J Plant Res
DOI 10.1007/s10265-007-0088-x

SHORT COMMUNICATION

Molecular phylogeny of horsetails (*Equisetum*) including chloroplast *atpB* sequences

Jean-Michel Guillon

Received: 9 November 2006 / Accepted: 21 March 2007

Abstract *Equisetum* is a genus of 15 extant species that are the sole surviving representatives of the class Sphenopsida. The generally accepted taxonomy of *Equisetum* recognizes two subgenera: *Equisetum* and *Hippochaete*. Two recent phylogenetical studies have independently questioned the monophyly of subgenus *Equisetum*. Here, I use original (*atpB*) and published (*rbcL*, *trnL-trnF*, *rps4*) sequence data to investigate the phylogeny of the genus. Analyses of *atpB* sequences give an unusual topology, with *E. bogotense* branching within *Hippochaete*. A Bayesian analysis based on all available sequences yields a tree with increased resolution, favoring the sister relationships of *E. bogotense* with subgenus *Hippochaete*.

Keywords *Equisetum* · Evolution · Horsetail · Phylogeny

Introduction

Horsetails (*Equisetum* L.) are the only extant members of the class Sphenopsida, free-sporing plants characterized by articulate stems bearing whorls of leaves at each node. Since the work of Hauke (1963, 1978), 15 species of *Equisetum* have been widely accepted, but many interspecific hybrids, involving all species except *E. bogotense*, are found in the wild. These hybrids are considered to be sterile (but see Krahulec et al. 1996) and to be dependent on vegetative reproduction for persistence and growth. The 15 species of *Equisetum* are grouped in two subgenera based on morphological characters such as the position of stomata: superficial in subgenus *Equisetum* (*E. arvense*, *E. bogotense*, *E. diffusum*, *E. fluviatile*, *E. palustre*, *E. pratense*, *E. sylvaticum*, and *E. telmateia*), sunken below the epidermal surface in subgenus *Hippochaete* (*E. giganteum*, *E. hyemale*, *E. laevigatum*, *E. myriochaetum*, *E. ramosissimum*, *E. scirpoides*, and *E. variegatum*). A barrier seems to prevent hybridization between plants of the subgenera *Equisetum* and *Hippochaete* (Duckett 1979).

Because characters found in the fossil record, such as large stems and persistent sheath teeth, are present in the sole *E. giganteum* among modern species, *E. giganteum* was proposed to be the most primitive living member of the genus (Schaffner 1925, 1930; Hauke 1963). The assumed primitive status of *E. giganteum* suggested that bisexuality could have been the primitive condition in *Equisetum* gametophytes (Hauke 1969, 1985). Conversely, Hauke (1968, 1969) proposed that strict unisexuality, as found in gametophytes of *E. bogotense*, was the most derived condition in the genus (Hauke 1968, 1969).

These conclusions have been called into question by two recent phylogenetical studies using chloroplast DNA data from all extant species of *Equisetum* (Des Marais et al. 2003; Guillon 2004). The two studies disagreed on the exact position of *Equisetum giganteum* but both found it nested within subgenus *Hippochaete*. This finding suggested that large size and bisexuality were derived characters rather than ancestral ones. Unexpectedly, *E. bogotense* was inferred to be either (1) basal to the whole genus [maximum parsimony (MP) analysis in Des Marais et al. 2003; MP, maximum likelihood (ML) and Bayesian analyses in Guillon 2004) or (2) sister to *Hippochaete* (ML

J.-M. Guillon (✉)
Laboratoire Ecologie Systématique et Evolution,
UMR 8079 CNRS-Université Paris-Sud, Bâtiment 362,
Université Paris-Sud, 91405 Orsay Cedex, France
e-mail: Jean-Michel.Guillon@u-psud.fr

and Bayesian analyses in Des Marais et al. 2003). Subgenus *Hippochaete* was found to be monophyletic and nested inside a paraphyletic subg. *Equisetum*.

