Cladistics

Bias and conflict in phylogenetic inference of myco-heterotrophic plants: a case study in Thismiaceae

Vincent Merckx[a,*], Freek T. Bakker[b], Suzy Huysmans[a] and Erik Smets[a,c]

[a]*Laboratory of Plant Systematics, K.U.Leuven, Kasteelpark Arenberg 31, PO Box 2437, BE-3001 Leuven, Belgium;* [b]*National Herbarium of the Netherlands, Wageningen University Branch, Gen. Foulkesweg 37, 6703 BL Wageningen, The Netherlands;* [c]*National Herbarium of the Netherlands, Leiden University Branch, PO Box 9514, NL-2300 RA, Leiden, The Netherlands*

Abstract

Due to morphological reduction and absence of amplifiable plastid genes, the identification of photosynthetic relatives of heterotrophic plants is problematic. Although nuclear and mitochondrial gene sequences may offer a welcome alternative source of phylogenetic markers, the presence of rate heterogeneity in these genes may introduce bias/systematic error in phylogenetic analyses. We examine the phylogenetic position of Thismiaceae based on nuclear 18S rDNA and mitochondrial *atpA* DNA sequence data, as well as using parsimony, likelihood and Bayesian inference methods. Significant differences in evolutionary rates of these genes between closely related taxa lead to conflicting results: while parsimony analyses of 18S rDNA and combined data strongly support the monophyly of Thismiaceae, Bayesian inference, with and without a relaxed molecular clock, as well as the Swofford–Olsen–Waddell–Hillis (SOWH) test confidently reject this hypothesis. We show that rate heterogeneity in our data leads to long-branch attraction artifacts in parsimony analysis. However, using model-based inference methods the question of whether Thismiaceae are monophyletic remains elusive. On the one hand maximum likelihood nonparametric bootstrapping and parametric hypothesis tests fail to support a paraphyletic Thismiaceae, on the other hand Bayesian inference methods (both without and with a relaxed clock) significantly reject a monophyletic Thismiaceae. These results show that an adequate sampling, the use of rate homogeneous data, and the application of different inference methods are important factors for developing phylogenetic hypotheses of myco-heterotrophic plants.

© The Willi Hennig Society 2009.

Introduction

Myco-heterotrophic plants (frequently wrongly referred to as ''saprophytes'') present a remarkable case of a parasitic nutrition strategy. Unlike plants that are directly parasitic on other plants (holo- or hemiparasites) myco-heterotrophic plants obtain carbon from adjacent photosynthetic plants through a shared common mycorrhizal network (Bidartondo et al., 2002; Leake, 2005; Selosse et al., 2006). This allows them to grow in shaded conditions, where autotrophic plants fail to reproduce (Bidartondo et al., 2004). The myco-heterotrophic mode of nutrition has evolved independently several times, and in non-related taxonomic groups. Many myco-heterotrophs show strong convergent evolution in particular adaptations to their mode of life, making the identification of the relatives of myco-heterotrophic plants in many cases a phylogenetic reconstruction challenge (Leake, 1994).

Particularly problematic are the myco-heterotrophic Thismiaceae, a monocot family that consists of more than 50 species in five genera (Maas-van de Kamer, 1998; Woodward et al., 2007). All Thismiaceae species are small achlorophyllous myco-heterotrophic herbs. Leaves are absent or reduced to scales. Most species occur in dense tropical rain forest and can only be spotted during the flowering period when aboveground organs appear through the leaf litter (Maas et al., 1986). The largest genus, *Thismia*, comprises about 32 species and occurs mainly in the tropics of Asia and South

Corresponding author:
 E-mail address: vincent.merckx@bio.kuleuven.be

America although some species extend into subtropical and even temperate regions (Maas et al., 1986). Remarkably, *Thismia* is absent from Africa while the genus *Afrothismia* with 11 species is endemic to tropical Africa. Other Thismiaceae genera are *Oxygyne*, with one species from Cameroon and two from Japan, *Haplothismia*, with one species endemic to the Western Ghats (India), and the recently discovered *Tiputinia*, with a single species from Ecuador (Sasidharan and Sujanapal, 2000; Woodward et al., 2007; Yokoyama et al., 2008). A remarkably common feature of most Thismiaceae is their extreme scarcity (Franke, 2004). The majority of species are known exclusively from the type collection, which in some cases dates back more than a century (Jonker, 1938; Maas et al., 1986). Many species are presumed extinct (Maas et al., 1986; Jarvie, 1996) while new taxa are discovered almost every year (e.g. Sainge et al., 2005; Cheek, 2006; Woodward et al., 2007).

Not surprisingly, due to the strong reduction of vegetative organs and the rarity of most species involved, Thismiaceae taxonomy has been the subject of much debate. Most classifications included Thismiaceae, as a subtribe "Thismieae", in a broadly defined Burmanniaceae (Miers, 1847; Jonker, 1938; Stevenson and Laconte, 1995; Maas-van de Kamer, 1998) while other authors favored the recognition of a separate Thismiaceae closely related to the myco-heterotrophic Burmanniaceae (Schlechter, 1921; Hutchinson, 1934, 1959; Dahlgren et al., 1985; Takhtajan, 1997). Thismiaceae or Burmanniaceae (including Thismieae) on their part were linked to various other families. Many authors proposed affinities with other myco-heterotrophic families such as Triuridaceae (Eichler, 1875; Baillon, 1894), Geosiridaceae (Cronquist, 1970), Corsiaceae (Hutchinson, 1959; Dahlgren et al., 1985), and Orchidaceae (Lindley, 1846; Karsten, 1858; Engler, 1888; Cronquist, 1970; Rübsamen, 1986). However, these relationships are now completely discredited based on convergence of character states involved, due to the myco-heterotrophic mode of life (Soltis et al., 2005). In addition, Thismiaceae were linked to a wide variety of monocot families including Haemodorocaceae, Iridaceae, Amaryllidaceae, Bromeliaceae, Taccaceae, and Velloziaceae (see Maas et al., 1986 for an overview) corroborating the lack of consensus about the relationships of these enigmatic plants. DNA-based phylogenetic analyses placed Thismiaceae in Dioscoreales (Caddick et al., 2000, 2002; Davis et al., 2004), but different results for their position within this order were obtained dependent on different sampling and data (Caddick et al., 2002; Merckx et al., 2006).

The use of DNA sequence data for the phylogenetic placement of heterotrophic plants is not without problems. Apart from poor sampling due to the general rarity of many taxa, the genomes of heterotrophic plants are often highly degraded, preventing the use of commonly used phylogenetic markers. In most holo-parasitic and myco-heterotrophic plants the commonly used plastid genes are highly divergent or even lacking entirely (dePamphilis and Palmer, 1990; dePamphilis et al., 1997; Nickrent et al., 1998; Wolfe and dePamphilis, 1998; Caddick et al., 2002; Bungard, 2004; Young and dePamphilis, 2005) and nuclear 18S rDNA of many species was found to evolve many times faster than in related photosynthetic plants (Nickrent and Starr, 1994; Nickrent et al., 1998; Barkman et al., 2004; Merckx et al., 2006). A strong bias in evolutionary rates of DNA sequences may mislead phylogenetic inference methods in producing artificial clades (Felsenstein, 1978; Hendy and Penny, 1989). Particularly when parsimony is chosen as the optimality criterion, strong nucleotide substitution rate differences among lineages has been shown to lead to systematic error known as "long branch attraction" (LBA; Swofford et al., 2001; Bergsten, 2005). Empirical studies have demonstrated that LBA can occur in real data sets, however, the pervasiveness of this problem remains unclear (Wiens and Hollingsworth, 2000; Clements et al., 2003; Anderson and Swofford, 2004; Wilcox et al., 2004; Bergsten, 2005).

Recently, the use of mitochondrial gene sequences has proven highly effective for inferring phylogenetic relationships of various achlorophyllous angiosperm groups (Nickrent et al., 2002; Barkman et al., 2004, 2007; Nickrent et al., 2004, 2005; Merckx et al., 2006; Merckx and Bidartondo, 2008). In many of these cases mitochondrial gene sequences showed less pronounced rate heterogeneity compared with that in 18S rDNA, and may therefore provide a useful alternative when plastid data is unavailable and nuclear data is plagued by rate heterogeneity. In any case congruence between nuclear and mitochondrial-derived topologies should be checked as it may be highly informative for the presence/possibility of (systematic) artefacts. For example, cytonuclear incongruence may suggest LBA due to rate heterogeneity in nuclear genes (Bergsten, 2005), or the presence of nonstandard processes such as RNA editing, introgression, lineage sorting, and horizontal gene transfer in one of the partitions (Wendel and Doyle, 1998; Davis and Wurdack, 2004; Mower et al., 2004; Nickrent et al., 2004; Barkman et al., 2007).

In this study we explore the use of mtDNA, detecting possible LBA artefacts, as well as expand upon previous phylogenetic hypotheses for Thismiaceae (Caddick et al., 2002; Merckx et al., 2006) by analyzing 18S rDNA and *atpA* mtDNA sequence data of 14 Thismiaceae taxa supplemented by a broad Dioscoreales sampling. Separate as well as combined gene tree topologies generated under different phylogenetic inference methods are discussed and compared with each other. Parametric and nonparametric bootstrapping is applied in order to critically evaluate the impact of LBA

and possible conflict between data sets and to formulate suggestions for "best practices" in similar experiments.

Materials and methods

Data

Fourty-one 18S rDNA sequences of Diosoreales were obtained from Caddick et al. (2002), Merckx et al. (2006), Merckx and Bidartondo (2008), and Merckx et al. (2008b). Five additional sequences were obtained with methods described in Merckx et al. (2006). Thirty-five *atpA* sequences were obtained from Davis et al. (2004) and Merckx and Bidartondo (2008). Eleven additional sequences were obtained with methods and primers from Eyre-Walker and Gaut (1997). The Appendix lists GenBank accessions of all included sequences.

Our sampling covers four out of five Thismiaceae genera (unsampled genus = *Oxygyne*) including six of twelve described *Afrothismia* species (50%) but only five of ca. 30 species of *Thismia* (approx. 17%). However, given the extreme rarity of the study group and the scarcity of herbarium material our sampling is the most complete currently available. Recently Yokoyama et al. (2008) reported on the phylogenetic position of *Oxygyne* using an 18S rDNA sequence of *O. shinzatoi*. At the time of writing of the current manuscript this sequence was not made available through a public database. All sequence data were manually aligned with MacClade 4.04 (Maddison and Maddison, 2001).

