







# Phylotranscriptomic Analyses of Mycoheterotrophic Monocots Show a Continuum of Convergent Evolutionary Changes in Expressed Nuclear Genes From Three Independent Nonphotosynthetic Lineages

Prakash Raj Timilsena<sup>1</sup>, Craig F. Barrett<sup>2</sup>, Alma Piñeyro-Nelson <sup>3,4</sup>, Eric K. Wafula<sup>1</sup>, Saravanaraj Ayyampalayam<sup>5</sup>, Joel R. McNeal<sup>6</sup>, Tomohisa Yukawa<sup>7</sup>, Thomas J. Givnish<sup>8</sup>, Sean W. Graham <sup>9</sup>, J. Chris Pires<sup>10</sup>, Jerrold I. Davis<sup>11</sup>, Cécile Ané <sup>8,12</sup>, Dennis W. Stevenson<sup>13</sup>, Jim Leebens-Mack <sup>5</sup>, Esteban Martínez-Salas<sup>14</sup>, Elena R. Álvarez-Buylla <sup>15,4</sup>, and Claude W. dePamphilis <sup>1,\*</sup>

<sup>1</sup>Department of Biology, The Pennsylvania State University, University Park, Pennsylvania

<sup>2</sup>Department of Biology, West Virginia University, Morgantown, West Virginia

<sup>3</sup>Departamento de Producción Agrícola y Animal, Universidad Autónoma Metropolitana-Xochimilco, Mexico City, Mexico

<sup>4</sup>Centro de Ciencias de la Complejidad, Universidad Nacional Autónoma de México, Mexico City, Mexico

<sup>5</sup>Department of Plant Biology, University of Georgia, Athens, Georgia, 3060

<sup>6</sup>Department of Ecology, Evolution, and Organismal Biology, Kennesaw State University, Georgia

<sup>7</sup>Tsukuba Botanical Garden, National Museum of Nature and Science, 1-1, Amakubo 4, Tsukuba, 305-0005, Japan

<sup>8</sup>Department of Botany, University of Wisconsin-Madison, Madison, Wisconsin

<sup>9</sup>Department of Botany, University of British Columbia, Vancouver, British Columbia, V6T 1Z4 Canada

<sup>10</sup>Division of Biological Sciences, University of Missouri–Columbia, Columbia, Missouri

<sup>11</sup>School of Integrative Plant Sciences and L.H. Bailey Hortorium, Cornell University, Ithaca, New York, 1485

<sup>12</sup>Department of Statistics, University of Wisconsin–Madison, Madison, Wisconsin

<sup>13</sup>New York Botanical Garden, New York, 1045

<sup>14</sup>Departamento de Botánica, Instituto de Biología, Universidad Nacional Autónoma de México, Mexico City, México

<sup>15</sup>Departamento de Ecología Funcional, Instituto de Ecología, Universidad Nacional Autónoma de México, Mexico City, Mexico

\*Corresponding author: E-mail: [cwd3@psu.edu](mailto:cwd3@psu.edu).

Accepted: 18 December 2022

## Abstract

Mycoheterotrophy is an alternative nutritional strategy whereby plants obtain sugars and other nutrients from soil fungi. Mycoheterotrophy and associated loss of photosynthesis have evolved repeatedly in plants, particularly in monocots. Although reductive evolution of plastomes in mycoheterotrophs is well documented, the dynamics of nuclear genome evolution remains largely unknown. Transcriptome datasets were generated from four mycoheterotrophs in three families (Orchidaceae, Burmanniaceae, Triuridaceae) and related green plants and used for phylogenomic analyses to resolve relationships among the mycoheterotrophs, their relatives, and representatives across the monocots. Phylogenetic trees based on 602 genes were mostly congruent with plastome phylogenies, except for an Asparagales + Liliales clade inferred in the nuclear trees. Reduction and loss of chlorophyll synthesis and photosynthetic gene expression and relaxation of purifying selection on retained genes were progressive, with greater loss in older nonphotosynthetic lineages. One hundred seventy-four of