Sequence data from the chloroplast *atpB* gene have proven to be useful for resolving relationships within ferns (Wolf 1997; Tsutsumi and Kato 2005; Korall et al. 2006), suggesting that the gene may also have phylogenetic utility for horsetails, especially for resolving deep nodes. I here report the analysis of a new data set obtained after sequencing part of *atpB* and the combined analysis of all sequences available for *Equisetum* in order to clarify the position of *E. bogotense*. Evolution of *Equisetum* characters is discussed in light of the resulting phylogenetic inferences.

Materials and methods

Sampling

All widely recognized extant species of horsetails are represented in this study. *Equisetum rps4* sequences used are those previously reported by Guillon (2004). *Equisetum rbcL* and *trnL-trnF* sequences used are those previously reported by Des Marais et al. (2003). The European Molecular Biology Laboratory (EMBL) accession details for *Equisetum atpB* data used in this study are listed in Table 1. Vouchers for *Equisetum atpB* sequences are as described for *rps4* (Guillon 2004).

Table 1 European Molecular Biology Laboratory (EMBL) information for *Equisetum atpB* sequences

Equisetum species	EMBL accession number
E. arvense L.	AM422389
E. bogotense Kunth	AM422390
E. diffusum D. Don	AM422391
E. fluviatile L	AM422392
E. giganteum L	AM422393
E. hyemale L.	AM422394
E. laevigatum A. Braun	AM422395
E. myriochaetum Cham. and Schltdl.	AM422396
E. palustre L.	AM422397
E. pratense Ehrh.	AM422398
E. ramosissimum Desf. subsp. *debile* (Roxb.) Hauke	AM422399
E. scirpoides Michx.	AM422400
E. sylvaticum L.	AM422401
E. telmateia Ehrh. subsp. *braunii*	AM422402
E. variegatum Schleicher	AM422403

DNA extraction, PCR, and nucleotide sequencing

Total genomic DNA was extracted from fresh or silica-gel-dried plant material using the DNAeasy Plant Mini Kit (Qiagen, Hilden, Germany). Part of the *atpB* gene (393 base pairs) was amplified using the two primers 5′-ATAATTGGGCCGGTTTTGGATGT-3′ and 5′-ACGACTTTGATGCCTGTTTCGAA-3′ under the following conditions: preincubation at 94°C for 5 min, followed by 34 cycles, each consisting of 45 s at 94°C, 45 s at 52°C, and 1 min at 72°C, and finally 5 min at 72°C. Amplification mixtures (25 µl) contained 50 mM KCl, 10 mM Tris–HCl (pH 9.0 at 25°C), 50 mM KCl, 0.1% Triton X-100, 2.5 mM $MgCl_2$, 200 µM of each of the four deoxynucleotide triphosphates (dNTPs), 0.4 µM of each primer, and 1 U of *Taq* DNA polymerase (Promega, Madison, USA). The polymerase chain reaction products were checked on agarose gel with ethidium bromide and purified by polyethylene glycol (PEG) precipitation (Rosenthal et al. 1993). Purified PCR products were sequenced using MWG Biotech (Ebersberg, Germany) custom sequencing service.

Outgroups, sequence alignment

Outgroups were chosen in fern major taxa because studies agree on grouping horsetails and ferns in a monophyletic clade sister to seed plants (Kenrick and Crane 1997; Nickrent et al. 2000; Renzaglia et al. 2000; Pryer et al. 2001; Renzaglia et al. 2002; Wykström and Pryer 2005). A representative of Psilophyta was also included, given its accepted inclusion within the fern clade (Wolf 1997; Nickrent et al. 2000; Renzaglia et al. 2000; Pryer et al. 2001; Wykström and Pryer 2005). Sequence data for outgroups were obtained from GenBank/EMBL: *Adiantum capillus-veneris* L.: *rps4* + *rbcL* + *atpB* = NC004766; *Angiopteris evecta* (J.R. Forst.) Hoffmann: *rps4* = AF313591; *Angiopteris lygodiifolia* Rosenst.: *rbcL* + *atpB* = X58429; *Ophioglossum reticulatum* L.: *rps4* = AF313594, *rbcL* = AF313582, *atpB* = U93825; *Psilotum nudum* (L.) P. Beauv.: *rps4* + *rbcL* + *atpB* = NC003386, *trnL-trnF* = AY241586.

