Naturalis Repository

A few-gene plastid phylogenetic framework for mycoheterotrophic monocots

Vivienne K.Y. Lam, Vincent S.F.T. Merckx, Sean W. Graham

Downloaded from:
https://doi.org/10.3732/ajb.1500412

Article 25fa Dutch Copyright Act (DCA) - End User Rights

This publication is distributed under the terms of Article 25fa of the Dutch Copyright Act (Auteurswet) with consent from the author. Dutch law entitles the maker of a short scientific work funded either wholly or partially by Dutch public funds to make that work publicly available following a reasonable period after the work was first published, provided that reference is made to the source of the first publication of the work.

This publication is distributed under the Naturalis Biodiversity Center 'Taverne implementation' programme. In this programme, research output of Naturalis researchers and collection managers that complies with the legal requirements of Article 25fa of the Dutch Copyright Act is distributed online and free of barriers in the Naturalis institutional repository. Research output is distributed six months after its first online publication in the original published version and with proper attribution to the source of the original publication.

You are permitted to download and use the publication for personal purposes. All rights remain with the author(s) and copyrights owner(s) of this work. Any use of the publication other than authorized under this license or copyright law is prohibited.

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the department of Collection Information know, stating your reasons. In case of a legitimate complaint, Collection Information will make the material inaccessible. Please contact us through email: collectie.informatie@naturalis.nl. We will contact you as soon as possible.

RESEARCH ARTICLE

AMERICAN JOURNAL OF BOTANY

A few-gene plastid phylogenetic framework for mycoheterotrophic monocots[1]

Vivienne K.Y. Lam[2,3], Vincent S.F.T. Merckx[4], and Sean W. Graham[2,3,5]

PREMISE OF THE STUDY: Few-gene studies with broad taxon sampling have provided major insights into phylogeny and underpin plant classification. However, they have typically excluded heterotrophic plants because of loss, pseudogenization, or rapid evolution of plastid genes. Here we performed a phylogenetic survey of three commonly retained plastid genes to assess their utility in placing mycoheterotrophs.

METHODS: We surveyed *accD*, *clpP*, and *matK* for 34 taxa from seven monocot families that include full mycoheterotrophs and a broad sampling of photosynthetic taxa. After screening for weak contaminants, we conducted phylogenetic analyses and characterized among-lineage rate variation.

KEY RESULTS: Likelihood analyses strongly supported local placements of fully mycoheterotrophic taxa for Corsiaceae, Iridaceae, Orchidaceae, and Petrosaviaceae, in positions consistent with other studies. Depression of likelihood bootstrap support values near mycoheterotrophic clades was alleviated when each mycoheterotrophic family was considered separately. Triuridaceae (*Sciaphila*) monophyly was recovered in a partitioned likelihood analysis, and the family then placed as sister to Cyclanthaceae-Pandanaceae. Burmanniaceae placed in Dioscoreales with weak to strong support depending on analysis details, and we inferred a plastid-based phylogeny for the family. Thismiaceae species may retain a plastid genome, based on *accD* retention. The inferred position of Thismiaceae is unstable, but was close to Taccaceae (Dioscoreales) in some analyses.

CONCLUSIONS: Long branches/elevated substitution rates, missing genes, and occasional contaminants are challenges for plastid-based phylogenetic inference with full mycoheterotrophs. However, most mycoheterotrophs can be readily integrated into the broad picture of plant phylogeny using several plastid genes and broad taxonomic sampling.

KEY WORDS achlorophyllous plants; Asparagales; Dioscoreales; DNA barcoding; gene retention; higher-order relationships; Liliales; Pandanales; Petrosaviales; plastomes

Mycoheterotrophic plants acquire carbon, water, and essential nutrients from fungal partners, typically those involved in mycorrhizal symbioses with green plants (Leake, 2004, 2005; Merckx et al., 2009a). As in holoparasitic plants, some mycoheterotrophs have become dependent on the host organism for all their nutritional needs and have lost the ability to photosynthesize: full mycoheterotrophs rely on fungal carbon for their entire life cycle (see Merckx, 2013). This major nutritional transition has evolved at least 47 times in land-plant evolution (Merckx et al., 2013a). In angiosperms the loss of photosynthesis may be associated with substantial changes in plant morphology, including reduced stature and foliage, and novel floral forms (Leake, 1994). Fully mycoheterotrophic plants are generally not green (although limited chlorophyll production is sometimes retained; Cummings and Welschmeyer, 1998; Barrett et al., 2014). Heterotrophic plants also often have substantially modified plastid genomes due to elevated rates of substitution, gene losses, and rearrangements (e.g., Wickett et al., 2008; Delannoy et al., 2011; Logacheva et al., 2011, 2014; Barrett and Davis, 2012; Lam et al., 2015; Mennes et al., 2015; Schelkunov et al., 2015; Bellot and Renner, 2016). The ultimate fate of the plastid genome in heterotrophs may be loss, although this may have occurred in only one embryophyte group (putatively in the holoparasite *Rafflesia*; Molina et al., 2014). The usual persistence of plastid genomes in heterotrophs is thought to reflect nonphotosynthetic functions performed by some plastid-encoded gene products (e.g., Bungard, 2004; Barbrook et al., 2006; Wicke et al., 2011). However, retained

---

[1] Manuscript received 22 September 2015; revision accepted 8 February 2016.

[2] Department of Botany, 6270 University Boulevard, University of British Columbia, Vancouver, British Columbia, V6T 1Z4, Canada;

[3] UBC Botanical Garden & Centre for Plant Research, 6804 Marine Drive SW, University of British Columbia, Vancouver, British Columbia, V6T 1Z4, Canada; and

[4] Naturalis Biodiversity Center, Leiden University, Leiden, The Netherlands

[5] Author for correspondence (e-mail: swgraham@interchange.ubc.ca)

doi:10.3732/ajb.1500412

functional genes may also experience elevated rates of evolution (e.g., Logacheva et al., 2011; Barrett et al., 2014). Gene losses, pseudogenization, and rate elevation may all contribute to uncertainty in inferring higher-order placement of mycoheterotrophs in plant phylogeny (Merckx and Freudenstein, 2010).

There are ~514 extant species of fully mycoheterotrophic plants (Merckx et al., 2013a). The majority of species (91%) and origins (~83%) are found in monocots, which seem to be particularly prone to this evolutionary transition (Imhof, 2010; Merckx et al., 2013a). In total, seven monocot families include fully mycoheterotrophic taxa (i.e., Burmanniaceae, Corsiaceae, Iridaceae, Orchidaceae, Petrosaviaceae, Thismiaceae, and Triuridaceae; Leake, 1994). Orchidaceae have at least 235 fully mycoheterotrophic species, which derive from an estimated 25 losses of photosynthesis (Merckx et al., 2013a, b), and the entire family is mycoheterotrophic during seedling establishment (e.g., Bernard, 1909; Leake, 1994; Rasmussen, 1995; Merckx, 2013), referred to as initial mycoheterotrophy. Burmanniaceae are a small family in Dioscoreales with approximately 13 genera and 130 species that include both autotrophic and fully mycoheterotrophic members, with multiple independent origins of full mycoheterotrophy (see Merckx et al., 2006 for a summary of the family's systematic history). Corsiaceae are a family of full mycoheterotrophs, comprising three small genera (*Arachnitis*, *Corsia*, and *Corsiopsis*; the latter is likely extinct; Zhang et al., 1999; Mennes et al., 2015). Historically, the family also has had a highly uncertain placement (see Mennes et al., 2015), although it was recently resolved as the sister group of Campynemataceae (Liliales) based on nuclear, mitochondrial, and whole plastome data (Mennes et al., 2015; Bodin et al., 2016). *Geosiris* (comprising three species) is the only fully mycoheterotrophic member of Iridaceae (Fay et al., 2000; Reeves et al., 2001; Goldblatt et al., 2008; Merckx et al., 2013b). Petrosaviaceae are a small family comprising the autotrophic *Japonolirion osense* and three species of fully mycoheterotrophic *Petrosavia* (Cameron et al., 2003). The family is the sole component of Petrosaviales, the sister group of all monocots except Alismatales and Acorales in most phylogenetic studies (e.g., Fuse and Tamura, 2000; Davis et al., 2004; Chase et al., 2006; Graham et al., 2006). Thismiaceae are a fully mycoheterotrophic family with ~five genera and 50 species. The family may be closely related to Burmanniaceae (e.g., APG, 2009), although recent molecular evidence suggests that it is instead most closely related to photosynthetic *Tacca* in Taccaceae (Merckx et al., 2009b, 2010). Finally, Triuridaceae are a fully mycoheterotrophic family with ~9 genera and 50 species. The family experienced an ancient loss of photosynthesis based on its inferred crown age (Mennes et al., 2013); late Cretaceous fossils are also known that may represent stem-lineage Triuridaceae (Gandolfo et al., 1998, 2002; Iles et al., 2015), but their trophic status is not known. The family's position in plant phylogeny has been highly uncertain. Molecular evidence from mitochondrial and nuclear genes strongly supports its membership in Pandanales (Mennes et al., 2013), an arrangement first reported by Chase et al. (2000) based on nuclear 18S rDNA. A recent plastome-based study placed Triuridaceae as the sister group of a clade comprising Cyclanthaceae and Pandanaceae with strong support across a broad variety of likelihood analyses (Lam et al., 2015).

Until recently, most of what is known about the broad picture of land-plant relationships came from pioneering phylogenetic surveys based on one to a few genes surveyed across a broad range of taxa. For example, many of the phylogenetic relationships underpinning Angiosperm Phylogeny Group (APG) classification schemes (APG, 1998, 2003, 2009) were based on two plastid genes (*atpB*, *rbcL*) and a single nuclear gene, 18S rDNA (e.g., Soltis et al., 2000). Individual genes can also have a strong phylogenetic signal concerning higher-order relationships; indeed, the broad outline of plant phylogeny is often apparent in single-gene analyses (e.g., Chase et al., 1993 for angiosperms using only *rbcL*; Givnish et al., 2005 for monocots using only *ndhF*). Analyses that included a few plastid genes have been used to infer the phylogenetic placement of several mycoheterotrophic lineages, including *Geosiris* (Fay et al., 2000) and *Petrosavia* (Petrosaviales; Fuse and Tamura, 2000; Cameron et al., 2003). However, the use of plastid markers in phylogenetic studies of heterotrophs has generally been regarded as problematic, reflecting expectations of elevated rates of evolution in retained genes (potentially problematic in phylogenetic inference if it leads to long-branch attraction, e.g., Felsenstein, 1978; Hendy and Penny, 1989) and the hypothesized loss of multiple photosynthetic or other genes or even entire plastid genomes (e.g., Cronquist, 1988, p. 467; Merckx et al., 2009b). Phylogenetic studies of mycoheterotrophs have therefore tended to focus on analyses of data sets comprising several mitochondrial and nuclear genes (e.g., Neyland and Hennigan, 2003; Merckx et al., 2006; Mennes et al., 2013). Despite these concerns, recent studies have validated the use of whole-plastid genomes to infer the placement of mycoheterotrophs (e.g., Logacheva et al., 2014; Lam et al., 2015; Mennes et al., 2015) by providing results congruent with other studies, often with strong branch support.

These recent results suggest that it would be useful to revisit the utility of plastid genes in broadly sampled, few-gene studies. The two photosynthesis-related genes *atpB* and *rbcL* have been among the most widely used plastid genes in higher-order plant phylogenetic studies. Both genes may have additional nonphotosynthetic roles (Bungard, 2004; Wicke et al., 2011, 2013), but with few exceptions (e.g., *Petrosavia*, Cameron et al., 2003; see also Logacheva et al., 2014) they have been lost or pseudogenized in fully heterotrophic lineages. The widely employed plastid *trnL-trnF* spacer region is also often missing in heterotrophic plastomes, in contrast to *matK*, another widely used gene that is commonly retained in heterotrophic lineages (see summaries in Barrett et al., 2014; Lam et al., 2015) and is often used in phylogenetic studies (e.g., Hilu et al., 2003). A portion of *matK* is also widely used as one of the core plant DNA barcoding markers (e.g., Hollingsworth et al., 2009). It is thus one of the few widely used plastid genes that may be suitable for inference of mycoheterotroph phylogeny. Obtaining plastid genes from heterotrophic plants can be difficult, and at least one study has suffered from the inclusion of erroneous plastid sequences from mycoheterotrophs (e.g., Kim et al., 2013; Kim, personal communication in Mennes et al., 2015): we suspect that PCR-based amplifications of plastid genes in heterotrophs may be prone to the recovery of contaminant sequences (cross-contaminants with other plant taxa) when authentic genes are not recoverable, for example, due to gene loss (also discussed later).

Here we assessed three plastid genes, *accD*, *clpP*, and *matK*, for their potential as phylogenetic markers in large-scale analyses and, in particular, to address whether they can help place mycoheterotrophs in monocot phylogeny with moderate to strong bootstrap support. The gene *accD* codes for the β-carboxyltransferase subunit of the acetyl-CoA carboxylase (ACCase), which is involved in fatty acid metabolism, and also regulates ACCase activity (Bungard, 2004); *clpP* codes for the catalytic subunit of plastid Clp protease (Shanklin et al., 1995; Wicke et al., 2011); *matK* encodes a maturase involved in splicing plastid group IIA plastid introns (Ems et al., 1995; Vogel et al., 1997; McNeal et al., 2009), although some group-IIA splicing may occur despite its loss (Delannoy et al., 2011). Thus, all three genes have functional roles that are not directly related to photosynthesis, and they are all typically retained in the plastid genomes of heterotrophic plants (holoparasites and full mycoheterotrophs; see Wicke et al., 2011; Barrett et al., 2014; Lam et al., 2015). They are therefore good candidate genes for plastid-based phylogenetic inferences that include these plants.

To test the utility of these three genes in broad-scale phylogenetic inference, we analyzed a monocot-wide data set in a range of analyses that consider all families containing mycoheterotrophic taxa. Several factors contributed to the complexity of this study. First, we had to assemble a phylogenetic framework for these three genes that represented a broad array of photosynthetic taxa, as two of the three genes (*accD* and *clpP*) had not previously been widely sampled in phylogenetic studies. Thus, we had to amplify and sequence these genes for most green taxa, which we aligned with those available from previous studies and new sequences from mycoheterotrophs. We used this alignment for phylogenetic inferences that included or excluded various sets of mycoheterotrophs. Second, we had to screen mycoheterotroph sequences for possible cross-contaminant sequences, which we occasionally encountered and removed from analysis. Third, genes recovered from individual species sometimes exhibited highly elevated substitution rates. We therefore based our phylogenetic inferences primarily on maximum-likelihood (ML) analyses, as likelihood methods are understood to be less prone to long-branch attraction problems than parsimony (e.g., Felsenstein, 1988; Yang, 1996; Huelsenbeck, 1997, 1998; Swofford et al., 2001; Yang and Rannala, 2012), and Bayesian methods may lead to inflated confidence in inferred phylogenetic relationships (e.g., Simmons et al., 2004; Yang and Rannala, 2005; Kolaczkowski and Thornton, 2009). We considered both unpartitioned and partitioned DNA substitution models to assess the degree to which the use of simple vs. complex models affects phylogenetic inferences; the partitioning scheme used accommodates differences in DNA substitution models (including model parameters) among genes and codon positions. We also included a parsimony analysis for comparison. Fourth, not all of the genes could be obtained experimentally from all mycoheterotrophic taxa that we surveyed, which inevitably led to somewhat patchy concatenated alignments.

Our analyses focused on non-orchid mycoheterotrophic monocots, those that interact with arbuscular mycorrhizal (AM) fungi (Leake, 2004; Waterman et al., 2013), although we also included data from recently published plastomes from several fully mycoheterotrophic orchids that associate with ectomycorrhizal fungi (i.e., *Rhizanthella gardneri*, Delannoy et al., 2011; *Neottia nidus-avis*, Logacheva et al., 2011; *Corallorhiza striata*, Barrett and Davis, 2012). Our study addressed two major objectives: (1) Does this few-gene data set allow placement of individual mycoheterotrophic taxa (species or families) with moderate to strong bootstrap support? (2) Are these placements congruent with recent results using other sources of molecular data, including nuclear and mitochondrial genes and recent whole-plastid genome studies? We also included a representative sampling of Burmanniaceae to compare with published phylogenetic inferences for this family based on nuclear and mitochondrial data (e.g., Caddick et al., 2002; Neyland, 2002; Davis et al., 2004; Merckx et al., 2006, 2009b). Finally, as plastid genes have not been reliably recovered from Thismiaceae to date, we were particularly interested in assessing whether we could recover any of the three plastid genes from this family for use in phylogenetic inference.

MATERIALS AND METHODS

DNA extraction, primer design, amplification, and sequencing—We obtained new DNA sequence data for 34 photosynthetic or fully mycoheterotrophic taxa from mycoheterotrophic lineages, and 61 additional photosynthetic taxa (59 of which are monocots; sources noted in Appendix S1, see Supplemental Data with the online version of this article). Total genomic DNAs were extracted from silica-dried material using a modified CTAB protocol (Doyle and Doyle, 1987; Rai et al., 2003) or were obtained from colleagues or DNA banks (Royal Botanic Gardens, Kew; Missouri Botanical Garden; SANBI, Kirstenbosch). We amplified and sequenced *accD*, *clpP*, and *matK*. DNA amplification and Sanger sequencing used previously published and new primers (Appendix S2, see online Supplemental Data). We designed additional novel primers by considering monocot-wide sequence alignments generated from available GenBank sequences, using visual inspection to identify conserved regions suitable for primer placement, and Oligo 7 (Rychlik, 2007) and Amplify 3x (Engels, 1993) to screen candidate primers based on their predicted success in DNA amplification (for criteria, see Graham and Olmstead, 2000). Exon boundaries for *clpP* were based on the complete plastid genome of *Dioscorea elephantipes* (GenBank accession NC_009601.1). DNA amplification and sequencing protocols generally followed Graham and Olmstead (2000), although in some cases we replaced *Taq* polymerase with Paq5000 (Agilent Technologies, Santa Clara, California [CA], USA). We purified amplification products using QIAquick PCR purification columns (Qiagen, Valencia, CA, USA) or ExoSAP-IT reagent (USB Corp., Cleveland, Ohio, USA), following manufacturer instructions, and used 1/26 reactions of BigDye Terminator v. 3.1 (Applied Biosystems, Foster City, CA, USA) to perform dideoxy sequencing reactions. We sequenced all regions at least twice, generally twice each in both forward and reverse directions. For some taxa, we were not able to sequence through the entire second intron in *clpP*, and so we represented these sequences in the alignment as two fragments. We obtained one or more genes for several taxa using assemblies of circular plastid genomes retrieved from genome skims using Illumina data, following Mennes et al. (2015); these genomes will be presented elsewhere.

Alignment construction—We performed base-calling and contig assembly using the program Sequencher 4.2.2. (Gene Codes Corp., Ann Arbor, Michigan, USA), and aligned finalized contigs using Se-Al 2.0a11 (Rambaut, 2002) and the criteria outlined by Graham et al. (2000); Kelchner (2000); and Simmons and Ochoterena (2000). We added new sequences for the three genes to those retrieved from published plastid genomes of 53 additional taxa (see Appendix S1). We offset hard-to-align noncoding regions (frequently limited to individual taxa) in the alignment in a staggered manner (Graham et al., 2006; Saarela and Graham, 2010). Unamplifiable genes, which in some cases are known to be lost (based on plastid genome evidence; Appendix S1), were treated as missing data. The final matrix is a concatenation of all three genes: *accD* and *matK* are exclusively protein-coding, and *clpP* includes short exon sequences and two introns (see Appendix S2). The concatenated matrix is 9595 bp long and includes representatives from 139 monocot taxa, spanning 11 monocot orders, including representatives from all seven monocot mycoheterotrophic families (38 taxa), and 11 outgroup species from 10 angiosperm families outside monocots (Appendix S1).

We identified several cases of possible contaminant sequences among individual genes recovered from mycoheterotrophs (see below), initially by using BLAST (Altschul et al., 1990). We suspect that these derive from residual nontarget DNA contamination that becomes apparent when the main amplification fails due to gene loss or rapid gene evolution away from the targeted priming sequences. We identified probable contaminants by examining maximum-likelihood (ML) trees inferred for each locus individually, to check for terminal taxa that are distantly related to congeneric or confamilial taxa. Typically these had short connecting branches to the nearest green taxon. One example is a likely contaminant *matK* sequence obtained from a *Thismia aseroe* extract (Appendix S3a, see online Supplemental Data) that shows 97% similarity to *Moraea riparia* (Iridaceae; GenBank accession JX903631.1). The behavior of this sequence contrasts with what we believe to be genuine Thismiaceae *accD* sequences (259–314 bp sequenced portions, compared with 511-bp portions sequenced for other taxa), found on much longer branches (these also often placed in unusual positions, see below), that have BLAST scores of 87–88% to other monocots, eudicots, and other angiosperms. Other putative contaminants that placed outside monocots were not so straightforward to place phylogenetically because of limited taxon sampling in other angiosperms here. An example of this is a *matK* sequence retrieved from a *Geomitra clavigera* extract, which we identified as a probable contaminant based on close BLAST-based similarity to a eudicot (99% to *Capparis spinosa*, Capparaceae; GenBank accession AY491650.1; in a phylogenetic analysis, this sequence grouped with two eudicots included as outgroups, Appendix S3b). We excluded all candidate contaminant sequences identified in this way, but tentatively included other genes from the same sample that passed our phylogenetic screen (see Appendix S1; note that none of the other sequences included in final alignments from these samples showed obvious double peaks, so contaminants are likely present at low concentration in the original DNA extracts).

Phylogenetic analysis—All subsequent analyses considered the final concatenated three-gene DNA sequence alignment, which is publicly available at figshare.com (doi:10.6084/m9.figshare.2062158).

We conducted initial analyses using parsimony and likelihood with all green and mycoheterotrophic taxa included. For the parsimony search, we used the program PAUP* v. 4.0b10 (Swofford, 2003), with tree-bisection-reconnection branch swapping and 1000 random stepwise addition replicates, holding 100 trees at each step, and otherwise used default settings in the search for shortest trees. For maximum-likelihood (ML) analysis we used the program RAxML v. 7.4.2 (Stamatakis, 2006) with a graphical interface (Silvestro and Michalak, 2012), considering partitioned and unpartitioned versions of the data, see below. For these and all subsequent likelihood analyses, we ran 20 independent heuristic searches using different starting points.

We then performed individual likelihood analyses considering each mycoheterotrophic family (or taxon) separately to try to minimize the potential for attraction between distantly related taxa, using the search method outlined above. For Burmanniaceae, we ran an additional analysis that included only green species for the family (i.e., likely photosynthetic taxa; Merckx et al., 2006), and for Thismiaceae, we also analyzed each species (three in total) separately, as we only retrieved a portion of *accD* for these taxa. We also ran an analysis including only photosynthetic angiosperms (considering photosynthetic mycoheterotrophs, green orchids were included and green Burmanniaceae excluded) to investigate the general utility of the three genes in inference of monocot higher-order relationships.

We ran two variant ML analyses in all cases, one considering the data unpartitioned, and another with the data partitioned by gene and codon position. For the latter, we defined 14 initial data partitions based on the three codon positions for coding regions (considering the two exons of *clpP* separately) and two introns (in *clpP*). We then used the program PartitionFinder v. 1.1.1 (Lanfear et al., 2012) to assess which of the initial partitions had significantly different DNA substitution models or model parameters in each case, using the Bayesian information criterion (BIC; Schwarz, 1978; Sullivan and Joyce, 2005) and the strict hierarchical clustering algorithm. We repeated these tests for each distinct likelihood analysis and used the final partitioning schemes for each ML analysis (summarized in online Appendix S4). We applied the GTR+G model to all data partitions, as this model or close variants were found for all unpartitioned data sets, and for most individual data partitions in partitioned analyses (Appendix S4; the GTR+G+I model was found in a few cases, but the I parameter for invariant sites may be accommodated by the gamma parameter G; see Yang, 2006).

We assessed branch support using bootstrap analysis (Felsenstein, 1985). For the likelihood analyses, we ran 500 bootstrap replicates using the rapid bootstrap option in RAxML, but considering the same partitioning schemes and models of DNA substitution described above. For the parsimony bootstrap analysis, we used 1000 bootstrap replicates, with 100 random addition replicates per bootstrap replicate, and otherwise used default settings. We considered 90% and above to be "strongly supported" (or "well supported"), 70–89% to be "moderately supported" and <70% to be "weakly supported" (Zgurski et al., 2008).

Constraint tests of monophyly—We tested for the monophyly of two groups of interest that were each not recovered as monophyletic in any shortest trees (see below): (1) a putative clade comprising Burmanniaceae and Thismiaceae corresponding to the circumscription of Burmanniaceae s.l. in recent versions of the Angiosperm Phylogeny Group classification system (APG, 2003, 2009); (2) a clade comprising nongreen Burmanniaceae, as these species are not expected to be grouped together in a clade, given the multiple independent losses of photosynthesis predicted in the family (Merckx et al., 2006). In both cases, we were interested in determining whether we had sufficient power to reject a hypothesis that placed the constrained taxa together. Each constraint analysis excluded other mycoheterotrophs that were not relevant to the hypothesis. We found the shortest likelihood trees that satisfied monophyly constraints that we set up for RAxML (using unpartitioned ML analysis). We then compared the resulting tree sets using the approximately unbiased (AU) (Shimodaira, 2002) and Shimodaira-Hasegawa (SH) tests in the program CONSEL (Shimodaira and Hasegawa, 2001), using site-likelihoods from unpartitioned ML analysis, to ask whether the constrained trees were significantly worse explanations of the data than the best (unconstrained) tree.

Characterizing rate elevation in heterotrophic lineages—We characterized relative differences in overall substitution rates in mycoheterotrophic and green lineages in a Bayesian framework, using the program BEAST v. 1.8.2 (Drummond et al., 2012). Our focus was on characterizing relative rate variation across lineages, and so we fixed the input tree topology based on the best likelihood tree recovered from the partitioned analysis that included all taxa (see Fig. 2), with the exception that we constrained Triuridaceae, represented by *Sciaphila*, to be monophyletic (see below). We specified a single GTR+G nucleotide substitution model across sites, a random local-clock model with a fixed mean rate of 1.0 substitution per site, and a Yule speciation model, and otherwise used default settings. We ran BEAST for a combined 800 million generations across 20 separate analyses, sampling trees every 1000 generations, assessing convergence using the program Tracer v. 1.6 (Rambaut et al., 2014). All parameters had final effective sample size (ESS) values of at least 200. We used the programs LogCombiner v. 1.8.2. and TreeAnnotator v.1.6.2 (Drummond et al., 2012) to combine trees from individual analyses, discarding the first 25% of sampled trees as burn-in and resampling at a lower frequency to yield 10,000 final trees. The combined tree was visualized using the program FigTree v. 1.31 (Rambaut, 2006).

