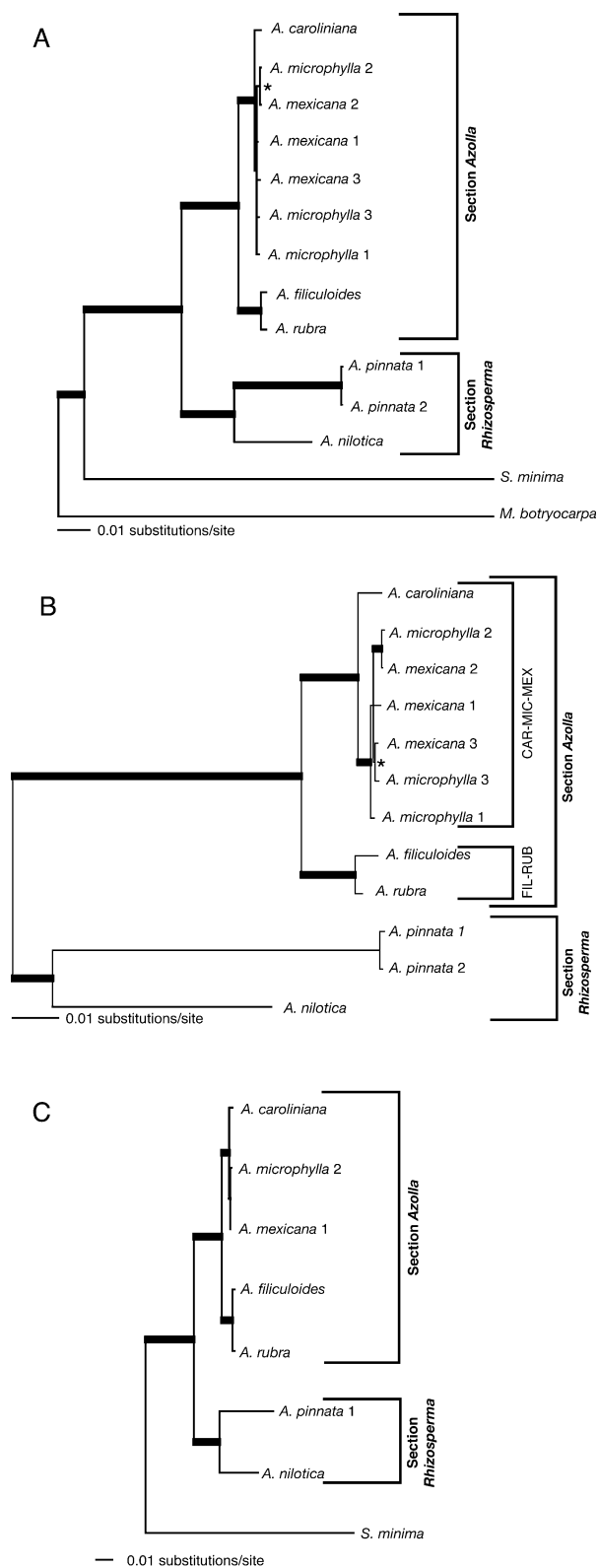
Fig. 1 *A*, Analysis 1: Bayesian phylogram depicting the topology and branch lengths in *Azolla* obtained from the three-locus coding data set (*atpB*, *rbcL*, and *rps4*); *Salvinia* and *Marsilea* are outgroups. An asterisk highlights the well-supported node for *Azolla mexicana* 2 plus *A. microphylla* 2. *B*, Analysis 2: Bayesian phylogram depicting

node 6). Accessions within the MIC-MEX complex diverged at 2.3 Ma (Pliocene; fig. 2, node 7). All estimates from the PL analysis were within 1 SD of their estimates from the bootstrap analysis (table 3).