Phylogenetic inference

Phylogenies were estimated under the parsimony (MP) and maximum likelihood (ML) criteria, as well as using a Bayesian approach without (BI) and with (BRC) a relaxed clock, using PAUP* ver. 4b10 (Swofford, 2002), GARLI ver. 0.951 (Zwickl, 2006), MrBayes ver. 3.1.2 (Huelsenbeck and Ronquist, 2001; Ronquist and Huelsenbeck, 2003), and BEAST ver. 1.4.6 (Drummond and Rambaut, 2007), respectively. Parsimony analyses were performed via heuristic searching using TBR branch swapping and 500 random addition sequence replicates saving multiple trees at each step. Branch support was calculated by nonparametric bootstrap analysis (Felsenstein, 1985) from heuristic searches on 1000 pseudo-replicate data sets. For ML, BI and BRC analyses the GTR + I + G model of evolution was identified on the ingroup as the best-fit model for both the 18S rDNA and *atpA* data sets and for the combined data using ModelTest ver. 3.06 (Posada and Crandall, 1998). The search for the tree with the highest likelihood was performed with GARLI estimating all model parameter values. The search was ended if no significantly improved topology (likelihood increase of 0.01) was found for 10 000 generations. Model parameters, branch lengths, and likelihood score of the final tree were optimized in PAUP* by performing a heuristic ML search with the GTR + I + G model of evolution but with the most optimal GARLI tree as constraint. Clade support was estimated by nonparametric bootstrap analysis on 500 pseudo-replicate data sets using the same search strategies and the GTR + I + G model of evolution. Bayesian analysis consisted of 5×10^6 generations starting from a random tree and sampling every 1000 generations. Two runs were conducted simultaneously, each consisting of four chains (one cold, three incrementally heated) with temperature T set to 0.2, the default value. A majority-rule consensus tree was calculated from the trees sampled during the last 1×10^6 generations. Convergence of the chains was checked using TRACER ver. 1.4 (Rambaut and Drummond, 2007) and the effective sampling size (ESS) parameter was found to exceed 500 for all parameters, which suggests acceptable mixing and sufficient sampling. Successful swap frequencies between chains (0.23–0.54) also suggested sufficient convergence. The Bayesian analysis of the combined data was performed with a partitioned model approach (GTR + I + G on each data set). Analyses were rerun with the codon option for *atpA* and with the 18S rDNA divided into stem and loop partitions, with a GTR + I + G model on each partition, but converged to similar results (not shown). To investigate the influence of taxon sampling on the difference between our results and those presented in Merckx et al. (2006) the BI analysis of the 18S rDNA dataset was rerun with all *Afrothismia* species excluded except *A. hydra*. Additionally we performed a Bayesian relaxed clock analysis with all three data sets using BEAST ver. 1.4.6 (Drummond and Rambaut, 2007). We applied a GTR + I + G model with four gamma categories on each data set. For the analysis of the combined data a GTR + I + G model was assigned to each gene partition following the "BEAST partitioning" manual (http://tlpcouvreur.googlepages.com/beastpartitioning). The uncorrelated lognormal clock model (Drummond et al., 2006) was selected for all analyses and two secondary calibration points with standard deviations were taken from a MULTIDIVTIME analysis (Thorne and Kishino, 2002) on monocot 18S rDNA and *atpA* data (Merckx and Bidartondo, 2008): a prior with a normal distribution of 120 ± 6 Ma was set for the root of the tree (stem node of Dioscoreales) and a prior with a normal distribution of 113 ± 8 Ma for the crown node of Dioscoreales. The distribution of all other priors was set to uniform. Posterior distributions of parameters were approximated using two independent Markov chain Monte Carlo analyses of 20 000 000 generations followed by a discarded burn-in of 2 000 000 generations (10%). Convergence of the chains was checked by evaluating the ESS values of each

parameter with TRACER ver. 1.4. Results were plotted on the tree in the posterior sample that has the maximum sum of posterior probabilities on its $n − 2$ internal nodes using the "Maximum clade credibility" option in TREEANNOTATOR ver. 1.4.6 (Drummond and Rambaut, 2007). The XML BEAST input files are available online as supporting information.

Alternative hypothesis testing

For all three data sets alternative hypotheses to the tree with the best likelihood inferred as outlined above were compared, using the approximately unbiased (AU) test (Shimodaira, 2002) and the Shimodaira–Hasegawa (SH) test (Shimodaira and Hasegawa, 1999). Alternative hypotheses included: (i) the highest likelihood topologies calculated from the other two data sets, (ii) the tree with the highest likelihood estimated with the constraint that Thismiaceae are monophyletic, (iii) all most parsimonious trees for each data set, (iv) 100 trees (due to computational limitations) randomly selected from the trees sampled during the Bayesian analysis of each data set once stationary was reached, (v) the tree resulting from the BRC analysis of each data set. Site-wise log-likelihoods for all trees were estimated using PAUP* and used as input for CONSEL ver. 0.1i (Shimodaira and Hasegawa, 2001). Multiscale bootstrap resampling was conducted with ten sets of 10 000 replicates each, with scale parameters ranging from 0.5 to 1.4.

Additionally we used parametric bootstrap analysis [Swofford–Olsen–Waddell–Hillis (SOWH) test] to test the monophyly of the Thismiaceae for each data set (Hillis et al., 1996; Huelsenbeck et al., 1996; Swofford et al., 1996; Goldman et al., 2000). A null model tree (Thismiaceae constrained to a clade) was estimated from the original data using GARLI. Likelihood score and model parameters of the null model were optimized in PAUP*. This model was then used to simulate 100 replicate data sets with SEQ-GEN ver. 1.3.2 (Rambaut and Grassly, 1997). On each of the data sets two ML tree searches were conducted using PAUP*, one to find the optimal tree, the other to find the optimal tree consistent with the null hypothesis. Due to computational limitations each heuristic search consisted of TBR branch-swapping on a single neighbour-joining tree with fixed model parameters. The distribution of differences in likelihood score between the "optimal" and "null" scores was determined and used to evaluate the significance of difference between the two topologies.

Testing for long-branch attraction

Parsimony analyses on 18S and 18S + *atpA* data sets recovered different Thismiaceae tree topologies (T_P) than maximum likelihood (T_{ML}) and Bayesian analyses (Figs 1 and 3). Since these differences involve long branches we used the method developed by Huelsenbeck (1997) to determine if long-branch attraction (LBA) is responsible for the observed conflict. Therefore we simulated 100 replicate data sets with the same dimensions as the original data sets on T_P and T_{ML}, using the GTR + I + G parameter values estimated previously by PAUP* for the ML trees (see "Phylogenetic inference"). Branch lengths and substitution model parameters for T_P were estimated by PAUP* prior to the data simulations. Both ML and MP heuristic searches were then performed on each replicate data set. Following Wilcox et al. (2004) our expectation was that if MP is not significantly influenced by LBA, then < 5% of the trees estimated by MP from data simulated on the ML topology should recover a monophyletic Thismiaceae. On the other hand, ML should not recover a paraphyletic Thismiaceae for data sets simulated on the MP tree (Wilcox et al., 2004).

Bayesian relative rates test

Bayesian relative rates tests on 18S rDNA and *atpA* data sets were conducted as described by Wilcox et al. (2004). Unlike the commonly used Wu and Li (1985) relative rates test, the Bayesian relative rates test accounts for tree structure and allows direct rate comparison between multiple taxa or clades. For each taxon the distribution of branch-length distance from the most-recent common ancestor (MRCA) of the ingroup was constructed based on the branch-lengths of 500 randomly sampled trees retained by the Bayesian analysis after stationary was reached. The ingroup (and corresponding MRCA) used for the branch length comparisons needs to be monophyletic in all of the sampled trees and therefore consisted of all Dioscoreales taxa except Nartheciaceae and Burmanniaceae for 18S rDNA trees and all Dioscoreales taxa except Nartheciaceae for *atpA* trees.

Results

Phylogenetic inference

MP analysis of 18S rDNA data recovered 19 equally most parsimonious trees of length 1201 steps. All most parsimonious trees contain a Thismiaceae clade with *Tacca* as sister clade. The monophyly of Thismiaceae is strongly supported (98% Bootstrap support; BS; Fig. 1a). ML analysis produced a similar topology (−ln L = 8248.14; Fig. 1b), differing mainly in the fact that Thismiaceae are a paraphyletic group, i.e. the genera *Thismia–Haplothismia–Tiputinia* are in a clade sister to *Tacca*, but this relationship is not supported, and *Afrothismia* is the sister group of this clade (68% BS). BI analysis obtains a tree similar to ML (Fig. 1b).
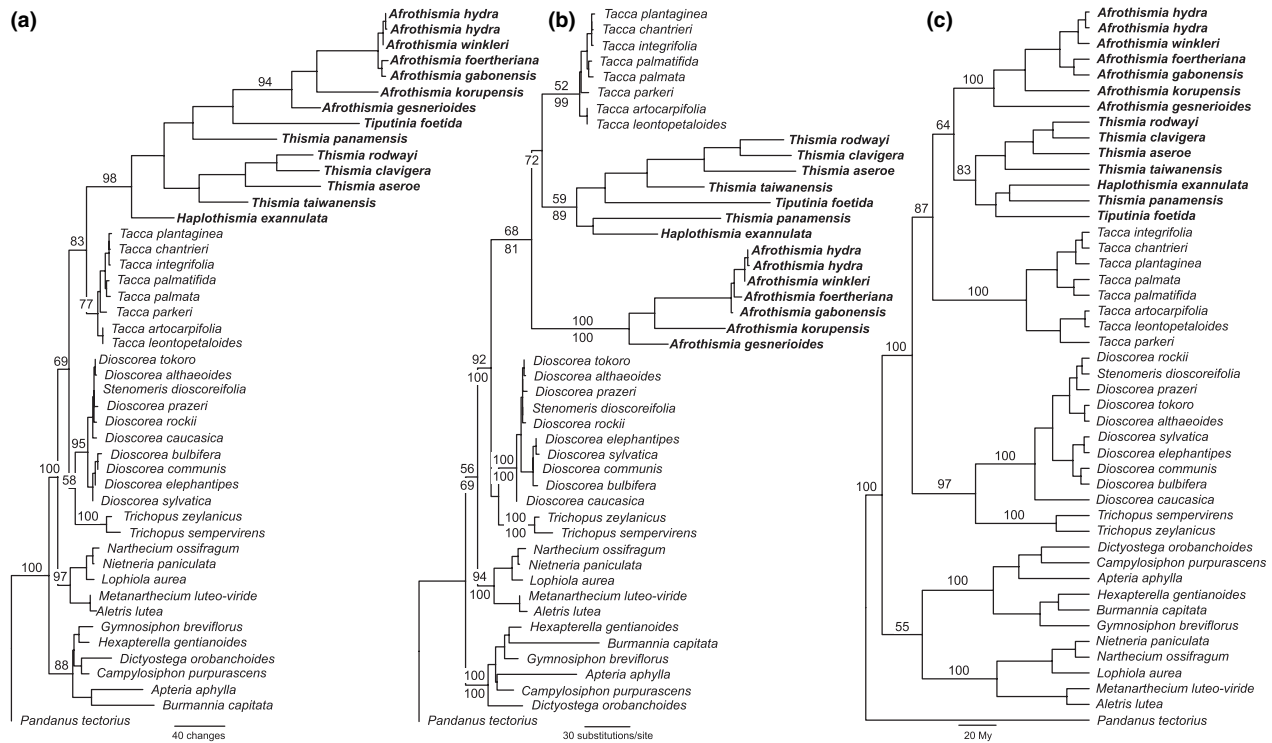
Fig. 1. Phylogenic relationships in Dioscoreales based on 18S rDNA sequences. Thismiaceae are indicated in bold. (a) One of 19 most parsimonious trees (1201 steps). Nonparametric bootstrap support values of major clades are shown above branches. (b) Most optimal maximum likelihood tree (−ln L = 8248.14). Non-parameteric bootstrap support values of major clades are shown above branches, Bayesian posterior probabilities below branches. (c) Chronogram obtained by Bayesian relaxed clock analysis with Bayesian posterior probabilities of major clades shown above branches.