© The Author(s) 2022. Published by Oxford University Press on behalf of Society for Molecular Biology and Evolution.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<https://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact [journals.permissions@oup.com](mailto:journals.permissions@oup.com)

1375 plant benchmark universally conserved orthologous genes were undetected in any mycoheterotroph transcriptome or the genome of the mycoheterotrophic orchid *Gastrodia* but were expressed in green relatives, providing evidence for massively convergent gene loss in nonphotosynthetic lineages. We designate this set of deleted or undetected genes *Missing in Mycoheterotrophs* (MIM). MIM genes encode not only mainly photosynthetic or plastid membrane proteins but also a diverse set of plastid processes, genes of unknown function, mitochondrial, and cellular processes. Transcription of a photosystem II gene (*psb29*) in all lineages implies a nonphotosynthetic function for this and other genes retained in mycoheterotrophs. Nonphotosynthetic plants enable novel insights into gene function as well as gene expression shifts, gene loss, and convergence in nuclear genomes.

**Key words:** mycoheterotrophy, photosynthesis, phylogenomics, transcriptomics, relaxed purifying selection, BUSCO.

## Significance

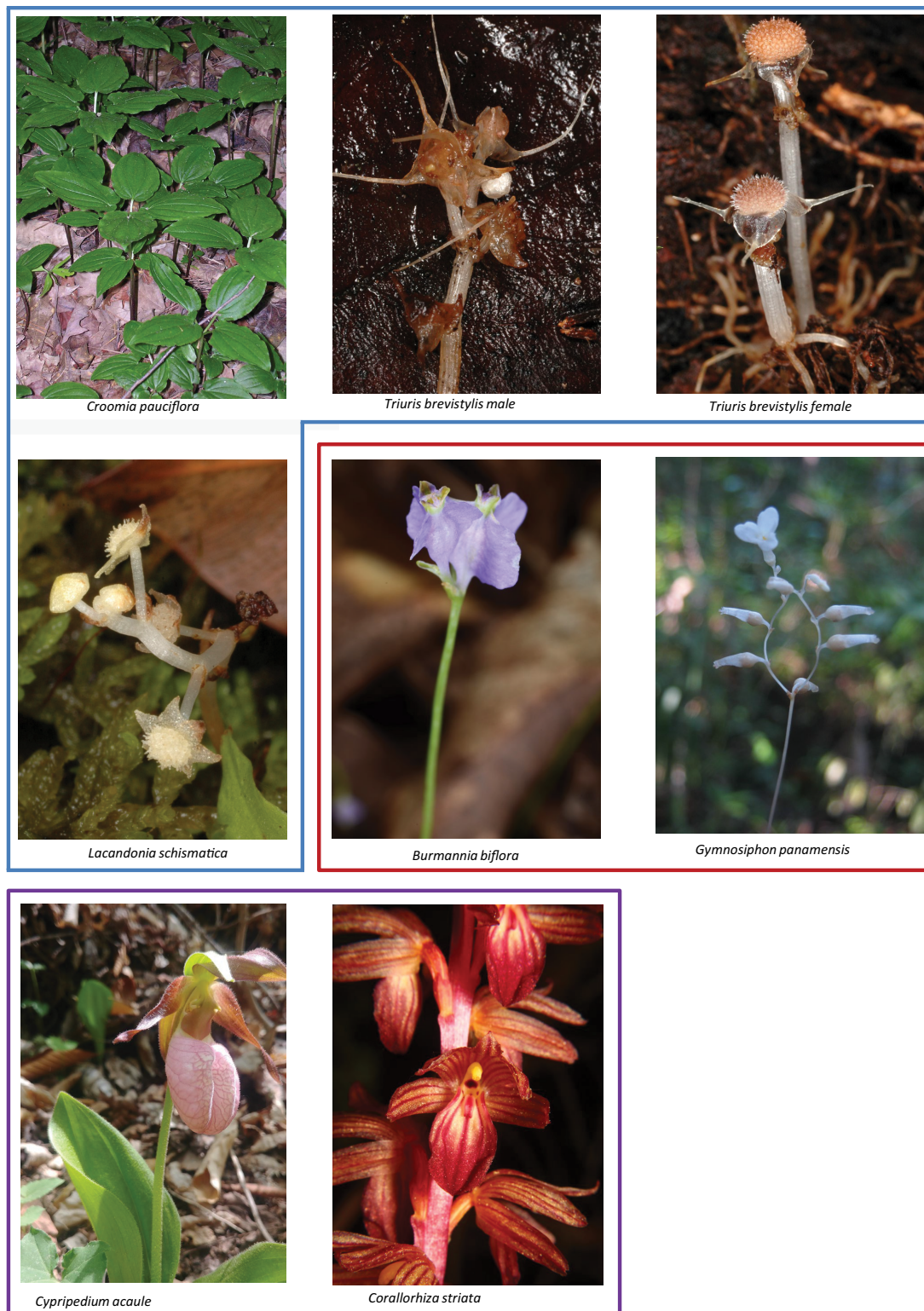
Although most plants are photosynthetic, species known as mycoheterotrophs obtain their sugars and other nutrients through parasitism of fungal symbionts. Little is known about the evolution of nuclear genomes and gene expression or their dynamics in mycoheterotrophs, despite abundant evidence of gene space downsizing and greatly increased rates of molecular evolution in their plastid genomes. Nuclear genes from transcriptomes of mycoheterotrophs provide a robust phylogenetic platform for comparative genomics, regardless of photosynthetic status. We find that five species in three independent lineages have progressively and convergently lost similar sets of expressed genes. *Missing in Mycoheterotroph* genes include photosynthetic genes and others not previously implicated in photosynthesis, while retained photosynthetic genes have additional function(s). We show that mycoheterotrophs are valuable evolutionary genetic transitions for understanding nuclear gene evolution and function.