RESULTS

***Green monocot phylogeny inferred from* accD, clpP, *and* matK**—The placements of autotrophic monocot families, orders and higher-level taxa in the ML analysis of autotrophic taxa (Fig. 1) are congruent with previous monocot-wide studies using more genes

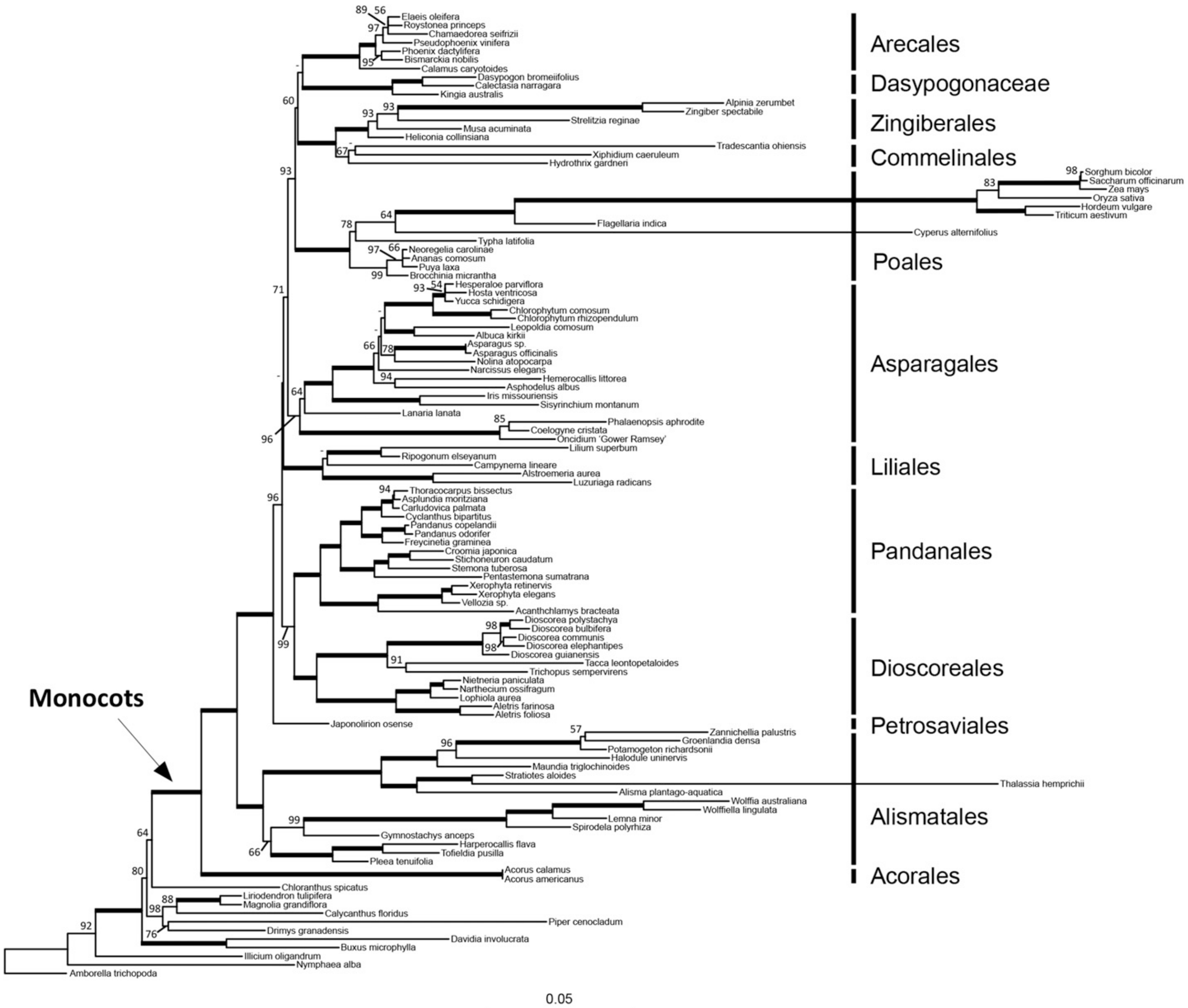


FIGURE 1 Three-gene phylogeny of photosynthetic monocots based on a partitioned three-gene maximum-likelihood analysis (for *accD*, *clpP*, and *matK*). The analysis includes green orchids but excludes green members of Burmanniaceae (see Materials and Methods section for details). Branches with 100% bootstrap support are shown as thick lines; other bootstrap values are indicated beside branches (<50% support indicated with a short dash). The scale bar indicates estimated number of substitutions per site.

(e.g., Graham et al., 2006; Hertweck et al., 2015). Most deep branches in the monocots were also strongly supported, with only minor exceptions (e.g., the placement of Dasypogonaceae, a sister-group relationship between Liliales and Asparagales-commelinids, both with <50% bootstrap support). There were no substantial differences in phylogenetic relationships of autotrophic monocots and their support values between the partitioned analysis and the unpartitioned analysis for this taxon set (data not shown).

Phylogenetic placement of mycoheterotrophs in monocot phylogeny—*(1) Placements with all mycoheterotrophs included*—Initial analyses (parsimony and likelihood) included all green and mycoheterotrophic lineages together. Parsimony trees (most parsimonious trees) grouped all or most of the fast-evolving lineages in a single large clade that may result from long-branch attraction, marked with an arrow in online Appendix S5 (all shortest trees either recovered this clade or had a slightly smaller version lacking Thismiaceae; see below on rates of evolution in mycoheterotrophs). This clade was poorly supported and had poorly supported and variable internal structure across shortest trees (note that black dots in this figure indicate branches that collapse in a strict consensus of the most-parsimonious trees). In contrast, full mycoheterotrophs that terminated shorter branches (i.e., *Geosiris*, orchids, *Petrosavia*, *Corsia*) placed with moderate to strong support in the parsimony analysis in Iridaceae, Orchidaceae, Petrosaviaceae, and Liliales, respectively. Corsiaceae were not recovered as monophyletic in this analysis, as *Arachnitis* grouped within the fast clade (Appendix S5).

In the likelihood analyses that included all mycoheterotrophs, fully mycoheterotrophic lineages on shorter branches again placed with moderate to strong likelihood bootstrap support in the positions observed in the parsimony analysis (see Fig. 2 and online Appendices S5, S6 for *Geosiris*, orchids, *Petrosavia*). However, multiple rapidly evolving lineages (see below) that grouped together in the parsimony analyses were instead inferred to be in dispersed positions

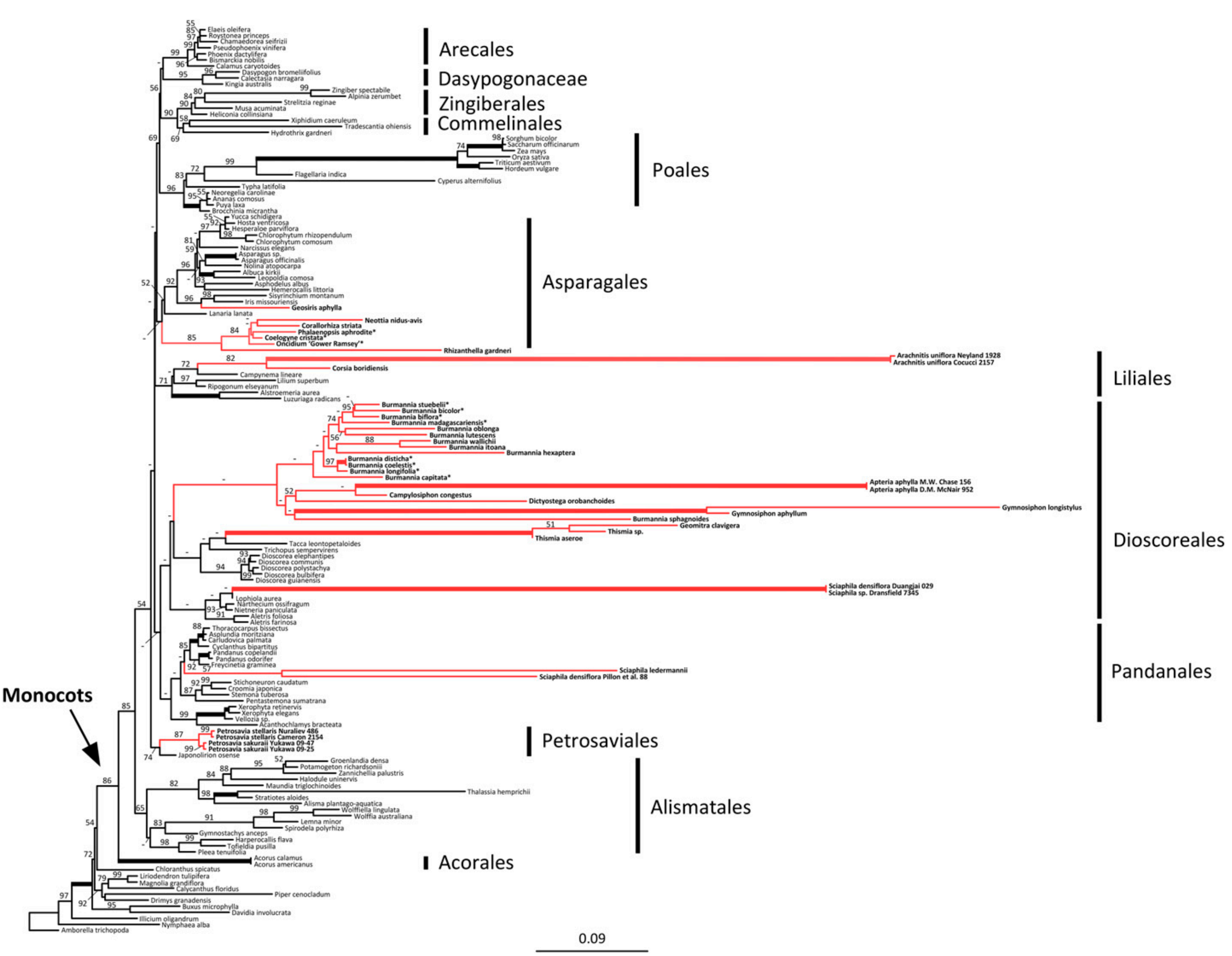


FIGURE 2 Three-gene phylogeny of photosynthetic and fully mycoheterotrophic monocots (all lineages considered simultaneously) based on a partitioned three-gene maximum-likelihood analysis (for *accD*, *clpP*, and *matK*). Lineages with mycoheterotrophs are indicated in red (asterisks indicate photosynthetic taxa in Burmanniaceae and Orchidaceae; the remainder are full mycoheterotrophs). Branches with 100% bootstrap support are shown as thick lines; other bootstrap values are indicated beside branches (<50% support indicated with a short dash). The scale bar indicates estimated number of substitutions per site.

in the likelihood analyses. Specifically, Burmanniaceae (the monophyly of which was poorly supported) were poorly supported as the sister group of a clade that included Taccaceae, Trichopodaceae, Thismiaceae, and Disocoreaceae, with Thismiaceae then the sister group of *Tacca* (Taccaceae). The latter arrangement was poorly supported, but the monophyly of Thismiaceae was well supported. *Sciaphila* (Triuridaceae) was divided between two locations, with one pair of taxa as the sister group of Cyclanthaceae and Pandanaceae in Pandanales, and the other nested in Nartheciaceae (Dioscoreales) as the sister group of *Lophiola*. This split placement was seen in both partitioned and unpartitioned ML analyses (Fig. 2; Appendix S6). The monophyly of Corsiaceae was well supported (i.e., *Arachnitis* no longer placed in the fast clade), and the entire family was supported as the sister group of Campynemataceae with weak to moderate support in both likelihood analyses (Fig. 2; Appendix S6). The inclusion of full mycoheterotrophs tended to depress likelihood bootstrap support for branches neighboring all of the mycoheterotrophic families, compared with analyses that included only green taxa. Neighboring branches often experienced at least 10% worse bootstrap support and often had substantially larger drops in support (e.g., for the monophyly of Asparagales and Dioscoreales, cf. Figs. 1 and 2).

(2) Placements with mycoheterotrophic taxa considered individually—In general, the likelihood analyses that examined each mycoheterotroph family separately recovered placements of mycoheterotrophic taxa consistent with the analyses that included them all simultaneously (cf. Figs. 2 and 3; online Appendix S7a–e), with exceptions outlined below for Thismiaceae and Triuridaceae. These local placements were also generally well supported (Fig. 3), a contrast with the likelihood analyses that included all mycoheterotrophic taxa simultaneously (Fig. 2, Appendix S6); there was less reduction in bootstrap support values for branches neighboring all of the mycoheterotrophic families when each of these was added individually. For example, the monophyly of both Asparagales and Dioscoreales was now well supported (cf. Figs. 1–3). There was now strong support for the monophyly of Corsiaceae, Iridaceae, Orchidaceae, and Petrosaviaceae (Fig. 3). Support for the monophyly of Burmanniaceae when other mycoheterotrophic families were excluded was also higher than when they were included, but was still weak (61–69% support, compared to <50% support; see Figs. 2, 3B; Appendix S6). However, when only photosynthetic Burmanniaceae were included, the bootstrap support for the family's monophyly improved substantially (to 100% in partitioned and unpartitioned ML analyses, see inset subtree in Fig. 4). The membership of Burmanniaceae in Dioscoreales was also well supported in this case, although its local position in the order still lacked strong support (Fig. 4). We could not reject the existence of a clade comprising Burmanniaceae and Thismiaceae (AU and SH tests comparing trees from best unconstrained vs. constrained ML analyses that included both families: $P = 0.104$ and 0.125, respectively; the unconstrained tree placed Burmanniaceae as the sister group of other Dioscoreales, and Thismiaceae as the sister group of *Tacca,* data not shown).

The behavior of Triuridaceae and Thismiaceae (in likelihood analyses, with each family included individually) was sensitive to the details of the analysis performed. In the partitioned likelihood analysis that included only *Sciaphila* (our representative of Triuridaceae) and that excluded other mycoheterotrophs, Triuridaceae were inferred to be monophyletic and they placed as the sister group of Cyclanthaceae-Pandanaceae (in Pandanales), all with weak support (Fig. 3E; Appendix S7f). The unpartitioned version of this analysis recovered two isolated lineages of *Sciaphila* in Pandanales and Dioscoreales (Appendix S7g) similar to the likelihood analyses that included all mycoheterotrophs (Fig. 2; Appendix S6), and again with weak support (Fig. 3E). Thismiaceae were unexpectedly inferred to be the sister group of *Japonolirion* in the partitioned analysis (Fig. 3C; Appendix S7h), but not in the unpartitioned analysis, where the family placed with Taccaceae and Trichopodaceae (Appendix S7i); note that Thismiaceae species are represented here only by *accD* (all analyses presented are concatenated three-gene analyses). Separate analyses that included individual species from the family placed the three surveyed species in divergent places (Fig. 5), but generally close to *Tacca* (Taccaceae) and/or *Trichopus* (Trichopodaceae) (i.e., *Geomitra clavigera* as the sister group of *Tacca* and *Trichopus* in both partitioned and unpartitioned likelihood analyses, *Thismia aseroe* as the sister group of *Tacca* for partitioned and unpartitioned analysis; *Thismia* sp. as the sister group of Maundiaceae in Alismatales in unpartitioned likelihood analysis, but as the sister group of *Tacca* in partitioned analysis; Fig. 5).

Relationships in Burmanniaceae—The relative positions of the photosynthetic members of Burmanniaceae were largely consistent among likelihood analyses. Various fully mycoheterotrophic lineages were recovered as nested among the photosynthetic taxa, but for the most part with poor support for the relative arrangements of green and nongreen taxa (Fig. 3B). Well-supported relationships in Burmanniaceae include a small clade comprising *B. bicolor*, *B. biflora*, and *B. stuebelii*, a sister-group relationship between *Burmannia itoana* and *B. wallichii*, between *B. coelestis* and *B. disticha*, between these two taxa and *B. longifolia*, and between *Gymnosiphon aphyllus* and *G. longistylus* (Fig. 3B). The intermingling of fully mycoheterotrophic members of Burmanniaceae with photosynthetic members of the family implies multiple evolutionary losses of photosynthesis (Fig. 3B; Appendix S7c). Consistent with a hypothesis of multiple losses of photosynthesis, the AU and SH tests rejected the existence of a constrained clade that comprises only nonphotosynthetic taxa of Burmanniaceae ($P < 0.02$ in both cases). When only green Burmanniaceae were included, the relative placement of *B. capitata* was different (but not well supported; cf. Figs. 3B, 4; Appendix S7c), and the sister-group relationship of *B. madagascariensis* and the clade comprising *B. bicolor*, *B. biflora*, and *B. stuebelii* was recovered with improved support (98–99%; Fig. 4).

Rate elevation in mycoheterotrophic monocots—Relative differences in the overall substitution rate are summarized in Fig. 6 (based on online Appendix S8). Three major rate bands are shown, with the highest rates (in red) at least twice as fast as the fastest slow ones (thin black lines); the thick black line is an intermediate rate. Many fully mycoheterotrophic lineages have intermediate to fast rates, although Petrosaviaceae, *Geosiris* (Iridaceae), *Corallorhiza striata* (Orchidaceae), and several fully mycoheterotrophic Burmanniaceae are exceptions (terminal photosynthetic lineages of Burmanniaceae are also in the lowest rate band; Appendix S8). Note that the rate band cut-offs used here (chosen to emphasize the most rapidly evolving lineages) include a broad range of rates (e.g., the lowest rate brand includes a nearly 9-fold range of rates, see Appendix S8 for more details). The most elevated rates (thick lines in Fig. 6) include several fully mycoheterotrophic lineages (*Arachnitis*

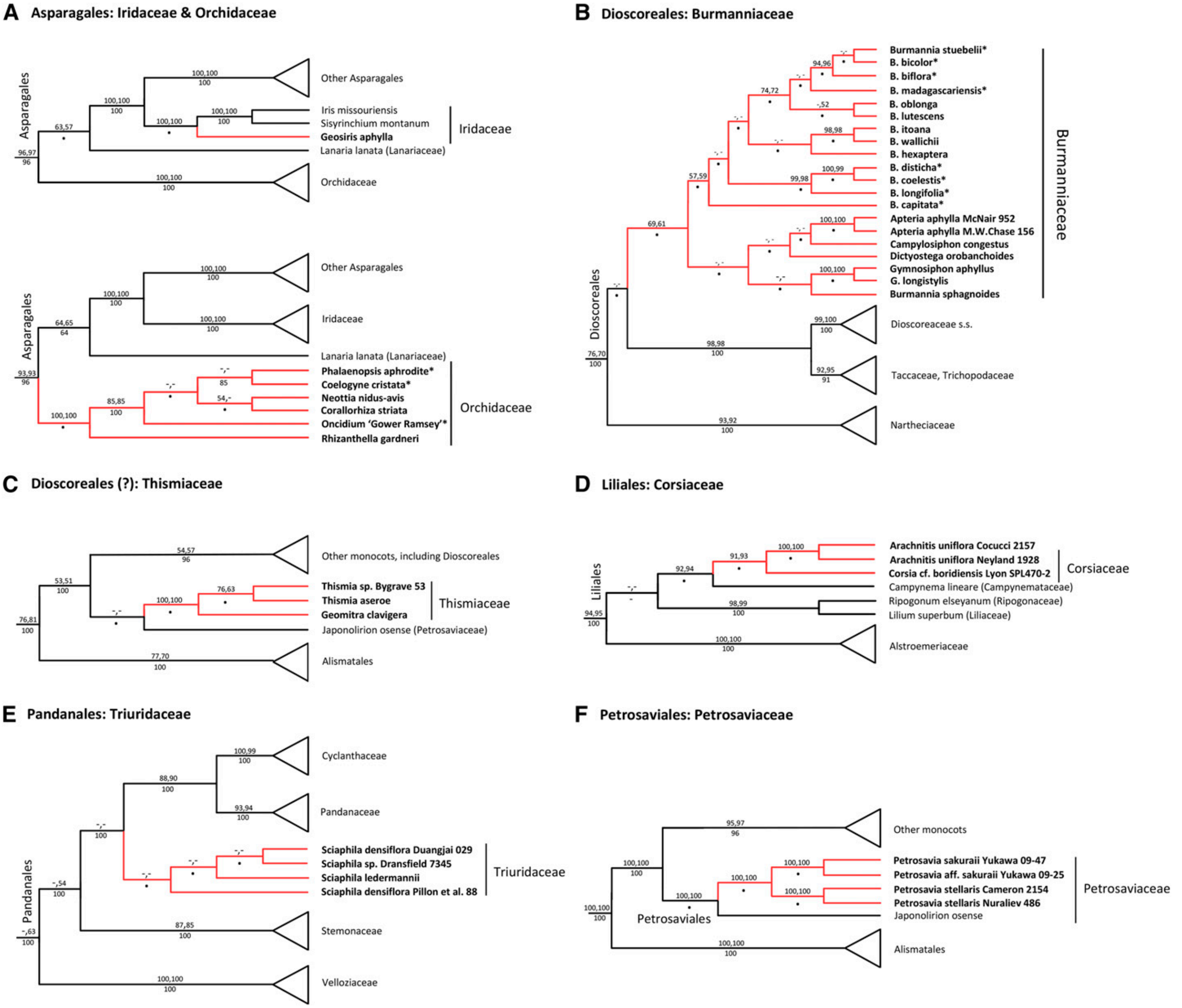


FIGURE 3 Summary of bootstrap support for local placements of monocot mycoheterotrophs in separate analysis of each family, based on partitioned maximum-likelihood (ML) analysis of *accD*, *clpP*, and *matK* (see Appendix S7a–f, h for branch lengths). (A) Iridaceae and Orchidaceae, (B) Burmanniaceae s.s., (C) Thismiaceae, (D) Corsiaceae, (E) Triuridaceae, (F) Petrosaviaceae. The ordinal placement of individual families according to APG (2009) is noted beside each figure (phylogenetic analyses for Iridaceae and Orchidaceae were done separately). Major clades collapsed for simplicity. Bootstrap support values noted beside branches: values above branches are with the mycoheterotrophic taxon included (left, partitioned ML analysis; right, unpartitioned ML analysis); those below branches are with the taxon excluded. Bootstrap support values <50% are indicated with a short dash; a dot indicates that the support value is not applicable (reflecting fewer taxa in analyses with photosynthetic taxa only). Lineages with mycoheterotrophs are indicated in red (asterisks indicate photosynthetic taxa in Burmanniaceae and Orchidaceae; the remainder are full mycoheterotrophs). Voucher names are included for species identified only to genus or when two samples were included for a species (see also Appendix S1).

in Corsiaceae, several lineages of Burmanniaceae, Orchidaceae, Thismiaceae, and Triuridaceae). Several photosynthetic lineages in Alismatales, Poales, and relatives are also rapidly evolving.

DISCUSSION

Using plastid genes to place mycoheterotrophic monocots—Compared with photosynthetic taxa, plastid genomes retrieved from heterotrophic lineages can be both rapidly evolving and reduced in terms of size and gene content, sometimes exceptionally so (e.g., Lam et al., 2015; Mennes et al., 2015; Bellot and Renner, 2016). Our few-gene approach thus provides general insights into the suitability of highly reduced, patchily sampled, rapidly evolving genomes for inferring the phylogenetic placement of heterotrophic plant lineages. Several key problems had to be addressed to do so. First, we lacked a phylogenetic framework (large-scale alignment of commonly retained genes) for photosynthetic taxa to help place the

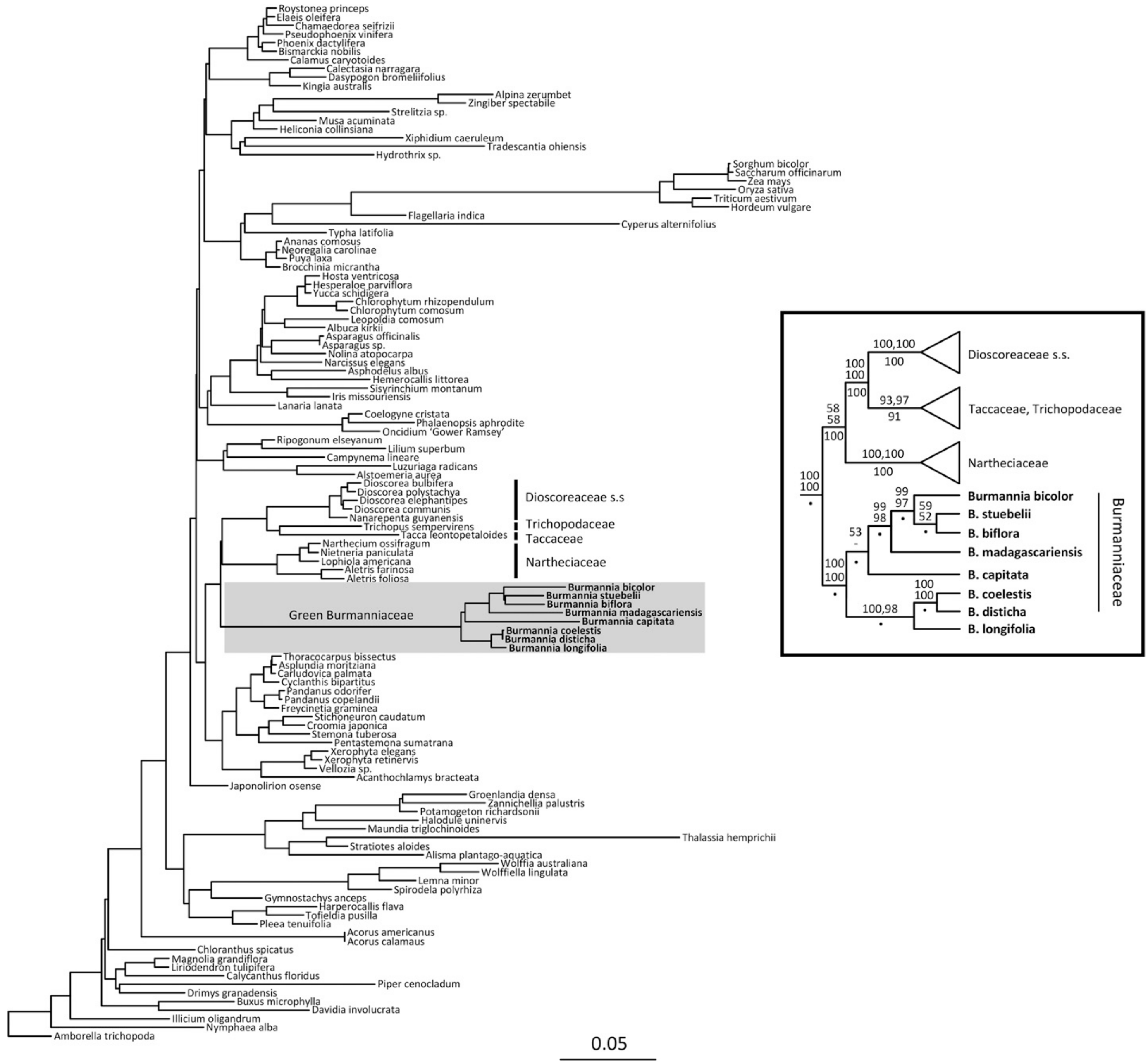


FIGURE 4 Phylogenetic placement of green Burmanniaceae in monocot-wide phylogeny, based on partitioned three-gene maximum-likelihood (ML) analysis. The phylogram indicates the placement of green Burmanniaceae (following Merckx et al., 2006; see Materials and Methods section for details). Families within Dioscoreales are indicated. The inset figure summarizes bootstrap support values within Dioscoreales (above branches: left, partitioned ML analysis; right, unpartitioned ML analysis. Below branches: Burmanniaceae excluded from analysis). Bootstrap support values <50% are indicated with a short dash; a dot indicates that the support value is not applicable (reflecting fewer taxa in analyses with Burmanniaceae excluded). The scale shows the estimated number of substitutions per site.

heterotrophic lineages. We addressed this by constructing a large-scale alignment of three commonly retained genes in heterotrophs (*accD*, *clpP*, *matK*) from photosynthetic monocots, which we sampled most heavily in the clades (orders) thought to have given rise to heterotrophic taxa. The higher-order relationships of photosynthetic monocots inferred using the three-gene data set are congruent with recent phylogenetic studies that employed a wide variety of taxonomic and gene samplings (e.g., Chase et al., 2006; Graham et al., 2006; Givnish et al., 2010; Soltis et al., 2011), and generally well-supported (Fig. 1). This congruence supports the general utility of using these genes for making inferences about high-order monocot relationships, at least for photosynthetic taxa.