The *Thismia–Haplothismia–Tiputinia–Tacca* clade is not significantly supported (72% BPP). BI analysis with relaxed clock (BRC) retrieves a monophyletic Thismiaceae without significant support (64% BPP; Fig. 1c). A rerun of the BI analysis with *Afrothismia* reduced to a single taxon (*A. hydra*) resulted in a monophyletic Thismiaceae with *A. hydra* placed as the sister group of *Thismia panamensis* (result not shown).

MP analysis of *atpA* data obtained 384 most parsimonious trees with a length of 478 steps (Fig. 2a). A *Thismia–Haplothismia–Tiputinia–Tacca* clade is present in the strict consensus of these trees, but receives no bootstrap support. The relationships between the main clades in the *atpA* strict consensus cladogram differ considerably from the strict consensus based on 18S rDNA data, but none of these differences are supported. The optimal ML tree based on *atpA* data (−ln L = 4502.71; Fig. 1b) is in agreement with the MP topology and also favors a paraphyletic Thismiaceae, but bootstrap support for internal relationships is lacking. In the Bayesian majority-rule consensus tree, which is highly similar to the ML topology, the *Thismia–Haplothismia–Tiputinia–Tacca* clade is significantly supported (96%

BPP; Fig. 1b). Significant support (95% BPP) for a *Thismia–Haplothismia–Tiputinia–Tacca* clade is also obtained by BRC analysis of *atpA* data (Fig. 2c). In comparison with MP, ML, and BI analyses, BRC analysis favors an alternative position of *Afrothismia* but without branch support.

MP analysis of the combined 18S rDNA and *atpA* data retains 1824 most optimal trees with 1704 steps. A Thismiaceae clade is present in all these trees and obtains 98% support in the bootstrap analysis (Fig. 3a). *Tacca* is sister group of this clade with strong support (93% BS). The optimal ML tree (−ln L = 13143.71; Fig. 3b) mainly differs from the MP strict consensus by the paraphyly of Thismiaceae. Although the *Thismia–Haplothismia–Tiputinia–Tacca* clade is not supported, this clade is maximally supported in the BI majority-rule consensus tree (100% BPP; Fig. 3b). BRC analysis retrieves a tree with relationships between the main clades similar to ML and BI analyses (Fig. 3c). The *Thismia–Haplothismia–Tiputinia–Tacca* clade is significantly supported (100% BPP), but the position of *Afrothismia* as sister group of this clade remains without strong support (62% BPP).
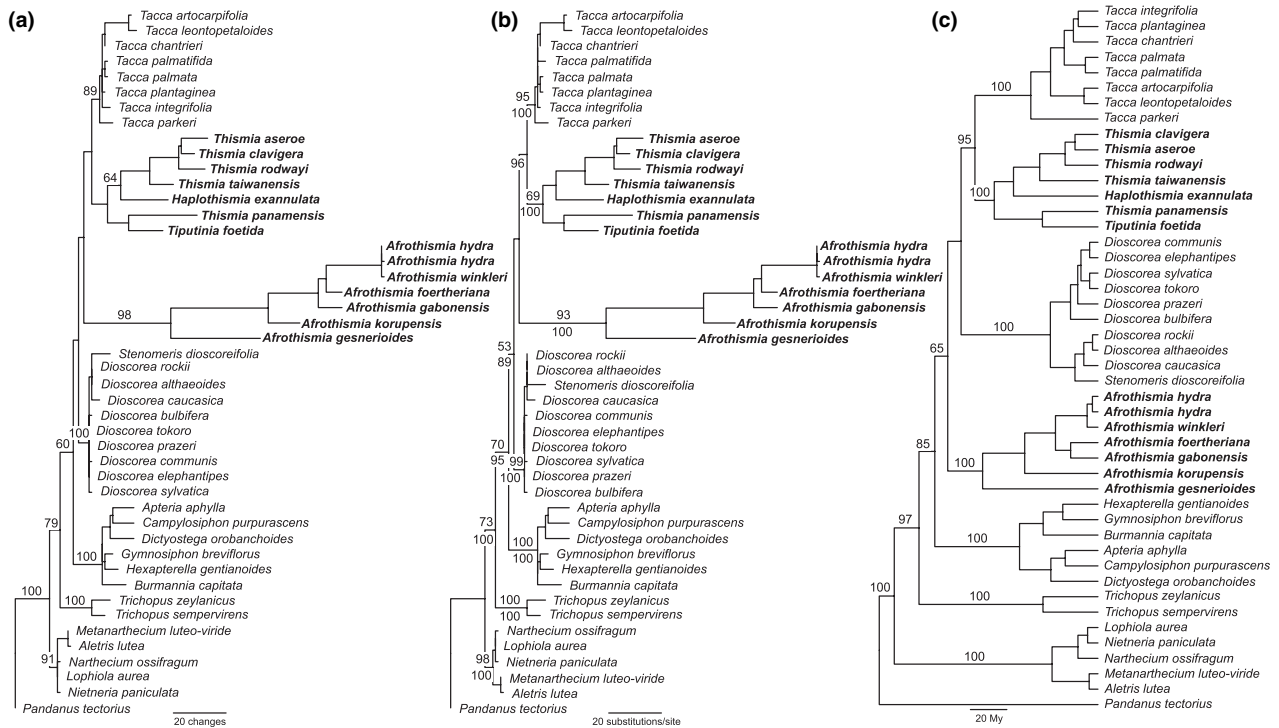
Fig. 2. Phylogeny of Dioscoreales based on *atpA* sequences. Thismiaceae are indicated in bold. (a) One of 384 most parsimonious trees (478 steps). Nonparametric bootstrap support values of major clades are shown above branches. (b) Most optimal maximum likelihood tree (−ln L = 4502.71). Nonparametric bootstrap support values of major clades are shown above branches, Bayesian posterior probabilities below branches. (c) Bayesian analysis with a relaxed clock. Bayesian posterior probabilities are shown above branches.

Alternative hypothesis testing

In four of 12 hypotheses obtained with different data sets and different optimality criteria (Figs 1–3), the paraphyletic status of Thismiaceae is significantly supported (BI and BRC analysis of *atpA* and 18S + *atpA* data). In contrast, MP analysis of 18S rDNA and of the 18S + *atpA* data retains a strongly supported monophyletic Thismiaceae. Table 1 lists the results of AU and SH tests to compare the topologies obtained under likelihood to each other, MP topologies, ML trees generated under a monophyletic Thismiaceae constraint, BI trees, and BRC trees. Both tests indicate that the topologies generated by ML with 18S rDNA and *atpA* data are significantly different (*P* < 0.01). In addition, the ML topology resulting from the 18S + *atpA* data set is significantly different from the *atpA* ML topology. Monophyly of Thismiaceae cannot be rejected by any of the parametric ML tests. When the most parsimonious trees, BI trees, and BRC trees are compared to the ML topologies significant between data set-differences are found, but not between inference methods/optimality criteria.

With ML analysis of *atpA*, the best tree with Thismiaceae monophyletic is 3.7 log-likelihood units worse than the optimal tree, containing a paraphyletic

Thismiaceae (Fig. 2b). Based on the SOWH test this difference is significant at 5% level and monophyly of Thismiaceae can thus be rejected using *atpA* data (Fig. 4). The log-likelihood difference between the 18S rDNA ML tree and the tree with Thismiaceae monophyletic is 1.2. Although this difference seems minor, the SOWH test reveals that the distribution of the log-likelihood differences under the null model (Thismiaceae monophyletic) is zero: analysis on data sets generated under the null model always recovers a Thismiaceae clade. For the SOWH test on 18S + *atpA* data similar results were obtained. All SOHW tests thus reject a monophyletic Thismiaceae.