Alignment of *atpB* and *rbcL* sequences was straightforward among all species, and alignments used for *rps4* and *trnL-trnF* were previously described (Guillon 2004; Des Marais et al. 2003). Only unambiguously alignable regions of *rps4* coding and adjacent noncoding sequences and *trnL-trnF* sequences were included in the conjoined matrix. Specifically, outgroups sequences for *rps4* adjacent noncoding sequences were not included, and only *Psilotum* was used as outgroup for *trnL-trnF* sequence. The number of characters in the conjoined matrix was 3,019, and the percentage of cells scored as missing data was 5%. The number of characters in the *atpB* matrix was 342, and the

percentage of cells scored as missing data was 1%. The data matrices are available upon request from the author.

Phylogenetic analyses

For MP analysis of *atpB* sequences, heuristic searches (starting from a tree constructed by random stepwise addition, and branch swapping with tree-bisection-reconnection with 100 addition sequence replicates) were conducted using PAUP* 4.0b10 (Swofford 2002), saving all the most parsimonious trees. Internal support for relationships was assessed using bootstrap analyses with 1,000 replicates. Homogeneity of the sequence data was assessed using the partition homogeneity test (Farris et al. 1995) with 1,000 branch and bound replicates, as implemented in PAUP* 4.0b10 (Swofford 2002).

For the Bayesian analysis of combined data, a distinct model was defined for each sequence based on Akaike Information Criterion: a general time-reversible (GTR) + I model for each of *trnL-trnF* and *rps4* adjacent noncoding sequences, a GTR + site-specific (SS) model for each of *rps4*, *rbcL* and *atpB* coding sequences. MrBayes version 3.1.2 (Huelsenbeck and Ronquist 2001; Ronquist and Huelsenbeck 2003) ran four Markov chains for 1,000,000 generations and sampled every 100 generations, starting with a random tree. Phylogenetic inferences were based on 7,500 trees sampled after that stability was reached.

Results and discussion

Among the 342 characters present in the *atpB* matrix, 151 sites (44%) were variable and 81 (24%) were phylogenetically informative. MP analysis of this data set yielded eight equally parsimonious trees, which differed in the positions of *E. arvense*, *E. diffusum*, *E. fluviatile*, *E. palustre*, *E. pratense*, *E. sylvaticum*, and *E. telmateia* (Fig. 1). Two major differences, compared with published phylogenies, were observed: (1) *E. bogotense* now branched within the *Hippochaete* clade [97% bootstrap (BS), decay index (DI) = 3]; and (2) *E. myriochaetum* grouped with *E. giganteum* in a monophyletic clade (98% BS, DI = 4). Support was much lower for other branches in the *Hippochaete* clade, except for the basal position of *E. scirpoides* (86% BS, DI = 2), and relationships among species of subgenus *Equisetum* were not resolved at all. ML and Bayesian analyses both confirmed these results (not shown).

After testing for homogeneity of data ($P = 0.089$; Farris et al. 1995), all DNA sequences available for the 15 *Equisetum* species (*rps4*, *rbcL*, *atpB* and *trnL-trnF*) were then combined. Among the 3,019 characters present in the conjoined matrix, 1,072 sites (36%) were variable, of