## Introduction

Mycorrhizal interactions are generally beneficial to fungal and green plant partners: plants provide photosynthate to the fungi, and mycorrhizal fungi enhance the uptake of water and nutrients by expanding the reach and density of the plant root system in the rhizosphere (Atul-Nayyar et al. 2009; Hoffmann et al. 2009; Smith and Read 2010). Mycoheterotrophic plants rely on fungi for all nutritional needs; some remain capable of photosynthesis (partial mycoheterotrophs), while others have lost photosynthetic function (full mycoheterotrophs) and are likely fully dependent upon their fungal associates. Many mycoheterotrophic species are “epiparasites,” obtaining carbon from pre-existing mycorrhizal mutualisms between fungi and photosynthetic plants (Leake 1994; Rasmussen 1995; McKendrick et al. 2000; Taylor et al. 2002; Bidartondo 2005; Taylor et al. 2013), although a few rely on saprophytic or pathogenic fungi. Fungal parasitism, mycorrhizal mutualism, and mycoheterotrophy have been described as a continuum along a spectrum of intimate ecological interactions between plants and fungi (Bronstein 1994; Waterman and Bidartondo 2008; Merckx, Bidartondo, et al. 2009), and fully mycoheterotrophic plant species have completely lost the ability to obtain energy through photosynthesis.

Mycoheterotrophy has evolved independently in diverse, land-plant lineages. Merckx, Mennes, et al. (2013, table 5.1) summarized the origins of mycoheterotrophy in land plants, concluding that full mycoheterotrophy has evolved independently at least 50 times, with representatives in five orders and seven families of monocots, three families of