Discussion

Phylogenetic Relationships

Our six-locus plastid data set resulted in a topology congruent with the results of Reid et al.'s (2006) analysis of nrITS, *trnL-trnF*, and *atpB-rbcL*. Previous studies suggested that *Azolla nilotica* belonged in a separate subgenus, *Tetrasporocarpia* (Saunders and Fowler 1993), due to its diploid chromosome number (2n=52 for *A. nilotica* vs. 2n=44 for all other *Azolla* species) and numerous unique morphological traits (e.g., sporocarps in groups of four and cauline multicellular trichomes). However, both Reid et al.'s (2006) study and our study confirm the monophyly of the traditional sections *Azolla* and *Rhizosperma* (Mettenius 1847; Tan et al. 1986; Wagner 1997), which form well-supported sister clades (fig. 1), each with clearly defined morphological characters (fig. 2). Morphological similarities between *A. pinnata* and section *Azolla* species are likely due to parallel evolution or to autapomorphic changes in *A. nilotica*.

Azolla rubra and *A. filiculoides* are closely related sister taxa (fig. 1*B*) and, as with Reid et al.'s (2006) work, little resolution is achieved within the CAR-MIC-MEX species complex, despite our data set having nearly twice as much data. Accessions of *A. microphylla* and *A. mexicana* are paraphyletic to one another and show little to no evidence of constituting separate evolutionary lineages (see fig. 1*B*). *Azolla caroliniana* is sister to a clade comprising both *A. microphylla* and *A. mexicana* accessions. Reid et al. (2006) showed this same relationship and also provided evidence for the monophyly of *A. caroliniana*. Based on the results of their three-locus study, Reid et al. (2006) recommended recognizing *A. caroliniana* as a separate species but suggested that *A. microphylla* and *A. mexicana* be treated as a single species. The results of our six-locus data support the conclusions of Reid et al. (2006) and do not confirm the morphological conclusion of Evrard and Van Hove (2004) that the entire CAR-MIC-MEX complex should be recognized as a single species. Future investigations into the MIC-MEX complex require greatly increased taxon sampling and the use of additional molecular tools, such as intersimple sequence repeats (Barker and Hauk 2003) and amplified fragment length polymorphisms (Perrie

