Available online at www.sciencedirect.com
SCIENCE DIRECT®
Molecular Phylogenetics and Evolution 36 (2005) 484–493
MOLECULAR PHYLOGENETICS AND EVOLUTION
www.elsevier.com/locate/ympev

Incongruence between primary sequence data and the distribution of a mitochondrial *atp1* group II intron among ferns and horsetails

Niklas Wikström [a,*], Kathleen M. Pryer [b]

[a] *Department of Systematic Botany, Evolutionary Biology Centre, Uppsala University, Norbyvägen 18D, 752 36 Uppsala, Sweden*
[b] *Department of Biology, Duke University, Durham, NC 27708, USA*

Received 1 June 2004; revised 17 February 2005
Available online 26 May 2005

Abstract

Using DNA sequence data from multiple genes (often from more than one genome compartment) to reconstruct phylogenetic relationships has become routine. Augmenting this approach with genomic structural characters (e.g., intron gain and loss, changes in gene order) as these data become available from comparative studies already has provided critical insight into some long-standing questions about the evolution of land plants. Here we report on the presence of a group II intron located in the mitochondrial *atp1* gene of leptosporangiate and marattioid ferns. Primary sequence data for the *atp1* gene are newly reported for 27 taxa, and results are presented from maximum likelihood-based phylogenetic analyses using Bayesian inference for 34 land plants in three data sets: (1) single-gene mitochondrial *atp1* (exon + intron sequences); (2) five combined genes (mitochondrial *atp1* [exon only]; plastid *rbcL*, *atpB*, *rps4*; nuclear SSU rDNA); and (3) same five combined genes plus morphology. All our phylogenetic analyses corroborate results from previous fern studies that used plastid and nuclear sequence data: the monophyly of euphyllophytes, as well as of monilophytes; whisk ferns (Psilotidae) sister to ophioglossoid ferns (Ophioglossidae); horsetails (Equisetopsida) sister to marattioid ferns (Marattiidae), which together are sister to the monophyletic leptosporangiate ferns. In contrast to the results from the primary sequence data, the genomic structural data (*atp1* intron distribution pattern) would seem to suggest that leptosporangiate and marattioid ferns are monophyletic, and together they are the sister group to horsetails—a topology that is rarely reconstructed using primary sequence data.

Keywords: *atp1*; *atpB*; Bayesian inference; Euphyllophytes; Ferns; Horsetails; Mitochondrial intron; Monilophytes; Morphology; *rbcL*; *rps4*; SSU rDNA

1. Introduction

The last 10 years of comparative genomics has seen an accelerated increase in the availability of genomic structural characters to phylogenetic studies. These data include the gain and loss of introns, loss or transfer of genes, gene duplication, changes in gene order, and the formation and disruption of gene clusters. The complex nature of these data has provided critical insight into some long-standing questions about the evolution of land plants. For example, the broad phylogenetic relationships that are indicated by the conservative distribution of plastid introns, such as the shared presence of two tRNA introns in the Charophyceae (green algae) and all land plants (Manhart and Palmer, 1990), and by structural rearrangements in the plastid genome, such as the 30-kb inversion shared by all vascular plants except lycophytes (Raubeson and Jansen, 1992a), are largely uncontested. However, recent phylogenetic studies that place Gnetales within conifers suggest that the absence of one of the two inverted repeats in plastids of conifers (Raubeson and Jansen, 1992b; Strauss et al., 1988) may

* Corresponding author. Fax: +46 18 471 6457.
E-mail addresses: Niklas.Wikstrom@ebc.uu.se (N. Wikström), pryer@duke.edu (K.M. Pryer).

1055-7903/$ - see front matter © 2005 Elsevier Inc. All rights reserved.
doi:10.1016/j.ympev.2005.04.008

be more complicated to interpret than initially perceived, given that Gnetales have both inverted repeat copies (Bowe et al., 2000; Chaw et al., 2000; Donoghue and Doyle, 2000; Qiu et al., 1999; Qiu and Palmer, 1999). Data on mitochondrial group I and group II introns (Beckert et al., 1999, 2001; Dombrovska and Qiu, 2004; Malek and Knoop, 1998; Pruchner et al., 2001; Qiu et al., 1998; Vangerow et al., 1999) is also proving critical, and although some mitochondrial patterns appear less conservative than those observed in plastids (Cho et al., 1998; Palmer et al., 2000), others do provide valuable information on relationships among major land plant groups (Dombrovska and Qiu, 2004; Qiu and Lee, 2000; Qiu and Palmer, 1999). Until now, the screening for mitochondrial introns has mainly focused on the relationships among liverworts, hornworts, and mosses (Beckert et al., 2001; Malek and Knoop, 1998; Pruchner et al., 2001; Qiu et al., 1998), and the question of which of these groups is most closely related to vascular plants.

Most phylogenetic analyses (Duff and Nickrent, 1999; Hedderson et al., 1998; Hiesel et al., 1994; Kenrick and Crane, 1997; Kranz and Huss, 1996; Raubeson and Jansen, 1992a) consistently recognize lycophytes (Lycophytina: clubmosses and relatives) as sister group to all remaining vascular plants, or euphyllophytes (Euphyllophytina, sensu Kenrick and Crane, 1997). Euphyllophytes comprise six major monophyletic lineages: Spermatophytata (seed plants), Equisetopsida (horsetails), Polypodiidae (leptosporangiate ferns), Psilotidae (whisk ferns), Marattiidae (marattioid ferns), and Ophioglossidae (ophioglossoid ferns). The relationships among these euphyllophyte lineages have been notoriously difficult to resolve. Phylogenetic analyses based on single genes (Duff and Nickrent, 1999; Hedderson et al., 1998; Hiesel et al., 1994; Kranz and Huss, 1996; Manhart, 1994, 1995), and/or morphology (Kenrick and Crane, 1997; Mishler et al., 1994; Pryer et al., 1995; Rothwell, 1999; Stevenson and Loconte, 1996) have either been incongruent, or only provided weak support for certain relationships. Pryer et al. (2001) partly resolved these difficulties. Their analyses, based on a combined data set of four genes and morphology, provided unequivocal support for the monophyly of all "seed-free" euphyllophyte lineages (Equisetopsida, Psilotidae, Marattiidae, Ophioglossidae, and Polypodiidae), and determined that together they formed the sister group to seed plants. Citing Beck and Stein (1993), Kenrick and Crane (1997) had already suggested the potential monophyly of this group based on a single anatomical character (shared protoxylem distribution), and in their provisional classification, this horsetail-fern clade was named Moniliformopses ("monilophytes," Judd et al., 2002; Pryer et al., 2004a,b). In addition to the morphological and primary sequence data supporting the monophyly of monilophytes, Pryer et al. (2001) documented a 9-bp length increase in the plastid protein coding gene *rps4* that was shared by all moniliophytes. Pryer et al. (2001) also provided strong support for grouping whisk ferns (Psilotidae) and ophioglossoid ferns (Ophioglossidae), a relationship indicated by earlier single-gene analyses (Hedderson et al., 1998; Manhart, 1994; Pryer et al., 1995), but never before unambiguously supported. The clade including Psilotidae and Ophioglossidae was resolved as the earliest-diverging monilophyte lineage, but relationships among horsetails (Equisetopsida), marattioid ferns (Marattiidae), and leptosporangiate ferns (Polypodiidae) were obscure, and there was conflict depending on which optimality criterion (maximum parsimony vs. maximum likelihood) they used (Pryer et al., 2001).