eudicots, one conifer, and one liverwort. Mycoheterotrophic species are morphologically distinct from autotrophs, growing to reproductive maturity as diminutive non-green plants with little to no chlorophyll; leaves are commonly absent or greatly reduced, plants typically exhibit little to no vegetative branching (fig. 1), and in some cases, loss or reductions of roots and stomata (Leake 1994; Bidartondo 2005; Merckx, Smets, et al. 2013). The majority (88%) of mycoheterotrophic species are monocots (Imhof 2010). Three monocot families (Corsiaceae, Thismiaceae, and Triuridaceae) include only full mycoheterotrophs, while Burmanniaceae, Orchidaceae, Iridaceae, and Petrosaviaceae contain both partially and fully mycoheterotrophic species. All Orchidaceae are “initial mycoheterotrophs,” living as mycoheterotrophic seedlings (i.e., “protocorms”), which eventually develop photosynthetic tissues in most species. Approximately 1–2% of the >20,000 currently described orchid species, however, retain juvenile mycoheterotrophy into reproductive maturity as leafless, putatively full or nearly full mycoheterotrophs (Freudenstein and Barrett 2010; Merckx, Mennes, et al. 2013; Jacquemyn et al. 2019). The orchids were estimated to have experienced a minimum of 30 independent losses of photosynthesis, although recent studies with dense intra-generic sampling suggest that this may be a gross underestimate (Barrett et al. 2014; Barrett and Kennedy 2018; Barrett et al. 2019). Burmanniaceae includes eight genera, all of which are heterotrophic except *Burmannia*, which includes fully mycoheterotrophic species, a few partial mycoheterotrophs, and some fully autotrophic members (Merckx et al. 2013). Thirty-seven of 56 species of



**FIG. 1.**—The seven focal monocot taxa used in the study. Blue box clockwise: Autotrophic *Croomia pauciflora* and related mycoheterotrophs, *Triuris brevistylis* (male and female) and *Lacandonia schismatica*. Red box: Autotrophic *Burmannia biflora* and related mycoheterotroph *Gymnosiphon panamensis*. Purple box: Autotrophic orchid *Cypripedium acaule* and related mycoheterotroph *Corallorhiza striata*. Photo credits: Joel R. McNeal (*Lacandonia*, *Triuris* and *Burmannia*), Craig F. Barrett (*Corallorhiza striata*, *Cypripedium acaule*), Esteban Martínez-Salas (*Gymnosiphon*), and Pete Pattavina (*Croomia pauciflora*).



*Burmannia* are full mycoheterotrophs, and this genus has experienced multiple losses of photosynthesis (Merckx et al. 2006, 2013). *Gymnosiphon* is the second largest genus in Burmanniaceae, with all 32 of its species mycoheterotrophic (Merckx et al. 2013).

Mycoheterotrophs are trophically distinct from parasitic plants in that they obtain nutrients from soil fungi, rather than parasitizing autotrophic green plants directly, but mycoheterotrophs and parasitic plants are similar in the sense that both forms of heterotrophy may eventually be accompanied by the loss of photosynthetic function. Reduction of gene content in the plastid genomes of parasitic and mycoheterotrophic plants and protists has been widely reported and reviewed, highlighting the value of nonphotosynthetic lineages of plants as valuable evolutionary genetic transitions for the study of plastid genome evolution and gene function (dePamphilis and Palmer 1990; Wolfe et al. 1992; Funk et al. 2007; McNeal et al. 2007, 2009; Delannoy et al. 2011; Wicke et al. 2011, 2013, 2016; Logacheva et al. 2014; Molina et al. 2014; Lam et al. 2015, 2018; Barrett et al. 2016; Angiosperm Phylogeny Group 2016; Feng et al. 2016; Lim et al. 2016; Lin et al. 2017; Graham et al. 2017; Hadariová et al. 2018; Joyce et al. 2018; Petersen et al. 2018; Wicke and Naumann 2018; Su et al. 2019). Conceptual models of plastome degradation have been proposed and tested, conforming to a more or less linear progression of gene loss by functional class that is remarkably consistent both within and between different parasitic and mycoheterotrophic lineages (Wicke et al. 2011; Barrett and Davis 2012; Barrett et al. 2014; Wicke et al. 2016; Graham et al. 2017; Wicke and Naumann 2018). Plastome reduction is first observed with the functional and/or physical loss of the NADH dehydrogenase-like (NDH-1) complex (*ndh* genes), which may precede loss of photosynthesis. This is followed by losses in “photosynthesis genes,” including: genes of Photosystems I and II; cytochrome (*pet* genes); Photosystem Assembly Factors (*ycf3* and *ycf4*, sometimes called *pag1* and *pag2*; [Wicke et al. 2011]); and other genes directly involved in photosynthesis (*cemA*, *ccsA*). The ribulose-1,5-bisphosphate carboxylase oxygenase (RuBisCo) large subunit gene (*rbcl*) is functionally retained in some nonphotosynthetic lineages (e.g., Delavault et al. 1996; Wolfe and dePamphilis 1998; Leebens-Mack and dePamphilis 2002; Wicke et al. 2011, 2016; Graham et al. 2017), supporting initial retention of a nonphotosynthetic function for this key photosynthesis gene in lipid biosynthesis (Schwender et al. 2004). Concurrently with photosynthesis genes, degradation occurs in the plastid-encoded RNA Polymerase (*rpo* genes), the main function of which appears to be transcription of photosynthesis-related genes in the chloroplast (Hajdukiewicz et al. 1997; Ishizaki et al. 2005; Borner et al. 2014; Yu et al. 2014). Following photosynthesis and *rpo* gene loss are members of the thylakoid