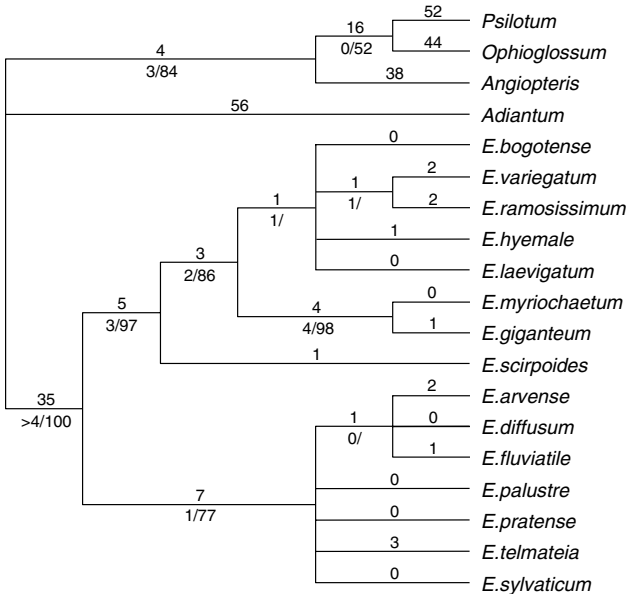

Fig. 1 One of eight shortest trees obtained in a parsimony analysis of *atpB* sequence data. *Numbers above branches* are branch lengths. Bremer decay indices (DI), followed by bootstrap (BS) percentages, are given *below branches*. A null decay index means that the branch collapses in the strict consensus tree. *Lines with no BS values below* denote nodes supported in < 50% of the replications. Tree length = 246 steps; consistency index (CI) = 0.83; retention index (RI) = 0.83

which 606 (20%) were phylogenetically informative. The Bayesian criterion was chosen for phylogeny reconstruction because it was shown to be more sensitive to phylogenetic signal and less sensitive to long-branch attraction (Alfaro et al 2003). Furthermore, it allowed the use of a distinct model for each analyzed sequence (see ''Materials and methods''), which greatly increased the likelihood of the model overall. Bayesian analysis of the conjoined matrix yielded a consensus tree that differed from published phylogenies (Fig. 2). Relationships inside subgenus *Hippochaete* were identical to those inferred from analyses of *rbcL* plus *trnL-trnF* sequences (Des Marais et al. 2003), with *E. bogotense* sister to *Hippochaete* [90% credibility value (CV)], whereas relationships within subgenus *Equisetum*, were compatible with those inferred from analyses of *rps4* coding sequence (Guillon 2004). Other analyses performed after substituting taxa for fern outgroups or including seed plant and Lycopodiopsida sequences gave identical results (not shown).

It is known that Bayesian confidence values do not correlate with BS values (Alfaro et al. 2003) and, as with BS or jackknife (JK) values, should not be considered as probabilities that clades are correctly resolved (Simmons et al. 2004). Here, the position of *E. bogotense* was further investigated by comparing the marginal likelihoods (approximated as harmonic means of the likelihoods values