the topology and branch lengths in *Azolla* sect. *Azolla* topology generated using the six-locus coding and noncoding data set (*atpB*, *rbcL*, *rps4*, *rps4-trnS*, *trnG-trnR*, and *trnL-trnF*); *Azolla* sect. *Rhizosperma* is used as an outgroup. An asterisk highlights the well-supported node for *A. mexicana* 3 plus *A. microphylla* 3. *C*, Summary topology used for divergence time estimation; branch lengths as calculated from the three-locus data set (*atpB*, *rbcL*, and *rps4*). For all topologies shown, thickened lines indicate well-supported branches from all three measures (maximum parsimony bootstrap percentage [$BP_{MP} \geq 70$]; maximum likelihood bootstrap percentage [$BP_{ML} \geq 70$]; posterior probability [$PP \geq 0.95$]).
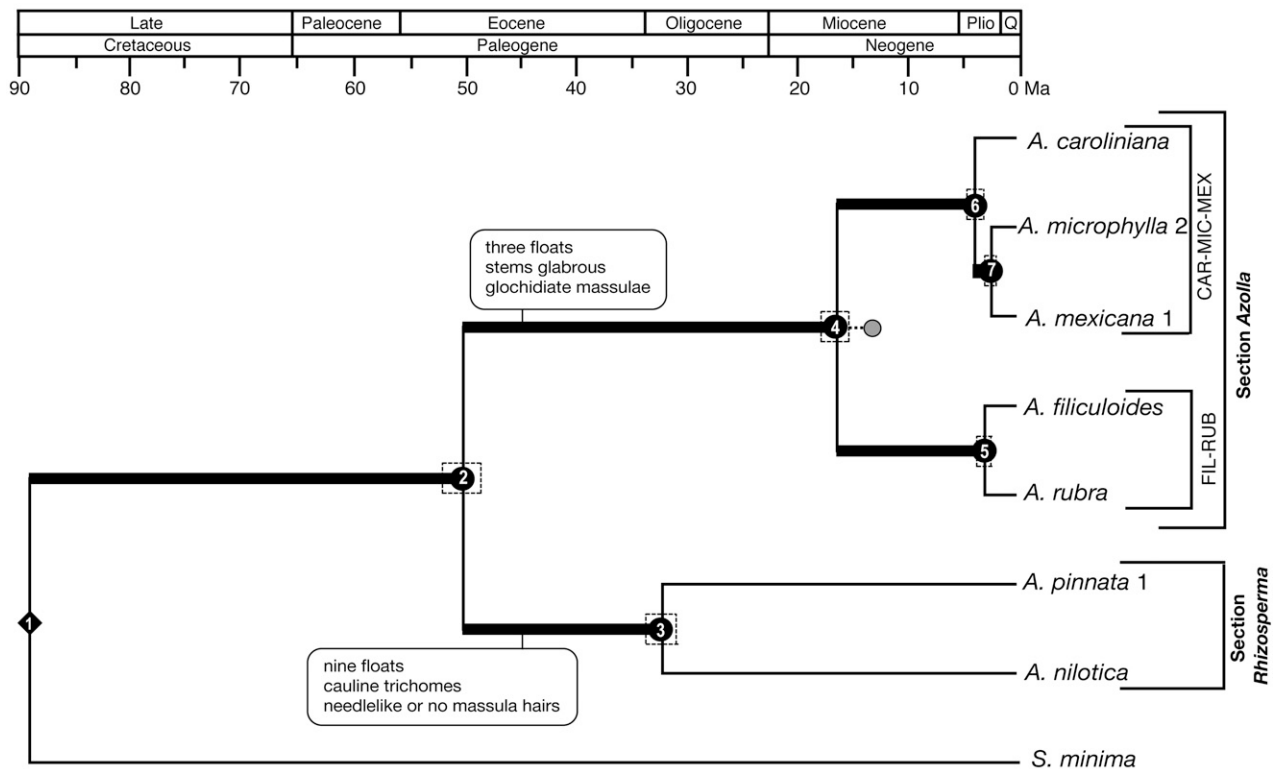
Fig. 2 Phylogenetic chronogram for *Azolla* plotted against the geologic timescale. Divergence time estimates were generated using penalized likelihood with two fossil constraints (see table 3). Divergence time estimates correspond to node positions; the gray circle indicates the minimum age constraint; the black diamond indicates the fixed-age calibration on the root node. Standard deviation for each divergence date estimate is represented by vertical lines of boxes at nodes. Thickened lines indicate highly supported branches according to all three measures (maximum parsimony bootstrap percentage [$BP_{MP} \geq 70$]; maximum likelihood bootstrap percentage [$BP_{ML} \geq 70$]; posterior probability [$PP \geq 0.95$]).

et al. 2003). It is important to note that several taxonomic and nomenclatural issues pertaining to this species complex have been identified in an unpublished thesis (Dunham 1986), as well as alluded to by Dunham and Fowler (1987). These complications will need to be carefully considered in any future study of this group.

Divergence Time Estimates and the Fossil Record

This study is the first to provide a temporal framework for evolution in *Azolla* based exclusively on extant lineages. *Azolla* possesses a tremendous fossil diversity, with more than 50 extinct species of *Azolla* described (Collinson 1980, 2001). Many are represented solely by megaspores and exhibit wide variation in morphology, particularly with regard to float number (Collinson 1980). The earliest stratigraphic occurrence (89 Ma) is the many-floated *A. montana* Hall in the Maastrichtian/Late Cretaceous (table 3; Collinson 1991).