ATP synthase (*atp* genes); in addition to being essential for photosynthesis, these genes likely play other roles such as maintaining a proton gradient across the thylakoid membrane, which is likely essential for plastid function (Kamikawa et al. 2015; Graham et al. 2017; Barrett and Kennedy 2018; Hahn et al. 2018; Barrett et al. 2019). Along with these losses, housekeeping genes experience degradation, including plastid ribosomal protein genes (*rpl*, *rps*), initiation factor (*infA*) and transfer RNA genes (*trn*), and in some lineages, *clpP*, *matK*, *ycf1*, *ycf2*, and *accD* (acetyl coA-carboxylase).

Plastomes of mycoheterotrophic Ericaceae have lost all photosynthesis genes and contain only housekeeping genes plus a protease subunit (*clpP*) and acetyl-CoA carboxylase subunit D (*accD*). Plastomes of the holoparasitic plant *Balanophora* have lost all genes except for a few housekeeping genes (Su et al. 2019), and no plastome has been detected in the extreme endoparasites *Rafflesia lagascae* (Molina et al. 2014) or the fully sequenced *Sapria himalayana* (Cai et al. 2021). All (or most) of the 11 plastid *ndh* genes have been reported to be lost from the sequenced plastomes of most mycoheterotrophs (Graham et al. 2017; Wicke and Naumann 2018), as well as from some green orchids and several other green, photosynthetic lineages (Lin et al. 2017) including carnivorous plants (Wicke et al. 2014, 2016; Silva et al. 2016; Nevill et al. 2019; Lin et al. 2021). Fully mycoheterotrophic monocots have also experienced losses of most plastid genes except some housekeeping genes (Petersen et al. 2018) and core bioenergetic genes (Lam et al. 2018). However, the long-term retention of one or more non-bioenergetic genes and bifunctional *trnE* (which is involved in heme biosynthesis in addition to its role in plastid translation) may explain the long-term evolutionary retention of the plastid genome (e.g., Graham et al. 2017; Klimpert et al. 2022). In both parasitic and mycoheterotrophic species, plastome degradation in heterotrophic plants is typically associated with the elevation of nucleotide substitutions (McNeal et al. 2007; Lam et al. 2015; Feng et al. 2016; Wicke et al. 2016; Barrett et al. 2018), genomic rearrangements, decrease in GC content, and sporadic weakening of purifying selection (e.g., Wicke et al. 2016; Barrett et al. 2018; Joyce et al. 2018), although not all of these phenomena have been reported in all cases.