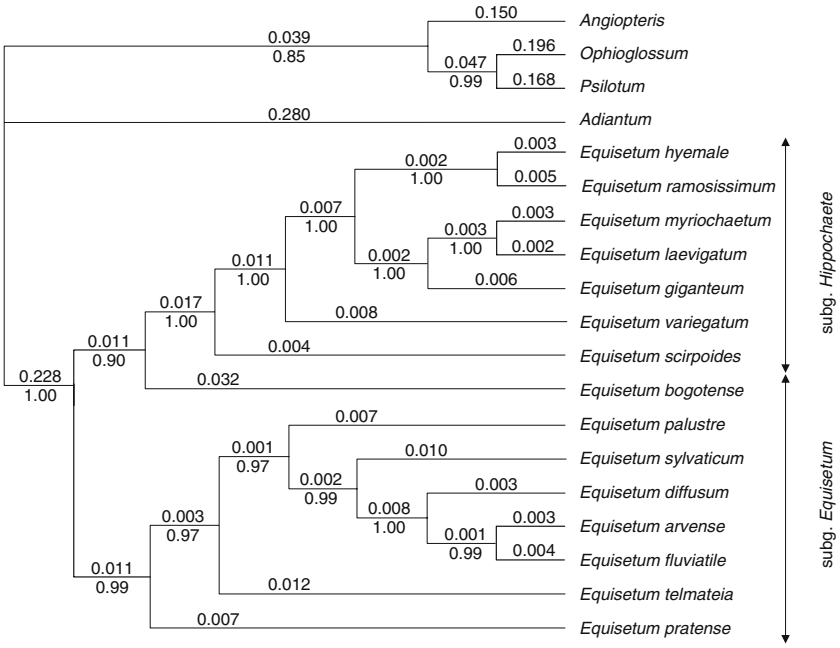

Fig. 2 The 50% majority-rule consensus tree obtained in a Bayesian analysis of the matrix combining *rbcL*, *atpB*, *trnL-trnF*, and *rps4* coding and adjacent noncoding sequences. *Numbers above branches* are mean changes per site. *Numbers below branches* are the credibility value (CV) calculated as the frequency of recovery of each clade

of trees sampled by MrBayes; Ronquist and Huelsenbeck 2003) under different topological constraints. The first constraint grouped *E. bogotense* and subgenus *Hippochaete* in a monophyletic clade, as in the tree obtained with the unconstrained analysis (Fig. 2), and yielded a ln likelihood of –11,681.38. The second constraint grouped all *Equisetum* species, except *E. bogotense* in a monophyletic clade, as in the alternative topology obtained by Des Marais et al. (2003; with MP) and Guillon (2004) and yielded a ln likelihood of –11,683.36. According to Kass and Raftery's table (1995), this difference provides positive evidence in favor of the first topology. Using the same method, very strong evidence was found against other potential branching positions for *E. bogotense*.

In addition to the monophyly of subgenus *Hippochaete* (100% CV) and subgenus *Equisetum* minus *E. bogotense* (99% CV), the following clades already recovered in separate analyses of *rps4* and *rbcL* + *trnL-trnF* obtained good support from the combined analysis: (1) *E. arvense*, *E. fluviatile*, and *E. diffusum* (100% CV), (2) *E. hyemale* and *E. ramosissimum* (100% CV), (3) *E. laevigatum* and *E. myriochaetum* (100% CV), (4) *E. hyemale*, *E. ramosissimum*, *E. laevigatum*, *E. myriochaetum*, and *E. giganteum* (100% CV), and (5) *E. hyemale*, *E. ramosissimum*, *E. laevigatum*, *E. myriochaetum*, *E. giganteum*, and *E. variegatum* (100% CV). In addition, (6) *Equisetum giganteum*, *E. laevigatum*, and *E. myriochaetum* (100% CV) and (7) *E. arvense* and *E. fluviatile* (99% CV) formed well-supported clades that were formerly recovered by Des Marais et al. (2003), and (8) *Equisetum arvense*, *E. fluviatile*, *E. diffusum* and *E. sylvaticum* (99% CV) formed another well-supported clade that was formerly recovered by Guillon (2004). Similar increase in resolution for combined analyses have been found in other studies including *rbcL* and *atpB* sequences (Wolf 1997; Tsutsumi and Kato 2005).

The sister relationship between *E. arvense* and *E. fluviatile* is congruent with the analysis of micromorphological data (Page 1972) and the observation that the corresponding natural hybrid *E.* x *litorale* is the most widespread in subgenus *Equisetum* (Duckett 1979). In contrast, the relationship of *E. sylvaticum* with *E. arvense*, *E. fluviatile,* and *E. diffusum* seems at odds with most classifications grouping *E. sylvaticum* with *E. pratense* on the basis of hybridization experiments (Duckett 1979), stem dimorphism (Hauke 1978), or antheridium morphology (Duckett 1973). The relationship of *E. giganteum* with *E. laevigatum* and *E. myriochaetum* is contrary to the view that *E. giganteum* would represent an early divergent lineage (Schaffner 1925, 1930; Hauke 1963). Indeed, taxonomists rather grouped *E. laevigatum* and *E. myriochaetum* with *E. ramosissimum* and *E. hyemale* in the section Primitiva (Schaffner 1930), or with *E. ramosissimum* in the section Ambigua (Hauke 1963). However, the close phylogenetic relationship of *E. giganteum* and *E. myriochaetum* is in keeping with the common occurrence of their hybrid *E. x schaffneri*. Overall, phylogenetic analyses based on DNA data (Des Marais et al. 2003; Guillon 2004; this study) do not support current infrageneric classifications of *Equisetum*. This incongruence may be the consequence of homoplasic evolution of morphological characters used by taxonomists. For instance, stem dimorphism led many authors to group *E. arvense*, *E. telmateia*, *E. pratense* and *E. sylvaticum* in the same section Heterophyadica (Hauke 1978). Yet the molecular phylog-