Our 89-Ma fixed-age calibration on the root node is identical to the minimum age constraint on the *Salvinia/Azolla* divergence used by Pryer et al. (2004) and is similar to their molecular age estimate of 89.17 ± 0.69 Ma for the Salviniaceae. We evaluated our molecular divergence time estimates for *Azolla* against the earliest reported occurrences of extant *Azolla* taxa. We estimate the diversification within *Azolla* to

have begun at 50.7 Ma (Eocene; fig. 2, node 2) with the divergence of sections *Azolla* and *Rhizosperma*. However, the first fossil record of an *Azolla* species with a collared megaspore, a trait shared by all extant *Azolla*, does not occur until 37.2–28.4 Ma (Oligocene; *A. colwellensis* Collinson; Collinson 1980).

Azolla pinnata is recorded from the Pliocene (1.8–5.3 Ma) (Florschutz 1945; Batten and Kovach 1990), and the oldest fossil record for *A. nilotica* is from only 550 AD (Birks 2002), whereas our results suggest that *A. pinnata* diverged from *A. nilotica* at 32.5 Ma (Oligocene; fig. 2, node 3). This discordance between the fossil and molecular ages suggests either (1) that the lineages leading to *A. pinnata* and *A. nilotica* diverged in the Oligocene while the extant species did not appear until much more recently (e.g., the Pliocene) or (2) that older fossils of extant species have not yet been found.

Azolla filiculoides is reported from the Langhian (13.65–15.97 Ma) (Mai 2001), and the oldest fossil records for the CAR-MIC-MEX complex are attributed to *A. caroliniana* from the Upper Miocene (5.3–11.6 Ma) (Gamerro 1981; Batten and Kovach 1990). We estimate 16.3 Ma (Miocene) for the divergence of CAR-MIC-MEX from the FIL-RUB lineage (fig. 2, node 4), which is in accordance with the *A. filiculoides* fossil from the Langhian (13.65–15.97 Ma).

Within the CAR-MIC-MEX complex, our analyses indicate the MIC-MEX lineage diverged from *A. caroliniana* at

3.9 Ma (fig. 2, node 6) and that divergences within the MIC-MEX lineage occurred as recently as 1.9 Ma (fig. 2, node 7). One fossil record for *A. microphylla* exists from 50,000 years ago (Schofield and Colinvaux 1969). Given the lack of morphological distinction between the members of the recently derived MIC-MEX lineages, we do not see any conflict between our age estimates and the fossil record for this group. No fossil specimens have been assigned to *A. mexicana*, so no comparisons were possible for this species. *Azolla rubra* also does not have any described fossils, but this is likely due to the overall similarity of *A. rubra* and *A. filiculoides* megaspores.

The literature on *Azolla* fossils consistently refers to a progressive reduction in the number of megaspore floats, from 24, through 18, 15, 12, and nine, to three. Unfortunately, our study of extant species alone cannot inform us about the possible pattern of reduction in megaspore float number in *Azolla* across geologic time (Saunders and Fowler 1993). We can simply determine that there was a basal split in the genus at about 50 Ma (fig. 2, node 2), with one lineage giving rise to extant taxa that have three floats and the other lineage to taxa with nine. Explaining this megaspore float pattern will require the combined phylogenetic analysis of fossil and extant species of *Azolla*.

Literature Cited

Altschul SF, TL Madden, AA Schäffer, J Zhang, Z Zhang, W Miller, DJ Lipman 1997 Gapped blast and psi-blast: a new generation of protein database search programs. Nucleic Acids Res 25:3389–3402.

Baker JA, B Entsch, DB McKay 2003 The cyanobiont in an *Azolla* fern is neither *Anabaena* nor *Nostoc*. FEMS Microbiol Lett 229:43–47.

Barker MS, WD Hauk 2003 An evaluation of *Sceptridium dissectum* (Ophioglossaceae) with ISSR markers: implications for *Sceptridium* systematics. Am Fern J 93:1–19.

Batten DJ, WL Kovach 1990 Catalog of Mesozoic and Tertiary megaspores. Am Assoc Stratigr Palynol Contrib Ser 24:1–227.

Birks HH 2002 The recent extinction of *Azolla nilotica* in the Nile Delta, Egypt. Acta Palaeobot 42:203–213.