As part of a broader effort to add mitochondrial data to attempt to resolve some of the problems indicated, we came across a previously undocumented group II intron located in the mitochondrial *atp1* gene of leptosporangiate and marattioid ferns. Here, we report on the primary sequence data for both the exon and intron of *atp1*, and we conduct phylogenetic analyses that combine the mitochondrial *atp1* data (newly reported here) with plastid *rbcL*, *atpB*, *rps4*, nuclear SSU rDNA, and morphological data previously reported (Pryer et al., 2001). Our best estimates of phylogeny are used for interpreting the *atp1* intron evolution.

2. Materials and methods

2.1. Taxon sampling

A total of 34 taxa (Table 1) were selected for this mitochondrial *atp1* study to match those taxa considered by Pryer et al. (2001) in the five data sets they made available (morphology; plastid *rbcL*, *atpB*, *rps4*; nuclear SSU rDNA). In many cases, the same DNA or voucher material used in that study was used here to sequence mitochondrial *atp1*. Despite several attempts, using various species, we were unable to amplify mitochondrial *atp1* for *Lygodium* (Schizaeaceae). For some families, a different representative genus was selected. Gleicheniaceae, for example, was represented here by *Sticherus* instead of *Gleichenia*, Salviniaceae by *Azolla* rather than *Salvinia*, and hornworts by *Phaeoceros*, not *Anthoceros*.

2.2. DNA extraction, amplification, and sequencing

Total cellular DNA was extracted using the DNeasy Plant Mini Kit from Qiagen following the protocol specified by the manufacturer. PCR fragments corresponding to bases 99–699 of the *Marchantia polymorpha* mitochondrial *atp1* gene were amplified for 27 species using forward primer F83-*atp1* (5′-ATGAGGTCGGTC GAGTGRT-3′) and reverse primer R725-*atp1* (5′-GGA

Table 1
Taxa included in this mitochondrial *atp1* study, with corresponding voucher or citation information and database accession numbers

Taxon	Voucher/citation	Database accession numbers	
		EMBL	Fern DNA DB[a]
Angiopteris evecta (G. Forst.) Hoffm.	N. Wikström 2001-81 (S)	AJ548861	2342
Austrobaileya scandens C. T. White	Barkman et al. (2000)	AY009410	—
Azolla filiculoides Lam.	N. Wikström 2001-86 (S)	AJ548863	2343
Blechnum occidentale L.	Mark W. Chase 11851 (K)	AJ548851	2344
Botrychium lunaria (L.) Sw.	F. Lind 78 (S)	AJ548870	2345
Chloranthus multistachys Pei	Qiu et al. (1999)	AF197665	—
Cyathea poeppigii (Hook.) Domin	R. Ornduff s.n. (UC)	AJ548854	80
Cycas revoluta Thunb.	Qiu et al. (1999)	AF197623	—
Danaea elliptica Sm.	J. Sharpe s.n. (UC)	AJ548864	451
Dicksonia antarctica Labill.	P. Kenrick 02-04-29 (BM)	AJ548853	2346
Equisetum × *ferrissii* Clute	K.M. Pryer 00-07 & D.M. Britton (DUKE)	AJ548866	1040
Equisetum telmateia Ehrh.	A. R. Smith 2575 (UC)	AJ548868	768
Ginkgo biloba L.	Qiu et al. (1999)	AF197625	—
Gnetum gnemon L.	Qiu et al. (1999)	AF197617	—
Haplomitrium hookeri (Sm.) Nees	F. Rumsey s.n. (BM)	AJ548876	2347
Hymenophyllum hirsutum (L.) Sw.	M. Kessler 9756, et al. (UC)	AJ548858	853
Huperzia lucidula (Michx.) Trevis.	T. Eriksson 664 (GH)	AJ548874	2348
Isoetes flaccida A. Braun	W.C. Taylor 5214 (MIL)	AJ548869	403
Marattia laxa Kunze	A.R. Smith 2566 (UC)	AJ548862	459
Marchantia polymorpha L.	Oda et al. (1992)	NC_001660	—
Marsilea minuta L.	B.J. Hoshizaki 900 (F)	AJ548860	984
Ophioglossum reticulatum L.	R. Moran 5644 (MO)	AJ548871	490
Osmunda regalis L.	N. Wikström 2001-78 (S)	AJ548859	2349
Phaeoceros laevis (L.) Prosk.	T. Hallingbäck 39503	AJ548875	2356
Phanerosorus sarmentosus (Baker) Copel.	M. Kato s.n. (TI)	AJ548856	866
Pinus sp.	Qiu et al. (1999)	AF197626	—
Plagiogyria japonica Nakai	M. Hasebe 27614 (TI)	AJ548855	501
Polytrichum formosum Hedw.	N. Wikström 2002-104 (S)	AJ548865	2350
Psilotum nudum (L.) P. Beauv.	N. Wikström 2001-75 (S)	AJ548873	2351
Pteridium aquilinum (L.) Kuhn	N. Wikström 2001-71 (S)	AJ548852	2352
Selaginella selaginoides (L.) P. Beauv.	T. Eriksson 704 (S)	AJ548877	2353
Sphagnum capillifolium (Ehrh.) Hedw.	N. Wikström 2002-103 (S)	AJ548867	2354
Sticherus palmatus (W. Schaffn. ex E. Fourn.) Copel.	A.R. Smith 2568 (UC)	AJ548857	177
Tmesipteris lanceolata P. A. Dang.	N. Wikström 230 (S)	AJ548872	2355

[a] Fern DNA Database <http://www.biology.duke.edu/pryerlab/ferndb> permanent record numbers (only for new sequence data reported here).

TCCGAAGCMGTGGCTGCTAC-3′). For three taxa (*Azolla*, *Haplomitrium*, and *Plagiogyria*), a shorter fragment was amplified using an alternative reverse primer R434-*atp1* (5′-TGGACAGATTTACGTGCAATAAT C-3′). PCRs were done in 25 μl aliquots using the Ready-To-Go PCR beads from Amersham Pharmacia. Sequencing was done in 5 μl aliquots using the ABI Prism BigDye terminator cycle sequencing ready reaction kit with Ampli*Taq* DNA polymerase FS from PE Applied Biosystems. Three additional primers were used for the sequencing reactions: F328-*atp1* (5′-GTTGATG CGTTRGGAGTACCY-3′), F411-*atp1* (5′-GATTAT TGCACGTAAATCTGTCCA-3′), and R348-*atp1* (5′-RGGTACTCCYAACGCATCAAC-3′). Reactions were cleaned as specified by the manufacturer and run on an ABI Prism 377 automated sequencer. Fragments were assembled using the Staden Package (Staden, 1996) and sequence data were aligned by hand using GDE (Smith et al., 1994). Seven *atp1* sequences included in our data set were already published.