Long-branch attraction

With the method of Huelsenbeck (1997) we found that MP recovered a monophyletic Thismiaceae for 14% of the data sets simulated along the 18S rDNA ML tree (with a paraphyletic Thismiaceae). ML analysis on the same simulated data sets never recovered a Thismiaceae clade. Simultaneously, ML and MP analyses on data sets simulated under Thismiaceae monophyly constraint (as recovered by MP with the original data) always recovered a monophyletic Thismiaceae. In a similar LBA test on the combined data set 7% of the MP trees were influenced by LBA.
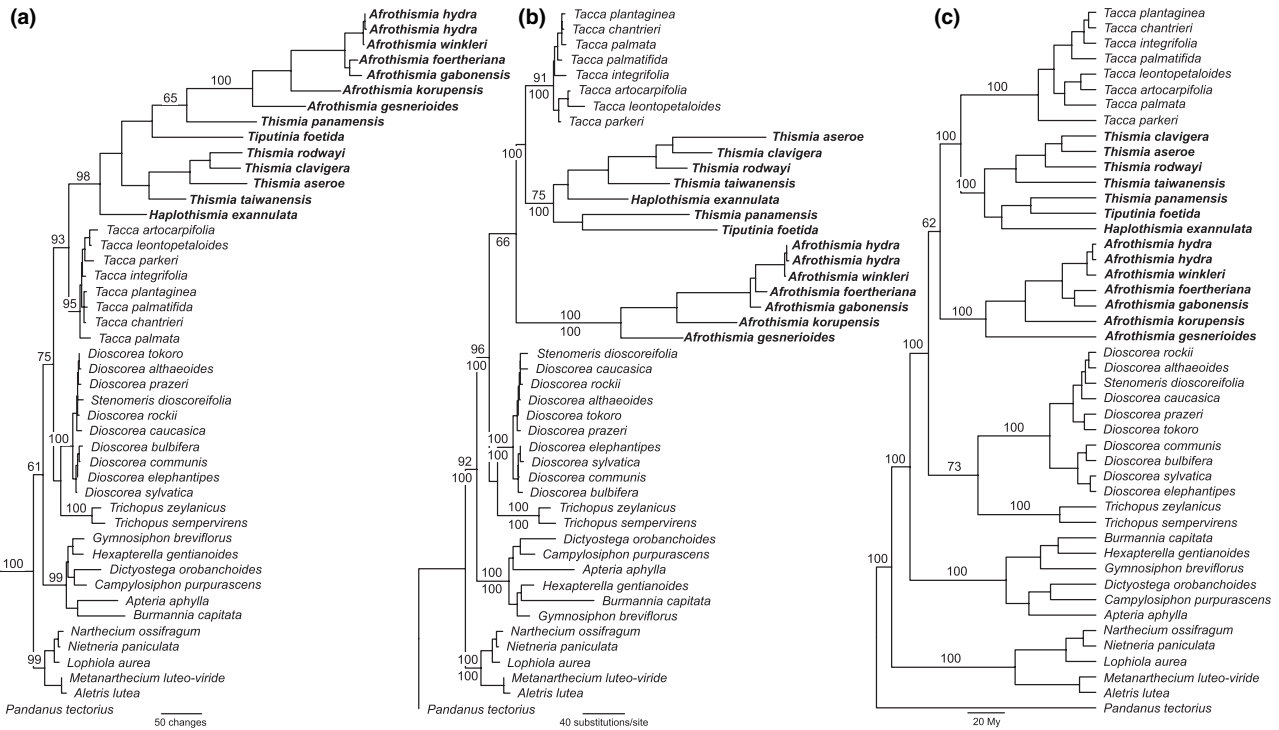
Fig. 3. Phylogeny of Dioscoreales based on combined analysis of 18S rDNA and *atpA* sequences. (a) One of 1824 most parsimonious trees (1704 steps). Nonparametric bootstrap support values of major clades are shown above branches. (b) Most optimal maximum likelihood tree (−ln L = 13143.71) with nonparametric bootstrap support values of major clades shown above branches. Bayesian posterior probabilities are indicated below branches. (c) Chronogram resulting from the Bayesian analysis with a relaxed molecular clock. Bayesian posterior probabilities of major clades are shown above branches.

Bayesian relative rates tests

The Bayesian relative rates plot of 18S rDNA (Fig. 5a) clearly shows a significant variation in branch lengths, and thus relative substitution rates, between Thismiaceae and its photosynthetic relatives. Compared to the average branch length of the longest *Tacca* branch (average of 0.33 from MRCA) the relative rate of 18S rDNA in Thismiaceae is more than 2.5-fold (in *Haplothismia*; average of 0.086 from MRCA) to more than 6.5-fold faster (in *Tiputinia*; average of 0.22 from MRCA). No significant differences in relative rates are observed between Thismiaceae species. The Bayesian relative rates plot of *aptA* shows a slightly different pattern (Fig. 5b). While all species of *Afrothismia* have significantly longer branches than any Dioscoreaceae species, the 95% confidence intervals of the other Thismiaceae are overlapping with those of Dioscoreaceae. The average branch length in *Afrothismia* is 2.5-fold (0.1 from MRCA in *A. korupensis*) to four fold (0.16 from MRCA in *A. hydra*) longer than the longest branch in *Tacca* (0.04 from MRCA). Except for *A. gesnerioides* and *A. korupensis* all *Afrothismia* species have significantly higher evolutionary rates than the remaining Thismiaceae.

Discussion

Conflicting topologies

Comparing the optimal trees resulting from 18S rDNA and 18S + *atpA* data, there is significant conflict between parsimony and most model-based analysis results. While parsimony recovers a monophyletic Thismiaceae, ML, BI and BRC (for 18S + *atpA* only) favors a paraphyletic Thismiaceae grouping. Support for this topological difference varies, but reaches maximal disagreement in the MP versus BI and BRC analyses of the combined data set. The SOWH test corroborates this by rejecting a monophyletic Thismiaceae for all data sets. Paradoxically, the AU and SH tests suggest that the topological conflicts between the different phylogenetic inference methods are not significant. The Bayesian relative rates test indicates that the 18S rDNA substitution rate is increased in all Thismiaceae taxa. Indeed, using the method of Huelsenbeck (1997), MP analysis on 18S rDNA and 18S + *atpA* data is shown to be significantly biased by LBA. MP analysis of these data sets groups the long-branch taxa, regardless what the true tree might be, leading to high bootstrap support for either correct or incorrect results.

Table 1
Comparison between the best ML trees (Figs 1–3), user constrained trees, and most parsimonious trees using 18S rDNA, *atpA*, and combined data sets by the approximately unbiased (AU) and Shimodaira–Hasegawa (SH) tests

Hypothesis	18S ML tree			*atpA* ML tree			18S + *atpA* ML tree		
	Δ ln L	AU	SH	Δ ln L	AU	SH	Δ ln L	AU	SH
18S ML topology	Best	0.664	0.935	67.7	<0.01	<0.01	31.1	0.237	0.368
atpA ML tree	162.2	<0.01	<0.01	Best	0.923	0.977	137.6	<0.01	<0.01
18S + *atpA* ML tree	55.9	<0.01	0.056	59.9	<0.01	0.012	Best	0.877	0.784
Thismiaceae monophyletic ML topology	1.2	0.596	0.908	3.7	0.077	0.748	1.18	0.767	0.687
18S most parsimonious trees (19)	13.2–25.7	0.076–0.391	0.640–0.356	84.7–92.9	<0.01	<0.01	47.2–59.4	0.208–0.023	0.157–0.051
atpA most parsimonious trees (384)	196.4–240.2	<0.01	<0.01	0–0.5	0.502–0.203	0.994–0.470	166.8–207.5	<0.01	<0.01
18S + *atpA* most parsimonious trees (1824)	13.9–32.5	0.605–0.012	0.492–0.087	41.8–72.8	<0.01	<0.01	20.1–41.7	0.628–<0.01	0.489–0.148
18S 100 randomly chosen BI topologies	0.4–13.0	0.763–0.083	0.999–0.062	37.7–106.6	<0.01	0.063–<0.01	19.4–83.3	0.771–<0.01	0.250–<0.01
atpA 100 randomly chosen BI topologies	87.5–247.7	<0.01	<0.01	0–17.0	0.892–<0.01	0.999–0.228	63.5–210.5	<0.01	0.139–<0.01
18S + *atpA* 100 randomly chosen BI topologies	51.7–78.6	<0.01	<0.01	49.2–75.5	<0.01	<0.01	0–11.5	0.717–0.157	0.999–0.658
18S most optimal BRC topology	3.0	0.389	0.745	67.6	<0.01	<0.01	32.1	0.139	0.304
atpA most optimal BRC topology	149.1	<0.01	<0.01	2.0	0.250	0.775	123.8	<0.01	<0.01
18S + *atpA* most optimal BRC topology	59.1	<0.01	<0.01	59.5	<0.01	<0.01	0.7	0.550	0.826

Significant differences are indicated in bold.

This bias makes MP an inappropriate method for analyzing our 18S rDNA and 18S + *atpA* data sets and the forthcoming results should be discarded. Interestingly, the absence of increased substitution rates in *atpA* sequences of *Thismia*, *Haplothismia*, and *Tiputinia* suggests a reduced impact of LBA, and indeed, MP produces a topology similar to that obtained with likelihood and Bayesian analysis.

Bayesian analysis with and without a relaxed clock of *atpA* and 18S + *atpA* data sets significantly rejects a monophyletic Thismiaceae. This contrasts with ML bootstrapping and AU and SH test results, which fails to support a paraphyletic Thismiaceae. Although bootstrap percentages and posterior probabilities are not directly comparable, several studies tried to assess their mutual relationship (Leaché and Reeder, 2002; Alfaro et al., 2003; Douady et al., 2003; Yang and Rannala, 2005). In general, ML bootstrap support values are considered as more conservative while BPP appear to have increased sensitivity to phylogenetic signal, which may result in high confidence in a correct result with fewer characters. However, BPP also appears to be more prone to assigning high confidence to incorrect short internodes (Alfaro et al., 2003; Yang and Rannala, 2005). Since we do not know the true tree, it is impossible to determine whether BI and BRC spuriously supports the paraphyly of Thismiaceae, or are just much more sensitive to the phylogenetic signal of the data sets. Inadequate modeling of data is an important factor that can contribute to misleading BPP values (Buckley, 2002; Alfaro et al., 2003; Douady et al., 2003; Lemmon and Moriarty, 2004). The most complex model available (GTR + I + G) was selected upon our data but this model may not be flexible enough to account for the observed complex rate heterogeneity. The application of a relaxed-clock offers a biologically more plausible approach to infer rate variation among organisms than a no-clock assumption (Pybus, 2006). In our analysis only slight differences were observed between a relaxed and a no-clock Bayesian approach, and none of these differences are significantly supported. For the 18S + *atpA* data BRC retained the same phylogenetic relationships between the Dioscoreales clades as BI and both methods obtained significant clade support (BPP ≥ 95%) for the same clades (Fig. 3b,c). Thus, the application of a more-realistic substitution model does not lead to different results, suggesting that the impact of model misspecification seems minimal for this data set. On the other hand, it is possible that a GTR + I + G model of substitution with a relaxed clock is also unable to account for the observed rate heterogeneity.