The often highly elevated rates of evolution of retained plastid genes in heterotrophic lineages may be problematic for phylogenetic inference. Although our analytical understanding of long-branch effects is still quite limited (Parks and Goldman, 2014),

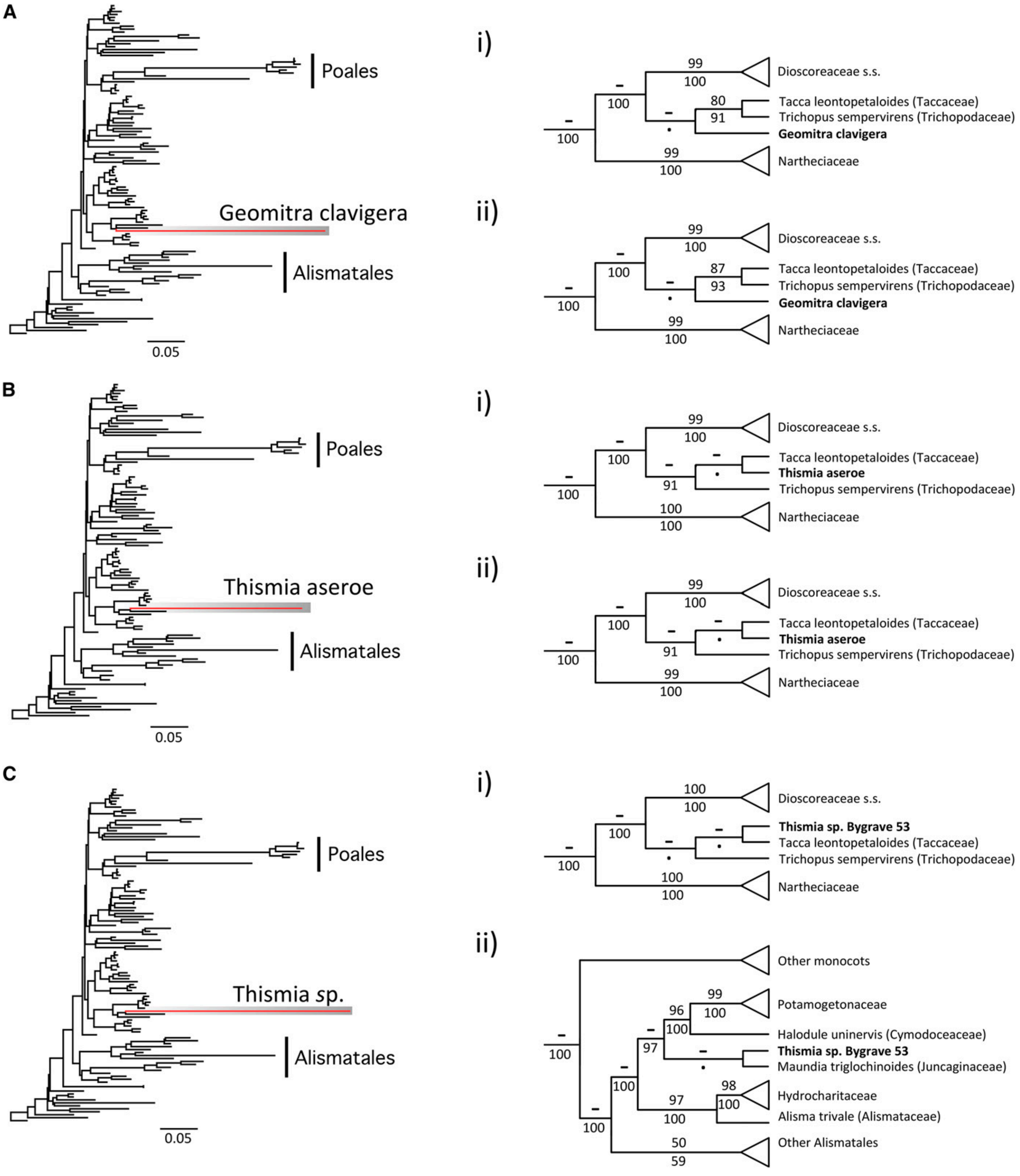


FIGURE 5 Phylogenetic placements inferred for individual species of Thismiaceae based on maximum-likelihood (ML) analysis (three-gene analyses that include only *accD* for this family, the only plastid gene recovered for it). Summary of results for (A) *Geomitra clavigera*, (B) *Thismia aseroe*, (C) *Thismia* sp. Left panel, phylograms of shortest trees for partitioned ML analysis (scale bars indicate estimated number of substitutions per site; mycoheterotrophic lineages indicated with blue branches). Right panel, summary of bootstrap support for placements of individual species based on: (i) partitioned ML analysis; (ii) unpartitioned ML analysis; values above branches are with Thismiaceae sequences included, those below are with them excluded. Bootstrap support values <50% are indicated with a short dash; a dot indicates that the support value is not applicable (reflecting fewer taxa in analyses with photosynthetic taxa only). The scale bar (A–C) indicates the estimated number of substitutions per site.

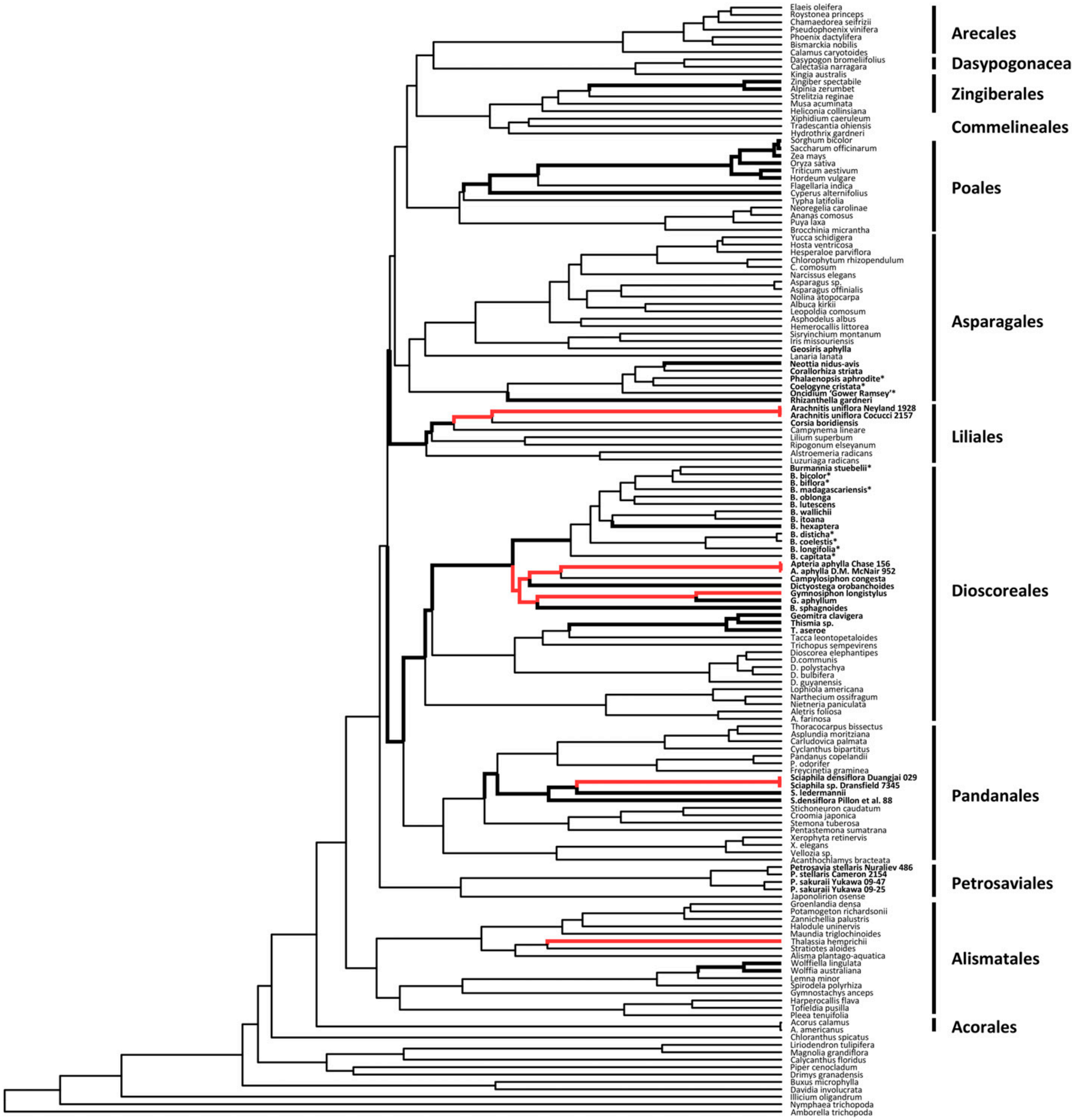


FIGURE 6 Relative substitution rates among green and fully mycoheterotrophic monocot lineages based on Bayesian analyses of a three-gene plastid (*accD*, *clpP*, and *matK*) data set, with a random-local-clock model and a constrained topology (see Materials and Methods for details). Thin branches indicate relative rates of <2.0 substitutions per site and below; thicker branches indicate intermediate relative rates of 2.0 to <4.0; thick red branches indicate those ≥4.0 (see online Appendix S8 for full details). Green species in Burmanniaceae and Orchidaceae are indicated with an asterisk (*); the others in these families are full mycoheterotrophs.

model-based methods like maximum likelihood are thought to be less sensitive than parsimony to problematic long branches (see also Felsenstein, 1978, 1988; Yang, 1996; Huelsenbeck, 1997, 1998; Swofford et al., 2001; Yang and Rannala, 2012). This lower sensitivity appears to be the case here, as our parsimony analysis depicted a collection of rapidly evolving mycoheterotrophic taxa as part of a poorly supported "fast" clade (Appendix S5) that is taxonomically highly heterogeneous (e.g., APG, 2009). This result is a strong contrast with the likelihood analyses (cf. Fig. 2 and Appendix S6). Despite substantial rate elevation, likelihood analyses placed mycoheterotrophic lineages in phylogenetic positions consistent with studies using nuclear and mitochondrial genes (e.g., Merckx et al., 2006, 2009b; Mennes et al., 2013, 2015). Thorough taxon sampling can help avoid long-branch artifacts (e.g., Pollock et al., 2002; Soltis and Soltis, 2004; Hedtke et al., 2006; Heath et al., 2008); however, adding taxa can introduce additional problems if the additional taxa are also on long branches (e.g., Kim, 1996; Hillis, 1998). The latter effect may also explain why including all mycoheterotrophic taxa simultaneously leads to substantially larger reductions in branch support for the branches neighboring heterotrophic lineages (those that were well supported in the analysis of green taxa only), than occurred when we included each mycoheterotrophic family individually (cf. Figs. 1–3).

Finally, missing data can present a challenge for phylogenetic inference and may be problematic in some cases here, as one or two of the three genes we considered were sometimes lost from the genome or were not readily recoverable for several taxa. Incompletely sampled taxa may still tend to improve the accuracy of phylogenetic inference when taxon sampling is otherwise limited (e.g., Wiens, 2006; Wiens and Tiu, 2012). However, the combination of missing data and long branches may nonetheless prove to be especially severe (e.g., Wiens, 2005, 2006), which is potentially problematic here for several patchily sampled taxa with elevated substitution rates (e.g., Figs. 2, 5; Appendix S1).

Utility and limits of the current approach—How well do our phylogenetic markers perform in likelihood analyses? In general, we did not see major differences in levels of bootstrap support values between unpartitioned and partitioned likelihood analyses for these taxa across the different data sets that we analyzed (Fig. 2 vs. Appendix S6; see also partitioned and unpartitioned comparisons in Fig. 3), supporting the idea that our results are not sensitive to whether the data are partitioned or not (partitioned likelihood analyses attempt to account for DNA substitution model or model parameter differences among different regions). This finding is also consistent with those of Lam et al. (2015) in their whole plastid-genome study of Triuridaceae. Because long branches in different mycoheterotroph families likely interfere with well-supported placement of individual taxa when considered simultaneously (Fig. 2), we focus the remaining discussion on the separate likelihood analyses of mycoheterotrophs (i.e., each family considered individually), summarized in Figs. 3–5.

We placed the majority of mycoheterotrophic taxa in monocot phylogeny with strong support, at least regarding their family-level placements (Fig. 3A, D, F). Our results are also consistent with current understanding of their phylogenetic placements based on other data sets (i.e., *Geosiris*, Fay et al., 2000; mycoheterotrophic Orchidaceae, Ruhfel et al., 2014; Corsiaceae, Mennes et al., 2015; Bodin et al., 2016; Petrosaviaceae, Cameron et al., 2003; Logacheva et al., 2014). The position of Corsiaceae has been unclear until very recently. The family was strongly supported here as the sister group of Campynemataceae in Liliales, which is consistent with recent whole plastome data and other data for the family (Mennes et al., 2015; Bodin et al., 2016). *Arachnitis* (Corsiaceae) has a particularly long branch due to highly elevated substitution rate (Figs. 2, 6; Appendix S7d). However, the other genus in the family, *Corsia*, has a moderately elevated rate of evolution (Fig. 6) and has retained all three surveyed genes, which may account for the family's clear and well-supported placement here in likelihood analyses (Fig. 3D). Our results do not support the findings of Neyland and Hennigan (2003) that Corsiaceae are polyphyletic, recovered in their analysis of a nuclear 26S rDNA data set (a probable artifact due to limited taxon sampling and their use of parsimony, see Mennes et al., 2015).

Three families were less confidently placed in monocot phylogeny by our three-gene data set: Burmanniaceae, Thismiaceae, and Triuridaceae (Figs. 3–5). With regards to the first, Burmanniaceae, our best likelihood trees are generally consistent with other recent studies based on nuclear and mitochondrial data (Merckx et al., 2006, 2008). We found strong support for Burmanniaceae being part of the order Dioscoreales when only green members of the family were included (Fig. 4) and moderate support for this arrangement when fully mycoheterotrophic taxa were included (Fig. 3B), although the family's position within the order was ambiguous in both cases. The green taxa of Burmanniaceae may be partial mycoheterotrophs based on vegetative reduction and reduction in chlorophyll (see Merckx et al., 2006), supported by isotopic evidence in *B. coelestis* (Bolin et al., in press). The green taxa often have lower substitution rates than full mycoheterotrophs in the family (Fig. 6; Appendix S8), but the family as a whole appears to have an elevated rate of evolution when only green taxa are considered (long branch subtending the green clade in Fig. 4, see also Fig. 6 and Appendix S8).

A diversity of phylogenetic studies based on molecular and morphological data have placed Triuridaceae in Pandanales (e.g., Chase et al., 2000; Rudall and Bateman, 2006; Mennes et al., 2013). The family was recently resolved as the sister group of Cyclanthaceae and Pandanaceae with strong support, based on a whole-plastome analysis (Lam et al., 2015). Here it was recovered as monophyletic only in the partitioned ML analysis (see shortest tree in Appendix S7f), but with poor support (Fig. 3E). The two samples of *S. densiflora* included here (vouchers *Pillon et al., 88* and *Duangjai 029* in Appendix S1) were distantly related to each other (Appendices S7f, g), perhaps pointing to within-genus misidentification of one of them, although we did not confirm this possibility here. The lack of monophyly of *Sciaphila* in some likelihood analyses is presumably a long-branch artifact. Triuridaceae were weakly supported as a member of Pandanales in the partitioned ML analysis (Fig. 3E). The family has some of the most elevated rates of evolution and some of the longest branches in our study (e.g., Figs. 2, 6; Appendix S8).

Particular caution seems warranted concerning our results for Thismiaceae. The data reported here for this family are especially intriguing because they may represent the first genuine plastid data to be recovered for it (i.e., for *accD*), from each of three species in the family (*Geomitra clavigera*, *Thismia aseroe*, and *Thismia* sp.; Fig. 5). Rates of molecular evolution are elevated for these taxa compared with photosynthetic taxa (Fig. 6; Appendix S8; note that these are three-gene analyses that include only *accD* from this family). The placement of Thismiaceae in analyses that include all three taxa was unstable (sister to Taccaceae or Petrosaviaceae; Appendices S7h, i) and very poorly supported (e.g., Fig. 3C). Analyses that included each species individually placed them in various positions with weak support that were also easily perturbable between analyses (e.g., by using partitioned vs. unpartitioned likelihood; Fig. 5A–C). The *Geomitra clavigera* and *Thismia* sp. samples placed close to or within Taccaceae and/or Trichopodaceae in some analyses (Fig. 5), consistent with previous studies by Merckx et al. (2006, 2009b), who presented mitochondrial and nuclear data that supported a placement of Thismiaceae in Dioscoreales, close to Taccaceae. The precise position of Thismiaceae in the order was strongly supported in their earlier study (Merckx et al., 2006), and moderately supported in a subsequent one (Merckx et al., 2009b), and they also provided evidence that *Afrothismia* (not included here) represents a lineage distinct from other Thismiaceae (Merckx et al., 2009b). The successful retrieval of *accD* sequences for *Geomitra*, *Thismia aseroe*, and *Thismia* sp. supports a retention of plastid genomes in at least some species in Thismiaceae. As such, we may expect to find small, cryptic plastomes in Thismiaceae that include, at the very least, the *accD* locus.

Are Burmanniaceae and Thismiaceae closely related?—The proposed taxonomic circumscription of Burmanniaceae to include Thismiaceae (APG, 2003, 2009) was based primarily on studies that surveyed several photosynthetic plastid genes (*atpB*, *rbcL*) from these families (Caddick et al., 2000, 2002) that may include contaminants (M. Chase, Royal Botanic Gardens, Kew, personal communication). Consistent with this, these genes have not been recovered in full circular genomes of any fully heterotrophic taxa in Burmanniaceae (V. Lam, unpublished data). Our inability here to reject a close relationship between these families in the AU and SH tests may simply reflect a lack of power to reject such hypotheses from the genes we surveyed (in particular, only *accD* was recoverable from Thismiaceae). We therefore propose that the broad treatment of Burmanniaceae to include members of Thismiaceae should be abandoned in future APG treatments until further evidence (ideally including plastid data and other sources of evidence) is obtained for where Thismiaceae fits in monocot phylogeny. Nonetheless, at least some of the plastid-based likelihood trees inferred here (Figs. 2, 5; Appendix S7i) are consistent with current evidence from mitochondrial and nuclear data, which support a position of Thismiaceae distinct from Burmanniaceae and closer to Taccaceae (Caddick et al., 2002; Merckx et al., 2006, 2009b).

Phylogenetic relationships within Burmanniaceae—Relationships within Burmanniaceae are generally poorly supported here. Not surprisingly, fully mycoheterotrophic taxa in this family often had highly elevated rates of evolution (Fig. 6; Appendix S8). Several taxa have one or two of the three genes that were not retrieved (Appendix S1), confirmed to be missing genes in several cases (V. Lam, unpublished data). Family relationships within Burmanniaceae were mostly strongly supported when only green Burmanniaceae are included in analysis (Fig. 4), as was family monophyly. Merckx et al. (2006) suggested that there were at least six independent losses of photosynthesis in Burmanniaceae, based on analyses of the mitochondrial *nad1 b-c* intron and nuclear 18S rDNA sequences. Although our sampling of Burmanniaceae has fewer taxa than theirs, the relationships inferred here also point to multiple independent losses of autotrophy among taxa, as there are three green lineages here (*B. capitata* and two small clades with three or four taxa) that are deeply nested among nongreen lineages (Fig. 3B; Appendix S7c). Although these "backbone" relationships were not well supported here (see Fig. 3B; note that our sampling of the family is more limited than that of Merckx et al., 2006 and Merckx et al., 2008, and we also used their scorings of photosynthetic vs. fully mycoheterotrophic taxa), the AU and SH tests indirectly support scenarios with more than one loss of photosynthesis, because a tree constraining all nongreen taxa as a clade is significantly longer than the best tree. Our analyses recovered several intrafamiliar relationships that were congruent with those from Merckx et al. (2006) including a clade comprising the achlorophyllous taxa *B. oblonga* and *B. lutescens*, another comprising achlorophyllous *B. itoana* and *B. wallichii*, and a clade comprising three autotrophic species (*Burmannia bicolor*, *B. stuebelii*, and *B. biflora*).

Retention of plastid genes and genomes in monocot mycoheterotrophs—Although some nonphotosynthetic plants are hypothesized to have lost their plastid genomes entirely (e.g., the eudicot holoparasite *Rafflesia*; Molina et al., 2014), which is demonstrated in multiple lineages of secondarily heterotrophic unicellular eukaryotes (Abrahamsen et al., 2004; Smith and Lee, 2014; Janouškovec et al., 2015), all other heterotrophic plant lineages surveyed to date have retained their plastid genomes (e.g., Wolfe et al., 1992; McNeal et al., 2007; Wickett et al., 2008; Delannoy et al., 2011; Logacheva et al., 2014; Lam et al., 2015; Mennes et al., 2015; Bellot and Renner, 2016). A small set of genes (including *accD*, *clpP*, *trnE*-UUC, I-CAU, and *f*M-CAU, the four rDNAs, some ribosomal proteins) are commonly retained across many heterotrophic lineages that have highly reduced plastomes (e.g., Wicke et al., 2013; Barrett et al., 2014; Lam et al., 2015; but see Bellot and Renner, 2016), suggesting that they are essential genes and may not be readily replaced by nuclear counterparts (either functionally transferred genes or replacement by analogous systems; Barbrook et al., 2006). We chose the three plastid genes (*accD*, *clpP*, and *matK*) for our phylogenetic survey because they have nonphotosynthetic roles and are frequently retained when photosynthetic genes are lost. Both *accD* and *clpP* have been lost occasionally in flowering plants, but this appears to be unrelated to the loss of photosynthesis (e.g., Jansen et al., 2007; Straub et al., 2011); *matK* loss may eventually occur when a sufficient number of plastid genes with group IIA introns have been lost, which may only occur in the later stages of plastid genome degradation in heterotrophic lineages (McNeal et al., 2009).

Some of the genes recovered here by amplification could conceivably represent nuclear or mitochondrial inserts of plastid genes. However, successful functional gene transfer from the plastid to nuclear genome is generally rare in land plants (e.g., Martin et al., 1998) and has not been confirmed in any plants for these three genes (although a potentially functional copy of *accD* may have been transferred to the nucleus in *Trachelium*; Rousseau-Gueutin et al., 2013). Nonfunctional copies would also be expected to degrade rapidly in the nuclear genome (e.g., Matsuo et al., 2005; but see Cusimano and Wicke, 2015). Mitochondrial inserts should also degrade (functional transfer of plastid protein-coding genes to mitochondria is unknown, e.g., Hao and Palmer, 2009), and the genes would generally be expected to evolve slowly after insertion, consistent with the lower mutation rate of this genome (e.g., Wolfe et al., 1987; Palmer and Herbon, 1988). The phylogenetically widespread retention here in mycoheterotrophs of rapidly evolving plastid genes that retain open reading frames is likely a hallmark of their retention in the plastid genome. In a few cases, we have also obtained plastid genomes from the same species (*Apteria aphylla*, *Arachnitis uniflora*, *Burmannia bicolor*, *Burmannia capitata*, *Geosiris aphylla*, *Petrosavia sakuraii*, *Petrosavia* aff. *sakuraii*, and *Sciaphila densiflora*; Appendix S1), confirming the retention of genes in plastid genomes that we also retrieved using PCR amplification.

Dealing with contamination in heterotrophic samples—Recovering contaminant sequences may be a frequent problem with samples from heterotrophic plants. We detected several possible instances here (seven of 34 taxa). We suspect these represent very weak contaminants that would not normally be evident, but which are apparent here when the target gene is no longer present or has evolved rapidly away from amplification or sequencing primer sites. It is possible that these represent instances of horizontal gene transfer, as has been demonstrated for parasitic plants (e.g., host to parasite, Davis and Wurdack, 2004; Yoshida et al., 2010; and parasite to host, Mower et al., 2004). However, this possibility may be unlikely, because mycoheterotrophs do not have plant-to-plant connections and any such transfers would have to involve a fungal intermediate. If not detected, the inclusion of contaminant genes would be highly problematic for phylogenetic inference (or in DNA barcoding studies, see below). While we propose that our approach for identifying contaminants is conservative, it may sometimes result in false positives, as sampling of outgroup sequences for *clpP* and *accD* is relatively sparse here and on GenBank, limiting our ability to distinguish the green taxa that contaminant sequences are most closely related to using tree-based methods. As a consequence, it was not always possible to confirm the probable identity of putative contaminating sequences. However, we have unpublished whole-plastid genome data in hand for several species that we also surveyed in the three-gene analysis here, which allowed us to verify authentic gene loss from the plastid genomes for several taxa where we obtained contaminant amplifications from conspecific samples (i.e., *clpP* and *matK* for *Apteria aphylla* in Burmanniaceae, and *matK* for *Arachnitis uniflora* in Corsiaceae; Appendix S1). These genomic data confirm the validity of including noncontaminant genes from samples that had residual contamination, at least for these two species, because the genes obtained by PCR amplification and Sanger sequencing here, which we included in analyses, were identical or very similar to those obtained using next-generation methods (Appendices S7c, d).

Suitability of the three genes as DNA barcoding markers—We did not specifically address the utility of the three plastid markers here as DNA barcoding markers, although two of them have been tested for their suitability in previous DNA barcoding studies (*accD* and *matK*, see below). In green plants, portions of *matK* and *rbcL* have been used as core barcoding markers, often supplemented with the *psbA-trnH* intergenic spacer and the nuclear ITS (internal transcribed spacer) region (e.g., Fazekas et al., 2008; Hollingsworth et al., 2009, 2011; Li et al., 2011). However, *rbcL* and *psbA-trnH* are not suitable as barcoding markers in fully heterotrophic plants because they either are, or involve, photosynthetic genes. The *matK* region used here is the DNA barcoding region for this gene (Appendix S2; Hollingsworth et al., 2009), and our study demonstrates that it can be recovered from fully mycoheterotrophic monocots with relative ease. However, we were unable to recover *matK* or *clpP* in multiple cases, and our amplification strategy sometimes recovered contaminants from these two genes that only become apparent after careful checking (Appendices S1, S3). In contrast, we were able to recover the *accD* locus from all taxa sampled here, without any contamination issues (note that *accD* is pseudogenized in some eudicot holoparasites; Wicke et al., 2013). The *accD* locus has only rarely been considered for DNA barcoding studies (Newmaster et al., 2008), and it would be useful to explore further its utility in DNA barcoding surveys that include mycoheterotrophs or holoparasites. It may be a useful supplementary barcoding region to consider including when the focus is not exclusively on green plants.

CONCLUSIONS

Mycoheterotrophic lineages can be placed in overall monocot phylogeny using our three-gene data set, generally with moderate to strong support for their placement, despite some issues with contamination, rate elevation, and missing genes. Our approach to inferring phylogenetic placement is a cost-effective alternative to next-generation sequencing and potentially allows phylogenetic surveys of many heterotrophic taxa. It is also a potentially useful screen in selecting DNAs that will be successful for next-generation library preparation. At least two of the markers considered here (*accD*, *matK*) may also be useful targets for including heterotrophic plants in DNA barcoding surveys. The underlying alignment we used could be improved upon further and may be particularly useful to do for other parasitic or mycoheterotrophic lineages, such as Orchidaceae, where we included only a few previously published sequences: this family includes numerous additional origins of full mycoheterotrophy (e.g., Freudenstein and Senyo, 2008; Merckx et al., 2013b; Barrett et al., 2014).

ACKNOWLEDGEMENTS

The authors thank Thassilo Franke, Stephanie Lyon, Daniel McNair, Ray Neyland, Tomohisa Yukawa, the Missouri Botanical Garden, the Royal Botanic Gardens, Kew and Kirstenbosch National Botanical Garden for assistance in obtaining plant material, and Craig Barrett for providing commelinid sequence data. Craig Barrett and an anonymous reviewer provided helpful comments for improving the manuscript. This work was supported by an NSERC (Natural Sciences and Engineering Research Council of Canada) Postgraduate Fellowship and UBC Four-Year Fellowship to V.K.Y.L., by NWO through a Veni Fellowship to V.S.F.T.M. (863.11.018), and by an NSERC Discovery grant to S.W.G.

LITERATURE CITED

Abrahamsen, M. S., T. J. Templeton, S. Enomoto, J. E. Abrahante, G. Zhu, C. A. Lancto, M. Deng, et al. 2004. Complete genome sequence of the apicomplexan, *Cryptosporidium parvum*. *Science* 304: 441–445.

Angiosperm Phylogeny Group. 1998. An ordinal classification for the families of flowering plants. *Annals of the Missouri Botanical Garden* 4: 531–553.

Angiosperm Phylogeny Group. 2003. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG II. *Botanical Journal of the Linnean Society* 141: 399–436.

Angiosperm Phylogeny Group. 2009. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG III. *Botanical Journal of the Linnean Society* 161: 105–121.

Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. 1990. Basic local alignment search tool. *Journal of Molecular Biology* 215: 403–410.

Barbrook, A. C., C. J. Howe, and S. Purton. 2006. Why are plastid genomes retained in non-photosynthetic organisms? *Trends in Plant Science* 11: 101–108.

Barrett, C. F., and J. I. Davis. 2012. The plastid genome of the mycoheterotrophic *Corallorhiza striata* (Orchidaceae) is in the relatively early stages of degradation. *American Journal of Botany* 99: 1513–1523.

Barrett, C. F., J. V. Freudenstein, J. Li, D. R. Mayfield-Jones, L. Perez, J. C. Pires, and C. Santos. 2014. Investigating the path of plastid genome degradation in early-transitional heterotrophic orchids, and implications for heterotrophic angiosperms. *Molecular Biology and Evolution* 31: 3095–3112.