2.3. cDNA analysis

To confirm the splicing sites of a mitochondrial *atp1* intron at the positions suggested by the multiple sequence alignment (Fig. 1), complementary DNA (cDNA) was generated for *Angiopteris evecta atp1* by reverse transcriptase amplification (RT-PCR), and subsequently sequenced. Total RNA extraction was done from fresh juvenile leaf tissue using the RNeasy Plant Mini Kit, and cDNA was generated using the Omniscript Reverse Transcriptase Kit, both kits supplied by Qiagen. Primer R725-*atp1* was used for synthesizing cDNA, and F83-*atp1* and R725-*atp1* were subsequently used for PCR amplification. PCR fragments were sequenced as specified above.

2.4. Phylogenetic analyses

Maximum likelihood-based phylogenetic analyses using a Bayesian inference procedure were conducted on

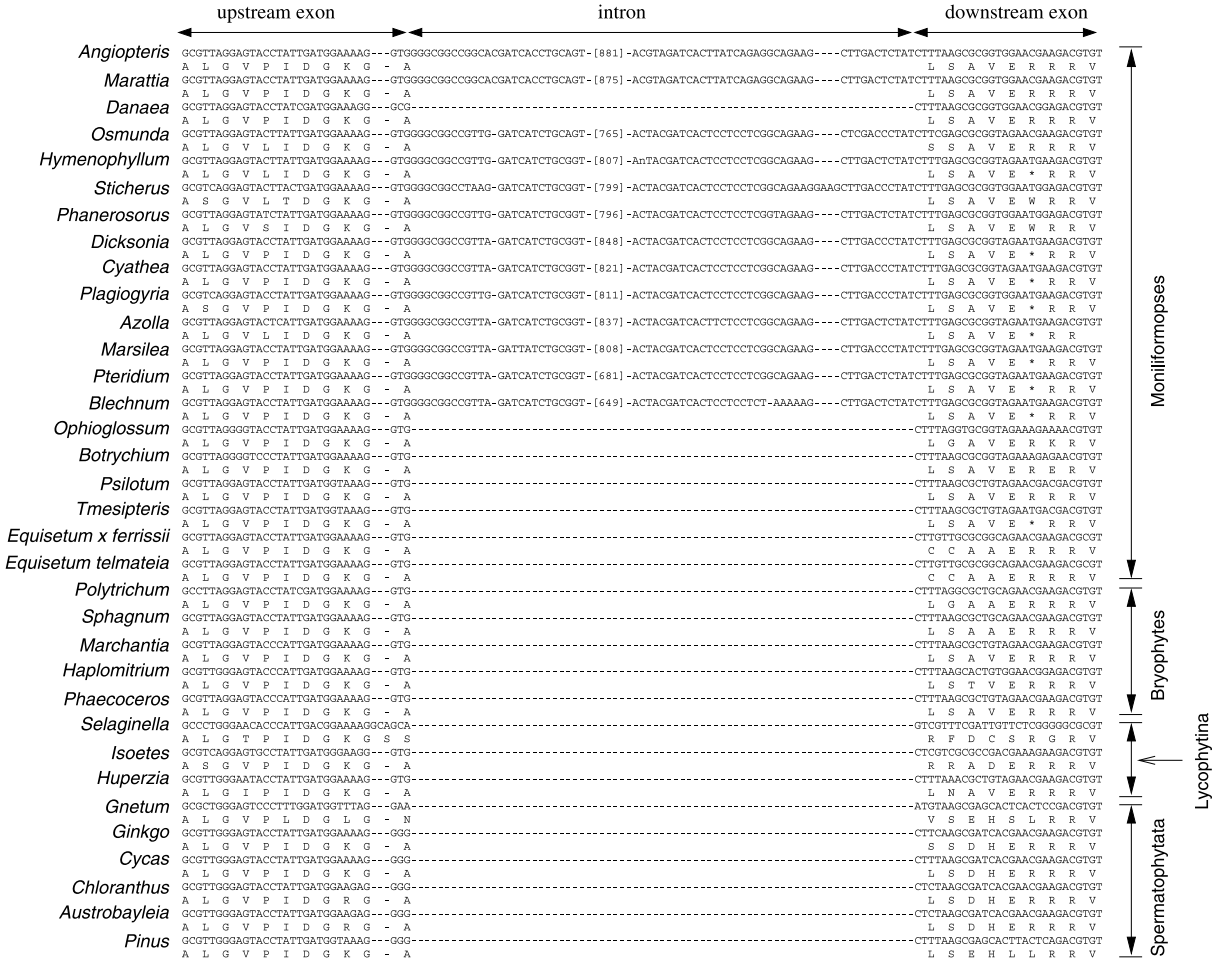

Fig. 1. Alignment of the mitochondrial *atp1* exon/intron boundaries for the land plants we sampled. Sequence data for the cDNA of *Angiopteris evecta atp1* confirmed the splice sites of the intron at the boundaries indicated (see also Fig. 2). Numbers in brackets within intron sequences indicate number of intron nucleotide sites not shown. Classification follows Kenrick and Crane (1997).

three data sets: (1) single-gene mitochondrial *atp1* (exon + intron); (2) five combined genes: mitochondrial *atp1* (exon only) + Pryer et al. (2001) data set (plastid *rbcL*, *atpB*, *rps4*; nuclear SSU rDNA); and (3) same five combined genes + Pryer et al. (2001) morphology data set. The morphology data set was revised here to include two additional molecular structural characters: (1) the absence/presence of a 3-amino acid insertion in the plastid *rps4* gene, as documented by Pryer et al. (2001); and (2) the absence/presence of the mitochondrial *atp1* intron, as documented here. The complete data sets are available in Nexus file format (Maddison et al., 1997) from the corresponding author.

In the single-gene *atp1* analyses, a total of two partitions were used, corresponding to the exon and intron data. Aligning the *atp1* intron, across all taxa in which it was found, was somewhat ambiguous and required that 250 intron characters be excluded from the single-gene *atp1* exon + intron data set, for a total of 1707 nucleotide characters (631 exon + 1076 intron characters). In the two combined data sets, the entire *atp1* intron was excluded from the analyses, for a total of 5697 nucleotide characters and 5697 nucleotide + 138 morphological characters, respectively. Bayesian analyses were done using MrBayes β-version 3.0B4 (Huelsenbeck and Ronquist, 2001). Unlike earlier versions of MrBayes, version 3.0 allows for different evolutionary models to be applied to different partitions of the data set, including models that enable phylogeny estimation using discrete morphological character data (Lewis, 2001; Tuffley and Steel, 1997). Taken together, these improvements allow for the simultaneous analysis of all available data, using appropriate models for each data type.