In our opinion the difference in support values between ML and Bayesian approaches do not represent conflict but rather reflect their different nature. Bayesian posterior probabilities are straightforward to interpret:
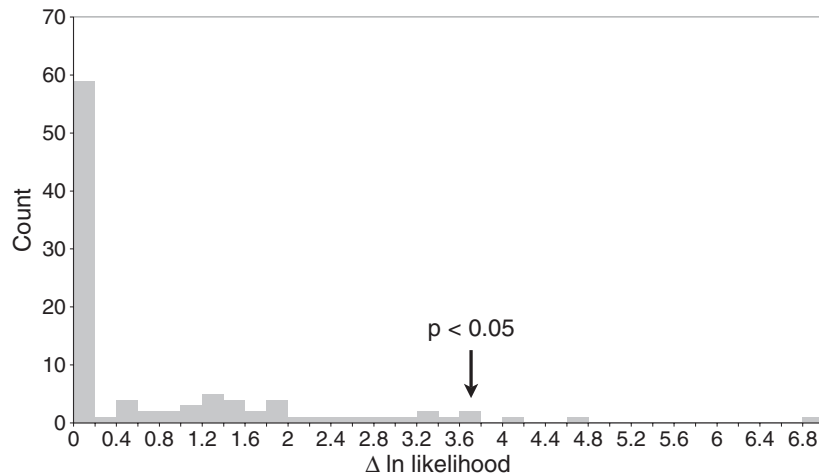
Fig. 4.   The distribution of differences in likelihood score between the "optimal" and "null" (i.e. Thismiaceae monophyly) tree scores for *atpA* data based on simulated sequences (see text). Arrow indicate the observed value of the test statistic and the probability of observing a value at least that large.

the posterior probability for a clade is the probability that the clade is true, given the data and the model. In contrast, nonparametric bootstrap values are much harder to interpret (Yang and Rannala, 2005). Using all our data both maximum likelihood and Bayesian approaches infer Thismiaceae as a paraphyletic group (Fig. 3). However, when using replacement datasets based on the original data (nonparametric bootstrap replicates), confidence in a non-monophyletic Thismiaceae fades. This suggests that the phylogenetic signal is vulnerable to resampling and that additional data is certainly needed to test the current hypothesis. Thus, our results indicate that based on our data (and substitution model) there is a high probability that Thismiaceae are paraphyletic. However, the short internal branch leading to the *Tacca-Thismia–Haplothismia–Tiputinia* clade and the lack of bootstrap support suggest that there are relatively few characters that support this group and additional data are clearly needed to test this hypothesis.

Our results are also consistent with previous studies that have indicated a tendency for the SOHW test to be prone to generating Type 1 errors because of model misspecification coupled with branch-length heterogeneity (Goldman et al., 2000; Strimmer and Rambaut, 2001; Buckley, 2002). High substitution rates and branch length heterogeneity are present in both nuclear and mitochondrial data set as shown by the Bayesian relative rates tests and may thus cause conflict between the SOWH test and the much more conservative AU and SH tests. Particularly our observations that the distribution of the null hypothesis (Thismiaceae monophyletic) for 18S rDNA and 18S + *atpA* data sets is equal to zero are disturbing. According to these results every single topology with a Thismiaceae clade can be

confidently rejected as an equally likely hypothesis no matter how small the difference in likelihood might be. There are two possible explanations for this apparent erroneous behavior of the SOWH test. First there is no guarantee that the most optimal topology was located for each replicate, although we used thorough heuristic searches (but fixed parameter values). When a search, both constrained and unconstrained, on a particular replicate converges on a suboptimal tree the general null distribution will tighten. However, it seems unlikely that this explanation is solely responsible for the observed null distribution (Goldman et al., 2000). Secondly, parametric tests, such as SOHW, may suffer from a dependence on a substitution model derived from the data to construct a null distribution of the test statistic. If this model is inadequate the test may become biased (Strimmer and Rambaut, 2001; Buckley, 2002). Particularly misspecified substitution models coupled with rate heterogeneity may cause the SOHW test to become too liberal (Huelsenbeck et al., 1996). Furthermore, Buckley (2002) also observed that in data sets where SOWH proved misleading, BPP were also misleading, which may indicate that model misspecification has caused spurious clade support in our Bayesian analyses (but see results obtained with Bayesian relaxed clock interference).

The 18S rDNA trees obtained with ML and BI analyses here are conflicting with the same analyses by Merckx et al. (2006) and Yokoyama et al. (2008) in which Thismiaceae, represented by a single *Afrothismia* 18S rDNA sequence, was monophyletic. As pruning *Afrothismia* to a single sequence in the present study leads to reproduce their results, this stresses once again the importance of a complete sampling to improve phylogenetic accuracy and avoid LBA artifacts (Graybeal,
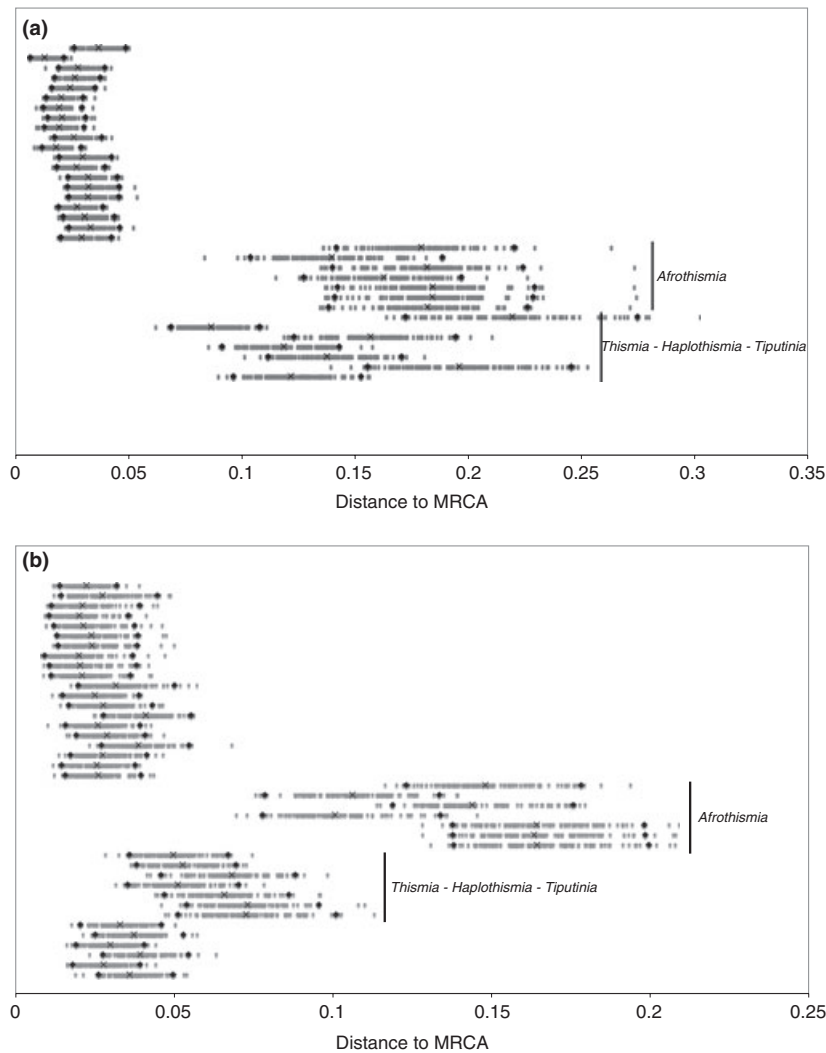
Fig. 5. Distribution of branch lengths from the most recent common ancestor (MRCA) of the included Thismiaceae and related Dioscoreales taxa for 18S rDNA (a) and *atpA* (b) data sets.

1998; Pollock et al., 2002; Zwickl and Hillis, 2002; Bergsten, 2005).

Conflicting data sets

While the AU and SH tests show that there is no significant difference between the results of the parsimony and model-based analyses, they detect strong conflict between the nuclear and mitochondrial data. This is reflected in the alternative positions of *Trichopus*, Burmanniaceae, and Nartheciaceae in the 18S rDNA and *atpA* topologies (Figs 1 and 2). Incongruence between *atpA* and non-mitochondrial genes has been observed frequently (e.g. Davis et al., 1998) but in nearly all studies a combined approach with nuclear and/or plastid genes was used mainly driven by the assumption that a larger number of characters improves

phylogenetic accuracy and resolution (Hillis, 1996; Davis et al., 2004; Chase et al., 2006; Fay et al., 2006; Petersen et al., 2006; Pires et al., 2006). Similarly, for 18S rDNA it has been hypothesized that its phylogenetic signal may be affected by constraints imposed by the secondary structure and it should be combined with other data to accommodate for these differences (Soltis et al., 1999). However, it is also known that a strong bias in branch-length heterogeneity may persist or even increase when more characters are added (Sanderson and Shaffer, 2002). In our combined analyses the 18S rDNA data set with 313 parsimony informative characters prevails over *atpA* with only 169 parsimony informative characters. The LBA artifact leading to a strongly supported Thismiaceae clade in the 18S rDNA MP analysis is not diminished by the addition of presumably unbiased *atpA* data and a nearly identical

topology is recovered by parsimony analysis of the combined data set. Only the position of Nartheciaceae as the first diverging lineage in Dioscoreales is innovative after the addition of *atpA* data.

For model-based methods the overall support and resolution improves when combining 18S rDNA and *atpA* data. Moreover, the differences that are observed between BI and BRC 18S rDNA trees disappear when *atpA* data is added. This empirical case shows that under the model-based approach combining data resolves conflicts between single-gene analyses and enhances phylogenetic resolution.

Are Thismiaceae monophyletic?