Bellot, S., and S. S. Renner. 2016. The plastomes of two species in the endoparasite genus *Pilostyles* (Apodanthaceae) each retain just five or six possibly functional genes. *Genome Biology and Evolution* 8: 189–201.

Bernard, N. 1909. L'évolution dans la symbiose. Les orchidées et leurs champignons commensaux. *Annales des Sciences Naturelles. Botanique* 9: 1–196.

Bodin, S. S., J. S. Kim, and J.-H. Kim. 2016. Phylogenetic inferences and evolution of plastid DNA in Campynemataceae and the mycoheterotrophic *Corsia dispar* D.L. Jones & B. Gray (Corsiaceae). *Plant Molecular Biology Reporter* 34: 192–210.

Bolin, J. F., K. U. Tennakoon, M. B. A. Majid, and D. D. Cameron. In press. Isotopic evidence of partial mycoheterotrophy in *Burmannia coelestis* (Burmanniaceae). *Plant Species Biology*.

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

Caddick, L. R., P. J. Rudall, P. Wilkin, and M. W. Chase. 2000. Yams and their allies: Systematics of Dioscoreales. *In* K. L. Wilson, D. A. Morrison [eds.], Monocots: Systematics and evolution, 475–487. CSIRO Publishing, Victoria, Australia.

Caddick, L. R., P. J. Rudall, P. Wilkin, T. A. J. Hedderson, and M. W. Chase. 2002. Phylogenetics of Dioscoreales based on analyses of morphological and molecular data. *Botanical Journal of the Linnean Society* 138: 123–144.

Cameron, K. M., M. W. Chase, and P. J. Rudall. 2003. Recircumscription of the monocotyledonous family Petrosaviaceae to include *Japonolirion. Brittonia* 55: 214–225.

Chase, M. W., M. F. Fay, D. S. Devey, O. Maurin, N. Rønsted, T. J. Davies, Y. Pillon, et al. 2006. Multigene analyses of monocot relationships: A summary. *Aliso* 22: 63–75.

Chase, M. W., D. E. Soltis, R. G. Olmstead, D. Morgan, D. H. Les, B. D. Mishler, M. R. Duvall, et al. 1993. Phylogenetics of seed plants: An analysis of nucleotide sequences from the plastid *rbc*L. *Annals of the Missouri Botanical Garden* 80: 528–580.

Chase, M. W., D. E. Soltis, P. S. Soltis, P. J. Rudall, M. F. Fay, W. H. Hahn, S. Sullivan, et al. 2000. Higher-level systematics of the monocotyledons: An assessment of current knowledge and a new classification. *In* K. L. Wilson and D. A. Morrison [eds.], Monocots: Systematics and evolution, 3–16. CSIRO Publishing, Victoria, Australia.

Cronquist, A. 1988. The evolution and classification of flowering plants, 2nd ed. New York Botanical Garden, Bronx, New York, USA.

Cummings, M. P., and N. A. Welschmeyer. 1998. Pigment composition of putatively achlorophyllous angiosperms. *Plant Systematics and Evolution* 210: 105–111.

Cusimano, N., and S. Wicke. 2015. Massive intracellular gene transfer during plastid genome reduction in nongreen Orobanchaceae. *New Phytologist*.

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

Davis, J. I., D. W. Stevenson, G. Petersen, O. Seberg, L. M. Campbell, J. V. Freudenstein, D. H. Goldman, et al. 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. *Systematic Botany* 29: 467–510.

Delannoy, E., S. Fujii, C. C. Des Francs-Small, M. Brundrett, and I. Small. 2011. Rampant gene loss in the underground orchid *Rhizanthella gardneri* highlight evolutionary constraints on plastid genomes. *Molecular Biology and Evolution* 28: 2077–2086.

Drummond, A. J., M. A. Suchard, D. Xie, and A. Rambaut. 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution* 29: 1969–1973.

Doyle, J. J., and J. L. Doyle. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin* 19: 11–15.

Ems, S. C., C. W. Morden, C. K. Dixon, K. H. Wolfe, C. W. dePamphilis, and J. D. Palmer. 1995. Transcription, splicing and editing of plastid RNAs in the nonphotosynthetic plant *Epifagus virginiana. Plant Molecular Biology* 29: 721–733.

Engels, B. 1993. Amplify3x [online]. Website http://engels.genetics.wisc.edu/amplify/ [accessed 15 May 2010].

Fay, M. F., P. J. Rudall, S. Sullivan, K. L. Stobart, A. Y. De Brujin, G. Reeves, F. Qamaruz-Zaman, et al. 2000. Phylogenetic studies of Asparagales based on four plastid DNA regions. *In* K. L. Wilson and D. A. Morrison [eds.], Monocots: Systematics and evolution, 360–371. CSIRO Publishing, Collingwood, Australia.

Fazekas, A. J., K. S. Burgess, P. R. Kesanakurti, S. W. Graham, S. G. Newmaster, B. C. Husband, D. M. Percy, et al. 2008. Multiple multilocus DNA barcodes from the plastid genome discriminate plant species equally well. *PLoS One* 3: e2802.

Felsenstein, J. 1978. Cases in which parsimony or compatibility methods will be positively misleading. *Systematic Zoology* 27: 401–410.

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

Felsenstein, J. 1988. Phylogenies from molecular sequences: Inference and reliability. *Annual Review of Genetics* 22: 521–565.

Freudenstein, J. V., and D. M. Senyo. 2008. Relationships and evolution of *matK* in a group of leafless orchids (*Corallorhiza* and Corallorhizinae; Orchidaceae: Epidendroideae). *American Journal of Botany* 95: 498–505.

Fuse, S., and M. N. Tamura. 2000. A phylogenetic analysis of the plastid *mat*K gene with emphasis on Melanthiaceae *sensu lato. Plant Biology* 2: 415–427.

Gandolfo, M. A., K. C. Nixon, and W. L. Crepet. 2002. Triuridaceae fossil flowers from the Upper Cretaceous of New Jersey. *American Journal of Botany* 89: 1940–1957.

Gandolfo, M. A., K. C. Nixon, W. L. Crepet, D. W. Stevenson, and E. M. Friis. 1998. Oldest known fossils of monocotyledons. *Nature* 394: 532–533.

Givnish, T. J., M. Ames, J. R. McNeal, M. R. McKain, P. R. Steele, C. W. DePamphilis, S. W. Graham, et al. 2010. Assembling the tree of the monocotyledons: Plastome sequence phylogeny and evolution of Poales. *Annals of the Missouri Botanical Garden* 97: 584–616.

Givnish, T. J., J. C. Pires, S. W. Graham, M. A. McPherson, L. M. Prince, T. B. Patterson, H. S. Rai, et al. 2005. Repeated evolution of net venation and fleshy fruits among monocots in shaded habitats confirms a priori predictions: Evidence from an *ndhF* phylogeny. *Proceedings of the Royal Society of London, B, Biological Sciences* 272: 1481–1490.

Goldblatt, P., A. Rodriguez, M. P. Powell, T. J. Davies, J. C. Manning, M. Van Der Bank, and V. Savolainen. 2008. Iridaceae 'Out of Australasia'? Phylogeny, biogeography, and divergence time based on plastid DNA sequences. *Systematic Botany* 3: 495–508.

Graham, S. W., and R. G. Olmstead. 2000. Utility of 17 chloroplast genes for inferring the phylogeny of the basal angiosperms. *American Journal of Botany* 87: 1712–1730.

Graham, S. W., P. A. Reeves, A. C. Burns, and R. G. Olmstead. 2000. Microstructural changes in noncoding chloroplast DNA: Interpretation, evolution, and utility of indels and inversions in basal angiosperm phylogenetic inference. *International Journal of Plant Sciences* 161: S83–S96.

Graham, S. W., J. M. Zgurski, M. A. McPherson, D. M. Cherniawsky, J. M. Saarela, E. S. C. Horne, et al. 2006. Robust inference of monocot deep phylogeny using an expanded multigene plastid data set. *Aliso* 22: 3–20.

Hao, W., and J. D. Palmer. 2009. Fine-scale mergers of chloroplast and mitochondrial genes create functional, transcompartmentally chimeric mitochondrial genes. *Proceedings of the National Academy of Sciences, USA* 106: 16728–16733.

Heath, T. A., S. M. Hedtke, and D. M. Hillis. 2008. Taxon sampling and the accuracy of phylogenetic analyses. *Journal of Systematics and Evolution* 46: 239–257.

Hedtke, S. M., T. M. Townsend, and D. M. Hillis. 2006. Resolution of phylogenetic conflict in large data sets by increased taxon sampling. *Systematic Biology* 55: 522–529.

Hendy, M. D., and D. Penny. 1989. A framework for quantitative study of evolutionary trees. *Systematic Zoology* 38: 297–309.

Hertweck, K. L., M. S. Kinney, S. A. Stuart, O. Maurin, S. Mathews, M. W. Chase, M. A. Gandaldo, and J. C. Pires. 2015. Phylogenetics, divergence times and diversification from three genomic partitions in monocots. *Botanical Journal of the Linnean Society* 178: 375–393.

Hillis, D. M. 1998. Taxonomic sampling, phylogenetic accuracy, and investigator bias. *Systematic Biology* 47: 3–8.

Hilu, K. W., T. Borsch, K. Müller, D. E. Soltis, P. S. Soltis, V. Savolainen, M. W. Chase, et al. 2003. Angiosperm phylogeny based on *matK* sequence information. *American Journal of Botany* 90: 1758–1776.

Hollingsworth, P. M., L. L. Forest, J. L. Spouge, M. Hajibabaei, S. Ratnasingham, M. van der Bank, M. W. Chase, et al. 2009. A DNA barcode for plants. *Proceedings of the National Academy of Sciences of the United States of America* 106: 12794–12797.

Hollingsworth, P. M., S. W. Graham, and D. P. Little. 2011. Choosing and using a plant DNA barcode. *PLoS One* 6: e19254.

Huelsenbeck, J. P. 1997. Is the Felsenstein zone a fly trap? *Systematic Biology* 46: 69–74.

Huelsenbeck, J. P. 1998. Systematic bias in phylogenetic analysis: Is the *Strepsiptera* problem solved? *Systematic Biology* 47: 519–537.

Iles, W. J. D., S. Y. Smith, M. A. Gandalfo, and S. W. Graham. 2015. Monocot fossils suitable for molecular dating analyses. *Botanical Journal of the Linnean Society* 178: 346–374.

Imhof, S. 2010. Are monocots particularly suited to develop mycoheterotrophy? *In* O. Seberg, G. Petersen, A. Barfod, and J. I. Davis [eds.], Diversity, phylogeny, and evolution in the monocotyledons, 11–23. Aarhus University Press, Copenhagen, Denmark.

Janouškovec, J., D. V. Tikhonenkov, F. Burki, A. T. Howe, M. Kolisko, A. P. Mylnikov, and P. J. Keeling. 2015. Factors mediating plastid dependency and the origins of parasitism in apicomplexans and their close relatives. *Proceedings of the National Academy of Sciences, USA* 112: 10200–10207.

Jansen, R. K., Z. Cai, L. A. Raubeson, H. Daniell, C. W. DePamphilis, J. Leebens-Mack, K. F. Müller, et al. 2007. Analysis of 81 genes from 64 plastid genomes resolves relationships in angiosperms and identifies genome-scale evolutionary patterns. *Proceedings of the National Academy of Sciences, USA* 104: 19369–19374.

Kelchner, S. A. 2000. The evolution of non-coding chloroplast DNA and its application in plant systematics. *Annals of the Missouri Botanical Garden* 87: 482–498.

Kim, J. 1996. General inconsistency conditions for maximum parsimony: Effects of branch lengths and increasing numbers of taxa. *Systematic Biology* 45: 363–374.

Kim, J. S., J.-K. Hong, M. W. Chase, M. F. Fay, and J.-H. Kim. 2013. Familial relationships of the monocot order Liliales based on a molecular phylogenetic analysis using four plastid loci: *matK*, *rbcL*, *atpB* and *atpF*. *Botanical Journal of the Linnean Society* 172: 5–21.

Kolaczkowski, B., and J. W. Thornton. 2009. Long-branch attraction bias and inconsistency in Bayesian phylogenetics. *PLoS One* 4: e7891.

Lam, V. K. Y., M. Soto Gomez, and S. W. Graham. 2015. The highly reduced plastome of mycoheterotrophic *Sciaphila* (Triuridaceae) is colinear with its green relative and is under strong purifying selection. *Genome Biology and Evolution* 7: 2220–2236.

Lanfear, R., B. Calcott, S. Y. W. Ho, and S. Guindon. 2012. PartitionFinder: Combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Molecular Biology and Evolution* 29: 1695–1701.

Leake, J. R. 1994. The biology of myco-heterotrophic ('saprophytic') plants. *New Phytologist* 127: 171–216.

Leake, J. R. 2004. Myco-heterotrophic/epiparasitic plant interactions with ectomycorrhizal and arbuscular mycorrhizal fungi. *Current Opinion in Plant Biology* 7: 422–428.

Leake, J. R. 2005. Plants parasitic on fungi: Unearthing the fungi in mycoheterotrophs and debunking the 'saprophytic' plant myth. *Mycologist* 19: 113–122.

Li, D.-Z., L.-M. Gao, H.-T. Li, H. Wang, X.-J. Ge, J.-Q. Liu, Z.-D. Chen, et al. 2011. Comparative analysis of a large dataset indicates that internal transcribed spacer (ITS) should be incorporated into the core barcode for seed plants. *Proceedings of the National Academy of Sciences, USA* 108: 19641–19646.

Logacheva, M. D., M. I. Schelkunov, M. S. Nuraliev, T. H. Samigullin, and A. A. Penin. 2014. The plastid genome of mycoheterotrophic monocot *Petrosavia stellaris* exhibits both gene losses and multiple rearrangements. *Genome Biology and Evolution* 6: 238–246.

Logacheva, M. D., M. I. Schelkunov, and A. A. Penin. 2011. Sequencing and analysis of plastid genome in mycoheterotrophic orchid *Neottia nidus-avis*. *Genome Biology and Evolution* 3: 1296–1303.

Martin, W., B. Stoebe, V. Goremykin, S. Hansmann, M. Hasegawa, and K. V. Kowallik. 1998. Gene transfer to the nucleus and the evolution of chloroplasts. *Nature* 393: 162–165.

Matsuo, M., Y. Ito, R. Yamauchi, and J. Obokata. 2005. The rice nuclear genome continuously integrates, shuffles, and eliminates the chloroplast genome to cause chloroplast-nuclear DNA flex. *Plant Cell* 17: 665–675.

McNeal, J. R., J. V. Kuehl, J. L. Boore, and C. W. DePamphilis. 2007. Complete plastid genome sequences suggest strong selection for retention of photosynthetic genes in the parasitic plant genus *Cuscuta*. *BMC Plant Biology* 7: 57.

McNeal, J. R., J. V. Kuehl, J. L. Boore, J. Leebens-Mack, and C. W. DePamphilis. 2009. Parallel loss of plastid introns and their maturase in the genus *Cuscuta*. *PLoS One* 4: e5982.

Mennes, C. B., V. K. Y. Lam, P. J. Rudall, S. P. Lyon, S. W. Graham, E. F. Smets, and V. S. F. T. Merckx. 2015. Ancient Gondwana break-up explains the distribution of the mycoheterotrophic family Corsiaceae (Liliales). *Journal of Biogeography* 69: 994–1004.

Mennes, C. B., E. F. Smets, S. N. Moses, and V. S. F. T. Merckx. 2013. New insights in the long-debated evolutionary history of Triuridaceae. *Molecular Phylogenetics and Evolution* 69: 994–1004.

Merckx, V. S. F. T. 2013. Mycoheterotrophy: An introduction. *In* V. Merckx [ed.], Mycoheterotrophy: The biology of plants living on fungi, 1–18. Springer-Verlag, New York, New York, USA.

Merckx, V., F. T. Bakker, S. Huysman, and E. Smets. 2009b. Bias and conflict in phylogenetic inference of myco-heterotrophic plants: A case study in Thismiaceae. *Cladistics* 25: 64–77.

Merckx, V., M. I. Bidartondo, and N. A. Hynson. 2009a. Myco-heterotrophy: When fungi host plants. *Annals of Botany* 104: 1255–1261.

Merckx, V., L. W. Chatrou, B. Lemaire, M. N. Sainge, S. Huysmans, and E. F. Smets. 2008. Diversification of myco-heterotrophic angiosperms: Evidence from Burmanniaceae. *BMC Evolutionary Biology* 8: 178.

Merckx, V., and J. V. Freudenstein. 2010. Evolution of mycoheterotrophy in plants: A phylogenetic perspective. *New Phytologist* 185: 605–609.

Merckx, V., S. Huysmans, and E. Smets. 2010. Cretaceous origins of mycoheterotrophic lineages in Dioscoreales. *In* O. Seberg, G. Petersen, A. Barfod, and J. I. Davis [eds.], Diversity, phylogeny, and evolution in the monocotyledons, 39–53. Aarhus University Press, Copenhagen, Denmark.

Merckx, V., P. Schols, H. Maas-Van De Kamer, P. Maas, S. Huysmans, and E. Smets. 2006. Phylogeny and evolution of Burmanniaceae (Dioscoreales) based on nuclear and mitochondrial data. *American Journal of Botany* 93: 1684–1698.

Merckx, V. S. F. T. 2013. Mycoheterotrophy: An introduction. *In* V. Merckx [ed.], Mycoheterotrophy: The biology of plants living on fungi, 1–18. Springer-Verlag, New York, New York, USA.

Merckx, V. S. F. T., J. V. Freudenstein, J. Kissling, M. J. M. Christenhusz, R. E. Stotler, B. Crandall-Stotler, N. Wickett, et al. 2013b. Taxonomy and classification. *In* V. Merckx [ed.], Mycoheterotrophy: The biology of plants living on fungi, 19–101. Springer-Verlag, New York, New York, USA.

Merckx, V. S. F. T., C. B. Mennes, K. G. Peay, and J. Geml. 2013a. Evolution and diversification. *In* V. Merckx [ed.], Mycoheterotrophy: The biology of plants living on fungi, 222–226. Springer-Verlag, New York, New York, USA.

Molina, J., K. M. Hazzouri, D. Nickrent, M. Geisler, R. S. Meyer, M. M. Pentony, J. M. Flowers, et al. 2014. Possible loss of the chloroplast genome in the parasitic flowering plant *Rafflesia lagascae* (Rafflesiaceae). *Molecular Biology and Evolution* 31: 793–803.

Mower, J. P., S. Stefanović, G. J. Young, and J. D. Palmer. 2004. Plant genetics: Gene transfer from parasitic to host plants. *Nature* 432: 165–166.

Newmaster, S. G., A. J. Fazekas, R. A. D. Steeves, and J. Janovec. 2008. Testing candidate plant barcode regions in the Myristicaceae. *Molecular Ecology Resources* 8: 480–490.

Neyland, R. 2002. A phylogeny inferred from large-subunit (26S) ribosomal DNA sequences suggest that Burmanniales are polyphyletic. *Australian Systematic Biology* 15: 19–28.

Neyland, R., and M. Hennigan. 2003. A phylogenetic analysis of large-subunit (26S) ribosome DNA sequences suggests that the Corsiaceae are polyphyletic. *New Zealand Journal of Botany* 41: 1–11.

Palmer, J. D., and L. A. Herbon. 1988. Plant mitochondrial DNA evolves rapidly in structure, but slowly in sequence. *Journal of Molecular Evolution* 28: 87–97.

Parks, S. L., and N. Goldman. 2014. Maximum likelihood inference of small trees in the presence of long branches. *Systematic Biology* 63: 798–811.

Pollock, D. D., D. J. Zwickl, J. A. McGuire, and D. M. Hillis. 2002. Increased taxon sampling is advantageous for phylogenetic inference. *Systematic Biology* 51: 664.

Rai, H. S., H. O'Brien, P. A. Reeves, R. G. Olmstead, and S. W. Graham. 2003. Inference of higher-order relationships in the cycads from a large chloroplast data set. *Molecular Phylogenetics and Evolution* 29: 350–359.

Rambaut, A. 2002. Se-Al v. 2.0a11: Sequence alignment program. Available from http://tree.bio.ed.ac.uk/software/seal/ [accessed 15 January 2016].

Rambaut, A. 2006. FigTree. Available from http://tree.bio.ed.ac.uk/software/figtree/ [accessed 15 January 2016].

Rambaut, A., M. A. Suchard, D. Xie, and A. J. Drummond. 2014. Tracer v1. 6. Available from http://beast.bio.ed.ac.uk/Tracer [accessed 15 January 2016].

Rasmussen, H. N. 1995. Terrestrial orchids: From seed to mycotrophic plant. Cambridge University Press, Cambridge, UK.

Reeves, G., M. W. Chase, P. Goldblatt, P. Rudall, M. F. Fay, A. V. Cox, B. Lejeune, and T. Souza-Chies. 2001. Molecular systematics of Iridaceae: Evidence from four plastid DNA regions. *American Journal of Botany* 88: 2074–2087.

Rousseau-Gueutin, M., X. Huang, E. Higginson, M. Ayliffe, A. Day, and J. N. Timmis. 2013. Potential functional replacement of the plastidic acetyl-CoA carboxylase subunit (*accD*) gene by recent transfer to the nucleus in some angiosperm lineages. *Plant Physiology* 161: 1918–1929.

Rudall, P. J., and R. M. Bateman. 2006. Morphological phylogenetic analysis of Pandanales: Testing contrasting hypotheses of floral evolution. *Systematic Botany* 31: 223–238.

Ruhfel, B. R., M. A. Gitzendanner, P. S. Soltis, D. E. Soltis, and J. G. Burleigh. 2014. From algae to angiosperms—Inferring the phylogeny of green plants (*Viridiplantae*) from 360 plastid genomes. *BMC Evolutionary Biology* 14: 23.

Rychlik, W. 2007. OLIGO 7 primer analysis software. *In* A. Yuryev [ed.], Methods in molecular biology, vol. 402, 35–59. Humana Press, New York, New York, USA.

Saarela, J. M., and S. W. Graham. 2010. Inference of phylogenetic relationships among the subfamilies of grasses (Poaceae: Poales) using meso-scale exemplar-based sampling of the plastid genome. *Botany* 88: 65–84.

Schelkunov, M. I., V. Y. Shtratnikova, M. S. Nuraliev, M.-A. Selosse, A. A. Penin, and M. D. Logacheva. 2015. Exploring the limits for reduction of plastid genomes: A case study of the mycoheterotrophic orchids *Epipogium aphyllum* and *Epipogium roseum*. *Genome Biology and Evolution* 7: 1179–1191.

Schwarz, G. 1978. Estimating the dimension of a model. *Annals of Statistics* 6: 461–464.

Shanklin, J., N. D. DeWitt, and J. M. Flanagan. 1995. The stroma of higher plant plastids contain ClpP and ClpC, functional homologs of *Escherichia coli* ClpP and ClpA: An archetypal two-component ATP-dependent protease. *Plant Cell* 7: 1713.

Shimodaira, H. 2002. An approximately unbiased test of phylogenetic tree selection. *Systematic Biology* 51: 492–508.

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

Silvestro, D., and I. Michalak. 2012. raxmlGUI: A graphical front-end for RaxML. *Organisms, Diversity & Evolution* 12: 335–337.

Simmons, M. P., and H. Ochoterena. 2000. Gaps as characters in sequence-based phylogenetic analyses. *Systematic Biology* 49: 369–381.

Simmons, M. P., K. M. Pickett, and M. Miya. 2004. How meaningful are Bayesian support values? *Molecular Biology and Evolution* 21: 188–199.

Smith, D. R., and R. W. Lee. 2014. A plastid without a genome: Evidence from the nonphotosynthetic green algal genus *Polytomella*. *Plant Physiology* 164: 1812–1819.

Soltis, D. E., S. A. Smith, N. Cellinese, K. J. Wurdack, D. C. Tank, S. F. Brockington, N. F. Refulio-Rodriguez, et al. 2011. Angiosperm phylogeny: 17 genes, 640 taxa. *American Journal of Botany* 98: 704–730.

Soltis, D. E., and P. S. Soltis. 2004. *Amborella* not a "basal angiosperm"? Not so fast. *American Journal of Botany* 91: 997–1001.

Soltis, D. E., P. S. Soltis, M. W. Chase, M. E. Mort, D. C. Albach, M. Zanis, V. Savolainen, et al. 2000. Angiosperm phylogeny inferred from 18S rDNA, *rbcL*, and *atpB* sequences. *Botanical Journal of the Linnean Society* 133: 381–461.

Stamatakis, A. 2006. RaxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22: 2688–2690.

Straub, S. C. K., M. Fishbein, T. Livshultz, Z. Foster, M. Parks, K. Weitmier, R. Cronn, and A. Liston. 2011. Building a model: Developing genome resources for common milkweed (*Asclepias syriaca*) with low coverage genome sequencing. *BMC Genomics* 12: 211.

Sullivan, J., and P. Joyce. 2005. Model selection in phylogenetics. *Annual Review of Ecology, Evolution and Systematics* 36: 445–466.

Swofford, D. L. 2003. PAUP*: Phylogenetic analysis using parsimony (*and other methods), version 4.0b10. Sinauer, Sunderland, Massachusetts, USA.

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

Vogel, J., T. Hübschmann, T. Börner, and W. R. Hess. 1997. Splicing and intron-internal RNA editing of *trnK-matK* transcripts in barley plastids: Support for *matK* as an essential splice factor. *Journal of Molecular Biology* 270: 179–187.

Waterman, R. J., M. R. Klooster, H. Hentrich, and M. I. Bidartondo. 2013. *In* V. Merckx [ed.], Mycoheterotrophy: The biology of plants living on fungi, 267–296. Springer-Verlag, New York, New York, USA.

Wicke, S., K. F. Müller, C. W. DePamphilis, D. Quandt, N. J. Wickett, Y. Zhang, S. S. Renner, and G. M. Schneeweiss. 2013. Mechanisms of functional and physical genome reduction in photosynthetic and nonphotosynthetic parasitic plants of the broomrape family. *Plant Cell* 25: 3711–3725.

Wicke, S., G. M. Schneeweiss, C. W. DePamphilis, K. F. Müller, and D. Quandt. 2011. The evolution of the plastid chromosome in land plants: gene content, gene order, gene function. *Plant Molecular Biology* 76: 273–297.

Wickett, N. J., Y. Zhang, S. K. Hansen, J. M. Roper, J. V. Kuehl, S. A. Plock, P. G. Wolfe, et al. 2008. Functional gene losses occur with minimal size reduction in the plastid genome of the parasitic liverwort *Aneura mirabilis*. *Molecular Biology and Evolution* 25: 393–401.

Wiens, J. J. 2005. Can incomplete taxa rescue phylogenetic analyses from long-branch attraction? *Systematic Biology* 54: 731–742.

Wiens, J. J. 2006. Missing data and the design of phylogenetic analyses. *Journal of Biomedical Informatics* 39: 34–42.

Wiens, J. J., and J. Tiu. 2012. Highly incomplete taxa can rescue phylogenetic analyses from the negative impacts of limited taxon sampling. *PLoS One* 7: e42925.

Wolfe, K. H., W.-H. Li, and P. M Sharp. 1987. Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. *Proceedings of the National Academy of Sciences, USA* 84: 9054-9058.

Wolfe, K. H., C. W. Morden, and J. D. Palmer. 1992. Function and evolution of a minimal plastid genome from a nonphotosynthetic parasitic plant. *Proceedings of the National Academy of Sciences, USA* 89: 10648–10652.

Yang, Z. 1996. Phylogenetic analysis using parsimony and likelihood methods. *Journal of Molecular Evolution* 42: 294–307.

Yang, Z. 2006. Computational molecular evolution. Oxford University Press, Oxford, UK.

Yang, Z., and B. Rannala. 2005. Branch-length prior influences Bayesian posterior probability of phylogeny. *Systematic Biology* 54: 455–470.

Yang, Z., and B. Rannala. 2012. Molecular phylogenetics: Principles and practice. *Nature Reviews. Genetics* 13: 303–314.

Yoshida, S., S. Maruyama, H. Nozaki, and K. Shirasu. 2010. Horizontal gene transfer by the parasitic plant *Striga hermonthica*. *Science* 328: 1128.