In the combined analyses, the data were divided into six partitions corresponding to the five genes (*rbcL*, *atpB*, *rps4*, SSU rDNA, and *atp1*) and morphology. MrModeltest 1.1b (Nylander, 2002), a simplified version of Modeltest 3.06 (Posada and Crandall, 1998), was used to select evolutionary models for the six different partitions. A general time reversible model (GTR + Γ + I) allowing for rate heterogeneity across sites, assuming a discrete gamma distribution for each partition and for a proportion of sites to be invariable, was selected for all nucleotide partitions. For the morphological partition,

the Mk model described by Lewis (2001) was used and rate heterogeneity across characters was accounted for by assuming a discrete gamma distribution. Five consecutive runs, each with four chains and a total of 1,000,000 generations were run in the Bayesian Markov chain Monte Carlo (MCMC) analyses. Following an initial "burn-in" phase of 100,000 generations to optimize the likelihood of tree-space, we sampled one tree every 100th generation and generated a majority-rule consensus tree of the 9000 trees sampled in each run. Convergence between the different runs was evaluated and confirmed, and the posterior probability values reported were averaged over the five consecutive runs. For the purpose of comparison with non-parametric bootstrapping (Alfaro et al., 2003), we chose to consider nodes with posterior probabilities equal to or greater than 0.95 (e.g., the node appears in at least 95% of the sampled trees) as being well-supported.

3. Results

3.1. Mitochondrial atp1 intron and cDNA analysis

Our amplification and sequencing of the mitochondrial *atp1* gene across land plants revealed the presence of an intron (ranging between 700 and 950 bp) within *atp1* for many of the monilophytes that we sampled. In Fig. 1 we show a multiple sequence alignment of the mitochondrial *atp1* exon/intron boundaries for these taxa. Fig. 2 shows the partial mitochondrial *atp1* DNA sequence reported here for *Angiopteris evecta*, including its 945 bp intron. By also sequencing the cDNA for *A. evecta* mitochondrial *atp1*, we were able to confirm the 5′ and 3′ splice sites of the intron at the positions indicated by the multiple sequence alignment in Fig. 1. The primary sequence reads for the 5′ intron end (GGGCGGC) and the 3′ intron end (AT) (Figs. 1 and 2) correspond with those previously reported for group II introns (Malek and Knoop, 1998), but do not strictly conform to the group II intron consensus sequence as indicated by Michel et al. (1989) and Bonen and Vogel (2001). Fig. 2 also highlights two RNA editing events in the flanking *atp1* exons that were confirmed by comparing the cDNA sequence with its corresponding genomic DNA sequence. Both editing events involve transitions from C to U (C–T at the cDNA level), changing the inferred amino acid serine to leucine.

3.2. Phylogenetic analyses

The 50% majority rule consensus trees resulting from bayesian analyses of the single-gene *atp1* exon + intron data set and of the combined five-genes + morphology data set are shown in Figs. 3 and 4, respectively. Both trees are rooted with *Marchantia*, branch lengths are drawn proportional to the inferred number of changes along each branch, as averaged

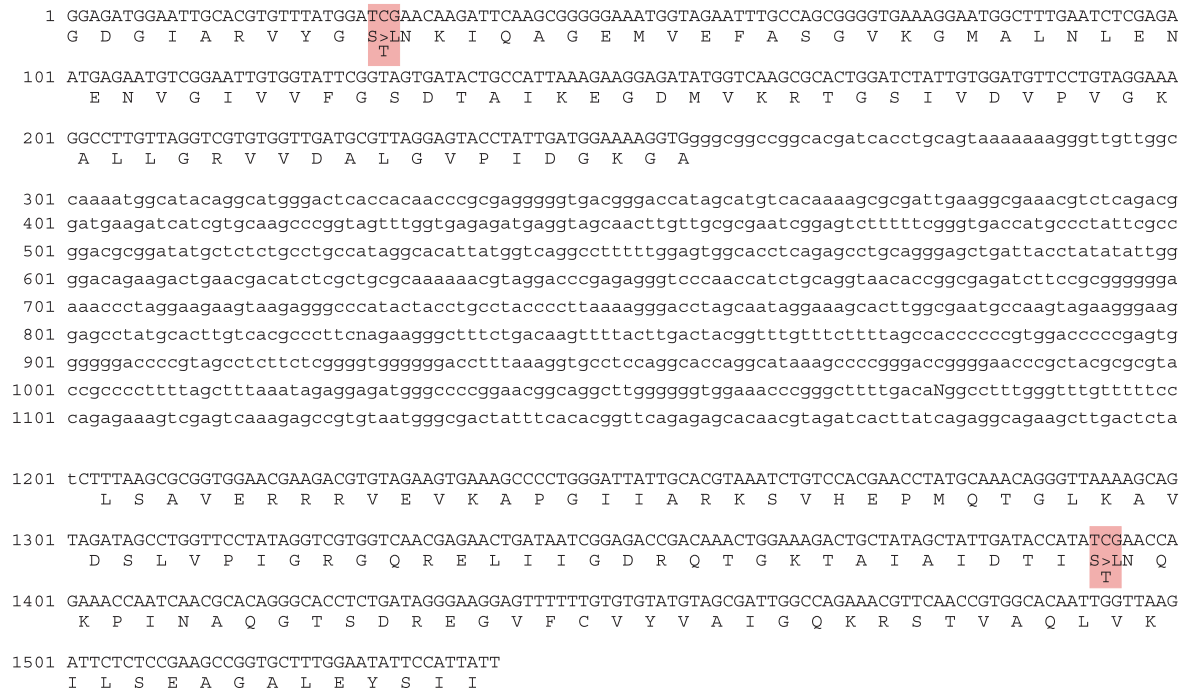

Fig. 2. Primary DNA sequence of the partial mitochondrial *atp1* gene reported here for *Angiopteris evecta* (intron sequence in lower case). Sequence data for the cDNA of *A. evecta atp1* identified the splice sites of the intron, and indicated two RNA editing events in the flanking *atp1* exons (indicated by the red boxes) that change the inferred amino acid serine to leucine.