The nuclear genomes of green plants contain the vast majority of photosynthetic and photosynthesis-related genes except for RuBisCo large subunit (*rbcl*) and around 28 photosystem I, photosystem II, and other thylakoid protein genes that are plastid-encoded (Wicke et al. 2011). Recent publication of the organellar and nuclear genomes of the obligate mycoheterotroph, *Gastrodia elata* (Orchidaceae) reveals extensive loss of photosynthetic genes in the plastid and nuclear genomes (Yuan et al. 2018). Genome sequences and annotations of species of



the parasitic eudicot *Cuscuta* (Convolvulaceae) found that nuclear genes involved in light absorption and higher photosynthetic activity are largely lost, as are many genes involved in nutrient uptake (Sun et al. 2018, Vogel et al. 2018). Complete genome sequence and annotation of an extreme nonphotosynthetic plant, the internal stem parasite *Sapria himalayana* (Rafflesiaceae), has lost 43% of conserved plant genes, and has apparently entirely lost its plastid genome (Cai et al. 2021). Comparative analysis of gene loss from the sequenced genomes of three parasites, *Sapria*, *Cuscuta*, and the photosynthetic hemiparasite *Striga hermonthica* (Orobanchaceae; Yoshida et al. 2019) reveal convergent losses of photosynthesis, defense, and stress response genes in the three independent parasite lineages (Cai et al. 2021).

Although convergent gene loss has been a predominant theme in understanding evolutionary shifts in parasite and mycoheterotroph gene sets, selective retention of some genes has been observed. Transcriptomic investigation of parasitic species within the eudicot genera *Triphysaria*, *Striga*, and *Phelipanche* (Orobanchaceae) indicated that chlorophyll biosynthesis pathway genes were still transcribed and evolutionarily constrained, while Photosystem I and II reaction center protein genes were missing from transcript profiles of the nonphotosynthetic parasite *Phelipanche aegyptiaca* (Wickett et al. 2011). These findings suggested that low levels of chlorophyll pigment might have a nonphotosynthetic function in at least this nonphotosynthetic plant (Wickett et al. 2011). A recent study investigating the impact of mycoheterotrophy in orchids found the retention of some genes involved in photosynthesis as well as reprogramming of the expression of genes involved in fatty acid and amino acid biosynthetic pathways (Jakalski et al. 2021).

Despite their many missing genes, relaxed selective constraints, and accelerated rates of sequence substitution resulting in sometimes very long branches on phylogenetic trees, plastid genomes have been used to resolve the phylogenetic positions of several problematic mycoheterotrophic taxa. For example, phylogenetic analyses of plastid genes and ribosomal DNA genes resolved the placement of Corsiaceae within Liliales (Mennes et al. 2015; Lam et al. 2018), and plastid genes and genomes were also used to resolve the placement of Triuridaceae in Pandanales (Lam et al. 2015, 2018). Three commonly retained plastid genes in mycoheterotrophic plants (*accD*, *clpP*, and *matK*) were used to generate broad-scale phylogenetic estimates of relationships in the monocots that included representative mycoheterotrophic species (Lam et al. 2016), but highly accelerated rates of molecular evolution in many (but not all) mycoheterotroph lineages, and low support values, were a significant challenge. However, Givnish et al. (2018) and Lam et al. (2018) successfully placed large numbers of mycoheterotrophic taxa across monocots using plastome

sequence data and maximum likelihood analyses. Still, plastomes may not always be the best candidates to be used for phylogenomic estimates of mycoheterotrophic or parasitic plants because the photosynthesis genes may be lost or evolving very rapidly relative to other lineages, potentially yielding misleading results (Lam et al. 2018). By contrast, the use of mitochondrial genomes, whose genes are both slowly evolving in general and less rate-elevated in mycoheterotrophic lineages in particular, can help place lineages that are susceptible to long-branch attraction issues for plastid data (e.g., Merckx, Bakker, et al. 2009; Klimpert et al. 2022; Lin et al. 2022).