Zgurski, J. M., H. S. Rai, Q. M. Fai, D. J. Bogler, J. Francisco-Ortega, and S. W. Graham. 2008. How well do we understand the overall backbone of cycad phylogeny? New insights from a large, multigene plastid data set. *Molecular Phylogenetics and Evolution* 47: 1232–1237.

Zhang, D., R. M. Saunders, and C.-M. Hu. 1999. *Corsiopsis chinensis* gen. et sp. nov. (Corsiaceae): First record of the family in Asia. *Systematic Botany* 24: 311–314.

RESEARCH ARTICLE

AMERICAN JOURNAL OF BOTANY

A few-gene plastid phylogenetic framework for mycoheterotrophic monocots[1]

Vivienne K.Y. Lam[2,3], Vincent S.F.T. Merckx[4], and Sean W. Graham[2,3,5]

PREMISE OF THE STUDY: Few-gene studies with broad taxon sampling have provided major insights into phylogeny and underpin plant classification. However, they have typically excluded heterotrophic plants because of loss, pseudogenization, or rapid evolution of plastid genes. Here we performed a phylogenetic survey of three commonly retained plastid genes to assess their utility in placing mycoheterotrophs.

METHODS: We surveyed *accD*, *clpP*, and *matK* for 34 taxa from seven monocot families that include full mycoheterotrophs and a broad sampling of photosynthetic taxa. After screening for weak contaminants, we conducted phylogenetic analyses and characterized among-lineage rate variation.

KEY RESULTS: Likelihood analyses strongly supported local placements of fully mycoheterotrophic taxa for Corsiaceae, Iridaceae, Orchidaceae, and Petrosaviaceae, in positions consistent with other studies. Depression of likelihood bootstrap support values near mycoheterotrophic clades was alleviated when each mycoheterotrophic family was considered separately. Triuridaceae (*Sciaphila*) monophyly was recovered in a partitioned likelihood analysis, and the family then placed as sister to Cyclanthaceae-Pandanaceae. Burmanniaceae placed in Dioscoreales with weak to strong support depending on analysis details, and we inferred a plastid-based phylogeny for the family. Thismiaceae species may retain a plastid genome, based on *accD* retention. The inferred position of Thismiaceae is unstable, but was close to Taccaceae (Dioscoreales) in some analyses.

CONCLUSIONS: Long branches/elevated substitution rates, missing genes, and occasional contaminants are challenges for plastid-based phylogenetic inference with full mycoheterotrophs. However, most mycoheterotrophs can be readily integrated into the broad picture of plant phylogeny using several plastid genes and broad taxonomic sampling.

KEY WORDS achlorophyllous plants; Asparagales; Dioscoreales; DNA barcoding; gene retention; higher-order relationships; Liliales; Pandanales; Petrosaviales; plastomes

Mycoheterotrophic plants acquire carbon, water, and essential nutrients from fungal partners, typically those involved in mycorrhizal symbioses with green plants (Leake, 2004, 2005; Merckx et al., 2009a). As in holoparasitic plants, some mycoheterotrophs have become dependent on the host organism for all their nutritional needs and have lost the ability to photosynthesize: full mycoheterotrophs rely on fungal carbon for their entire life cycle (see Merckx, 2013). This major nutritional transition has evolved at least 47 times in land-plant evolution (Merckx et al., 2013a). In angiosperms the loss of photosynthesis may be associated with substantial changes in plant morphology, including reduced stature and foliage, and novel floral forms (Leake, 1994). Fully mycoheterotrophic plants are generally not green (although limited chlorophyll production is sometimes retained; Cummings and Welschmeyer, 1998; Barrett et al., 2014). Heterotrophic plants also often have substantially modified plastid genomes due to elevated rates of substitution, gene losses, and rearrangements (e.g., Wickett et al., 2008; Delannoy et al., 2011; Logacheva et al., 2011, 2014; Barrett and Davis, 2012; Lam et al., 2015; Mennes et al., 2015; Schelkunov et al., 2015; Bellot and Renner, 2016). The ultimate fate of the plastid genome in heterotrophs may be loss, although this may have occurred in only one embryophyte group (putatively in the holoparasite *Rafflesia*; Molina et al., 2014). The usual persistence of plastid genomes in heterotrophs is thought to reflect nonphotosynthetic functions performed by some plastid-encoded gene products (e.g., Bungard, 2004; Barbrook et al., 2006; Wicke et al., 2011). However, retained

---

[1] Manuscript received 22 September 2015; revision accepted 8 February 2016.

[2] Department of Botany, 6270 University Boulevard, University of British Columbia, Vancouver, British Columbia, V6T 1Z4, Canada;

[3] UBC Botanical Garden & Centre for Plant Research, 6804 Marine Drive SW, University of British Columbia, Vancouver, British Columbia, V6T 1Z4, Canada; and

[4] Naturalis Biodiversity Center, Leiden University, Leiden, The Netherlands

[5] Author for correspondence (e-mail: swgraham@interchange.ubc.ca)

doi:10.3732/ajb.1500412

functional genes may also experience elevated rates of evolution (e.g., Logacheva et al., 2011; Barrett et al., 2014). Gene losses, pseudogenization, and rate elevation may all contribute to uncertainty in inferring higher-order placement of mycoheterotrophs in plant phylogeny (Merckx and Freudenstein, 2010).

There are ~514 extant species of fully mycoheterotrophic plants (Merckx et al., 2013a). The majority of species (91%) and origins (~83%) are found in monocots, which seem to be particularly prone to this evolutionary transition (Imhof, 2010; Merckx et al., 2013a). In total, seven monocot families include fully mycoheterotrophic taxa (i.e., Burmanniaceae, Corsiaceae, Iridaceae, Orchidaceae, Petrosaviaceae, Thismiaceae, and Triuridaceae; Leake, 1994). Orchidaceae have at least 235 fully mycoheterotrophic species, which derive from an estimated 25 losses of photosynthesis (Merckx et al., 2013a, b), and the entire family is mycoheterotrophic during seedling establishment (e.g., Bernard, 1909; Leake, 1994; Rasmussen, 1995; Merckx, 2013), referred to as initial mycoheterotrophy. Burmanniaceae are a small family in Dioscoreales with approximately 13 genera and 130 species that include both autotrophic and fully mycoheterotrophic members, with multiple independent origins of full mycoheterotrophy (see Merckx et al., 2006 for a summary of the family's systematic history). Corsiaceae are a family of full mycoheterotrophs, comprising three small genera (*Arachnitis*, *Corsia*, and *Corsiopsis*; the latter is likely extinct; Zhang et al., 1999; Mennes et al., 2015). Historically, the family also has had a highly uncertain placement (see Mennes et al., 2015), although it was recently resolved as the sister group of Campynemataceae (Liliales) based on nuclear, mitochondrial, and whole plastome data (Mennes et al., 2015; Bodin et al., 2016). *Geosiris* (comprising three species) is the only fully mycoheterotrophic member of Iridaceae (Fay et al., 2000; Reeves et al., 2001; Goldblatt et al., 2008; Merckx et al., 2013b). Petrosaviaceae are a small family comprising the autotrophic *Japonolirion osense* and three species of fully mycoheterotrophic *Petrosavia* (Cameron et al., 2003). The family is the sole component of Petrosaviales, the sister group of all monocots except Alismatales and Acorales in most phylogenetic studies (e.g., Fuse and Tamura, 2000; Davis et al., 2004; Chase et al., 2006; Graham et al., 2006). Thismiaceae are a fully mycoheterotrophic family with ~five genera and 50 species. The family may be closely related to Burmanniaceae (e.g., APG, 2009), although recent molecular evidence suggests that it is instead most closely related to photosynthetic *Tacca* in Taccaceae (Merckx et al., 2009b, 2010). Finally, Triuridaceae are a fully mycoheterotrophic family with ~9 genera and 50 species. The family experienced an ancient loss of photosynthesis based on its inferred crown age (Mennes et al., 2013); late Cretaceous fossils are also known that may represent stem-lineage Triuridaceae (Gandolfo et al., 1998, 2002; Iles et al., 2015), but their trophic status is not known. The family's position in plant phylogeny has been highly uncertain. Molecular evidence from mitochondrial and nuclear genes strongly supports its membership in Pandanales (Mennes et al., 2013), an arrangement first reported by Chase et al. (2000) based on nuclear 18S rDNA. A recent plastome-based study placed Triuridaceae as the sister group of a clade comprising Cyclanthaceae and Pandanaceae with strong support across a broad variety of likelihood analyses (Lam et al., 2015).

Until recently, most of what is known about the broad picture of land-plant relationships came from pioneering phylogenetic surveys based on one to a few genes surveyed across a broad range of taxa. For example, many of the phylogenetic relationships underpinning Angiosperm Phylogeny Group (APG) classification schemes (APG, 1998, 2003, 2009) were based on two plastid genes (*atpB*, *rbcL*) and a single nuclear gene, 18S rDNA (e.g., Soltis et al., 2000). Individual genes can also have a strong phylogenetic signal concerning higher-order relationships; indeed, the broad outline of plant phylogeny is often apparent in single-gene analyses (e.g., Chase et al., 1993 for angiosperms using only *rbcL*; Givnish et al., 2005 for monocots using only *ndhF*). Analyses that included a few plastid genes have been used to infer the phylogenetic placement of several mycoheterotrophic lineages, including *Geosiris* (Fay et al., 2000) and *Petrosavia* (Petrosaviales; Fuse and Tamura, 2000; Cameron et al., 2003). However, the use of plastid markers in phylogenetic studies of heterotrophs has generally been regarded as problematic, reflecting expectations of elevated rates of evolution in retained genes (potentially problematic in phylogenetic inference if it leads to long-branch attraction, e.g., Felsenstein, 1978; Hendy and Penny, 1989) and the hypothesized loss of multiple photosynthetic or other genes or even entire plastid genomes (e.g., Cronquist, 1988, p. 467; Merckx et al., 2009b). Phylogenetic studies of mycoheterotrophs have therefore tended to focus on analyses of data sets comprising several mitochondrial and nuclear genes (e.g., Neyland and Hennigan, 2003; Merckx et al., 2006; Mennes et al., 2013). Despite these concerns, recent studies have validated the use of whole-plastid genomes to infer the placement of mycoheterotrophs (e.g., Logacheva et al., 2014; Lam et al., 2015; Mennes et al., 2015) by providing results congruent with other studies, often with strong branch support.

These recent results suggest that it would be useful to revisit the utility of plastid genes in broadly sampled, few-gene studies. The two photosynthesis-related genes *atpB* and *rbcL* have been among the most widely used plastid genes in higher-order plant phylogenetic studies. Both genes may have additional nonphotosynthetic roles (Bungard, 2004; Wicke et al., 2011, 2013), but with few exceptions (e.g., *Petrosavia*, Cameron et al., 2003; see also Logacheva et al., 2014) they have been lost or pseudogenized in fully heterotrophic lineages. The widely employed plastid *trnL-trnF* spacer region is also often missing in heterotrophic plastomes, in contrast to *matK*, another widely used gene that is commonly retained in heterotrophic lineages (see summaries in Barrett et al., 2014; Lam et al., 2015) and is often used in phylogenetic studies (e.g., Hilu et al., 2003). A portion of *matK* is also widely used as one of the core plant DNA barcoding markers (e.g., Hollingsworth et al., 2009). It is thus one of the few widely used plastid genes that may be suitable for inference of mycoheterotroph phylogeny. Obtaining plastid genes from heterotrophic plants can be difficult, and at least one study has suffered from the inclusion of erroneous plastid sequences from mycoheterotrophs (e.g., Kim et al., 2013; Kim, personal communication in Mennes et al., 2015): we suspect that PCR-based amplifications of plastid genes in heterotrophs may be prone to the recovery of contaminant sequences (cross-contaminants with other plant taxa) when authentic genes are not recoverable, for example, due to gene loss (also discussed later).

Here we assessed three plastid genes, *accD*, *clpP*, and *matK*, for their potential as phylogenetic markers in large-scale analyses and, in particular, to address whether they can help place mycoheterotrophs in monocot phylogeny with moderate to strong bootstrap support. The gene *accD* codes for the β-carboxyltransferase subunit of the acetyl-CoA carboxylase (ACCase), which is involved in fatty acid metabolism, and also regulates ACCase activity (Bungard, 2004); *clpP* codes for the catalytic subunit of plastid Clp protease (Shanklin et al., 1995; Wicke et al., 2011); *matK* encodes a maturase involved in splicing plastid group IIA plastid introns (Ems et al., 1995; Vogel et al., 1997; McNeal et al., 2009), although some group-IIA splicing may occur despite its loss (Delannoy et al., 2011). Thus, all three genes have functional roles that are not directly related to photosynthesis, and they are all typically retained in the plastid genomes of heterotrophic plants (holoparasites and full mycoheterotrophs; see Wicke et al., 2011; Barrett et al., 2014; Lam et al., 2015). They are therefore good candidate genes for plastid-based phylogenetic inferences that include these plants.

To test the utility of these three genes in broad-scale phylogenetic inference, we analyzed a monocot-wide data set in a range of analyses that consider all families containing mycoheterotrophic taxa. Several factors contributed to the complexity of this study. First, we had to assemble a phylogenetic framework for these three genes that represented a broad array of photosynthetic taxa, as two of the three genes (*accD* and *clpP*) had not previously been widely sampled in phylogenetic studies. Thus, we had to amplify and sequence these genes for most green taxa, which we aligned with those available from previous studies and new sequences from mycoheterotrophs. We used this alignment for phylogenetic inferences that included or excluded various sets of mycoheterotrophs. Second, we had to screen mycoheterotroph sequences for possible cross-contaminant sequences, which we occasionally encountered and removed from analysis. Third, genes recovered from individual species sometimes exhibited highly elevated substitution rates. We therefore based our phylogenetic inferences primarily on maximum-likelihood (ML) analyses, as likelihood methods are understood to be less prone to long-branch attraction problems than parsimony (e.g., Felsenstein, 1988; Yang, 1996; Huelsenbeck, 1997, 1998; Swofford et al., 2001; Yang and Rannala, 2012), and Bayesian methods may lead to inflated confidence in inferred phylogenetic relationships (e.g., Simmons et al., 2004; Yang and Rannala, 2005; Kolaczkowski and Thornton, 2009). We considered both unpartitioned and partitioned DNA substitution models to assess the degree to which the use of simple vs. complex models affects phylogenetic inferences; the partitioning scheme used accommodates differences in DNA substitution models (including model parameters) among genes and codon positions. We also included a parsimony analysis for comparison. Fourth, not all of the genes could be obtained experimentally from all mycoheterotrophic taxa that we surveyed, which inevitably led to somewhat patchy concatenated alignments.

Our analyses focused on non-orchid mycoheterotrophic monocots, those that interact with arbuscular mycorrhizal (AM) fungi (Leake, 2004; Waterman et al., 2013), although we also included data from recently published plastomes from several fully mycoheterotrophic orchids that associate with ectomycorrhizal fungi (i.e., *Rhizanthella gardneri*, Delannoy et al., 2011; *Neottia nidus-avis*, Logacheva et al., 2011; *Corallorhiza striata*, Barrett and Davis, 2012). Our study addressed two major objectives: (1) Does this few-gene data set allow placement of individual mycoheterotrophic taxa (species or families) with moderate to strong bootstrap support? (2) Are these placements congruent with recent results using other sources of molecular data, including nuclear and mitochondrial genes and recent whole-plastid genome studies? We also included a representative sampling of Burmanniaceae to compare with published phylogenetic inferences for this family based on nuclear and mitochondrial data (e.g., Caddick et al., 2002; Neyland, 2002; Davis et al., 2004; Merckx et al., 2006, 2009b). Finally, as plastid genes have not been reliably recovered from Thismiaceae to date, we were particularly interested in assessing whether we could recover any of the three plastid genes from this family for use in phylogenetic inference.

MATERIALS AND METHODS

DNA extraction, primer design, amplification, and sequencing—We obtained new DNA sequence data for 34 photosynthetic or fully mycoheterotrophic taxa from mycoheterotrophic lineages, and 61 additional photosynthetic taxa (59 of which are monocots; sources noted in Appendix S1, see Supplemental Data with the online version of this article). Total genomic DNAs were extracted from silica-dried material using a modified CTAB protocol (Doyle and Doyle, 1987; Rai et al., 2003) or were obtained from colleagues or DNA banks (Royal Botanic Gardens, Kew; Missouri Botanical Garden; SANBI, Kirstenbosch). We amplified and sequenced *accD*, *clpP*, and *matK*. DNA amplification and Sanger sequencing used previously published and new primers (Appendix S2, see online Supplemental Data). We designed additional novel primers by considering monocot-wide sequence alignments generated from available GenBank sequences, using visual inspection to identify conserved regions suitable for primer placement, and Oligo 7 (Rychlik, 2007) and Amplify 3x (Engels, 1993) to screen candidate primers based on their predicted success in DNA amplification (for criteria, see Graham and Olmstead, 2000). Exon boundaries for *clpP* were based on the complete plastid genome of *Dioscorea elephantipes* (GenBank accession NC_009601.1). DNA amplification and sequencing protocols generally followed Graham and Olmstead (2000), although in some cases we replaced *Taq* polymerase with Paq5000 (Agilent Technologies, Santa Clara, California [CA], USA). We purified amplification products using QIAquick PCR purification columns (Qiagen, Valencia, CA, USA) or ExoSAP-IT reagent (USB Corp., Cleveland, Ohio, USA), following manufacturer instructions, and used 1/26 reactions of BigDye Terminator v. 3.1 (Applied Biosystems, Foster City, CA, USA) to perform dideoxy sequencing reactions. We sequenced all regions at least twice, generally twice each in both forward and reverse directions. For some taxa, we were not able to sequence through the entire second intron in *clpP*, and so we represented these sequences in the alignment as two fragments. We obtained one or more genes for several taxa using assemblies of circular plastid genomes retrieved from genome skims using Illumina data, following Mennes et al. (2015); these genomes will be presented elsewhere.

Alignment construction—We performed base-calling and contig assembly using the program Sequencher 4.2.2. (Gene Codes Corp., Ann Arbor, Michigan, USA), and aligned finalized contigs using Se-Al 2.0a11 (Rambaut, 2002) and the criteria outlined by Graham et al. (2000); Kelchner (2000); and Simmons and Ochoterena (2000). We added new sequences for the three genes to those retrieved from published plastid genomes of 53 additional taxa (see Appendix S1). We offset hard-to-align noncoding regions (frequently limited to individual taxa) in the alignment in a staggered manner (Graham et al., 2006; Saarela and Graham, 2010). Unamplifiable genes, which in some cases are known to be lost (based on plastid genome evidence; Appendix S1), were treated as missing data. The final matrix is a concatenation of all three genes: *accD* and *matK* are exclusively protein-coding, and *clpP* includes short exon sequences and two introns (see Appendix S2). The concatenated matrix is 9595 bp long and includes representatives from 139 monocot taxa, spanning 11 monocot orders, including representatives from all seven monocot mycoheterotrophic families (38 taxa), and 11 outgroup species from 10 angiosperm families outside monocots (Appendix S1).

We identified several cases of possible contaminant sequences among individual genes recovered from mycoheterotrophs (see below), initially by using BLAST (Altschul et al., 1990). We suspect that these derive from residual nontarget DNA contamination that becomes apparent when the main amplification fails due to gene loss or rapid gene evolution away from the targeted priming sequences. We identified probable contaminants by examining maximum-likelihood (ML) trees inferred for each locus individually, to check for terminal taxa that are distantly related to congeneric or confamilial taxa. Typically these had short connecting branches to the nearest green taxon. One example is a likely contaminant *matK* sequence obtained from a *Thismia aseroe* extract (Appendix S3a, see online Supplemental Data) that shows 97% similarity to *Moraea riparia* (Iridaceae; GenBank accession JX903631.1). The behavior of this sequence contrasts with what we believe to be genuine Thismiaceae *accD* sequences (259–314 bp sequenced portions, compared with 511-bp portions sequenced for other taxa), found on much longer branches (these also often placed in unusual positions, see below), that have BLAST scores of 87–88% to other monocots, eudicots, and other angiosperms. Other putative contaminants that placed outside monocots were not so straightforward to place phylogenetically because of limited taxon sampling in other angiosperms here. An example of this is a *matK* sequence retrieved from a *Geomitra clavigera* extract, which we identified as a probable contaminant based on close BLAST-based similarity to a eudicot (99% to *Capparis spinosa*, Capparaceae; GenBank accession AY491650.1; in a phylogenetic analysis, this sequence grouped with two eudicots included as outgroups, Appendix S3b). We excluded all candidate contaminant sequences identified in this way, but tentatively included other genes from the same sample that passed our phylogenetic screen (see Appendix S1; note that none of the other sequences included in final alignments from these samples showed obvious double peaks, so contaminants are likely present at low concentration in the original DNA extracts).

Phylogenetic analysis—All subsequent analyses considered the final concatenated three-gene DNA sequence alignment, which is publicly available at figshare.com (doi:10.6084/m9.figshare.2062158).

We conducted initial analyses using parsimony and likelihood with all green and mycoheterotrophic taxa included. For the parsimony search, we used the program PAUP* v. 4.0b10 (Swofford, 2003), with tree-bisection-reconnection branch swapping and 1000 random stepwise addition replicates, holding 100 trees at each step, and otherwise used default settings in the search for shortest trees. For maximum-likelihood (ML) analysis we used the program RAxML v. 7.4.2 (Stamatakis, 2006) with a graphical interface (Silvestro and Michalak, 2012), considering partitioned and unpartitioned versions of the data, see below. For these and all subsequent likelihood analyses, we ran 20 independent heuristic searches using different starting points.

We then performed individual likelihood analyses considering each mycoheterotrophic family (or taxon) separately to try to minimize the potential for attraction between distantly related taxa, using the search method outlined above. For Burmanniaceae, we ran an additional analysis that included only green species for the family (i.e., likely photosynthetic taxa; Merckx et al., 2006), and for Thismiaceae, we also analyzed each species (three in total) separately, as we only retrieved a portion of *accD* for these taxa. We also ran an analysis including only photosynthetic angiosperms (considering photosynthetic mycoheterotrophs, green orchids were included and green Burmanniaceae excluded) to investigate the general utility of the three genes in inference of monocot higher-order relationships.

We ran two variant ML analyses in all cases, one considering the data unpartitioned, and another with the data partitioned by gene and codon position. For the latter, we defined 14 initial data partitions based on the three codon positions for coding regions (considering the two exons of *clpP* separately) and two introns (in *clpP*). We then used the program PartitionFinder v. 1.1.1 (Lanfear et al., 2012) to assess which of the initial partitions had significantly different DNA substitution models or model parameters in each case, using the Bayesian information criterion (BIC; Schwarz, 1978; Sullivan and Joyce, 2005) and the strict hierarchical clustering algorithm. We repeated these tests for each distinct likelihood analysis and used the final partitioning schemes for each ML analysis (summarized in online Appendix S4). We applied the GTR+G model to all data partitions, as this model or close variants were found for all unpartitioned data sets, and for most individual data partitions in partitioned analyses (Appendix S4; the GTR+G+I model was found in a few cases, but the I parameter for invariant sites may be accommodated by the gamma parameter G; see Yang, 2006).

We assessed branch support using bootstrap analysis (Felsenstein, 1985). For the likelihood analyses, we ran 500 bootstrap replicates using the rapid bootstrap option in RAxML, but considering the same partitioning schemes and models of DNA substitution described above. For the parsimony bootstrap analysis, we used 1000 bootstrap replicates, with 100 random addition replicates per bootstrap replicate, and otherwise used default settings. We considered 90% and above to be "strongly supported" (or "well supported"), 70–89% to be "moderately supported" and <70% to be "weakly supported" (Zgurski et al., 2008).

Constraint tests of monophyly—We tested for the monophyly of two groups of interest that were each not recovered as monophyletic in any shortest trees (see below): (1) a putative clade comprising Burmanniaceae and Thismiaceae corresponding to the circumscription of Burmanniaceae s.l. in recent versions of the Angiosperm Phylogeny Group classification system (APG, 2003, 2009); (2) a clade comprising nongreen Burmanniaceae, as these species are not expected to be grouped together in a clade, given the multiple independent losses of photosynthesis predicted in the family (Merckx et al., 2006). In both cases, we were interested in determining whether we had sufficient power to reject a hypothesis that placed the constrained taxa together. Each constraint analysis excluded other mycoheterotrophs that were not relevant to the hypothesis. We found the shortest likelihood trees that satisfied monophyly constraints that we set up for RAxML (using unpartitioned ML analysis). We then compared the resulting tree sets using the approximately unbiased (AU) (Shimodaira, 2002) and Shimodaira-Hasegawa (SH) tests in the program CONSEL (Shimodaira and Hasegawa, 2001), using site-likelihoods from unpartitioned ML analysis, to ask whether the constrained trees were significantly worse explanations of the data than the best (unconstrained) tree.

Characterizing rate elevation in heterotrophic lineages—We characterized relative differences in overall substitution rates in mycoheterotrophic and green lineages in a Bayesian framework, using the program BEAST v. 1.8.2 (Drummond et al., 2012). Our focus was on characterizing relative rate variation across lineages, and so we fixed the input tree topology based on the best likelihood tree recovered from the partitioned analysis that included all taxa (see Fig. 2), with the exception that we constrained Triuridaceae, represented by *Sciaphila*, to be monophyletic (see below). We specified a single GTR+G nucleotide substitution model across sites, a random local-clock model with a fixed mean rate of 1.0 substitution per site, and a Yule speciation model, and otherwise used default settings. We ran BEAST for a combined 800 million generations across 20 separate analyses, sampling trees every 1000 generations, assessing convergence using the program Tracer v. 1.6 (Rambaut et al., 2014). All parameters had final effective sample size (ESS) values of at least 200. We used the programs LogCombiner v. 1.8.2. and TreeAnnotator v.1.6.2 (Drummond et al., 2012) to combine trees from individual analyses, discarding the first 25% of sampled trees as burn-in and resampling at a lower frequency to yield 10,000 final trees. The combined tree was visualized using the program FigTree v. 1.31 (Rambaut, 2006).

RESULTS

***Green monocot phylogeny inferred from* accD, clpP, *and* matK**—The placements of autotrophic monocot families, orders and higher-level taxa in the ML analysis of autotrophic taxa (Fig. 1) are congruent with previous monocot-wide studies using more genes

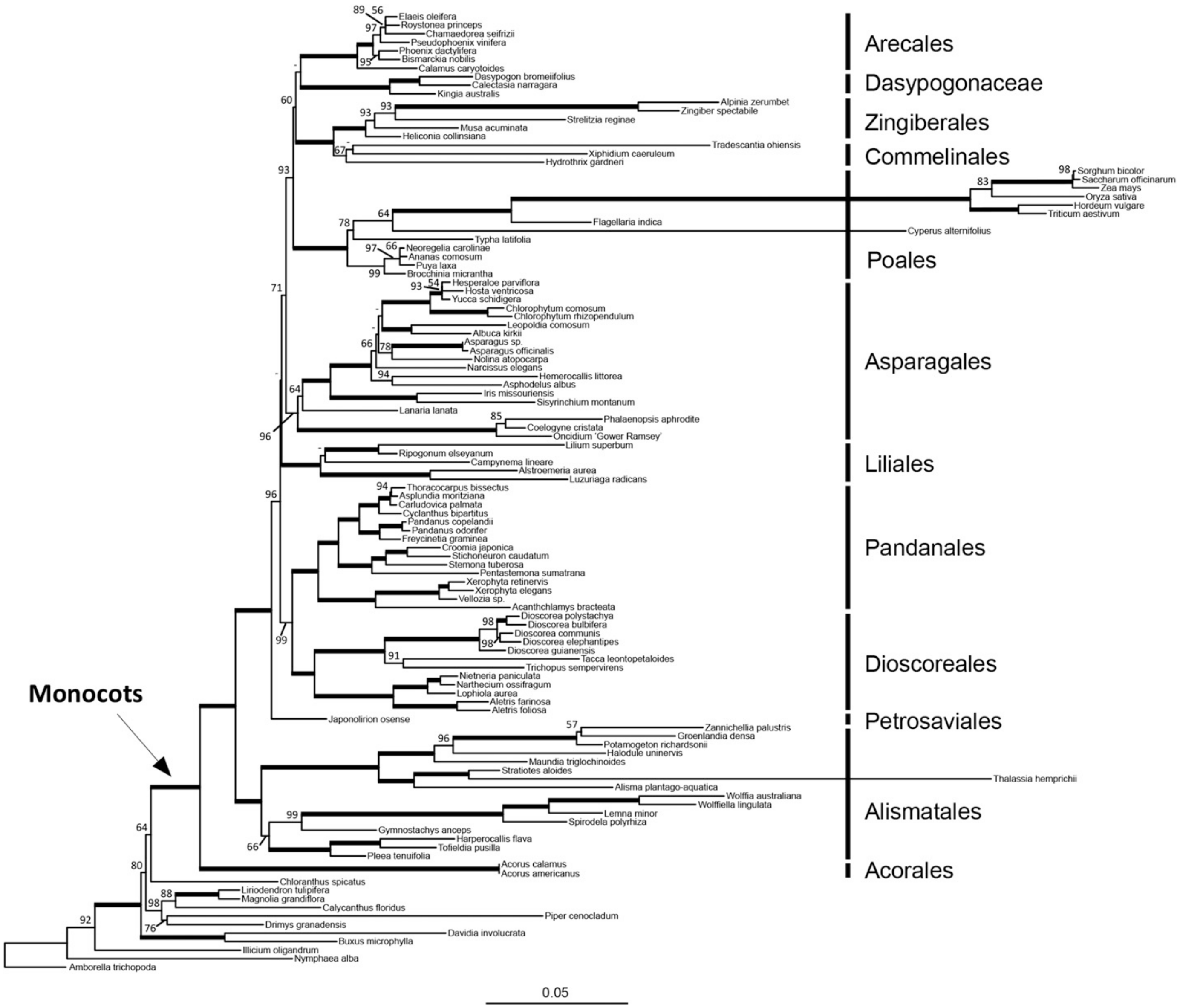


FIGURE 1 Three-gene phylogeny of photosynthetic monocots based on a partitioned three-gene maximum-likelihood analysis (for *accD*, *clpP*, and *matK*). The analysis includes green orchids but excludes green members of Burmanniaceae (see Materials and Methods section for details). Branches with 100% bootstrap support are shown as thick lines; other bootstrap values are indicated beside branches (<50% support indicated with a short dash). The scale bar indicates estimated number of substitutions per site.