Given the current data and taxon sampling, we suggest that the paraphyly of Thismiaceae, given current molecular matrices, is empirically adequate. However, our dataset is limited in size and may not represent the phylogenetic relationships between the taxa accurately. Thus future studies should focus on additional data—preferably not prone to rate heterogeneity—as well as additional sampling to test this hypothesis. However, our data suggests that *Tacca* and the *Thismia–Haplothismia–Tiputinia* clade are linked by a short branch. Short branches may be very difficult to confidentially resolve, even with large, multilocus data sets (Wiens et al., 2008). Based on morphology a non-monophyletic Thismiaceae would be highly surprising and a case of remarkable morphological convergence. The first specimen of *Afrothismia* was described as ''*Thismia winkleri*'' by Engler (1905) only to be placed in a new genus by Schlechter (1906) who argued that a variety of distinctive characters justified the description of a new genus. Although *Thismia* and *Afrothismia* have very similar habits, there are indeed several features that distinguish these genera from each other and *Afrothismia* from other Thismiaceae. *Afrothismia* flowers are easily identified by their two-chambered perianth tube, the upper and the lower compartments are separated by an annulus, and the anthers are adnate to the gynoecium. But their most distinctive feature is an underground rhizome, which consists of clusters of bulbils (Cheek, 2003). Other Thismiaceae have corraloid, vermiform or tuberous roots (Maas-van de Kamer, 1998; Woodward et al., 2007). From a biogeographical perspective, *Afrothismia* has an exclusively African distribution, while *Thismia* is absent from Africa. Only the cryptic Thismiaceae genus *Oxygyne*, also occurs in Africa (Cheek and Williams, 1999). Nevertheless, no author ever considered *Afrothismia* separate from Thismiaceae. An independent origin for the loss of chlorophyll in *Afrothismia*, however, would not be unexpected. The absence of chlorophyll has shown to be an invalid synapomorphy in groups were loss of chlorophyll events occurs (Merckx et al., 2006). Particularly in myco-heterotrophic groups a switch from autotrophy to heterotrophy has often occurred multiple times independently in related lineages (e.g. Burmanniaceae, Merckx et al., 2006, 2008a; Orchidaceae, Molvray et al., 2000) and subsequent morphological convergence, i.e. in characters such as leaves and root structures, may conceal the relationships of closely and even distantly related groups (Leake, 1994).

Conclusions

Our results show that phylogenetic inference of myco-heterotrophic plants can prove challenging for various reasons. Due to the general rarity of taxa, adequate taxon sampling is often difficult to achieve. In addition, degeneration of both morphology and plastid genes may leave few reliable phylogenetic characters with which to infer phylogenetic relationships. When plastid gene sequences are unavailable or highly divergent, nuclear and mitochondrial DNA regions will be the data of choice. However, the presence of elevated substitution rates, commonly observed in mitochondrial and nuclear genes of heterotrophic plants, may lead to bias in phylogenetic inference.

Acknowledgements

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.

Anderson, F.E., Swofford, D.L., 2004. Should we be worried about long-branch attraction in real data sets? Investigations using metazoan 18S rDNA. Mol. Phylogenet. Evol. 33, 440–451.

Baillon, H.E., 1894. Histoire des Plantes. L. Hachette, Paris.

Barkman, T.J., Lim, S.-H., Salleh, K.M., Nais, J., 2004. Mitochondrial DNA sequences reveal the photosynthetic relatives of *Rafflesia*, the world's largest flower. Proc. Natl Acad. Sci. USA 101, 787–792.

Barkman, T.J., McNeal, J.R., Lim, S.H., Coat, G., Croom, H.B., Young, N.D., dePamphilis, C.W., 2007. Mitochondrial DNA suggests at least 11 origins of parasitism in angiosperms and reveals genomic chimerism in parasitic plants. BMC Evol. Biol. 7, e248.

Bergsten, J., 2005. A review of long-branch attraction. Cladistics 21, 163–193.

Bidartondo, M.I., Redecker, D., Hijri, I., Wiemken, A., Bruns, T.D., Domìnguez, L., Sérsic, A., Leake, J.R., Read, D.J., 2002. Epiparasitic plants specialized on arbuscular mycorrhizal fungi. Nature 419, 389–392.

Bidartondo, M.I., Burghardt, B., Gebauer, G., Bruns, T.D., Read, D.J., 2004. Changing partners in the dark: isotopic and molecular evidence of ectomycorrhizal liaisons between forest orchids and trees. Proc. R. Soc. B 271, 1799–1806.

Buckley, T.R., 2002. Model misspecification and probabilistic test of topology: evidence from empirical data sets. Syst. Biol. 51, 509–523.

Bungard, R.A., 2004. Photosynthetic evolution in parasitic plants: insight from the chloroplast genome. Bioessays 26, 235–247.

Caddick, L.R., Rudall, P.J., Wilkin, P., Chase, M.W., 2000. Yams and their allies: systematics of Dioscoreales. In: Wilson, K.J., Morrison, D.A. (Eds.), Monocots: Systematics and Evolution. CSIRO, Melbourne, pp. 475–487.

Caddick, L.R., Rudall, P.J., Wilkin, P., Hedderson, T.A., Chase, M.W., 2002. Phylogenetics of Dioscoreales based on combined analyses of morphological and molecular data. Bot. J. Linn. Soc. 138, 123–144.

Chase, M.W., Fay, M.F., Devey, D.S., Maurin, O., Ronsted, N., Davis, T.J., Pillon, Y., Petersen, G., Seberg, O., Tamura, M.N., Asmussen, C.B., Hilu, K., Borsch, T., Davis, J.I., Stevenson, D.W., Pires, J.C., Givnish, T.J., Sytsma, K.J., McPherson, M.A., Graham, S.W., Rai, H.S., 2006. Multigene analyses of monocot relationships: a summary. In: Columbus, J.T., Friar, E.A., Porter, J.M., Prince, L.M., Simpson, M.G. (Eds.), Monocots: Comparative Biology and Evolution (Excluding Poales). Rancho Santa Ana Botanic Garden, Claremont, pp. 63–75.

Cheek, M., 2003. A new species of *Afrothismia* (Burmanniaceae) from Kenya. Kew Bull. 58, 951–955.

Cheek, M., 2006. *Afrothismia amietii* (Burmanniaceae), a new species from Cameroon. Kew Bull. 61, 605–607.

Cheek, M., Williams, S., 1999. A review of African saprophytic flowering plants. In: Timberlake, J., Kativu, S. (Eds.), African Plants: Biodiversity, Taxonomy and Uses. Royal Botanic Gardens, Kew, pp. 39–49.

Clements, K.D., Gray, R.D., Choat, J.H., 2003. Rapid evolutionary divergences in reef fishes of the family Acanthuridae (Perciformes: Teleostei). Mol. Phylogenet. Evol. 26, 190–201.

Cronquist, A., 1970. The Evolution and Classification of Flowering Plants. Nelson, London.

Dahlgren, R.M.T., Clifford, H.T., Yeo, P.F., 1985. The Families of the Monocotyledons. Springer-Verlag, Berlin.

Davis, C.C., Wurdack, K.J., 2004. Host-to-parasite gene transfer in flowering plants: phylogenetic evidence from Malpighiales. Science 305, 676–678.

Davis, J.I., Simmons, M.P., Stevenson, D.W., Wendel, J.F., 1998. Data decisiveness, data quality, and incongruence in phylogenetic analysis: an example from the monocotyledons using mitochondrial *atpA* sequences. Syst. Biol. 47, 282–310.

Davis, J.I., Stevenson, D.W., Petersen, G., Seberg, O., Campbell, L.M., Freudenstein, J.V., Goldman, D.H., Hardy, C.R., Michelangeli, F.A., Simmons, M.P., Specht, C.D., Vergara-Silva, F., Gandolfo, M.A., 2004. A phylogeny of the monocots, as inferred from *rbcL* and *atpA* sequence variation, and a comparison of methods for calculating jackknife and bootstrap values. Syst. Bot. 29, 467–510.

Douady, C.J., Delsuc, F., Boucher, Y., Doolittle, W.F., Douzery, E.J.P., 2003. Comparison of Bayesian and maximum likelihood bootstrap measures of phylogenetic reliability. Mol. Biol. Evol. 20, 248–254.

Drummond, A.J., Rambaut, A., 2007. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol. Biol. 7, e214.

Drummond, A.J., Ho, S.Y.W., Phillips, M.J., Rambaut, A., 2006. Relaxed phylogenetics and dating with confidence. PLoS Biol. 4, e88.

Eichler, A.W., 1875. Blüthendiagramme 1. W. Engelmann, Leipzig.

Engler, H.G.A., 1888. Burmanniaceae. In: Engler, H.G.A., Prantl, K. (Eds.), Die Natütlichen Pflanzenfamilien, 1st edn. WEngelmann, Leipzig, pp. 44–51.

Engler, A., 1905. *Thismia winkleri* Engl., eine neue afrikanische Burmanniacee. Bot. Jahrb. Syst. 38, 89–91.

Eyre-Walker, A., Gaut, B.S., 1997. Correlated rates of synonymous site evolution across plant genomes. Mol. Biol. Evol. 14, 455–460.

Fay, M.F., Chase, M.W., Ronsted, N., Devey, D.S., Pillon, Y., Pires, J.C., Petersen, G., Seberg, O., Davis, J.I., 2006. Phylogenetics of Liliales: summarized evidence from combined analyses of five plastid and one mitochondrial loci. In: Columbus, J.T., Friar, E.A., Porter, J.M., Prince, L.M., Simpson, M.G. (Eds.), Monocots: Comparative Biology and Evolution (Excluding Poales). Rancho Santa Ana Botanic Garden, Claremont, pp. 559–565.

Felsenstein, J., 1978. Cases in which parsimony and compatibility methods will be positively misleading. Syst. Zool. 27, 401–410.

Felsenstein, J., 1985. Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39, 783–791.

Franke, T., 2004. *Afrothismia saingei* (Burmanniaceae, Thismieae), a new myco-heterotrophic plant from Cameroon. Syst. Geogr. Pl. 74, 27–33.

Goldman, N., Anderson, J.P., Rodrigo, A.G., 2000. Likelihood-based test of topologies in phylogenetics. Syst. Biol. 49, 652–670.

Graybeal, A., 1998. Is it better to add taxa or characters to a difficult phylogenetic problem? Syst. Biol. 47, 9–17.

Hendy, M.D., Penny, D., 1989. A framework for the quantitative study of evolutionary trees. Syst. Zool. 38, 297–309.

Hillis, D.M., 1996. Inferring complex phylogenies. Nature 383, 130–131.