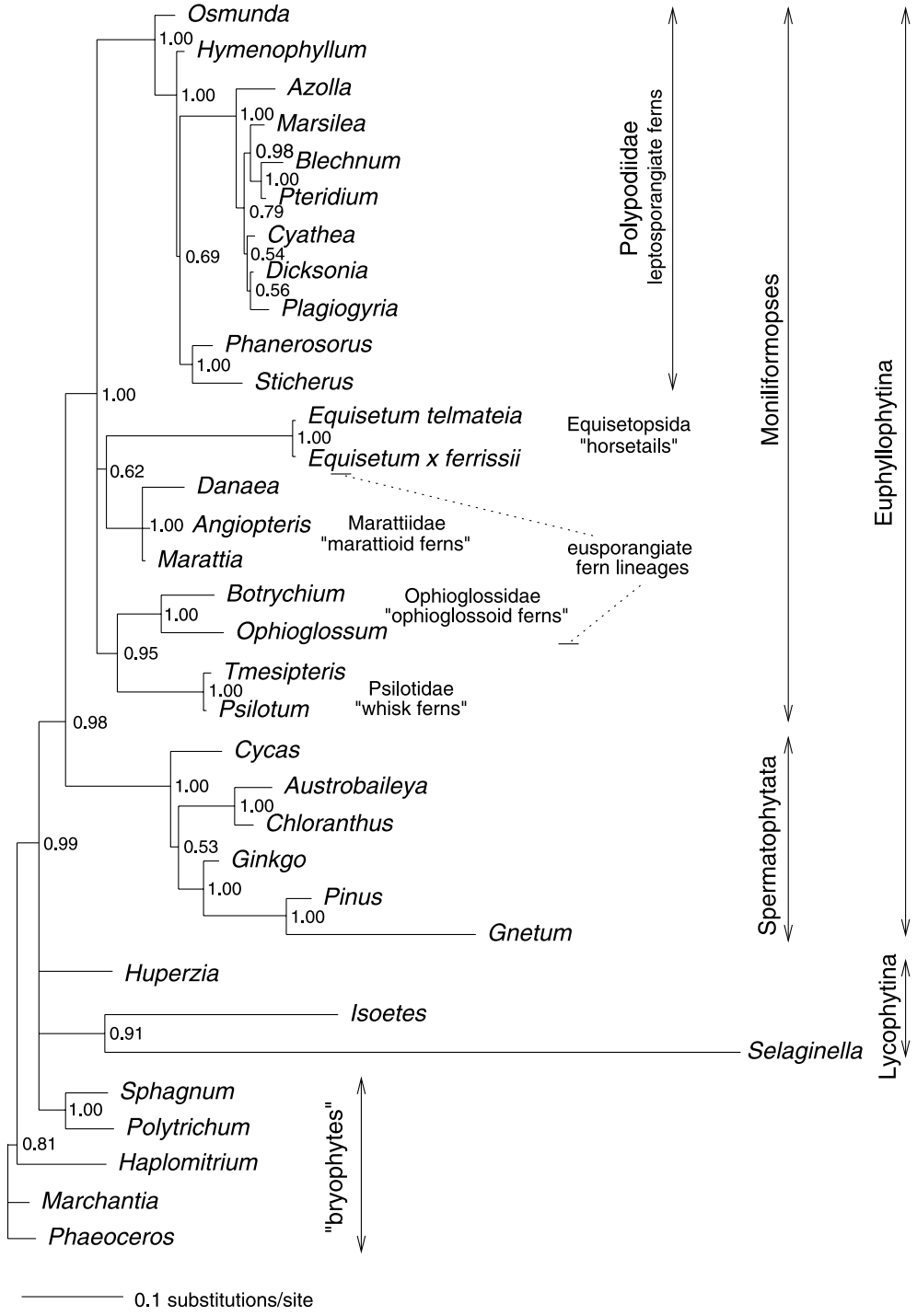

Fig. 3. The 50% majority-rule consensus tree, rooted with *Marchantia*, resulting from bayesian analyses of mitochondrial *atp1* exon + intron data. Branch lengths are proportional to the inferred number of changes, as averaged across the posterior distribution, and numbers at nodes are bayesian posterior probabilities. Classification follows Kenrick and Crane (1997).

across the posterior distribution, and numbers at nodes indicate branch support as measured by bayesian posterior probabilities. In both trees, horsetails (Equisetopsida) are grouped with marattioid ferns, although the *atp1* exon + intron data alone (Fig. 3) only provide weak support (62% posterior probability) for this result compared to the five-genes + morphology data (93% posterior probability) (Fig. 4). Euphyllophytina and Moniliformopses are both always well supported (⩾98% posterior probability), as is the grouping of whisk ferns (Psilotidae) with ophioglossoid ferns (Ophioglossidae) (⩾95% posterior probability). In both

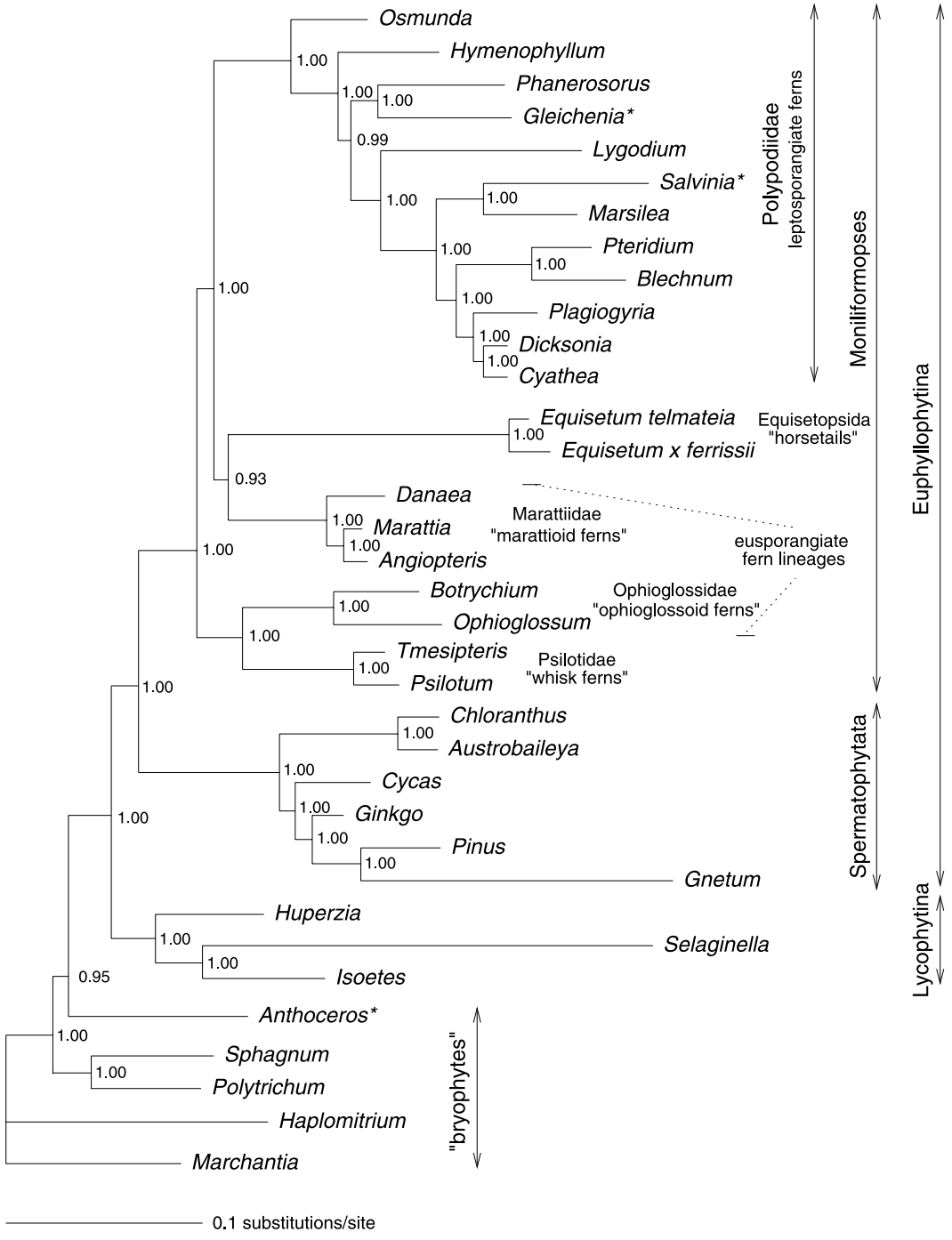

Fig. 4. The 50% majority-rule consensus tree, rooted with *Marchantia*, resulting from bayesian analyses of the combined data (five-genes + morphology). Branch lengths are proportional to the inferred number of changes, as averaged across the posterior distribution, and numbers at nodes are bayesian posterior probabilities. Classification follows Kenrick and Crane (1997). *In the *atp1* partition, *Gleichenia*, *Salvinia*, and *Anthoceros* were represented by *Sticherus*, *Azolla*, and *Phaeoceros*, respectively.

these analyses, the overall topology among leptosporangiate ferns corresponds for the most part with that obtained by Pryer et al. (2001).