Collinson ME 1980 A new multiple-floated *Azolla* from the Eocene of Britain with a brief review of the genus. Palaeontology 23:213–229.

——— 1991 Diversification of modern heterosporous pteridophytes. Pages 119–150 *in* S Blackmore, SH Barnes, eds. Pollen and spores. Clarendon, Oxford.

——— 2001 Cainozoic ferns and their distribution. Brittonia 53:173–235.

Dunham DG 1986 Taxonomic re-evaluation and species recognition of *Azolla* Lam., with particular reference to section *Azolla*. PhD thesis. Portsmouth Polytechnic.

Dunham DG, K Fowler 1987 Taxonomy and species recognition in *Azolla* Lam. Pages 7–16 *in Azolla* utilization. International Rice Research Institute, Manila.

Evrard C, C Van Hove 2004 Taxonomy of the American *Azolla* species (Azollaceae): a critical review. Syst Geogr Plants 74:301–318.

Florschutz F 1945 *Azolla* vit het Nederlandsche Palaeoceen en Pleistoceen. Verh Geol-Mijnb Gen Ned Kol 14:19–197.

Gamerro JC 1981 *Azolla* y *Salvinia* (Pteridophyta, Salviniales) en la formacion Parana (Mioceno superior), Santa Fe, Republica Argentina. IV Symp Argent Paleobot Palinol 3:12–13.

Hasebe M, PG Wolf, KM Pryer, K Ueda, M Ito, R Sano, GJ Gastony, et al 1995 Fern phylogeny based on *rbcL* nucleotide sequences. Am Fern J 85:134–181.

Korall P, DS Conant, JS Metzgar, H Schneider, KM Pryer 2007 A molecular phylogeny of scaly tree ferns (Cyatheaceae). Am J Bot 94:873–886.

Korall P, KM Pryer, JS Metzgar, H Schneider, DS Conant 2006 Tree ferns: monophyletic groups and their relationships as revealed by four protein-coding plastid loci. Mol Phylogenet Evol 39:830–845.

Lumpkin TA 1993 *Azolla*. Pages 338–342 *in* Flora of North America Editorial Committee, eds. Flora of North America. Vol 2. Oxford University Press, New York.

Maddison WP, DR Maddison 2003 MacClade: analysis of phylogeny and character evolution. Version 4.06. Sinauer, Sunderland, MA.

Mai DH 2001 Die mittelmiozaenen und obermiozaenen Floren aus der Meuroer und Raunoer Folge in der Lausitz. III. Fundstellen und Palaeobiologie. Palaeontogr Abt B 258:1–85.

Mettenius GH 1847 Über *Azolla*. Linnaea 20:259–282.

Moore AW 1969 *Azolla*: biology and agronomic significance. Bot Rev 35:17–34.

Nadot S, G Bittar, L Carter, R LaCroix, B Lejeune 1995 A phylogenetic analysis of monocotyledons based on the chloroplast gene *rps4*, using parsimony and a new numerical phenetics method. Mol Phylogenet Evol 4:257–282.

Nagalingum N, H Schneider, KM Pryer 2007 Molecular phylogenetic relationships and morphological evolution in the heterosporous fern genus *Marsilea*. Syst Bot 32:16–25.

Perrie LR, PJ Brownsey, PJ Lockhart, EA Brown, MF Large 2003 Biogeography of temperate Australasian *Polystichum* ferns as inferred from chloroplast sequence and AFLP. J Biogeogr 30:1729–1736.

Peters GA, JC Meeks 1989 The *Azolla-Anabaena* symbiosis: basic biology. Annu Rev Plant Physiol Plant Mol Biol 40:193–210.

Plazinski J, Q Zheng, R Taylor, L Croft, BG Rolfe 1990 DNA probes show genetic variation in cyanobacterial symbionts of the *Azolla* fern and a closer relationship to free-living *Nostoc* strains than to free-living *Anabaena* strains. Appl Environ Microbiol 56:1263–1270.