Hillis, D.M., Mable, B.K., Moritz, C., 1996. Application of molecular systematics and the future of the field. In: Hilllis, D.M., Mortiz, C., Mable, B.K. (Eds.), Molecular Systematics, 2nd edn, Sinauer Associates, Sunderland, pp. 515–543.

Huelsenbeck, J.P., 1997. Is the Felsenstein zone a fly trap? Syst. Biol. 46, 69–74.

Huelsenbeck, J.P., Ronquist, F., 2001. MrBayes: Bayesian inference of phylogenetic trees. Bioinformatics 17, 754–755.

Huelsenbeck, J.P., Hillis, M.D., Nielsen, R., 1996. A likelihood ratio test of monophyly. Syst. Biol. 45, 546–558.

Hutchinson, J., 1934. The Families of Flowering Plants. Monocotyledons. Clarendon Press, Oxford.

Hutchinson, J., 1959. The Families of Flowering Plants. Monocotyledons, 2nd edn. Clarendon Press, Oxford.

Jarvie, J.K., 1996. *Thismia lauriana* (Burmanniaceae), a new species from central Kalimantan. Blumea 41, 257–259.

Jonker, F.P., 1938. A monograph of the Burmanniaceae. Meded. Bot. Mus. Herb. Rijksuniv. Utrecht 51, 1–279.

Karsten, H., 1858. Über die stellung einiger familien parasitischer pflanzen im natürlichen system. Nov. Actorum Acad. Caes. Leop.-Carol. Nat. Cur. 26, 885–927.

Leaché, A.D., Reeder, T.W., 2002. Molecular systematics of the Eastern Fence Lizard (*Sceloporus undulatus*): a comparison of parsimony, likelihood, and Bayesian approaches. Syst. Biol. 51, 44–68.

Leake, J.R., 1994. The biology of myco-heterotrophic ("saprophytic") plants. New Phytol. 127, 171–216.

Leake, J.R., 2005. Plants parasitic on fungi: unearthing the fungi in myco-heterotrophs and debunking the "saprophytic" plant myth. Mycologist 19, 113–122.

Lemmon, A.R., Moriarty, E.C., 2004. The importante of proper model assumption in Bayesian phylogenetics. Syst. Biol. 53, 265–277.

Lindley, J., 1846. The Vegetable Kingdom. Bradbury & Evans, London.

Maas, P.J.M., Maas-van de Kamer, H., van Benthem, J., Snelders, H.C.M., Rübsamen, T., 1986. Burmanniaceae. Flora Neotropica 42, 1–189.

Maas-van de Kamer, H., 1998. Burmanniaceae. In: Kubitzki, K., (Ed.), Families and Genera of Vascular Plants, Monocotyledons, Lilianae (Except Orchidaceae). Springer, Berlin, pp. 154–164.

Maddison, D.R., Maddison, W.P., 2001. MacClade 4: Analysis of Phylogeny and Character Evolution, ver. 4.01. Sinauer Associates, Sunderland, MA.

Merckx, V., Bidartondo, M.I., 2008. Breakdown and cospeciation in the arbuscular mycorrhizal mutualism. Proc. R. Soc. B. 275, 1029–1035.

Merckx, V., Schols, P., Maas-van de Kamer, H., Maas, P., Huysmans, S., Smets, E., 2006. Phylogeny and evolution of Burmanniaceae (Dioscoreales) based on nuclear and mitochondrial data. Am. J. Bot. 93, 1684–1998.

Merckx, V., Chatrou, L.W., Lemaire, B., Sainge, M., Huysmans, S., Smets, E., 2008a. Influence of tectonic and climatic events on diversification and biogeography of myco-heterotrophic angiosperms: evidence from Burmanniaceae. BMC Evol. Biol. 8, 178.

Merckx, V., Schols, P., Geuten, K., Huysmans, S., Smets, E., 2008b. Phylogenetic relationships in Nartheciaceae (Dioscoreales), with focus on pollen and orbicule morphology. Belg. J. Bot. 141, 64–77.

Miers, J., 1847. On a new genus of plants of the family Burmanniaceae. Proc. Linn. Soc. Lond. 1, 328–329.

Molvray, M., Kores, P.J., Chase, M.W., 2000. Polyphyly of mycoheterotrophic orchids and functional influences on floral and molecular characters. In: Wilson, K.L., Morrison, D.A. (Eds.), Monocots: Systematics and Evolution, CSIRO, Melbourne, pp. 441–448.

Mower, J.P., Stefanovic, S., Young, G.J., Palmer, J.D., 2004. Gene transfer from parasitic to host plants. Nature 432, 165–166.

Nickrent, D.L., Starr, E.M., 1994. High rates of nucleotide substitution in nuclear small-subunit (18S) rDNA from holoparasitic flowering plants. J. Mol. Evol. 39, 62–70.

Nickrent, D.L., Duff, R.J., Colwell, A.E., Wolfe, A.D., Young, N.D., Steiner, K.E., dePamphilis, C.W., 1998. Molecular phylogenetic and evolutionary studies of parasitic plants. In: Soltis, D., Soltis, P., Doyle, J. (Eds.), Molecular Systematics of Plants II. DNA Sequencing. Kluwer Academic Publishers, Boston, MA, pp. 211–241.

Nickrent, D.L., Blarer, A., Qiu, Y.-L., Soltis, D.E., Soltis, P.S., Zanis, M., 2002. Molecular data place Hydnoraceae with Aristolochiaceae. Am. J. Bot. 89, 1809–1817.

Nickrent, D.L., Blarer, A., Qiu, Y.-L., Vidal-Russell, R., Anderson, F.E., 2004. Phylogenetic inference in Rafflesiales: the influence of rate heterogeneity and horizontal gene transfer. BMC Evol. Biol. 4, e40.

Nickrent, D.L., Der, J.P., Anderson, F.E., 2005. Discovery of the photosynthetic relatives of the "Maltese mushroom" Cynomorium. BMC Evol. Biol. 5, e38.

dePamphilis, C.W., Palmer, J.D., 1990. Loss of photosynthetic and chlororespiratory genes from the plastid genome of a parasitic flowering plant. Nature 348, 337–339.

dePamphilis, C.W., Young, N.D., Wolfe, A.D., 1997. Evolution of plastid gene rps2 in a lineage of hemiparasitic and holoparastitic plants: many losses of photosynthesis and complex patterns of rate variation. Proc. Natl Acad. Sci. USA 94, 7367–7372.

Petersen, G., Seberg, O., Davis, J.I., Goldman, D.H., Stevenson, D.W., Campbell, L.M., Michelangeli, F.A., Specht, C.D., Chase, C.W., Fay, M.F., Pires, J.C., Freudenstein, J.V., Hardy, C.R., Simmons, M.P., 2006. Mitochondrial data in monocot phylogenetics. In: Columbus, J.T., Friar, E.A., Porter, J.M., Prince, L.M., Simpson, M.G. (Eds.), Monocots: Comparative Biology and Evolution (Excluding Poales). Rancho Santa Ana Botanic Garden, Claremont, CA, USA, pp. 52–62.

Pires, J.C., Maureira, I.J., Givnish, T.J., Sytsma, K.J., Seberg, O., Petersen, G., Davis, J.I., Stevenson, D.W., Rudall, P.J., Fay, M.F., Chase, M.W., 2006. Phylogeny, genome size, and chromosome evolution of Asparagales. In: Columbus, J.T., Friar, E.A., Porter, J.M., Prince, L.M., Simpson, M.G., (Eds.), Monocots: Comparative Biology and Evolution (Excluding Poales). Rancho Santa Ana Botanic Garden, Claremont, CA, USA, pp. 287–304.

Pollock, D.D., Zwickl, D.J., McGuire, J.A., Hillis, D.M., 2002. Increased taxon sampling is advantageous for phylogenetic inference. Syst.Biol. 51, 664–671.

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

Pybus, O.G., 2006. Model selection and the molecular clock. PLoS Biol. 4, e151.

Rambaut, A., Drummond, A.J., 2007. Tracer ver. 1.4, Available at: http://beast.bio.ed.ac.uk/Tracer.

Rambaut, A., Grassly, N.C., 1997. Seq-Gen: an application for the Monte Carlo simulation of DNA sequence evolution along phylogenetic trees. Comput. Applic. Biosci. 13, 235–238.

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

Rübsamen, T., 1986. Morphologische, Embtyologische und Systematische Untersuchungen an Burmanniaceae und Corsiaceae mit Ausblick auf die Orchidaceae-Apostasioideae. Dissertationes Botanicae, Band 92. J. Cramer, Berlin.

Sainge, M.N., Franke, T., Agerer, R., 2005. A new species of Afrothismia (Burmanniaceae, tribe Thismieae) from Korup National Park, Cameroon. Willdenowia 35, 287–291.

Sanderson, M.J., Shaffer, H.B., 2002. Troubleshooting molecular phylogenetic analyses. Annu. Rev. Ecol. Syst. 33, 49–72.

Sasidharan, N., Sujanapal, P., 2000. Rediscovery of Haplothismia exannulata Airy Shaw (Burmanniaceae) from its type locality. Rheedea 10, 131–134.

Schlechter, R., 1906. Burmanniacae africanae. Bot. Jahrb. Syst. 38, 137–143.

Schlechter, R., 1921. Die Thismieae. Notizbl. Bot. Gart. Berlin-Dahlem 8, 31–45.

Selosse, M.-A., Richard, F., He, X., Simard, S.W., 2006. Mycorrhizal networks: des liaisons dangereuses? Trends Ecol. Evol. 21, 621–628.

Shimodaira, H., 2002. An approximately unbiased test of phylogenetic tree selection. Syst. Biol. 51, 492–508.

Shimodaira, H., Hasegawa, M., 1999. Multiple comparisons of log-likelihoods with applications to phylogenetic inference. Mol. Biol. Evol. 16, 114–116.

Shimodaira, H., Hasegawa, M., 2001. CONSEL: for assessing the confidence of phylogenetic tree selection. Bioinformatics 17, 1246–1247.

Soltis, P.S., Soltis, D.E., Wolf, P.G., Nickrent, D.L., Chaw, S.-M., Chapman, R.L., 1999. The phylogeny of land plants inferred from 18S rDNA sequences: pushing the limits of rDNA signal? Mol. Biol. Evol. 16, 1774–1784.