(e.g., Graham et al., 2006; Hertweck et al., 2015). Most deep branches in the monocots were also strongly supported, with only minor exceptions (e.g., the placement of Dasypogonaceae, a sister-group relationship between Liliales and Asparagales-commelinids, both with <50% bootstrap support). There were no substantial differences in phylogenetic relationships of autotrophic monocots and their support values between the partitioned analysis and the unpartitioned analysis for this taxon set (data not shown).

Phylogenetic placement of mycoheterotrophs in monocot phylogeny—*(1) Placements with all mycoheterotrophs included*—Initial analyses (parsimony and likelihood) included all green and mycoheterotrophic lineages together. Parsimony trees (most parsimonious trees) grouped all or most of the fast-evolving lineages in a single large clade that may result from long-branch attraction, marked with an arrow in online Appendix S5 (all shortest trees either recovered this clade or had a slightly smaller version lacking Thismiaceae; see below on rates of evolution in mycoheterotrophs). This clade was poorly supported and had poorly supported and variable internal structure across shortest trees (note that black dots in this figure indicate branches that collapse in a strict consensus of the most-parsimonious trees). In contrast, full mycoheterotrophs that terminated shorter branches (i.e., *Geosiris*, orchids, *Petrosavia*, *Corsia*) placed with moderate to strong support in the parsimony analysis in Iridaceae, Orchidaceae, Petrosaviaceae, and Liliales, respectively. Corsiaceae were not recovered as monophyletic in this analysis, as *Arachnitis* grouped within the fast clade (Appendix S5).

In the likelihood analyses that included all mycoheterotrophs, fully mycoheterotrophic lineages on shorter branches again placed with moderate to strong likelihood bootstrap support in the positions observed in the parsimony analysis (see Fig. 2 and online Appendices S5, S6 for *Geosiris*, orchids, *Petrosavia*). However, multiple rapidly evolving lineages (see below) that grouped together in the parsimony analyses were instead inferred to be in dispersed positions

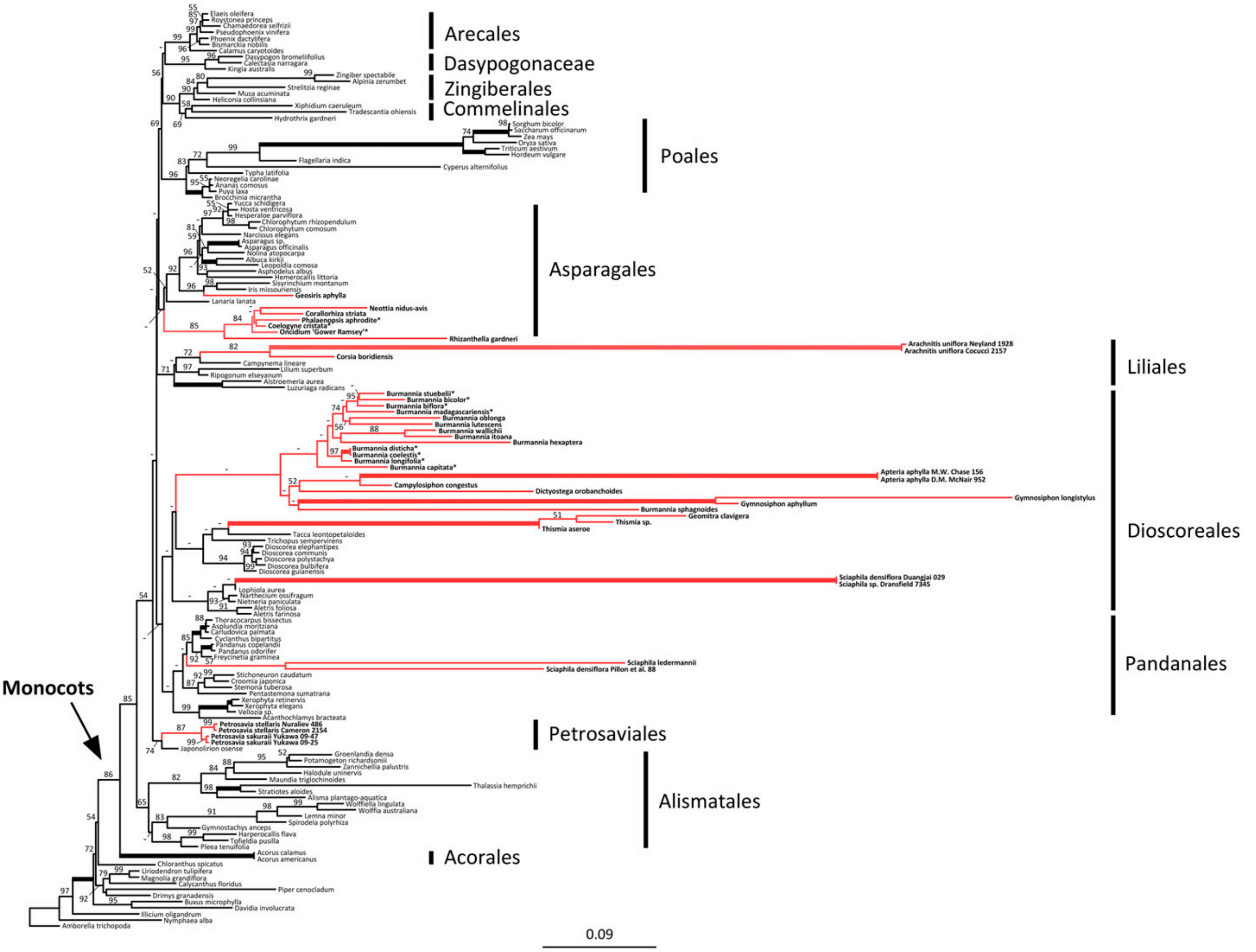


FIGURE 2 Three-gene phylogeny of photosynthetic and fully mycoheterotrophic monocots (all lineages considered simultaneously) based on a partitioned three-gene maximum-likelihood analysis (for *accD*, *clpP*, and *matK*). Lineages with mycoheterotrophs are indicated in red (asterisks indicate photosynthetic taxa in Burmanniaceae and Orchidaceae; the remainder are full mycoheterotrophs). Branches with 100% bootstrap support are shown as thick lines; other bootstrap values are indicated beside branches (<50% support indicated with a short dash). The scale bar indicates estimated number of substitutions per site.

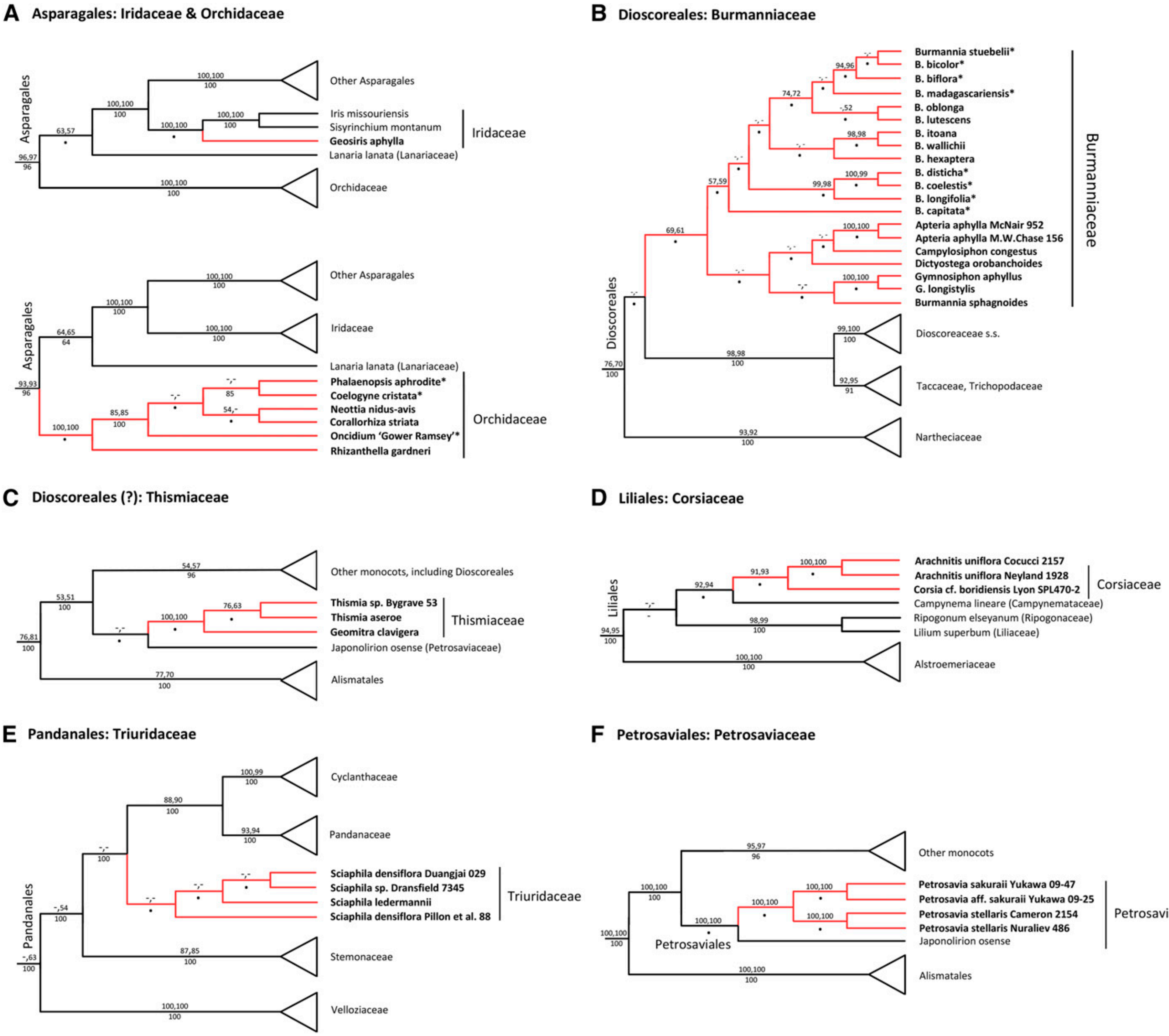
in the likelihood analyses. Specifically, Burmanniaceae (the monophyly of which was poorly supported) were poorly supported as the sister group of a clade that included Taccaceae, Trichopodaceae, Thismiaceae, and Disocoreaceae, with Thismiaceae then the sister group of *Tacca* (Taccaceae). The latter arrangement was poorly supported, but the monophyly of Thismiaceae was well supported. *Sciaphila* (Triuridaceae) was divided between two locations, with one pair of taxa as the sister group of Cyclanthaceae and Pandanaceae in Pandanales, and the other nested in Nartheciaceae (Dioscoreales) as the sister group of *Lophiola*. This split placement was seen in both partitioned and unpartitioned ML analyses (Fig. 2; Appendix S6). The monophyly of Corsiaceae was well supported (i.e., *Arachnitis* no longer placed in the fast clade), and the entire family was supported as the sister group of Campynemataceae with weak to moderate support in both likelihood analyses (Fig. 2; Appendix S6). The inclusion of full mycoheterotrophs tended to depress likelihood bootstrap support for branches neighboring all of the mycoheterotrophic families, compared with analyses that included only green taxa. Neighboring branches often experienced at least 10% worse bootstrap support and often had substantially larger drops in support (e.g., for the monophyly of Asparagales and Dioscoreales, cf. Figs. 1 and 2).

(2) Placements with mycoheterotrophic taxa considered individually—In general, the likelihood analyses that examined each mycoheterotroph family separately recovered placements of mycoheterotrophic taxa consistent with the analyses that included them all simultaneously (cf. Figs. 2 and 3; online Appendix S7a–e), with exceptions outlined below for Thismiaceae and Triuridaceae. These local placements were also generally well supported (Fig. 3), a contrast with the likelihood analyses that included all mycoheterotrophic taxa simultaneously (Fig. 2, Appendix S6); there was less reduction in bootstrap support values for branches neighboring all of the mycoheterotrophic families when each of these was added individually. For example, the monophyly of both Asparagales and Dioscoreales was now well supported (cf. Figs. 1–3). There was now strong support for the monophyly of Corsiaceae, Iridaceae, Orchidaceae, and Petrosaviaceae (Fig. 3). Support for the monophyly of Burmanniaceae when other mycoheterotrophic families were excluded was also higher than when they were included, but was still weak (61–69% support, compared to <50% support; see Figs. 2, 3B; Appendix S6). However, when only photosynthetic Burmanniaceae were included, the bootstrap support for the family's monophyly improved substantially (to 100% in partitioned and unpartitioned ML analyses, see inset subtree in Fig. 4). The membership of Burmanniaceae in Dioscoreales was also well supported in this case, although its local position in the order still lacked strong support (Fig. 4). We could not reject the existence of a clade comprising Burmanniaceae and Thismiaceae (AU and SH tests comparing trees from best unconstrained vs. constrained ML analyses that included both families: $P = 0.104$ and 0.125, respectively; the unconstrained tree placed Burmanniaceae as the sister group of other Dioscoreales, and Thismiaceae as the sister group of *Tacca,* data not shown).

The behavior of Triuridaceae and Thismiaceae (in likelihood analyses, with each family included individually) was sensitive to the details of the analysis performed. In the partitioned likelihood analysis that included only *Sciaphila* (our representative of Triuridaceae) and that excluded other mycoheterotrophs, Triuridaceae were inferred to be monophyletic and they placed as the sister group of Cyclanthaceae-Pandanaceae (in Pandanales), all with weak support (Fig. 3E; Appendix S7f). The unpartitioned version of this analysis recovered two isolated lineages of *Sciaphila* in Pandanales and Dioscoreales (Appendix S7g) similar to the likelihood analyses that included all mycoheterotrophs (Fig. 2; Appendix S6), and again with weak support (Fig. 3E). Thismiaceae were unexpectedly inferred to be the sister group of *Japonolirion* in the partitioned analysis (Fig. 3C; Appendix S7h), but not in the unpartitioned analysis, where the family placed with Taccaceae and Trichopodaceae (Appendix S7i); note that Thismiaceae species are represented here only by *accD* (all analyses presented are concatenated three-gene analyses). Separate analyses that included individual species from the family placed the three surveyed species in divergent places (Fig. 5), but generally close to *Tacca* (Taccaceae) and/or *Trichopus* (Trichopodaceae) (i.e., *Geomitra clavigera* as the sister group of *Tacca* and *Trichopus* in both partitioned and unpartitioned likelihood analyses, *Thismia aseroe* as the sister group of *Tacca* for partitioned and unpartitioned analysis; *Thismia* sp. as the sister group of Maundiaceae in Alismatales in unpartitioned likelihood analysis, but as the sister group of *Tacca* in partitioned analysis; Fig. 5).

Relationships in Burmanniaceae—The relative positions of the photosynthetic members of Burmanniaceae were largely consistent among likelihood analyses. Various fully mycoheterotrophic lineages were recovered as nested among the photosynthetic taxa, but for the most part with poor support for the relative arrangements of green and nongreen taxa (Fig. 3B). Well-supported relationships in Burmanniaceae include a small clade comprising *B. bicolor*, *B. biflora*, and *B. stuebelii*, a sister-group relationship between *Burmannia itoana* and *B. wallichii*, between *B. coelestis* and *B. disticha*, between these two taxa and *B. longifolia*, and between *Gymnosiphon aphyllus* and *G. longistylus* (Fig. 3B). The intermingling of fully mycoheterotrophic members of Burmanniaceae with photosynthetic members of the family implies multiple evolutionary losses of photosynthesis (Fig. 3B; Appendix S7c). Consistent with a hypothesis of multiple losses of photosynthesis, the AU and SH tests rejected the existence of a constrained clade that comprises only nonphotosynthetic taxa of Burmanniaceae ($P < 0.02$ in both cases). When only green Burmanniaceae were included, the relative placement of *B. capitata* was different (but not well supported; cf. Figs. 3B, 4; Appendix S7c), and the sister-group relationship of *B. madagascariensis* and the clade comprising *B. bicolor*, *B. biflora*, and *B. stuebelii* was recovered with improved support (98–99%; Fig. 4).

Rate elevation in mycoheterotrophic monocots—Relative differences in the overall substitution rate are summarized in Fig. 6 (based on online Appendix S8). Three major rate bands are shown, with the highest rates (in red) at least twice as fast as the fastest slow ones (thin black lines); the thick black line is an intermediate rate. Many fully mycoheterotrophic lineages have intermediate to fast rates, although Petrosaviaceae, *Geosiris* (Iridaceae), *Corallorhiza striata* (Orchidaceae), and several fully mycoheterotrophic Burmanniaceae are exceptions (terminal photosynthetic lineages of Burmanniaceae are also in the lowest rate band; Appendix S8). Note that the rate band cut-offs used here (chosen to emphasize the most rapidly evolving lineages) include a broad range of rates (e.g., the lowest rate brand includes a nearly 9-fold range of rates, see Appendix S8 for more details). The most elevated rates (thick lines in Fig. 6) include several fully mycoheterotrophic lineages (*Arachnitis*



A Asparagales: Iridaceae & Orchidaceae

B Dioscoreales: Burmanniaceae

C Dioscoreales (?): Thismiaceae

D Liliales: Corsiaceae

E Pandanales: Triuridaceae

F Petrosaviales: Petrosaviaceae

FIGURE 3 Summary of bootstrap support for local placements of monocot mycoheterotrophs in separate analysis of each family, based on partitioned maximum-likelihood (ML) analysis of *accD*, *clpP*, and *matK* (see Appendix S7a–f, h for branch lengths). (A) Iridaceae and Orchidaceae, (B) Burmanniaceae s.s., (C) Thismiaceae, (D) Corsiaceae, (E) Triuridaceae, (F) Petrosaviaceae. The ordinal placement of individual families according to APG (2009) is noted beside each figure (phylogenetic analyses for Iridaceae and Orchidaceae were done separately). Major clades collapsed for simplicity. Bootstrap support values noted beside branches: values above branches are with the mycoheterotrophic taxon included (left, partitioned ML analysis; right, unpartitioned ML analysis); those below branches are with the taxon excluded. Bootstrap support values <50% are indicated with a short dash; a dot indicates that the support value is not applicable (reflecting fewer taxa in analyses with photosynthetic taxa only). Lineages with mycoheterotrophs are indicated in red (asterisks indicate photosynthetic taxa in Burmanniaceae and Orchidaceae; the remainder are full mycoheterotrophs). Voucher names are included for species identified only to genus or when two samples were included for a species (see also Appendix S1).

in Corsiaceae, several lineages of Burmanniaceae, Orchidaceae, Thismiaceae, and Triuridaceae). Several photosynthetic lineages in Alismatales, Poales, and relatives are also rapidly evolving.

DISCUSSION

Using plastid genes to place mycoheterotrophic monocots—Compared with photosynthetic taxa, plastid genomes retrieved from heterotrophic lineages can be both rapidly evolving and reduced in terms of size and gene content, sometimes exceptionally so (e.g., Lam et al., 2015; Mennes et al., 2015; Bellot and Renner, 2016). Our few-gene approach thus provides general insights into the suitability of highly reduced, patchily sampled, rapidly evolving genomes for inferring the phylogenetic placement of heterotrophic plant lineages. Several key problems had to be addressed to do so. First, we lacked a phylogenetic framework (large-scale alignment of commonly retained genes) for photosynthetic taxa to help place the

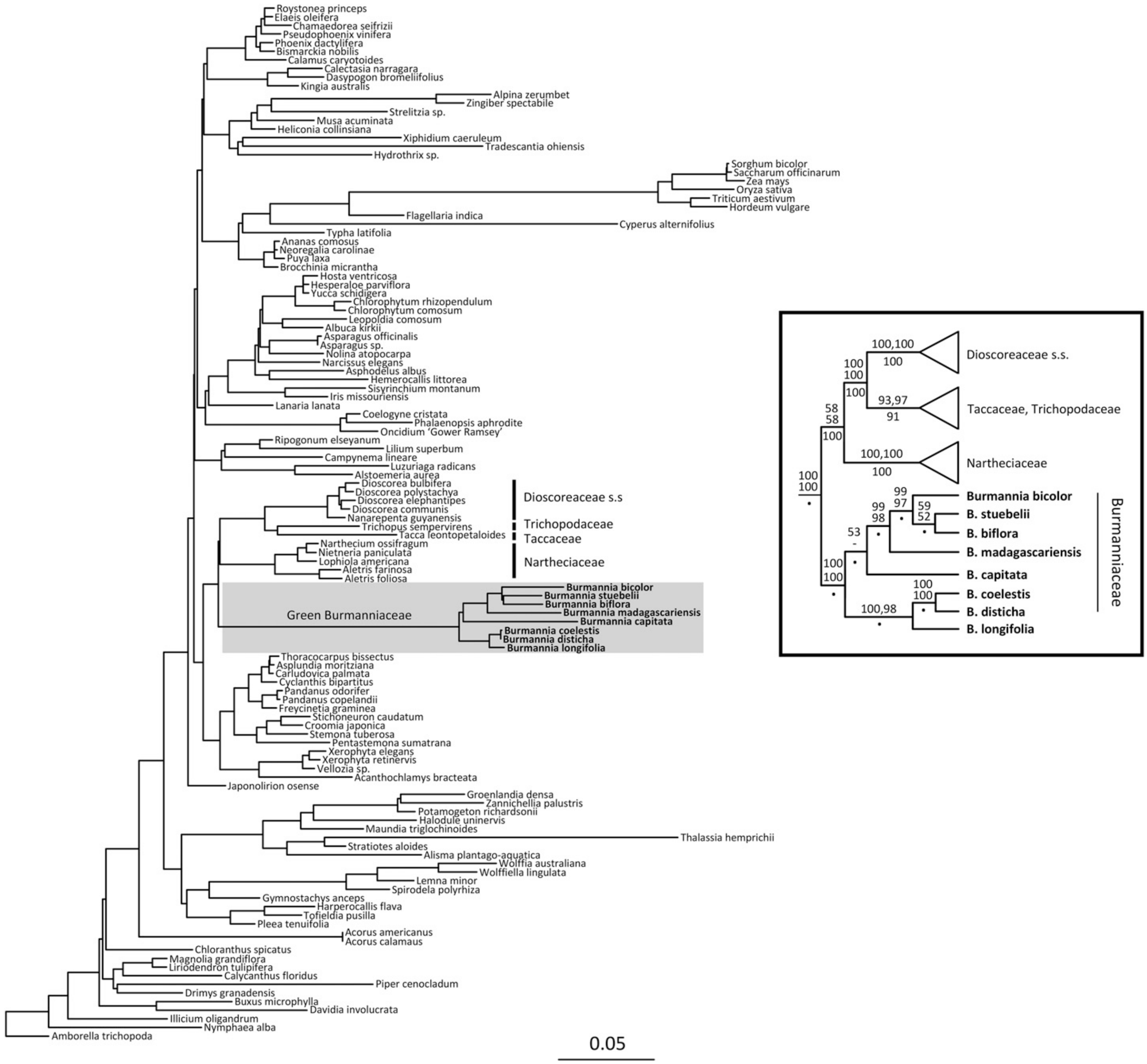


FIGURE 4 Phylogenetic placement of green Burmanniaceae in monocot-wide phylogeny, based on partitioned three-gene maximum-likelihood (ML) analysis. The phylogram indicates the placement of green Burmanniaceae (following Merckx et al., 2006; see Materials and Methods section for details). Families within Dioscoreales are indicated. The inset figure summarizes bootstrap support values within Dioscoreales (above branches: left, partitioned ML analysis; right, unpartitioned ML analysis. Below branches: Burmanniaceae excluded from analysis). Bootstrap support values <50% are indicated with a short dash; a dot indicates that the support value is not applicable (reflecting fewer taxa in analyses with Burmanniaceae excluded). The scale shows the estimated number of substitutions per site.

heterotrophic lineages. We addressed this by constructing a large-scale alignment of three commonly retained genes in heterotrophs (*accD*, *clpP*, *matK*) from photosynthetic monocots, which we sampled most heavily in the clades (orders) thought to have given rise to heterotrophic taxa. The higher-order relationships of photosynthetic monocots inferred using the three-gene data set are congruent with recent phylogenetic studies that employed a wide variety of taxonomic and gene samplings (e.g., Chase et al., 2006; Graham et al., 2006; Givnish et al., 2010; Soltis et al., 2011), and generally well-supported (Fig. 1). This congruence supports the general utility of using these genes for making inferences about high-order monocot relationships, at least for photosynthetic taxa.