eny strongly suggests that stem dimorphism in *E. arvense*, and possibly in *E. sylvaticum*, is homoplasic.

Our analysis supports an early divergence of *E. bogotense* as sister to subgenus *Hippochaete* and the occurrence of two monophyletic clades (subgenus *Equisetum* minus *E. bogotense* and subgenus *Hippochaete*). This conclusion, based on available sequence data, should be restrained by the result of the separate analysis of *atpB*, which shows *E. bogotense* branching higher within the *Hippochaete* clade (Fig. 1). An error or a contamination is not responsible for this incongruence because it has been reproduced with DNA from two *E. bogotense* specimens of different origins (data not shown). Furthermore, it is difficult to figure out how inferences based on *atpB* (a chloroplast gene) can result from a process of discord (Maddison 1997). Nonrandom homoplasy (Lecointre and Deleporte 2005) cannot be totally ruled out, but no such bias as a difference in guanine cytosine (GC) content, in evolution rate or in nonsynonymous versus synonymous substitutions has been found likely to account for the observed pattern of incongruence. Hence, simple stochasticity may still be retained as the favored explanation, given the relatively low amount of *atpB* data. Future work on *Equisetum* phylogeny should help by sequencing additional chloroplast genes in order to find if the proposed basal position of *E. bogotense* can be confirmed from separate analyses.

An early divergence of *E. bogotense* has interesting implications for the evolution within *Equisetum*, as characters shared among subgenus *Equisetum* might be ancestral for horsetails. However, some of these characters (stem-branching pattern, cone shape, endodermal pattern) are homoplasic when compared among extant *Equisetum* species (Des Marais et al. 2003; Guillon 2004), so that inferences about their ancestral states are likely to be obscured by an excessive lability. Besides, some unique features of *E. bogotense*, such as filamentous and strictly unisexual gametophytes (Hauke 1968, 1969), might reflect a long history of evolution independent from other extant *Equisetum* lineages. Future research focused on *E. bogotense* are warranted in order to test these hypotheses.

Acknowledgments I thank Marc Girondot, Bernard Lejeune, and Sophie Nadot for useful advice and discussion. Odile Jonot, Paola Bertolino, and Catherine Colombeau are acknowledged for their help in DNA extractions and amplifications. Chad Husby provided plants from his personal collections and commented on the manuscript. Chad Husby and Franz Tod kindly collected many wild plants for me. I am also grateful to many people who sent me plants and, in some cases, vouchered them for me: Peter Atkinson at Cambridge University Botanic Garden; Holly Forbes at University of California Botanic Garden; Mike Gross at Georgian Court College; Philip O. Ashby, Fiona Inches, and Adele Smith at Royal Botanic Garden, Edinburgh; Sophie Dunand-Martin et Philippe Clerc at Conservatoire et Jardin Botanique de la Ville de Genève; Clinton Morse and Andrew Doran at University of Connecticut; and Nathalie Machon at Conservatoire National Botanique du Bassin Parisien. Tatiana Giraud, Claire Lavigne, Roxana Yockteng, and Odylle Cudelou are also acknowledged for their help.