Soltis, D.E., Soltis, P.S., Endress, P.K., Chase, M.W., 2005. Phylogeny and Evolution of Angiosperms. Sinauer Associates, Sunderland, MA.

Stevenson, D.W., Laconte, H., 1995. Cladistic analysis of monocot families. In: Rudall, P.J., Cribb, P.J., Cutler, D.F., Humphries, C.J. (Eds.), Monocotyledons: Systematics and Evolution, Royal Botanic Gardens, Kew, pp. 685–735.

Strimmer, K., Rambaut, A., 2001. Inferring confidence sets of possible misspecified gene trees. Proc. R. Soc. B 269, 137–142.

Swofford, D.L., 2002. PAUP* Phylogenetic Analysis Using Parsimony (* and other methods), ver 4. Sinauer Associates, Sunderland, MA.

Swofford, D.L., Olsen, G.J., Waddel, P.J., Hillis, D.M., 1996. Phylogenetic inference. In: Hilllis, D.M., Mortiz, C., Mable, B.K., (Eds.), Molecular Systematics, 2nd edn, Sinauer Associates, Sunderland, Mass, pp. 407–514.

Swofford, D.L., Waddell, P.J., Huelsenbeck, J.P., Foster, P.G., Lewis, P.O., Rogers, J.S., 2001. Bias in phylogenetic estimation and its relevance to the choice between parsimony and likelihood methods. Syst. Biol. 50, 525–539.

Takhtajan, A., 1997. Diversity and Classification of Flowering Plants. Columbia University Press, New York.

Thorne, J.L., Kishino, H., 2002. Divergence time estimation and rate evolution with multi-locus data sets. Syst. Biol. 51, 689–702.

Wendel, J.F., Doyle, J.J., 1998. Phylogenetic incongruence: window into genome history and molecular evolution. In: Soltis, D., Soltis, P., Doyle, J., (Eds.), Molecular Systematics of Plants II. DNA Sequencing. Kluwer Academic Publishers, Boston, MA, pp. 265–296.

Wiens, J.J., Hollingsworth, B.D., 2000. War of the Iguanas: conflicting molecular and morphological phylogenies and long-branch attraction in Iguanid lizards. Syst. Biol. 49, 143–159.

Wiens, J.J., Kuczynski, C.A., Smith, S.A., Mulcahy, D.G., Sites, J.W., Townsend, T.M., Reeder, T.W., 2008. Branch lengths, support, and congruence: testing the phylogenomic approach with 20 nuclear loci in snakes. Syst. Biol. 57, 420–431.

Wilcox, T.P., Garcia de Leon, F.J., Hendrickson, D.A., Hillis, D.M., 2004. Convergence among cave catfishes: long-branch attraction and a Bayesian relative rate test. Mol. Phylogenet. Evol. 31, 1101–1113.

Wolfe, A.D., dePamphilis, C.W., 1998. The effect of relaxed functional constraints on the photosynthetic gene *rbcL* in photosynthetic and nonphotosynthetic parasitic plants. Mol. Biol. Evol. 15, 1243–1258.

Woodward, C.L., Berry, P.E., Maas-van de Kamer, H., Swing, K., 2007. *Tiputinia foetida*, a new mycoheterotrophic genus of Thismiaceae from Amazonian Ecuador, and a likely case of deceit pollination. Taxon 56, 157–162.

Wu, C.-I., Li, W.-H., 1985. Evidence for higher rates of nucleotide substitution in rodents than in man. Proc. Natl Acad. Sci. USA 82, 1741–1745.

Yang, Z., Rannala, B., 2005. Branch-length prior influences Bayesian posterior probability of phylogeny. Syst. Biol. 54, 455–470.

Yokoyama, J., Koizumi, Y., Yokota, M., Tsukaya, H., 2008. Phylogenetic relationships of *Oxygyne shinzatoi* (Burmanniaceae) inferred from 18S rDNA sequences. J. Plant. Res. 121, 27–32.

Young, N.D., dePamphilis, C.W., 2005. Rate variation in parasitic plants: correlated and uncorrelated patterns among plastid genes of different function. BMC Evol. Biol. 5, 16.

Zwickl, D.J., 2006. Genetic algorithm approaches for the phylogenetic analysis of large biological sequence data sets under the maximum likelihood criterion. PhD dissertation, The University of Texas at Austin, TX.

Zwickl, D.J., Hillis, D.M., 2002. Increased taxon sampling greatly reduces phylogenetic error. Syst. Biol. 51, 588–598.

Supporting information

Additional Supporting Information may be found in the online version of this article:

18S.xml
BEAST ver. 1.4.6 input file for analysis of 18S rDNA data

atpA.xml
BEAST ver. 1.4.6 input file for analysis of atpA data

18S + atpA.xml
BEAST ver. 1.4.6 input file for analysis of combined 18S rDNA and atpA data

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

Appendix

Taxon—Genbank accession: 18S rDNA, *atpA*; Voucher; Locality. Burmanniaceae: *Apteria aphylla* (Nutt.) Barnh. ex Small—DQ786035, EU421007; Chase 156, NCU; USA. *Burmannia capitata* (Walt. ex Gmel.) Mart.—EU816732, EU421009; Neyland 958, MCN; USA. *Campylosiphon purpurascens* Benth.—EU420996, EU421024; Banki 1257, U; Guyana. *Dictyostega orobanchoides* (Hook.) Miers—DQ786056, EU421026; Maas et al. 9620, U; French Guiana. *Gymnosiphon breviflorus* Gleason—DQ786041, EU421032; Ek 1577, U; French Guiana. Dioscoreaceae: *Dioscorea althaeoides* R. Knuth—EU420997, EU421027; RBGE 19940649; China. *Dioscorea bulbifera* L.—AF069203, FJ215775; RBGE 19821960; Cultivated. *Dioscorea caucasica* Lipsky—FJ215769, FJ215779; RBGE 19110024; Cultivated. *Dioscorea communis* (L.) Caddick & Wilkin—EU186223, AY277804; V. Merckx 2, LV; cultivated. *Dioscorea elephantipes* (L'Hér.) Engl.—FJ215767, FJ215777; RBGE 19280228; South Africa. *Dioscorea prazeri* Prain & Burkill—DQ786089, EU421028; Wilkin 878, K; Thailand. *Dioscorea rockii* Prain & Burkill—DQ786090, EU421029; Chase 21052, K; Sri Lanka. *Dioscorea sylvatica* (Knuth) Eckl.—FJ215768, FJ215778; RBGE 19803437; South Africa. *Dioscorea tokoro* Makino ex Myabe—DQ786088, FJ215776; Merckx 01, LV; Cultivated. *Stenomeris dioscoreifolia* Planch.—DQ786087, EU421042; Risdale 550, ISU; Philippines. Nartheciaceae: *Aletris lutea* Small—DQ786092, FJ215780; Anderson 36, LV; USA. *Lophiola aurea* Ker-Gawl.—DQ786091, EU421039; Newell 23/8, K; USA. *Metanarthecium luteo-viride* Maxim.—AF309410, EU421040; Inoue s.n., K; Japan. *Narthecium ossifragum* Huds.—AF309411, AY299809; Jaquemart 46-9, LV; Belgium. *Nietneria paniculata* Steyerm.—EU186219, EU421041; O. Hokche & P.J.M. Maas 849, U; Venezuela. Pandanaceae (outgroup): *Pandanus tectorius* Parkinson ex Du Roi (outgroup)—AY952391, EU421052; Merckx 201, LV; Cultivated. Taccaceae: *Tacca artocarpifolia* Seem.—AF309397, EU421043; Caddick 305, K; Madagascar. *Tacca chantrieri* André—DQ786086, EU421044; Chase 175, NCU; Cultivated. *Tacca integrifolia* Ker-Gawl.—DQ786085, EU421045; Boyce 1074, K; Malaysia. *Tacca leontopetaloides* (L.) Kuntze—EU420999, AF039252; Wilkin 817, K; Thailand. *Tacca palmata* Blume—EU421000, EU421046; Chase 6201, K; Cultivated. *Tacca palmatifida* Baker—DQ786084, FJ215774; Chase 1377, K; Indonesia. *Tacca parkeri* Seem.—EU421001, AY299849; Berry 5620, MO; Venezuela. *Tacca plantaginea* (Hance) Drenth—U42063, FJ215773; BG Leiden 520520, Cultivated. Thismiaceae: *Afrothismia foertheriana* T. Franke, Sainge & Agerer—EU420988, EU421002; Merckx et al. 126, LV; Cameroon. *Afrothismia gabonensis* Dauby & Stévart—FJ215766, FJ215772; Dauby 167, BRLU; Gabon. *Afrothismia gesnerioides* H. Maas—EU420989, EU421003; Merckx et al. 110, LV; Cameroon. *Afrothismia hydra* Sainge and Franke—EU420990, EU421004; Merckx et al. 115, LV; Cameroon. *Afrothismia hydra* Sainge and Franke—FJ215765, FJ215771; Merckx et al. 113, LV; Cameroon. *Afrothismia korupensis* Sainge & T. Franke—EU420991, EU421005; Merckx et al. 114, LV; Cameroon. *Afrothismia winkleri* Schltr.—EU420992, EU421006; Merckx et al. 106, LV; Cameroon. *Haplothismia exannulata* Airy Shaw—DQ786082, EU421037; Sasidharan and Sujanapal 30476, KFRI; India. *Thismia aseroe* Becc.—AF309404, EU421048; Caddick 349, K; Malaysia. *Thismia clavigera* (Becc.) F. Muell.—AF309405, EU421049; Caddick 354, K; Malaysia. *Thismia panamensis* (Standley) Jonk.—DQ786081, EU421050; Aizprua 2946, LV; Panama. *Thismia rodwayi* F. Muell.—AF309403, AY299849; Wapstra s.n., HO; Australia. *Thismia taiwanensis* Yang, Saunders and Hsu—DQ786080, EU421051; Yang et al. 28981, PPI ; Taiwan. *Tiputinia foetida* P.E. Berry & C.L. Woodward—FJ215764, FJ215770; Alvaro Javier Perez Castaneda s.n., LV; Ecuador. Trichopodaceae: *Trichopus sempervirens* (H. Perrier) Caddick & Wilkin—AF309395, AY299724; Wilkin et al. 948, K; Madagascar. *Trichopus zeylanicus* Gaertn.—AF309394, AY277805; Chase 16354, K; Sri Lanka.