Results from the combined five-gene data set (morphology excluded) differ from those of the five-gene + morphology data set in two-ways (tree not shown). First, *Cyathea* is not grouped as sister to *Dicksonia*, but as sister to a clade including *Dicksonia* and *Plagiogyria*. The *Dicksonia–Plagiogyria* clade, however, is weakly supported (54% posterior probability). Second, the grouping of horsetails with marattioid ferns shows increased support (from 93 to 99% posterior probability) when only the primary sequence data of the genes are considered and the morphological data is excluded.

4. Discussion

4.1. Phylogenetic relationships

Adding mitochondrial *atp1* sequence data to the Pryer et al. (2001) data set strengthened support for the results they reported. Although a strict comparison is complicated because different analytical protocols were used, results from our single-gene (mitochondrial *atp1* exon + intron; Fig. 3) analyses are largely in agreement with their maximum likelihood-based results using one nuclear and three plastid genes. The monophyly of euphyllophytes and monilophytes is well supported (98 and 100% posterior probability, Fig. 3), as is the grouping of whisk ferns (Psilotidae) with ophioglossoid ferns (Ophioglossidae) (95% posterior probability), and horsetails (Equisetopsida) with marattioid ferns (Marattiidae)—though with weak support (62% posterior probability). All analyses (single-gene *atp1* exon + intron, five-genes, and five-genes + morphology) corroborate the major results of Pryer et al. (2001), including the grouping of horsetails with all other seed-free euphyllophyte lineages, supporting the monophyly of Moniliformopses (sensu Kenrick and Crane, 1997).

The precise relationship of horsetails within monilophytes remains equivocal. Previous analyses by Pryer et al. (2001, 2004b) of primary sequence data indicated their placement as sister to marattioid ferns, and the addition of mitochondrial *atp1* sequence data reinforces this pattern. However, the *atp1* intron distribution pattern alone (Figs. 1 and 5) does not lend support to this grouping. Moreover, we note that the inclusion of the morphological data set (which includes intron distribution data) in our analyses lowers the support for grouping horsetails and marattioid ferns from 99 to 93% posterior probability. However, the topology that result from the analysis of the morphological partition alone (result not shown) places horsetails as sister to whisk ferns (with weak support), a relationship that is never resolved with primary sequence data.

The incongruence between primary sequence data on the one hand and intron distribution data on the other may simply reflect a more general pattern that mitochondrial introns sometimes are lost (Beckert et al., 1999, 2001; Dombrovska and Qiu, 2004; Malek and Knoop, 1998; Pruchner et al., 2001; Qiu et al., 1998; Vangerow et al., 1999), but it may also result from a bias in our tree reconstruction methods. Monte Carlo simulations using slightly different topologies indicate that some alternative placements of horsetails are more problematic to reconstruct than the topology (horsetails + marattioid ferns) currently supported by primary sequence data. Although the differences in posterior probabilities (93 vs. 99%), when we do and when we do not consider the morphological data, are minor, the possibility that horsetails group with marattioid ferns as a result of an analytical bias should be considered further.

Qiu and Palmer (1999), in their review of early land plant relationships, listed various concerns associated with primary sequence-based phylogenetics, and they argued for the importance of augmenting these studies with comparative genomic structural characters. The presence/absence distribution pattern provided by the mitochondrial *atp1* intron that we document here is the first report of genomic structural data within the monilophyte clade that can be used to assist in interpreting these deep relationships. Although this particular intron pattern suggests some incongruence with results from primary sequence-based analyses, it is our expectation that the mitochondrial, as well as the nuclear and

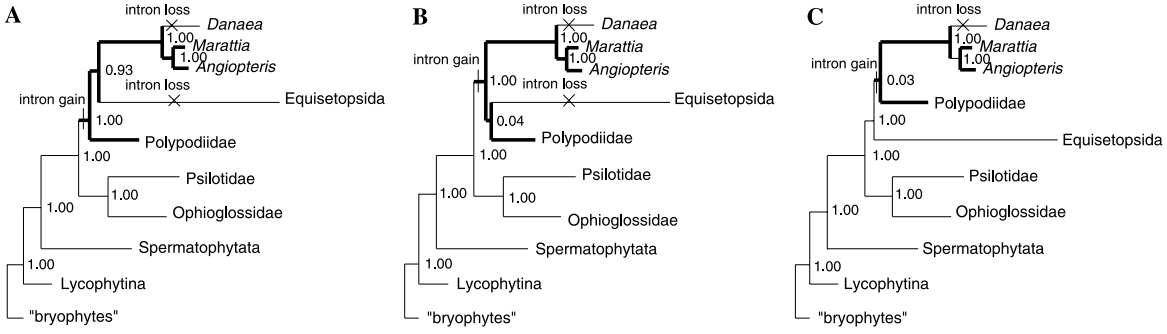

Fig. 5. Inferring mitochondrial *atp1* intron gain and loss on three topologies with alternative placements of horsetails (Equisetopsida). Each placement has been hypothesized previously based on molecular and/or morphological data, and corresponds with the best supported (A) and two less supported (B and C) topologies that resulted from our analyses of the combined data (five-genes + morphology). (A) Horsetails sister to marattioid ferns (*Danaea*, *Marattia*, *Angiopteris*; 93% posterior probability): implies that the *atp1* intron has been lost at least twice. (B) Horsetails sister to leptosporangiate ferns (Polypodiidae; 4% posterior probability): implies that the *atp1* intron has been lost at least twice; this grouping supported by Nickrent et al. (2000). (C) Horsetails sister to a clade that includes leptosporangiate and marattioid ferns (3% posterior probability): implies a single loss of the *atp1* intron; this placement suggested by the *atp1* intron distribution pattern, and the one most consistent with the morphology-based analyses of Bremer (1985) and Rothwell (1999) in that it implies monophyly of marattioid and leptosporangiate ferns. Classification follows Kenrick and Crane (1997).

chloroplast, genomes of these plants will yield additional structural examples that will help to resolve the long-standing question of horsetail relationships.