References

Alfaro ME, 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

Des Marais DL, Smith AR, Britton DM, Pryer KM (2003) Phylogenetic relationships and evolution of extant horsetails, *Equisetum*, based on chloroplast DNA sequence data (*rbcL* and *trnL-F*). Int J Plant Sci 164:737–751

Duckett JG (1973) Comparative morphology of the gametophytes of the genus *Equisetum*, subgenus *Equisetum*. Bot J Linn Soc 66:1–22

Duckett JG (1979) An experimental study of the reproductive biology and hybridization in the European and North American species of *Equisetum*. Bot J Linn Soc 79:205–229

Farris JS, Källersjo AG, Kluge AG, Bult C (1995) Testing significance of incongruence. Cladistics 10:315–319

Guillon J-M (2004) Phylogeny of horsetails (*Equisetum*) based on the chloroplast *rps4* gene and adjacent noncoding sequences. Syst Bot 29:251–259

Hauke RL (1963) A taxonomical monograph of the genus *Equisetum* subgenus *Hippochaete*. Beheifte zur Nova Hedwigia 8:1–123

Hauke R L (1968) Gametangia of *Equisetum bogotense*. Bull Torrey Bot Club 95:341–345

Hauke RL (1969) Gametophyte development in Latin American horsetails. Bull Torrey Bot Club 96:568–577

Hauke RL (1978) A taxonomic monograph of *Equisetum* subgenus *Equisetum*. Nova Hedwigia 30:385–455

Hauke RL (1985) Gametophytes of *Equisetum giganteum*. Am Fern J 75:132

Huelsenbeck JP, Ronquist F (2001) MRBAYES: Bayesian inference of phylogeny. Bioinformatics 17:754–755

Kass RE, Raftery AE (1995) Bayes factors. J Am Stat Assoc 90:773–795

Kenrick P, Crane PR (1997) The origin and early evolution of plants on land. Nature 389:33–39

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

Krahulec F, Hrouda L, Kovarova M (1996). Production of gametophytes by *Hippochaete* (Equisetaceae) hybrids. Preslia 67:213–218

Lecointre G, Deleporte P (2005) Total evidence requires exclusion of phylogenetically misleading data. Zool Scr 34:101–117

Maddison WP (1997) Gene trees in species trees. Syst Biol 46:523–536

Nickrent DL, Parkinson CL, Palmer JD, Duff RJ (2000) Multigene phylogeny of land plants with special reference to bryophytes and the earliest land plants. Mol Biol Evol 17:1885–1895

Page CN (1972) An assessment of inter-specific relationships in *Equisetum* subgenus *Equisetum*. New Phytol 71:355–369

Pryer KM, Schneider H, Smith AR, Cranfill R, Wolf PG, Hunt JS, Sipes SD (2001) Horsetails and ferns are a monophyletic group and the closest living relatives to seed plants. Nature 409:618–621

Renzaglia KS, Duff RJ, Nickrent DL, Garbary DJ (2000) Vegetative and reproductive innovations of early land plants: implications for a unified phylogeny. Philos Trans R Soc B 355:769–793

Renzaglia KS, Dengate SB, Schmitt SJ, Duckett JG (2002) Novel features of Equisetum arvense spermatozoids: insights into pteridophyte evolution. New Phytol 154:159–174

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

Rosenthal A, Coutelle O, Craxton M (1993) Large-scale production of DNA sequencing templates by microtitre format PCR. Nucleic Acids Res 21:173–174

Schaffner JH (1925) Main lines of evolution in *Equisetum*. Am Fern J 15:8–12

Schaffner JH (1930) Geographical distribution of the species of *Equisetum* in relation to their phylogeny. Am Fern J 20:89–106

Simmons MP, Pickett KM, Miya M (2004) How meaningful are Bayesian support values ? Mol Biol Evol 21:188–199

Swofford DL (2002) PAUP*: phylogenetic analyses using parsimony (* and other methods), version 4.0b10. Sinauer Associates, Sunderland

Tsutsumi C, Kato M (2005) Molecular phylogenetic study on Davalliaceae. Fern Gaz 17:147–162

Wikström KM, Pryer KM (2005) Incongruence between primary sequence data and the distribution of a mitochondrial *atp1* group II intron among ferns and horsetails. Mol Phylogenet Evol 36:484–493

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