The often highly elevated rates of evolution of retained plastid genes in heterotrophic lineages may be problematic for phylogenetic inference. Although our analytical understanding of long-branch effects is still quite limited (Parks and Goldman, 2014),

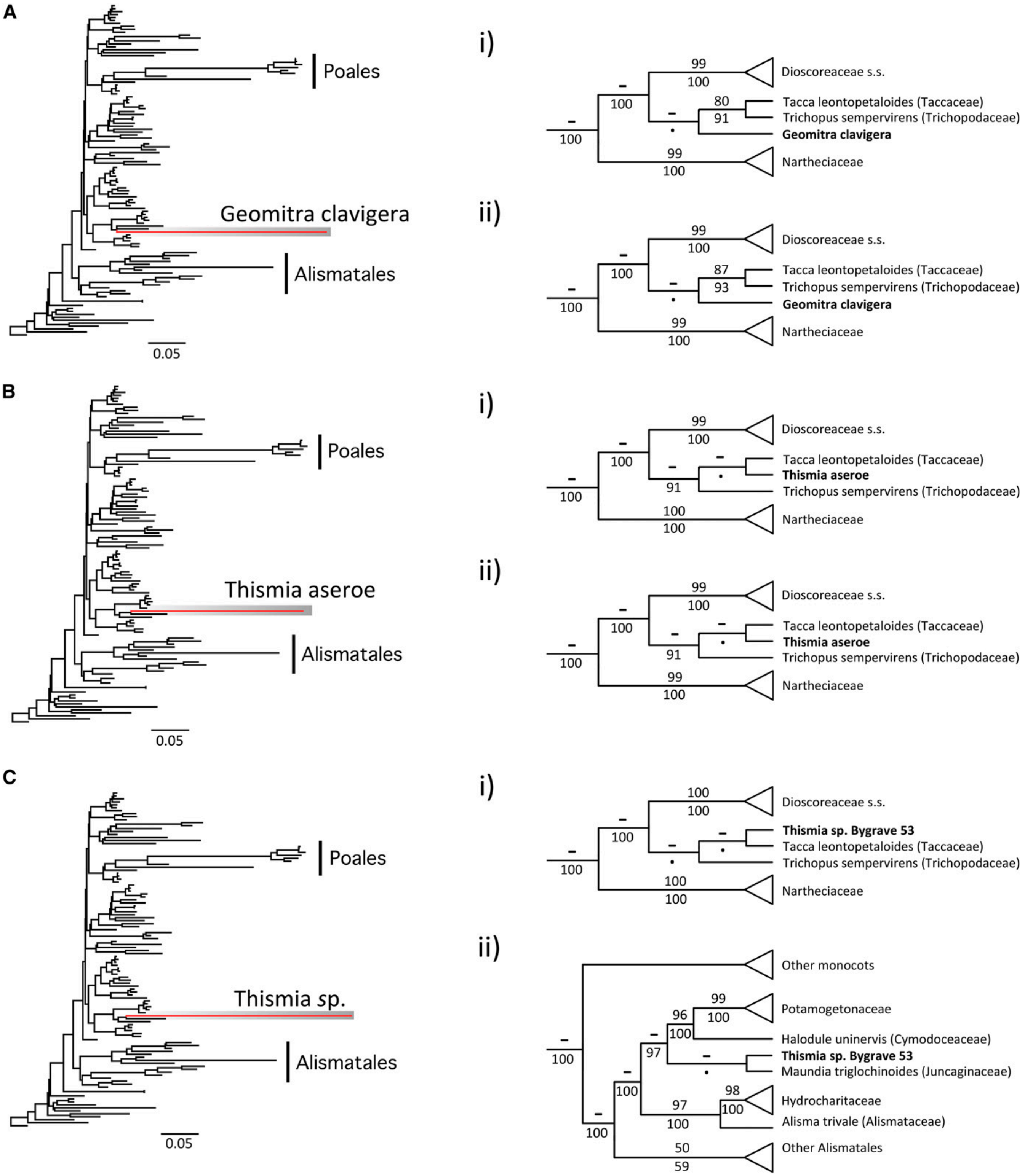


FIGURE 5 Phylogenetic placements inferred for individual species of Thismiaceae based on maximum-likelihood (ML) analysis (three-gene analyses that include only *accD* for this family, the only plastid gene recovered for it). Summary of results for (A) *Geomitra clavigera*, (B) *Thismia aseroe*, (C) *Thismia* sp. Left panel, phylograms of shortest trees for partitioned ML analysis (scale bars indicate estimated number of substitutions per site; mycoheterotrophic lineages indicated with blue branches). Right panel, summary of bootstrap support for placements of individual species based on: (i) partitioned ML analysis; (ii) unpartitioned ML analysis; values above branches are with Thismiaceae sequences included, those below are with them excluded. Bootstrap support values <50% are indicated with a short dash; a dot indicates that the support value is not applicable (reflecting fewer taxa in analyses with photosynthetic taxa only). The scale bar (A–C) indicates the estimated number of substitutions per site.

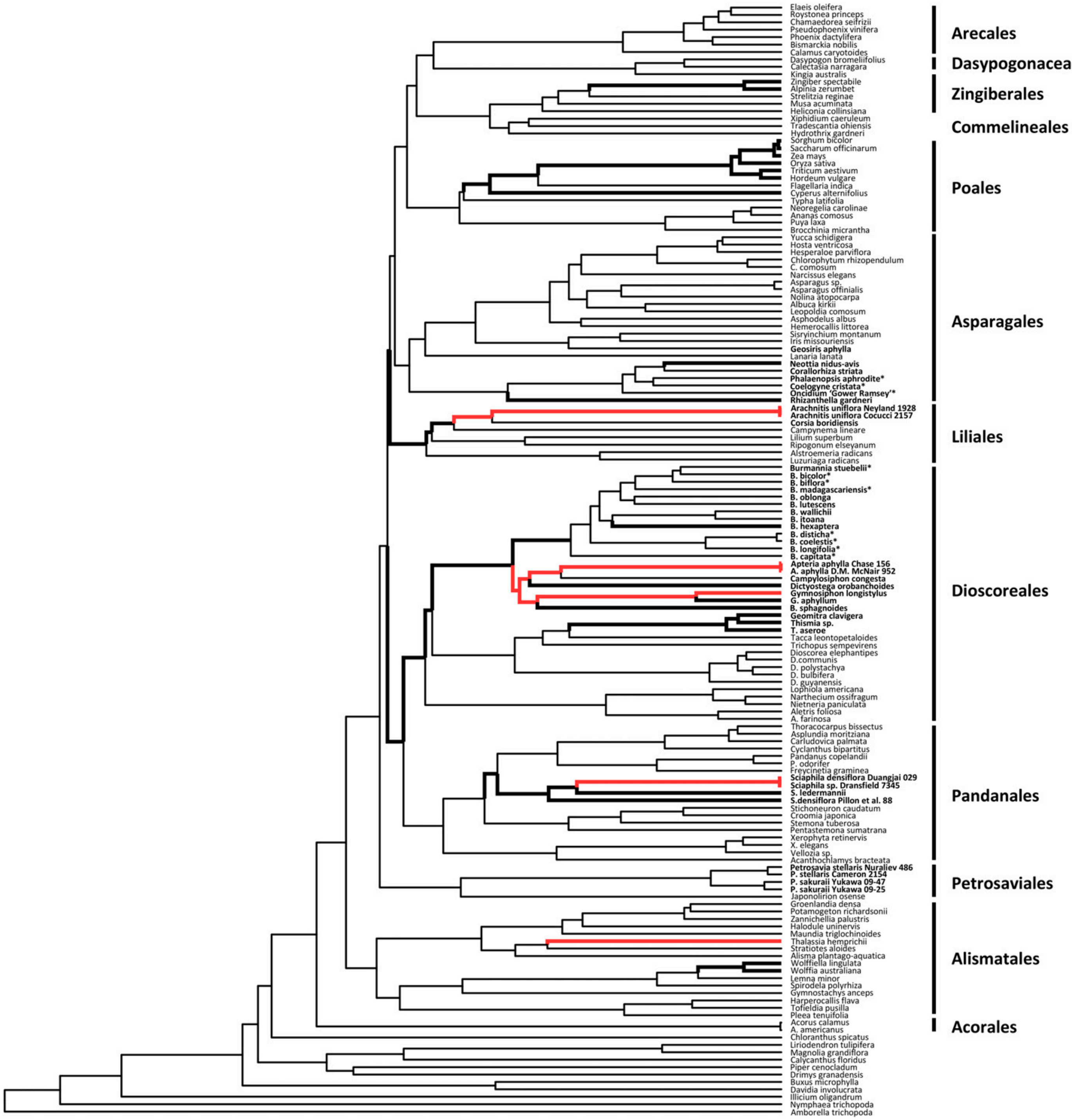


FIGURE 6 Relative substitution rates among green and fully mycoheterotrophic monocot lineages based on Bayesian analyses of a three-gene plastid (*accD*, *clpP*, and *matK*) data set, with a random-local-clock model and a constrained topology (see Materials and Methods for details). Thin branches indicate relative rates of <2.0 substitutions per site and below; thicker branches indicate intermediate relative rates of 2.0 to <4.0; thick red branches indicate those ≥4.0 (see online Appendix S8 for full details). Green species in Burmanniaceae and Orchidaceae are indicated with an asterisk (*); the others in these families are full mycoheterotrophs.

model-based methods like maximum likelihood are thought to be less sensitive than parsimony to problematic long branches (see also Felsenstein, 1978, 1988; Yang, 1996; Huelsenbeck, 1997, 1998; Swofford et al., 2001; Yang and Rannala, 2012). This lower sensitivity appears to be the case here, as our parsimony analysis depicted a collection of rapidly evolving mycoheterotrophic taxa as part of a poorly supported "fast" clade (Appendix S5) that is taxonomically highly heterogeneous (e.g., APG, 2009). This result is a strong contrast with the likelihood analyses (cf. Fig. 2 and Appendix S6). Despite substantial rate elevation, likelihood analyses placed mycoheterotrophic lineages in phylogenetic positions consistent with studies using nuclear and mitochondrial genes (e.g., Merckx et al., 2006, 2009b; Mennes et al., 2013, 2015). Thorough taxon sampling can help avoid long-branch artifacts (e.g., Pollock et al., 2002; Soltis and Soltis, 2004; Hedtke et al., 2006; Heath et al., 2008); however, adding taxa can introduce additional problems if the additional taxa are also on long branches (e.g., Kim, 1996; Hillis, 1998). The latter effect may also explain why including all mycoheterotrophic taxa simultaneously leads to substantially larger reductions in branch support for the branches neighboring heterotrophic lineages (those that were well supported in the analysis of green taxa only), than occurred when we included each mycoheterotrophic family individually (cf. Figs. 1–3).

Finally, missing data can present a challenge for phylogenetic inference and may be problematic in some cases here, as one or two of the three genes we considered were sometimes lost from the genome or were not readily recoverable for several taxa. Incompletely sampled taxa may still tend to improve the accuracy of phylogenetic inference when taxon sampling is otherwise limited (e.g., Wiens, 2006; Wiens and Tiu, 2012). However, the combination of missing data and long branches may nonetheless prove to be especially severe (e.g., Wiens, 2005, 2006), which is potentially problematic here for several patchily sampled taxa with elevated substitution rates (e.g., Figs. 2, 5; Appendix S1).

Utility and limits of the current approach—How well do our phylogenetic markers perform in likelihood analyses? In general, we did not see major differences in levels of bootstrap support values between unpartitioned and partitioned likelihood analyses for these taxa across the different data sets that we analyzed (Fig. 2 vs. Appendix S6; see also partitioned and unpartitioned comparisons in Fig. 3), supporting the idea that our results are not sensitive to whether the data are partitioned or not (partitioned likelihood analyses attempt to account for DNA substitution model or model parameter differences among different regions). This finding is also consistent with those of Lam et al. (2015) in their whole plastid-genome study of Triuridaceae. Because long branches in different mycoheterotroph families likely interfere with well-supported placement of individual taxa when considered simultaneously (Fig. 2), we focus the remaining discussion on the separate likelihood analyses of mycoheterotrophs (i.e., each family considered individually), summarized in Figs. 3–5.

We placed the majority of mycoheterotrophic taxa in monocot phylogeny with strong support, at least regarding their family-level placements (Fig. 3A, D, F). Our results are also consistent with current understanding of their phylogenetic placements based on other data sets (i.e., *Geosiris*, Fay et al., 2000; mycoheterotrophic Orchidaceae, Ruhfel et al., 2014; Corsiaceae, Mennes et al., 2015; Bodin et al., 2016; Petrosaviaceae, Cameron et al., 2003; Logacheva et al., 2014). The position of Corsiaceae has been unclear until very recently. The family was strongly supported here as the sister group of Campynemataceae in Liliales, which is consistent with recent whole plastome data and other data for the family (Mennes et al., 2015; Bodin et al., 2016). *Arachnitis* (Corsiaceae) has a particularly long branch due to highly elevated substitution rate (Figs. 2, 6; Appendix S7d). However, the other genus in the family, *Corsia*, has a moderately elevated rate of evolution (Fig. 6) and has retained all three surveyed genes, which may account for the family's clear and well-supported placement here in likelihood analyses (Fig. 3D). Our results do not support the findings of Neyland and Hennigan (2003) that Corsiaceae are polyphyletic, recovered in their analysis of a nuclear 26S rDNA data set (a probable artifact due to limited taxon sampling and their use of parsimony, see Mennes et al., 2015).

Three families were less confidently placed in monocot phylogeny by our three-gene data set: Burmanniaceae, Thismiaceae, and Triuridaceae (Figs. 3–5). With regards to the first, Burmanniaceae, our best likelihood trees are generally consistent with other recent studies based on nuclear and mitochondrial data (Merckx et al., 2006, 2008). We found strong support for Burmanniaceae being part of the order Dioscoreales when only green members of the family were included (Fig. 4) and moderate support for this arrangement when fully mycoheterotrophic taxa were included (Fig. 3B), although the family's position within the order was ambiguous in both cases. The green taxa of Burmanniaceae may be partial mycoheterotrophs based on vegetative reduction and reduction in chlorophyll (see Merckx et al., 2006), supported by isotopic evidence in *B. coelestis* (Bolin et al., in press). The green taxa often have lower substitution rates than full mycoheterotrophs in the family (Fig. 6; Appendix S8), but the family as a whole appears to have an elevated rate of evolution when only green taxa are considered (long branch subtending the green clade in Fig. 4, see also Fig. 6 and Appendix S8).

A diversity of phylogenetic studies based on molecular and morphological data have placed Triuridaceae in Pandanales (e.g., Chase et al., 2000; Rudall and Bateman, 2006; Mennes et al., 2013). The family was recently resolved as the sister group of Cyclanthaceae and Pandanaceae with strong support, based on a whole-plastome analysis (Lam et al., 2015). Here it was recovered as monophyletic only in the partitioned ML analysis (see shortest tree in Appendix S7f), but with poor support (Fig. 3E). The two samples of *S. densiflora* included here (vouchers *Pillon et al., 88* and *Duangjai 029* in Appendix S1) were distantly related to each other (Appendices S7f, g), perhaps pointing to within-genus misidentification of one of them, although we did not confirm this possibility here. The lack of monophyly of *Sciaphila* in some likelihood analyses is presumably a long-branch artifact. Triuridaceae were weakly supported as a member of Pandanales in the partitioned ML analysis (Fig. 3E). The family has some of the most elevated rates of evolution and some of the longest branches in our study (e.g., Figs. 2, 6; Appendix S8).

Particular caution seems warranted concerning our results for Thismiaceae. The data reported here for this family are especially intriguing because they may represent the first genuine plastid data to be recovered for it (i.e., for *accD*), from each of three species in the family (*Geomitra clavigera*, *Thismia aseroe*, and *Thismia* sp.; Fig. 5). Rates of molecular evolution are elevated for these taxa compared with photosynthetic taxa (Fig. 6; Appendix S8; note that these are three-gene analyses that include only *accD* from this family). The placement of Thismiaceae in analyses that include all three taxa was unstable (sister to Taccaceae or Petrosaviaceae; Appendices S7h, i) and very poorly supported (e.g., Fig. 3C). Analyses that included each species individually placed them in various positions with weak support that were also easily perturbable between analyses (e.g., by using partitioned vs. unpartitioned likelihood; Fig. 5A–C). The *Geomitra clavigera* and *Thismia* sp. samples placed close to or within Taccaceae and/or Trichopodaceae in some analyses (Fig. 5), consistent with previous studies by Merckx et al. (2006, 2009b), who presented mitochondrial and nuclear data that supported a placement of Thismiaceae in Dioscoreales, close to Taccaceae. The precise position of Thismiaceae in the order was strongly supported in their earlier study (Merckx et al., 2006), and moderately supported in a subsequent one (Merckx et al., 2009b), and they also provided evidence that *Afrothismia* (not included here) represents a lineage distinct from other Thismiaceae (Merckx et al., 2009b). The successful retrieval of *accD* sequences for *Geomitra*, *Thismia aseroe*, and *Thismia* sp. supports a retention of plastid genomes in at least some species in Thismiaceae. As such, we may expect to find small, cryptic plastomes in Thismiaceae that include, at the very least, the *accD* locus.

Are Burmanniaceae and Thismiaceae closely related?—The proposed taxonomic circumscription of Burmanniaceae to include Thismiaceae (APG, 2003, 2009) was based primarily on studies that surveyed several photosynthetic plastid genes (*atpB*, *rbcL*) from these families (Caddick et al., 2000, 2002) that may include contaminants (M. Chase, Royal Botanic Gardens, Kew, personal communication). Consistent with this, these genes have not been recovered in full circular genomes of any fully heterotrophic taxa in Burmanniaceae (V. Lam, unpublished data). Our inability here to reject a close relationship between these families in the AU and SH tests may simply reflect a lack of power to reject such hypotheses from the genes we surveyed (in particular, only *accD* was recoverable from Thismiaceae). We therefore propose that the broad treatment of Burmanniaceae to include members of Thismiaceae should be abandoned in future APG treatments until further evidence (ideally including plastid data and other sources of evidence) is obtained for where Thismiaceae fits in monocot phylogeny. Nonetheless, at least some of the plastid-based likelihood trees inferred here (Figs. 2, 5; Appendix S7i) are consistent with current evidence from mitochondrial and nuclear data, which support a position of Thismiaceae distinct from Burmanniaceae and closer to Taccaceae (Caddick et al., 2002; Merckx et al., 2006, 2009b).

Phylogenetic relationships within Burmanniaceae—Relationships within Burmanniaceae are generally poorly supported here. Not surprisingly, fully mycoheterotrophic taxa in this family often had highly elevated rates of evolution (Fig. 6; Appendix S8). Several taxa have one or two of the three genes that were not retrieved (Appendix S1), confirmed to be missing genes in several cases (V. Lam, unpublished data). Family relationships within Burmanniaceae were mostly strongly supported when only green Burmanniaceae are included in analysis (Fig. 4), as was family monophyly. Merckx et al. (2006) suggested that there were at least six independent losses of photosynthesis in Burmanniaceae, based on analyses of the mitochondrial *nad1 b-c* intron and nuclear 18S rDNA sequences. Although our sampling of Burmanniaceae has fewer taxa than theirs, the relationships inferred here also point to multiple independent losses of autotrophy among taxa, as there are three green lineages here (*B. capitata* and two small clades with three or four taxa) that are deeply nested among nongreen lineages (Fig. 3B; Appendix S7c). Although these "backbone" relationships were not well supported here (see Fig. 3B; note that our sampling of the family is more limited than that of Merckx et al., 2006 and Merckx et al., 2008, and we also used their scorings of photosynthetic vs. fully mycoheterotrophic taxa), the AU and SH tests indirectly support scenarios with more than one loss of photosynthesis, because a tree constraining all nongreen taxa as a clade is significantly longer than the best tree. Our analyses recovered several intrafamiliar relationships that were congruent with those from Merckx et al. (2006) including a clade comprising the achlorophyllous taxa *B. oblonga* and *B. lutescens*, another comprising achlorophyllous *B. itoana* and *B. wallichii*, and a clade comprising three autotrophic species (*Burmannia bicolor*, *B. stuebelii*, and *B. biflora*).

Retention of plastid genes and genomes in monocot mycoheterotrophs—Although some nonphotosynthetic plants are hypothesized to have lost their plastid genomes entirely (e.g., the eudicot holoparasite *Rafflesia*; Molina et al., 2014), which is demonstrated in multiple lineages of secondarily heterotrophic unicellular eukaryotes (Abrahamsen et al., 2004; Smith and Lee, 2014; Janouškovec et al., 2015), all other heterotrophic plant lineages surveyed to date have retained their plastid genomes (e.g., Wolfe et al., 1992; McNeal et al., 2007; Wickett et al., 2008; Delannoy et al., 2011; Logacheva et al., 2014; Lam et al., 2015; Mennes et al., 2015; Bellot and Renner, 2016). A small set of genes (including *accD*, *clpP*, *trnE*-UUC, I-CAU, and *f*M-CAU, the four rDNAs, some ribosomal proteins) are commonly retained across many heterotrophic lineages that have highly reduced plastomes (e.g., Wicke et al., 2013; Barrett et al., 2014; Lam et al., 2015; but see Bellot and Renner, 2016), suggesting that they are essential genes and may not be readily replaced by nuclear counterparts (either functionally transferred genes or replacement by analogous systems; Barbrook et al., 2006). We chose the three plastid genes (*accD*, *clpP*, and *matK*) for our phylogenetic survey because they have nonphotosynthetic roles and are frequently retained when photosynthetic genes are lost. Both *accD* and *clpP* have been lost occasionally in flowering plants, but this appears to be unrelated to the loss of photosynthesis (e.g., Jansen et al., 2007; Straub et al., 2011); *matK* loss may eventually occur when a sufficient number of plastid genes with group IIA introns have been lost, which may only occur in the later stages of plastid genome degradation in heterotrophic lineages (McNeal et al., 2009).

Some of the genes recovered here by amplification could conceivably represent nuclear or mitochondrial inserts of plastid genes. However, successful functional gene transfer from the plastid to nuclear genome is generally rare in land plants (e.g., Martin et al., 1998) and has not been confirmed in any plants for these three genes (although a potentially functional copy of *accD* may have been transferred to the nucleus in *Trachelium*; Rousseau-Gueutin et al., 2013). Nonfunctional copies would also be expected to degrade rapidly in the nuclear genome (e.g., Matsuo et al., 2005; but see Cusimano and Wicke, 2015). Mitochondrial inserts should also degrade (functional transfer of plastid protein-coding genes to mitochondria is unknown, e.g., Hao and Palmer, 2009), and the genes would generally be expected to evolve slowly after insertion, consistent with the lower mutation rate of this genome (e.g., Wolfe et al., 1987; Palmer and Herbon, 1988). The phylogenetically widespread retention here in mycoheterotrophs of rapidly evolving plastid genes that retain open reading frames is likely a hallmark of their retention in the plastid genome. In a few cases, we have also obtained plastid genomes from the same species (*Apteria aphylla*, *Arachnitis uniflora*, *Burmannia bicolor*, *Burmannia capitata*, *Geosiris aphylla*, *Petrosavia sakuraii*, *Petrosavia* aff. *sakuraii*, and *Sciaphila densiflora*; Appendix S1), confirming the retention of genes in plastid genomes that we also retrieved using PCR amplification.

Dealing with contamination in heterotrophic samples—Recovering contaminant sequences may be a frequent problem with samples from heterotrophic plants. We detected several possible instances here (seven of 34 taxa). We suspect these represent very weak contaminants that would not normally be evident, but which are apparent here when the target gene is no longer present or has evolved rapidly away from amplification or sequencing primer sites. It is possible that these represent instances of horizontal gene transfer, as has been demonstrated for parasitic plants (e.g., host to parasite, Davis and Wurdack, 2004; Yoshida et al., 2010; and parasite to host, Mower et al., 2004). However, this possibility may be unlikely, because mycoheterotrophs do not have plant-to-plant connections and any such transfers would have to involve a fungal intermediate. If not detected, the inclusion of contaminant genes would be highly problematic for phylogenetic inference (or in DNA barcoding studies, see below). While we propose that our approach for identifying contaminants is conservative, it may sometimes result in false positives, as sampling of outgroup sequences for *clpP* and *accD* is relatively sparse here and on GenBank, limiting our ability to distinguish the green taxa that contaminant sequences are most closely related to using tree-based methods. As a consequence, it was not always possible to confirm the probable identity of putative contaminating sequences. However, we have unpublished whole-plastid genome data in hand for several species that we also surveyed in the three-gene analysis here, which allowed us to verify authentic gene loss from the plastid genomes for several taxa where we obtained contaminant amplifications from conspecific samples (i.e., *clpP* and *matK* for *Apteria aphylla* in Burmanniaceae, and *matK* for *Arachnitis uniflora* in Corsiaceae; Appendix S1). These genomic data confirm the validity of including noncontaminant genes from samples that had residual contamination, at least for these two species, because the genes obtained by PCR amplification and Sanger sequencing here, which we included in analyses, were identical or very similar to those obtained using next-generation methods (Appendices S7c, d).

Suitability of the three genes as DNA barcoding markers—We did not specifically address the utility of the three plastid markers here as DNA barcoding markers, although two of them have been tested for their suitability in previous DNA barcoding studies (*accD* and *matK*, see below). In green plants, portions of *matK* and *rbcL* have been used as core barcoding markers, often supplemented with the *psbA-trnH* intergenic spacer and the nuclear ITS (internal transcribed spacer) region (e.g., Fazekas et al., 2008; Hollingsworth et al., 2009, 2011; Li et al., 2011). However, *rbcL* and *psbA-trnH* are not suitable as barcoding markers in fully heterotrophic plants because they either are, or involve, photosynthetic genes. The *matK* region used here is the DNA barcoding region for this gene (Appendix S2; Hollingsworth et al., 2009), and our study demonstrates that it can be recovered from fully mycoheterotrophic monocots with relative ease. However, we were unable to recover *matK* or *clpP* in multiple cases, and our amplification strategy sometimes recovered contaminants from these two genes that only become apparent after careful checking (Appendices S1, S3). In contrast, we were able to recover the *accD* locus from all taxa sampled here, without any contamination issues (note that *accD* is pseudogenized in some eudicot holoparasites; Wicke et al., 2013). The *accD* locus has only rarely been considered for DNA barcoding studies (Newmaster et al., 2008), and it would be useful to explore further its utility in DNA barcoding surveys that include mycoheterotrophs or holoparasites. It may be a useful supplementary barcoding region to consider including when the focus is not exclusively on green plants.

CONCLUSIONS

Mycoheterotrophic lineages can be placed in overall monocot phylogeny using our three-gene data set, generally with moderate to strong support for their placement, despite some issues with contamination, rate elevation, and missing genes. Our approach to inferring phylogenetic placement is a cost-effective alternative to next-generation sequencing and potentially allows phylogenetic surveys of many heterotrophic taxa. It is also a potentially useful screen in selecting DNAs that will be successful for next-generation library preparation. At least two of the markers considered here (*accD*, *matK*) may also be useful targets for including heterotrophic plants in DNA barcoding surveys. The underlying alignment we used could be improved upon further and may be particularly useful to do for other parasitic or mycoheterotrophic lineages, such as Orchidaceae, where we included only a few previously published sequences: this family includes numerous additional origins of full mycoheterotrophy (e.g., Freudenstein and Senyo, 2008; Merckx et al., 2013b; Barrett et al., 2014).

ACKNOWLEDGEMENTS

The authors thank Thassilo Franke, Stephanie Lyon, Daniel McNair, Ray Neyland, Tomohisa Yukawa, the Missouri Botanical Garden, the Royal Botanic Gardens, Kew and Kirstenbosch National Botanical Garden for assistance in obtaining plant material, and Craig Barrett for providing commelinid sequence data. Craig Barrett and an anonymous reviewer provided helpful comments for improving the manuscript. This work was supported by an NSERC (Natural Sciences and Engineering Research Council of Canada) Postgraduate Fellowship and UBC Four-Year Fellowship to V.K.Y.L., by NWO through a Veni Fellowship to V.S.F.T.M. (863.11.018), and by an NSERC Discovery grant to S.W.G.

LITERATURE CITED

Abrahamsen, M. S., T. J. Templeton, S. Enomoto, J. E. Abrahante, G. Zhu, C. A. Lancto, M. Deng, et al. 2004. Complete genome sequence of the apicomplexan, *Cryptosporidium parvum. Science* 304: 441–445.

Angiosperm Phylogeny Group. 1998. An ordinal classification for the families of flowering plants. *Annals of the Missouri Botanical Garden* 4: 531–553.

Angiosperm Phylogeny Group. 2003. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG II. *Botanical Journal of the Linnean Society* 141: 399–436.

Angiosperm Phylogeny Group. 2009. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG III. *Botanical Journal of the Linnean Society* 161: 105–121.

Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. 1990. Basic local alignment search tool. *Journal of Molecular Biology* 215: 403–410.

Barbrook, A. C., C. J. Howe, and S. Purton. 2006. Why are plastid genomes retained in non-photosynthetic organisms? *Trends in Plant Science* 11: 101–108.

Barrett, C. F., and J. I. Davis. 2012. The plastid genome of the mycoheterotrophic *Corallorhiza striata* (Orchidaceae) is in the relatively early stages of degradation. *American Journal of Botany* 99: 1513–1523.

Barrett, C. F., J. V. Freudenstein, J. Li, D. R. Mayfield-Jones, L. Perez, J. C. Pires, and C. Santos. 2014. Investigating the path of plastid genome degradation in early-transitional heterotrophic orchids, and implications for heterotrophic angiosperms. *Molecular Biology and Evolution* 31: 3095–3112.

Bellot, S., and S. S. Renner. 2016. The plastomes of two species in the endoparasite genus *Pilostyles* (Apodanthaceae) each retain just five or six possibly functional genes. *Genome Biology and Evolution* 8: 189–201.

Bernard, N. 1909. L'évolution dans la symbiose. Les orchidées et leurs champignons commensaux. *Annales des Sciences Naturelles. Botanique* 9: 1–196.