4.2. Mitochondrial atp1 intron evolution

As with other studies that document the absence/presence of mitochondrial introns (Beckert et al., 1999, 2001; Malek and Knoop, 1998; Qiu et al., 1998; Vangerow et al., 1999), the *atp1* intron distribution across euphyllophyte taxa cannot be interpreted without inferring secondary losses. The topology that is best supported by the analyses of the combined data suggests that the intron was gained prior to the Equisetopsida + Marattiidae + Polypodiidae split, and was subsequently lost twice, once in the horsetail stem lineage and once in *Danaea* (Fig. 5A). Because the primary sequence of the intron shows a high degree of similarity across all taxa in which it is present, the alternative possibility of two independent gains of the intron (once in Polypodiidae, and once in marattioid ferns—followed by a loss in *Danaea*) is considered unlikely.

Fig. 5 also considers two other alternative placements of horsetails—topologies that are very rarely obtained in our combined five-genes + morphology analyses (Figs. 5B and C), but that have been indicated by other analyses of molecular and/or morphological data (Bremer, 1985; Nickrent et al., 2000; Rothwell, 1999). One of the alternative placements (Fig. 5B) groups horsetails with leptosporangiate ferns (Polypodiidae). Although support for this grouping in our analyses is weak (4% posterior probability), a previous analysis using a combined four-gene data set (plastid *rbcL* and SSU rDNA from all three plant genomes; Nickrent et al., 2000) supported this grouping. In this scenario (Fig. 5B), we also infer the *atp1* intron to have been gained once and lost at least twice, once in the horsetail stem lineage and once in *Danaea*. The mitochondrial *atp1* intron distribution pattern (Fig. 1) appears to lend support to horsetails as sister to a monophyletic clade including both leptosporangiate and marattioid ferns (Polypodiidae + Marattiidae; Fig. 5C). The support for this placement of horsetails in our analyses is weak (3% posterior probability), although it would imply that the *atp1* intron had been gained once, and lost once in the lineage leading to *Danaea* (Fig. 5C).

Under any of these topologies (Fig. 5), it is certain that the *atp1* intron has been lost in the lineage leading to *Danaea*, and this indicates that *atp1* intron loss is a plausible evolutionary event. This pattern of occasional intron loss conforms to that seen for other mitochondrial introns (Beckert et al., 1999, 2001; Dombrovska and Qiu, 2004; Malek and Knoop, 1998; Pruchner et al., 2001; Qiu et al., 1998; Vangerow et al., 1999). Therefore, inferring one additional loss in the horsetail stem lineage, required under the two alternative placements of horsetails shown in Figs. 5A and B, could therefore be considered only a marginally more complicated interpretation. Additional sampling among leptosporangiate ferns has revealed a further loss in *Ceratopteris* (Wikström and Pryer, unpublished data), lending more support to the interpretation that the potential of this particular intron is limited in so far as being a useful indicator of broad phylogenetic relationships among monilophytes.

Acknowledgments

This research was supported by a Marie Curie Fellowship of the European Community Programme Improving Human Potential and the Socio-Economic Knowledge under contract No. HPMFCT-2000-00631 to N. Wikström. Financial support to K.M. Pryer from the National Science Foundation (DEB-0347840) is gratefully acknowledged. We are thankful to Eric Schuettpelz for constructive comments provided on an earlier draft of the manuscript and to the Duke Systematics Discussion Group for useful contributions.

References

Alfaro, M.E., Zoller, S., Lutzoni, F., 2003. Bayes or bootstrap? A simulation study comparing the performance of Bayesian Markov chain Monte Carlo sampling and bootstrapping in assessing phylogenetic confidence. Mol. Biol. Evol. 20, 255–266.

Barkman, T.J., Chenery, G., McNeal, J.R., Lyons-Weiler, J., Ellisens, W.J., Moore, G., Wolfe, A.D., dePamphilis, C.W., 2000. Independent and combined analyses of sequences from all three genomic compartments converge on the root of flowering plant phylogeny. Proc. Natl. Acad. Sci. USA 97, 13166–13171.

Beck, C.B., Stein, W.E., 1993. *Crossia virginiana* gen. et sp. nov., a new member of the Stenokoleales from the Middle Devonian of southwestern Virginia. Palaeontograph. B 229, 115–134.

Beckert, S., Muhle, H., Pruchner, D., Knoop, V., 2001. The mitochondrial *nad2* gene as a novel marker locus for phylogenetic analysis of early land plants: a comparative analysis in mosses. Mol. Phylogenet. Evol. 18, 117–126.

Beckert, S., Steinhauser, S., Muhle, H., Knoop, V., 1999. A molecular phylogeny of bryophytes based on nucleotide sequences of the mitochondrial *nad5* gene. Plant Syst. Evol. 218, 179–192.

Bonen, L., Vogel, J., 2001. The ins and outs of group II introns. Trends Genet. 17, 322–331.

Bowe, M.L., Coat, G., dePamphilis, C.W., 2000. Phylogeny of seed plants based on all three plant genomic compartments: extant gymnosperms are monophyletic and Gnetales are derived conifers. Proc. Natl. Acad. Sci. USA 97, 4092–4097.

Bremer, K., 1985. Summery of green plant phylogeny and classification. Cladistics 1, 369–385.

Chaw, S.-M., Parkinson, C.L., Cheng, Y., Vincent, T.M., Palmer, J.D., 2000. Seed plant phylogeny inferred from all three plant genomes: monophyly of extant gymnosperms and origin of Gnetales from conifers. Proc. Natl. Acad. Sci. USA 97, 4086–4091.

Cho, Y., Qiu, Y.-L., Kuhlman, P., Palmer, J.D., 1998. Explosive invasion of plant mitochondria by a group I intron. Proc. Natl. Acad. Sci. USA 95, 14244–14249.

Dombrovska, O., Qiu, Y.-L., 2004. Distribution of introns an the mitochondrial gene *nad1* in land plants: phylogenetic and molecular evolutionary implications. Mol. Phylogenet. Evol. 32, 246–263.

Donoghue, M.J., Doyle, J.A., 2000. Seed plant phylogeny: demise of the anthophyte hypothesis? Curr. Biol. 10, R106–R109.

Duff, R.J., Nickrent, D.L., 1999. Phylogenetic relationships of land plants using mitochondrial small-subunit rDNA sequences. Am. J. Bot. 86, 372–386.

Hedderson, T.A., Chapman, R., Cox, C.J., 1998. Bryophytes and the origin and diversification of land plants: new evidence from molecules. In: Bates, J.W., Ashton, N.W., Duckett, J.G. (Eds.), Bryology for the Twenty-First Century. Maney Publishing and the British Bryological Society, Leeds, pp. 65–77.

Hiesel, R., von Haeseler, A., Brennicke, A., 1994. Plant mitochondrial nucleic acid sequences as a tool for phylogenetic analysis. Proc. Natl. Acad. Sci. USA 91, 634–638.

Huelsenbeck, J.P., Ronquist, F., 2001. MRBAYES: Bayesian inference of phylogeny. Bioinformatics 17, 754–755.