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

Pryer KM 1999 Phylogeny of marsileaceous ferns and relationships of the fossil *Hydropteris pinnata* reconsidered. Int J Plant Sci 160: 931–954.

Pryer KM, E Schuettpelz, PG Wolf, H Schneider, AR Smith, R Cranfill 2004 Phylogeny and evolution of ferns (monilophytes) with a focus on the early leptosporangiate divergences. Am J Bot 91: 1582–1598.

Pryer KM, AR Smith, JE Skog 1995 Phylogenetic relationships of extant ferns based on evidence from morphology and *rbcL* sequences. Am Fern J 85:205–282.

Reid JD, GM Plunkett, GA Peters 2006 Phylogenetic relationships in the heterosporous fern genus *Azolla* (Azollaceae) based on DNA sequence data from three noncoding sequence regions. Int J Plant Sci 167:529–538.

Ronquist F, JP Huelsenbeck 2003 MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19:1572–1574.

Rothwell GW, RA Stockey 1994 The role of *Hydropteris pinnata* gen. et sp. nov. in reconstructing the cladistics of heterosporous ferns. Am J Bot 81:479–492.

Sanderson MJ 2002 Estimating absolute rates of molecular evolution and divergence times: a penalized likelihood approach. Mol Biol Evol 19:101–109.

——— 2003 r8s: inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. Bioinformatics 19:301–302.

Saunders RMK, K Fowler 1992 A morphological taxonomic revision of *Azolla* Lam. section *Rhizosperma* (Mey.) Mett. (Azollaceae). Bot J Linn Soc 109:329–357.

——— 1993 The supraspecific taxonomy and evolution of the fern genus *Azolla* (Azollaceae). Plant Syst Evol 184:175–193.

Schofield EK, PA Colinvaux 1969 Fossil *Azolla* from the Galapagos Islands. Bull Torrey Bot Club 96:623–628.

Schuettpelz E, P Korall, KM Pryer 2006 Plastid *atpA* data provide improved support for deep relationships among ferns. Taxon 55: 897–906.

Schuettpelz E, KM Pryer 2007 Fern phylogeny inferred from 400 leptosporangiate species and three plastid genes. Taxon 56 (forthcoming).

Smith AR, RB Cranfill 2002 Intrafamilial relationships of the thelypteroid ferns (Thelypteridaceae). Am Fern J 92:131–149.

Smith AR, KM Pryer, E Schuettpelz, P Korall, H Schneider, R Cranfill, PG Wolf 2006 A classification for extant ferns. Taxon 55:705–731.

Strasburger E 1873 Über *Azolla*. Dabis, Jena.

Svenson HK 1944 The New World species of *Azolla*. Am Fern J 34: 69–84.

Swofford DL 2002 PAUP*: phylogenetic analysis using parsimony (*and other methods). Version 4.0b10. Sinauer, Sunderland, MA.

Taberlet P, L Gielly, G Pautou, J Bouvet 1991 Universal primers for amplification of three non-coding regions of chloroplast DNA. Plant Mol Biol 17:1105–1109.

Tan BC, P Payawal, I Watanabe, N Lacdan, C Ramirez 1986 Modern taxonomy of *Azolla*: a review. Philipp Agric 69:491–512.

Wagner GM 1997 *Azolla*: a review of its biology and utilization. Bot Rev 63:1–26.

Wolf PG 1997 Evaluation of *atpB* nucleotide sequences for phylogenetic studies of ferns and other pteridophytes. Am J Bot 84:1429–1440.

Zimmerman WJ, TA Lumpkin, I Watanabe 1989 Classification of *Azolla* spp., section *Azolla*. Euphytica 43:223–232.

Zimmerman WJ, I Watanabe, T Ventura, P Payawal, TA Lumpkin 1991 Aspects of the genetic and botanical status of Neotropical *Azolla* species. New Phytol 119:561–566.