Bodin, S. S., J. S. Kim, and J.-H. Kim. 2016. Phylogenetic inferences and evolution of plastid DNA in Campynemataceae and the mycoheterotrophic *Corsia dispar* D.L. Jones & B. Gray (Corsiaceae). *Plant Molecular Biology Reporter* 34: 192–210.

Bolin, J. F., K. U. Tennakoon, M. B. A. Majid, and D. D. Cameron. In press. Isotopic evidence of partial mycoheterotrophy in *Burmannia coelestis* (Burmanniaceae). *Plant Species Biology*.

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

Caddick, L. R., P. J. Rudall, P. Wilkin, and M. W. Chase. 2000. Yams and their allies: Systematics of Dioscoreales. *In* K. L. Wilson, D. A. Morrison [eds.], Monocots: Systematics and evolution, 475–487. CSIRO Publishing, Victoria, Australia.

Caddick, L. R., P. J. Rudall, P. Wilkin, T. A. J. Hedderson, and M. W. Chase. 2002. Phylogenetics of Dioscoreales based on analyses of morphological and molecular data. *Botanical Journal of the Linnean Society* 138: 123–144.

Cameron, K. M., M. W. Chase, and P. J. Rudall. 2003. Recircumscription of the monocotyledonous family Petrosaviaceae to include *Japonolirion. Brittonia* 55: 214–225.

Chase, M. W., M. F. Fay, D. S. Devey, O. Maurin, N. Rønsted, T. J. Davies, Y. Pillon, et al. 2006. Multigene analyses of monocot relationships: A summary. *Aliso* 22: 63–75.

Chase, M. W., D. E. Soltis, R. G. Olmstead, D. Morgan, D. H. Les, B. D. Mishler, M. R. Duvall, et al. 1993. Phylogenetics of seed plants: An analysis of nucleotide sequences from the plastid *rbc*L. *Annals of the Missouri Botanical Garden* 80: 528–580.

Chase, M. W., D. E. Soltis, P. S. Soltis, P. J. Rudall, M. F. Fay, W. H. Hahn, S. Sullivan, et al. 2000. Higher-level systematics of the monocotyledons: An assessment of current knowledge and a new classification. *In* K. L. Wilson and D. A. Morrison [eds.], Monocots: Systematics and evolution, 3–16. CSIRO Publishing, Victoria, Australia.

Cronquist, A. 1988. The evolution and classification of flowering plants, 2nd ed. New York Botanical Garden, Bronx, New York, USA.

Cummings, M. P., and N. A. Welschmeyer. 1998. Pigment composition of putatively achlorophyllous angiosperms. *Plant Systematics and Evolution* 210: 105–111.

Cusimano, N., and S. Wicke. 2015. Massive intracellular gene transfer during plastid genome reduction in nongreen Orobanchaceae. *New Phytologist*.

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

Davis, J. I., D. W. Stevenson, G. Petersen, O. Seberg, L. M. Campbell, J. V. Freudenstein, D. H. Goldman, et al. 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. *Systematic Botany* 29: 467–510.

Delannoy, E., S. Fujii, C. C. Des Francs-Small, M. Brundrett, and I. Small. 2011. Rampant gene loss in the underground orchid *Rhizanthella gardneri* highlight evolutionary constraints on plastid genomes. *Molecular Biology and Evolution* 28: 2077–2086.

Drummond, A. J., M. A. Suchard, D. Xie, and A. Rambaut. 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution* 29: 1969–1973.

Doyle, J. J., and J. L. Doyle. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin* 19: 11–15.

Ems, S. C., C. W. Morden, C. K. Dixon, K. H. Wolfe, C. W. dePamphilis, and J. D. Palmer. 1995. Transcription, splicing and editing of plastid RNAs in the nonphotosynthetic plant *Epifagus virginiana. Plant Molecular Biology* 29: 721–733.

Engels, B. 1993. Amplify3x [online]. Website http://engels.genetics.wisc.edu/amplify/ [accessed 15 May 2010].

Fay, M. F., P. J. Rudall, S. Sullivan, K. L. Stobart, A. Y. De Brujin, G. Reeves, F. Qamaruz-Zaman, et al. 2000. Phylogenetic studies of Asparagales based on four plastid DNA regions. *In* K. L. Wilson and D. A. Morrison [eds.], Monocots: Systematics and evolution, 360–371. CSIRO Publishing, Collingwood, Australia.

Fazekas, A. J., K. S. Burgess, P. R. Kesanakurti, S. W. Graham, S. G. Newmaster, B. C. Husband, D. M. Percy, et al. 2008. Multiple multilocus DNA barcodes from the plastid genome discriminate plant species equally well. *PLoS One* 3: e2802.

Felsenstein, J. 1978. Cases in which parsimony or compatibility methods will be positively misleading. *Systematic Zoology* 27: 401–410.

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

Felsenstein, J. 1988. Phylogenies from molecular sequences: Inference and reliability. *Annual Review of Genetics* 22: 521–565.

Freudenstein, J. V., and D. M. Senyo. 2008. Relationships and evolution of *matK* in a group of leafless orchids (*Corallorhiza* and Corallorhizinae; Orchidaceae: Epidendroideae). *American Journal of Botany* 95: 498–505.

Fuse, S., and M. N. Tamura. 2000. A phylogenetic analysis of the plastid *mat*K gene with emphasis on Melanthiaceae *sensu lato. Plant Biology* 2: 415–427.

Gandolfo, M. A., K. C. Nixon, and W. L. Crepet. 2002. Triuridaceae fossil flowers from the Upper Cretaceous of New Jersey. *American Journal of Botany* 89: 1940–1957.

Gandolfo, M. A., K. C. Nixon, W. L. Crepet, D. W. Stevenson, and E. M. Friis. 1998. Oldest known fossils of monocotyledons. *Nature* 394: 532–533.

Givnish, T. J., M. Ames, J. R. McNeal, M. R. McKain, P. R. Steele, C. W. DePamphilis, S. W. Graham, et al. 2010. Assembling the tree of the monocotyledons: Plastome sequence phylogeny and evolution of Poales. *Annals of the Missouri Botanical Garden* 97: 584–616.

Givnish, T. J., J. C. Pires, S. W. Graham, M. A. McPherson, L. M. Prince, T. B. Patterson, H. S. Rai, et al. 2005. Repeated evolution of net venation and fleshy fruits among monocots in shaded habitats confirms a priori predictions: Evidence from an *ndhF* phylogeny. *Proceedings of the Royal Society of London, B, Biological Sciences* 272: 1481–1490.

Goldblatt, P., A. Rodriguez, M. P. Powell, T. J. Davies, J. C. Manning, M. Van Der Bank, and V. Savolainen. 2008. Iridaceae 'Out of Australasia'? Phylogeny, biogeography, and divergence time based on plastid DNA sequences. *Systematic Botany* 3: 495–508.

Graham, S. W., and R. G. Olmstead. 2000. Utility of 17 chloroplast genes for inferring the phylogeny of the basal angiosperms. *American Journal of Botany* 87: 1712–1730.

Graham, S. W., P. A. Reeves, A. C. Burns, and R. G. Olmstead. 2000. Microstructural changes in noncoding chloroplast DNA: Interpretation, evolution, and utility of indels and inversions in basal angiosperm phylogenetic inference. *International Journal of Plant Sciences* 161: S83–S96.

Graham, S. W., J. M. Zgurski, M. A. McPherson, D. M. Cherniawsky, J. M. Saarela, E. S. C. Horne, et al. 2006. Robust inference of monocot deep phylogeny using an expanded multigene plastid data set. *Aliso* 22: 3–20.

Hao, W., and J. D. Palmer. 2009. Fine-scale mergers of chloroplast and mitochondrial genes create functional, transcompartmentally chimeric mitochondrial genes. *Proceedings of the National Academy of Sciences, USA* 106: 16728–16733.

Heath, T. A., S. M. Hedtke, and D. M. Hillis. 2008. Taxon sampling and the accuracy of phylogenetic analyses. *Journal of Systematics and Evolution* 46: 239–257.

Hedtke, S. M., T. M. Townsend, and D. M. Hillis. 2006. Resolution of phylogenetic conflict in large data sets by increased taxon sampling. *Systematic Biology* 55: 522–529.

Hendy, M. D., and D. Penny. 1989. A framework for quantitative study of evolutionary trees. *Systematic Zoology* 38: 297–309.

Hertweck, K. L., M. S. Kinney, S. A. Stuart, O. Maurin, S. Mathews, M. W. Chase, M. A. Gandaldo, and J. C. Pires. 2015. Phylogenetics, divergence times and diversification from three genomic partitions in monocots. *Botanical Journal of the Linnean Society* 178: 375–393.

Hillis, D. M. 1998. Taxonomic sampling, phylogenetic accuracy, and investigator bias. *Systematic Biology* 47: 3–8.

Hilu, K. W., T. Borsch, K. Müller, D. E. Soltis, P. S. Soltis, V. Savolainen, M. W. Chase, et al. 2003. Angiosperm phylogeny based on *matK* sequence information. *American Journal of Botany* 90: 1758–1776.

Hollingsworth, P. M., L. L. Forest, J. L. Spouge, M. Hajibabaei, S. Ratnasingham, M. van der Bank, M. W. Chase, et al. 2009. A DNA barcode for plants. *Proceedings of the National Academy of Sciences of the United States of America* 106: 12794–12797.

Hollingsworth, P. M., S. W. Graham, and D. P. Little. 2011. Choosing and using a plant DNA barcode. *PLoS One* 6: e19254.

Huelsenbeck, J. P. 1997. Is the Felsenstein zone a fly trap? *Systematic Biology* 46: 69–74.

Huelsenbeck, J. P. 1998. Systematic bias in phylogenetic analysis: Is the *Strepsiptera* problem solved? *Systematic Biology* 47: 519–537.

Iles, W. J. D., S. Y. Smith, M. A. Gandalfo, and S. W. Graham. 2015. Monocot fossils suitable for molecular dating analyses. *Botanical Journal of the Linnean Society* 178: 346–374.

Imhof, S. 2010. Are monocots particularly suited to develop mycoheterotrophy? *In* O. Seberg, G. Petersen, A. Barfod, and J. I. Davis [eds.], Diversity, phylogeny, and evolution in the monocotyledons, 11–23. Aarhus University Press, Copenhagen, Denmark.

Janouškovec, J., D. V. Tikhonenkov, F. Burki, A. T. Howe, M. Kolisko, A. P. Mylnikov, and P. J. Keeling. 2015. Factors mediating plastid dependency and the origins of parasitism in apicomplexans and their close relatives. *Proceedings of the National Academy of Sciences, USA* 112: 10200–10207.

Jansen, R. K., Z. Cai, L. A. Raubeson, H. Daniell, C. W. DePamphilis, J. Leebens-Mack, K. F. Müller, et al. 2007. Analysis of 81 genes from 64 plastid genomes resolves relationships in angiosperms and identifies genome-scale evolutionary patterns. *Proceedings of the National Academy of Sciences, USA* 104: 19369–19374.

Kelchner, S. A. 2000. The evolution of non-coding chloroplast DNA and its application in plant systematics. *Annals of the Missouri Botanical Garden* 87: 482–498.

Kim, J. 1996. General inconsistency conditions for maximum parsimony: Effects of branch lengths and increasing numbers of taxa. *Systematic Biology* 45: 363–374.

Kim, J. S., J.-K. Hong, M. W. Chase, M. F. Fay, and J.-H. Kim. 2013. Familial relationships of the monocot order Liliales based on a molecular phylogenetic analysis using four plastid loci: *matK*, *rbcL*, *atpB* and *atpF*. *Botanical Journal of the Linnean Society* 172: 5–21.

Kolaczkowski, B., and J. W. Thornton. 2009. Long-branch attraction bias and inconsistency in Bayesian phylogenetics. *PLoS One* 4: e7891.

Lam, V. K. Y., M. Soto Gomez, and S. W. Graham. 2015. The highly reduced plastome of mycoheterotrophic *Sciaphila* (Triuridaceae) is colinear with its green relative and is under strong purifying selection. *Genome Biology and Evolution* 7: 2220–2236.

Lanfear, R., B. Calcott, S. Y. W. Ho, and S. Guindon. 2012. PartitionFinder: Combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Molecular Biology and Evolution* 29: 1695–1701.

Leake, J. R. 1994. The biology of myco-heterotrophic ('saprophytic') plants. *New Phytologist* 127: 171–216.

Leake, J. R. 2004. Myco-heterotrophic/epiparasitic plant interactions with ectomycorrhizal and arbuscular mycorrhizal fungi. *Current Opinion in Plant Biology* 7: 422–428.

Leake, J. R. 2005. Plants parasitic on fungi: Unearthing the fungi in mycoheterotrophs and debunking the 'saprophytic' plant myth. *Mycologist* 19: 113–122.

Li, D.-Z., L.-M. Gao, H.-T. Li, H. Wang, X.-J. Ge, J.-Q. Liu, Z.-D. Chen, et al. 2011. Comparative analysis of a large dataset indicates that internal transcribed spacer (ITS) should be incorporated into the core barcode for seed plants. *Proceedings of the National Academy of Sciences, USA* 108: 19641–19646.

Logacheva, M. D., M. I. Schelkunov, M. S. Nuraliev, T. H. Samigullin, and A. A. Penin. 2014. The plastid genome of mycoheterotrophic monocot *Petrosavia stellaris* exhibits both gene losses and multiple rearrangements. *Genome Biology and Evolution* 6: 238–246.

Logacheva, M. D., M. I. Schelkunov, and A. A. Penin. 2011. Sequencing and analysis of plastid genome in mycoheterotrophic orchid *Neottia nidus-avis*. *Genome Biology and Evolution* 3: 1296–1303.

Martin, W., B. Stoebe, V. Goremykin, S. Hansmann, M. Hasegawa, and K. V. Kowallik. 1998. Gene transfer to the nucleus and the evolution of chloroplasts. *Nature* 393: 162–165.

Matsuo, M., Y. Ito, R. Yamauchi, and J. Obokata. 2005. The rice nuclear genome continuously integrates, shuffles, and eliminates the chloroplast genome to cause chloroplast-nuclear DNA flex. *Plant Cell* 17: 665–675.

McNeal, J. R., J. V. Kuehl, J. L. Boore, and C. W. DePamphilis. 2007. Complete plastid genome sequences suggest strong selection for retention of photosynthetic genes in the parasitic plant genus *Cuscuta*. *BMC Plant Biology* 7: 57.

McNeal, J. R., J. V. Kuehl, J. L. Boore, J. Leebens-Mack, and C. W. DePamphilis. 2009. Parallel loss of plastid introns and their maturase in the genus *Cuscuta*. *PLoS One* 4: e5982.

Mennes, C. B., V. K. Y. Lam, P. J. Rudall, S. P. Lyon, S. W. Graham, E. F. Smets, and V. S. F. T. Merckx. 2015. Ancient Gondwana break-up explains the distribution of the mycoheterotrophic family Corsiaceae (Liliales). *Journal of Biogeography* 69: 994–1004.

Mennes, C. B., E. F. Smets, S. N. Moses, and V. S. F. T. Merckx. 2013. New insights in the long-debated evolutionary history of Triuridaceae. *Molecular Phylogenetics and Evolution* 69: 994–1004.

Merckx, V. S. F. T. 2013. Mycoheterotrophy: An introduction. *In* V. Merckx [ed.], Mycoheterotrophy: The biology of plants living on fungi, 1–18. Springer-Verlag, New York, New York, USA.

Merckx, V., F. T. Bakker, S. Huysman, and E. Smets. 2009b. Bias and conflict in phylogenetic inference of myco-heterotrophic plants: A case study in Thismiaceae. *Cladistics* 25: 64–77.

Merckx, V., M. I. Bidartondo, and N. A. Hynson. 2009a. Myco-heterotrophy: When fungi host plants. *Annals of Botany* 104: 1255–1261.

Merckx, V., L. W. Chatrou, B. Lemaire, M. N. Sainge, S. Huysmans, and E. F. Smets. 2008. Diversification of myco-heterotrophic angiosperms: Evidence from Burmanniaceae. *BMC Evolutionary Biology* 8: 178.

Merckx, V., and J. V. Freudenstein. 2010. Evolution of mycoheterotrophy in plants: A phylogenetic perspective. *New Phytologist* 185: 605–609.

Merckx, V., S. Huysmans, and E. Smets. 2010. Cretaceous origins of mycoheterotrophic lineages in Dioscoreales. *In* O. Seberg, G. Petersen, A. Barfod, and J. I. Davis [eds.], Diversity, phylogeny, and evolution in the monocotyledons, 39–53. Aarhus University Press, Copenhagen, Denmark.

Merckx, V., P. Schols, H. Maas-Van De Kamer, P. Maas, S. Huysmans, and E. Smets. 2006. Phylogeny and evolution of Burmanniaceae (Dioscoreales) based on nuclear and mitochondrial data. *American Journal of Botany* 93: 1684–1698.

Merckx, V. S. F. T. 2013. Mycoheterotrophy: An introduction. *In* V. Merckx [ed.], Mycoheterotrophy: The biology of plants living on fungi, 1–18. Springer-Verlag, New York, New York, USA.

Merckx, V. S. F. T., J. V. Freudenstein, J. Kissling, M. J. M. Christenhusz, R. E. Stotler, B. Crandall-Stotler, N. Wickett, et al. 2013b. Taxonomy and classification. *In* V. Merckx [ed.], Mycoheterotrophy: The biology of plants living on fungi, 19–101. Springer-Verlag, New York, New York, USA.

Merckx, V. S. F. T., C. B. Mennes, K. G. Peay, and J. Geml. 2013a. Evolution and diversification. *In* V. Merckx [ed.], Mycoheterotrophy: The biology of plants living on fungi, 222–226. Springer-Verlag, New York, New York, USA.

Molina, J., K. M. Hazzouri, D. Nickrent, M. Geisler, R. S. Meyer, M. M. Pentony, J. M. Flowers, et al. 2014. Possible loss of the chloroplast genome in the parasitic flowering plant *Rafflesia lagascae* (Rafflesiaceae). *Molecular Biology and Evolution* 31: 793–803.

Mower, J. P., S. Stefanović, G. J. Young, and J. D. Palmer. 2004. Plant genetics: Gene transfer from parasitic to host plants. *Nature* 432: 165–166.

Newmaster, S. G., A. J. Fazekas, R. A. D. Steeves, and J. Janovec. 2008. Testing candidate plant barcode regions in the Myristicaceae. *Molecular Ecology Resources* 8: 480–490.

Neyland, R. 2002. A phylogeny inferred from large-subunit (26S) ribosomal DNA sequences suggest that Burmanniales are polyphyletic. *Australian Systematic Biology* 15: 19–28.

Neyland, R., and M. Hennigan. 2003. A phylogenetic analysis of large-subunit (26S) ribosome DNA sequences suggests that the Corsiaceae are polyphyletic. *New Zealand Journal of Botany* 41: 1–11.

Palmer, J. D., and L. A. Herbon. 1988. Plant mitochondrial DNA evolves rapidly in structure, but slowly in sequence. *Journal of Molecular Evolution* 28: 87–97.

Parks, S. L., and N. Goldman. 2014. Maximum likelihood inference of small trees in the presence of long branches. *Systematic Biology* 63: 798–811.

Pollock, D. D., D. J. Zwickl, J. A. McGuire, and D. M. Hillis. 2002. Increased taxon sampling is advantageous for phylogenetic inference. *Systematic Biology* 51: 664.

Rai, H. S., H. O'Brien, P. A. Reeves, R. G. Olmstead, and S. W. Graham. 2003. Inference of higher-order relationships in the cycads from a large chloroplast data set. *Molecular Phylogenetics and Evolution* 29: 350–359.

Rambaut, A. 2002. Se-Al v. 2.0a11: Sequence alignment program. Available from http://tree.bio.ed.ac.uk/software/seal/ [accessed 15 January 2016].

Rambaut, A. 2006. FigTree. Available from http://tree.bio.ed.ac.uk/software/figtree/ [accessed 15 January 2016].

Rambaut, A., M. A. Suchard, D. Xie, and A. J. Drummond. 2014. Tracer v1. 6. Available from http://beast.bio.ed.ac.uk/Tracer [accessed 15 January 2016].

Rasmussen, H. N. 1995. Terrestrial orchids: From seed to mycotrophic plant. Cambridge University Press, Cambridge, UK.

Reeves, G., M. W. Chase, P. Goldblatt, P. Rudall, M. F. Fay, A. V. Cox, B. Lejeune, and T. Souza-Chies. 2001. Molecular systematics of Iridaceae: Evidence from four plastid DNA regions. *American Journal of Botany* 88: 2074–2087.

Rousseau-Gueutin, M., X. Huang, E. Higginson, M. Ayliffe, A. Day, and J. N. Timmis. 2013. Potential functional replacement of the plastidic acetyl-CoA carboxylase subunit (*accD*) gene by recent transfer to the nucleus in some angiosperm lineages. *Plant Physiology* 161: 1918–1929.

Rudall, P. J., and R. M. Bateman. 2006. Morphological phylogenetic analysis of Pandanales: Testing contrasting hypotheses of floral evolution. *Systematic Botany* 31: 223–238.

Ruhfel, B. R., M. A. Gitzendanner, P. S. Soltis, D. E. Soltis, and J. G. Burleigh. 2014. From algae to angiosperms—Inferring the phylogeny of green plants (*Viridiplantae*) from 360 plastid genomes. *BMC Evolutionary Biology* 14: 23.

Rychlik, W. 2007. OLIGO 7 primer analysis software. *In* A. Yuryev [ed.], Methods in molecular biology, vol. 402, 35–59. Humana Press, New York, New York, USA.

Saarela, J. M., and S. W. Graham. 2010. Inference of phylogenetic relationships among the subfamilies of grasses (Poaceae: Poales) using meso-scale exemplar-based sampling of the plastid genome. *Botany* 88: 65–84.

Schelkunov, M. I., V. Y. Shtratnikova, M. S. Nuraliev, M.-A. Selosse, A. A. Penin, and M. D. Logacheva. 2015. Exploring the limits for reduction of plastid genomes: A case study of the mycoheterotrophic orchids *Epipogium aphyllum* and *Epipogium roseum*. *Genome Biology and Evolution* 7: 1179–1191.

Schwarz, G. 1978. Estimating the dimension of a model. *Annals of Statistics* 6: 461–464.

Shanklin, J., N. D. DeWitt, and J. M. Flanagan. 1995. The stroma of higher plant plastids contain ClpP and ClpC, functional homologs of *Escherichia coli* ClpP and ClpA: An archetypal two-component ATP-dependent protease. *Plant Cell* 7: 1713.

Shimodaira, H. 2002. An approximately unbiased test of phylogenetic tree selection. *Systematic Biology* 51: 492–508.

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

Silvestro, D., and I. Michalak. 2012. raxmlGUI: A graphical front-end for RaxML. *Organisms, Diversity & Evolution* 12: 335–337.

Simmons, M. P., and H. Ochoterena. 2000. Gaps as characters in sequence-based phylogenetic analyses. *Systematic Biology* 49: 369–381.

Simmons, M. P., K. M. Pickett, and M. Miya. 2004. How meaningful are Bayesian support values? *Molecular Biology and Evolution* 21: 188–199.

Smith, D. R., and R. W. Lee. 2014. A plastid without a genome: Evidence from the nonphotosynthetic green algal genus *Polytomella*. *Plant Physiology* 164: 1812–1819.

Soltis, D. E., S. A. Smith, N. Cellinese, K. J. Wurdack, D. C. Tank, S. F. Brockington, N. F. Refulio-Rodriguez, et al. 2011. Angiosperm phylogeny: 17 genes, 640 taxa. *American Journal of Botany* 98: 704–730.

Soltis, D. E., and P. S. Soltis. 2004. *Amborella* not a "basal angiosperm"? Not so fast. *American Journal of Botany* 91: 997–1001.

Soltis, D. E., P. S. Soltis, M. W. Chase, M. E. Mort, D. C. Albach, M. Zanis, V. Savolainen, et al. 2000. Angiosperm phylogeny inferred from 18S rDNA, *rbcL*, and *atpB* sequences. *Botanical Journal of the Linnean Society* 133: 381–461.

Stamatakis, A. 2006. RaxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22: 2688–2690.

Straub, S. C. K., M. Fishbein, T. Livshultz, Z. Foster, M. Parks, K. Weitmier, R. Cronn, and A. Liston. 2011. Building a model: Developing genome resources for common milkweed (*Asclepias syriaca*) with low coverage genome sequencing. *BMC Genomics* 12: 211.

Sullivan, J., and P. Joyce. 2005. Model selection in phylogenetics. *Annual Review of Ecology, Evolution and Systematics* 36: 445–466.

Swofford, D. L. 2003. PAUP*: Phylogenetic analysis using parsimony (*and other methods), version 4.0b10. Sinauer, Sunderland, Massachusetts, USA.

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

Vogel, J., T. Hübschmann, T. Börner, and W. R. Hess. 1997. Splicing and intron-internal RNA editing of *trnK-matK* transcripts in barley plastids: Support for *matK* as an essential splice factor. *Journal of Molecular Biology* 270: 179–187.

Waterman, R. J., M. R. Klooster, H. Hentrich, and M. I. Bidartondo. 2013. *In* V. Merckx [ed.], Mycoheterotrophy: The biology of plants living on fungi, 267–296. Springer-Verlag, New York, New York, USA.

Wicke, S., K. F. Müller, C. W. DePamphilis, D. Quandt, N. J. Wickett, Y. Zhang, S. S. Renner, and G. M. Schneeweiss. 2013. Mechanisms of functional and physical genome reduction in photosynthetic and nonphotosynthetic parasitic plants of the broomrape family. *Plant Cell* 25: 3711–3725.

Wicke, S., G. M. Schneeweiss, C. W. DePamphilis, K. F. Müller, and D. Quandt. 2011. The evolution of the plastid chromosome in land plants: gene content, gene order, gene function. *Plant Molecular Biology* 76: 273–297.

Wickett, N. J., Y. Zhang, S. K. Hansen, J. M. Roper, J. V. Kuehl, S. A. Plock, P. G. Wolfe, et al. 2008. Functional gene losses occur with minimal size reduction in the plastid genome of the parasitic liverwort *Aneura mirabilis*. *Molecular Biology and Evolution* 25: 393–401.

Wiens, J. J. 2005. Can incomplete taxa rescue phylogenetic analyses from long-branch attraction? *Systematic Biology* 54: 731–742.

Wiens, J. J. 2006. Missing data and the design of phylogenetic analyses. *Journal of Biomedical Informatics* 39: 34–42.

Wiens, J. J., and J. Tiu. 2012. Highly incomplete taxa can rescue phylogenetic analyses from the negative impacts of limited taxon sampling. *PLoS One* 7: e42925.

Wolfe, K. H., W.-H. Li, and P M Sharp. 1987. Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. *Proceedings of the National Academy of Sciences, USA* 84: 9054-9058.

Wolfe, K. H., C. W. Morden, and J. D. Palmer. 1992. Function and evolution of a minimal plastid genome from a nonphotosynthetic parasitic plant. *Proceedings of the National Academy of Sciences, USA* 89: 10648–10652.

Yang, Z. 1996. Phylogenetic analysis using parsimony and likelihood methods. *Journal of Molecular Evolution* 42: 294–307.

Yang, Z. 2006. Computational molecular evolution. Oxford University Press, Oxford, UK.

Yang, Z., and B. Rannala. 2005. Branch-length prior influences Bayesian posterior probability of phylogeny. *Systematic Biology* 54: 455–470.

Yang, Z., and B. Rannala. 2012. Molecular phylogenetics: Principles and practice. *Nature Reviews. Genetics* 13: 303–314.

Yoshida, S., S. Maruyama, H. Nozaki, and K. Shirasu. 2010. Horizontal gene transfer by the parasitic plant *Striga hermonthica*. *Science* 328: 1128.

Zgurski, J. M., H. S. Rai, Q. M. Fai, D. J. Bogler, J. Francisco-Ortega, and S. W. Graham. 2008. How well do we understand the overall backbone of cycad phylogeny? New insights from a large, multigene plastid data set. *Molecular Phylogenetics and Evolution* 47: 1232–1237.

Zhang, D., R. M. Saunders, and C.-M. Hu. 1999. *Corsiopsis chinensis* gen. et sp. nov. (Corsiaceae): First record of the family in Asia. *Systematic Botany* 24: 311–314.