Judd, W.S., Campbell, C.S., Kellogg, E.A., Stevens, P.F., Donoghue, M.J., 2002. Plant Systematics: A Phylogenetic Approach, second ed. Sinauer, Sunderland, MA.

Kenrick, P., Crane, P.R., 1997. The Origin and Early Diversification of Land Plants: A Cladistic Study. Smithsonian Institution, Washington, DC.

Kranz, H.D., Huss, V.A.R., 1996. Molecular evolution of pteridophytes and their relationship to seed plants: evidence from complete 18S rRNA gene sequences. Plant Syst. Evol. 202, 1–11.

Lewis, P.O., 2001. A likelihood approach to estimating phylogeny from discrete morphological character data. Syst. Biol. 50, 913–925.

Maddison, D.R., Swofford, D.L., Maddison, W.P., 1997. NEXUS: an extensible file format for systematic information. Syst. Biol. 46, 590–621.

Malek, O., Knoop, V., 1998. *Trans*-splicing group II introns in plant mitochondria: the complete set of *cis*-arranged homologs in ferns, fern-allies, and a hornwort. RNA 4, 1599–1609.

Manhart, J.R., 1994. Phylogenetic analysis of green plant *rbcL* sequences. Mol. Phylogenet. Evol. 3, 114–127.

Manhart, J.R., 1995. Chloroplast 16S rDNA sequences and phylogenetic relationships of fern allies and ferns. Am. Fern J. 85, 182–192.

Manhart, J.R., Palmer, J.D., 1990. The gain of two chloroplast tRNA introns marks the green algal ancestors of land plants. Nature 345, 268–270.

Michel, F., Umesono, K., Ozeki, H., 1989. Comparative and functional anatomy of group II catalytic introns—a review. Gene 82, 5–30.

Mishler, B.D., Lewis, L.A., Buchheim, M.A., Renzaglia, K.S., Garbary, D.J., Delwiche, C.F., Zechman, F.W., Kantz, T.S., Chapman, R.L., 1994. Phylogenetic relationships of the "green algae" and "bryophytes. Ann. Mo. Bot. Gard. 81, 451–483.

Nickrent, D.L., Parkinson, C.L., Palmer, J.D., Duff, R.J., 2000. Multigene phylogeny of land plants with special reference to bryophytes and the earliest land plants. Mol. Biol. Evol. 17, 1885–1895.

Nylander J.A.A., 2002. MrModeltest ver. 1.1b. Distributed by the author (johan.nylander@ebc.uu.se). Department of Systematic Zoology, EBC, Uppsala University, Sweden..

Oda, K., Yamato, K., Ohta, E., Nakamura, Y., Takemura, M., Nozato, N., Akashi, K., Kanegae, T., Ogura, Y., Kohchi, T., Ohyama, K., 1992. Gene organization deduced from the complete sequence of liverwort *Marchantia polymorpha* mitochondrial DNA. A primitive form of plant mitochondrial genome. J. Mol. Biol. 223, 1–7.

Palmer, J.D., Adams, K.L., Cho, Y., Parkinson, C.L., Qiu, Y.-L., Song, K., 2000. Dynamic evolution of plant mitochondrial genomes: mobile genes and introns and highly variable mutation rates. Proc. Natl. Acad. Sci. USA 97, 6960–6966.

Posada, D., Crandall, K.A., 1998. MODEL TEST: testing the model of DNA substitution. Bioinformatics 14, 817–818.

Pruchner, D., Nassal, B., Schindler, M., Knoop, V., 2001. Mosses share mitochondrial group II introns with flowering plants, not with liverworts. Mol. Genet. Genomics 266, 608–613.

Pryer, K.M., Schneider, H., Smith, A.R., Cranfill, R., Wolf, P.G., Hunt, J.S., Sipes, S.D., 2001. Horsetails and ferns are a monophyletic group and the closest living relatives to seed plants. Nature 409, 618–621.

Pryer, K.M., Schneider, H., Magallón, S., 2004a. The radiation of vascular plants. In: Cracraft, J., Donoghue, M.J. (Eds.), Assembling the Tree of Life. Oxford University Press, London, pp. 138–153.

Pryer, K.M., Schuettpelz, E., Wolf, P.G., Schneider, H., Smith, A.R., Cranfill, R., 2004b. Phylogeny and evolution of ferns (monilophytes) with a focus on the early leptosporangiate divergences. Am. J. Bot. 91, 1582–1598.

Pryer, K.M., Smith, A.R., Skog, J.E., 1995. Phylogenetic relationships of extant ferns based on evidence from morphology and *rbcL* sequences. Am. Fern J. 85, 205–282.

Qiu, Y.-L., Cho, Y., Cox, C.J., Palmer, J.D., 1998. The gain of three mitochondrial introns identifies liverworts as the earliest land plants. Nature 394, 671–674.

Qiu, Y.-L., Lee, J., 2000. Transition to a land flora: a molecular phylogenetic perspective. J. Phycol. 36, 799–802.

Qiu, Y.-L., Lee, J., Bernasconi-Quadroni, F., Soltis, D.E., Soltis, P.S., Zanis, M., Chen, Z., Savolainen, V., Chase, M.W., 1999. The earliest angiosperms: evidence from mitochondrial, plastid and nuclear genomes. Nature 402, 404–407.

Qiu, Y.-L., Palmer, J.D., 1999. Phylogeny of early land plants: insights from genes and genomes. Trends Plant Sci. 4, 26–30.

Raubeson, L.A., Jansen, R.K., 1992a. Chloroplast DNA evidence on the ancient evolutionary split in vascular land plants. Science 255, 1697–1699.

Raubeson, L.A., Jansen, R.K., 1992b. A rare chloroplast-DNA structural mutation is shared by all conifers. Biochem. Syst. Ecol. 20, 17–24.

Rothwell, G.W., 1999. Fossils and ferns in the resolution of land plant phylogeny. Bot. Rev. 65, 188–218.

Smith, S.W., Overbeek, R., Woese, C.R., Gilbert, W., Gillevet, P.M., 1994. The Genetic Data Environment, an expandable GUI for multiple sequence analyses. Cabios 10, 671–675.

Staden, R., 1996. The Staden sequence analysis package. Mol. Biotechnol. 5, 233–241.

Stevenson, D.W., Loconte, H., 1996. Ordinal and familial relationships of pteridophyte genera. In: Camus, J.M., Gibby, M., Johns, R.J. (Eds.), Pteridology in Perspective. Royal Botanic Gardens, Kew, pp. 435–467.

Strauss, S.H., Palmer, J.D., Howe, G.T., Doerksen, A.H., 1988. Chloroplast genomes of two conifers lack a large inverted repeat and are extensively rearranged. Proc. Natl. Acad. Sci. USA 85, 3898–3902.

Tuffley, C., Steel, M., 1997. Links between maximum likelihood and maximum parsimony under a simple model of site substitution. Bull. Math. Biol. 59, 581–607.

Vangerow, S., Teerkorn, T., Knoop, V., 1999. Phylogenetic information in the mitochondrial *nad5* gene of pteridophytes: RNA editing and intron sequences. Plant Biol. 1, 235–243.