Nuclear rDNA genes are universally available markers for phylogenetic inference of both photosynthetic and nonphotosynthetic plants (Nickrent, Duff, et al. 1997, Nickrent, Yan, et al. 1997; Lemaire et al. 2011; Nickrent 2020). Large- and small-subunit rDNAs and intergenic spacers have often provided critical evidence for phylogenetic placement of nonphotosynthetic plants (e.g., Merckx et al. 2006; Merckx, Bakker, et al. 2009; Lemaire et al. 2011; Mennes et al. 2013; Nickrent 2020). However, the number of available loci is limited, and extreme rate heterogeneity and compositional bias are also common in such plants (Nickrent, Duff, et al. 1997; Nickrent, Yan, et al. 1997; Lemaire et al. 2011; Su et al. 2019). Single-copy nuclear genes, which are represented by hundreds of unlinked loci distributed throughout the genome, have shown tremendous potential for phylogenomic analyses of nonphotosynthetic plants (Duarte et al. 2010; Zimmer and Wen 2012; Naumann et al. 2013; Wickett et al. 2014; Baker et al. 2022), but such analyses have only recently been introduced for mycoheterotrophic species (Yuan et al. 2018; One Thousand Plant Transcriptomes Initiative 2019).

Here, we demonstrate that a phylo-comparative, transcriptomic approach provides unprecedented power to resolve traditionally difficult higher-level monocot relationships (including placement of mycoheterotrophic lineages) and to concurrently infer patterns of genome evolution in association with the mycoheterotrophic lifestyle. We focus on a large collection of inferred single-copy nuclear genes obtained from available genomes and new transcriptome resources generated by the Monocot Tree of Life Project (MonAToL) (Givnish et al. 2010, 2018) to reconstruct a robust phylogenetic hypothesis of relationships among monocot orders, with a sampling focused on three mycoheterotrophic lineages within Pandanales and Dioscoreales and one mycoheterotrophic orchid species. Our sampling includes all five families of Pandanales and three families of Dioscoreales, including four fully mycoheterotrophic species (i.e., *Triuris brevistylis* and *Lacandonia schismatica* in Triuridaceae, *Gymnosiphon panamensis* in Burmanniaceae, and *Corallorhiza stricta* in Orchidaceae; fig. 1) and related photosynthetic species. We then use these phylogenetic trees to investigate changes in the

representation of a set of widely conserved single-copy genes, as well as nuclear-encoded genes for chlorophyll biosynthesis, photosynthesis, and light harvesting antennae proteins in the mycoheterotrophs relative to their autotrophic relatives. In addition, we conduct molecular evolutionary analyses to test for shifts in selective constraint acting on chlorophyll biosynthesis and photosynthesis genes associated with independent origins of mycoheterotrophy.

## Results

### Overview of Genome Scaffolds, Transcriptome Datasets, and Assemblies

From the 12 genomes used for orthogroup circumscription, we cleaned and validated between 19,623 to 39,049 coding regions (CDRs, average 30,957 per species) out of which 16,120 to 27,929 were sorted into gene families of size two or larger. A total of 24,873 orthogroups were obtained, including 2,856 that were unsupported singletons (only in one species) and 602 that were single-copy in all 12 of the scaffold species. On average, 71.52% of the CDRs were classified into gene families of two or more genes (supplementary tables S1–S3, Supplementary Material online).

The Illumina RNASeq dataset consisted of a total of 761,330,504 pairs of reads with an average of 33,101,326 reads per transcriptome. Following diagnostic cleaning, and removal of adaptors, contaminants, and low quality and short fragments, a total of 687,995,060 forward reads and 663,934,180 reverse reads were recovered for assembly. An average of 88.8% of the reads passed the diagnostic tests to be used for de novo assembly with Trinity. A total of 2,263,791 primary contigs were obtained from the assembly with the number of contigs per transcriptome ranging from 30,689 to 308,671. The average  $N_{50}$  length ( $L_{50}$ ) of the transcripts (largest contig length in an assembly at which 50% the bases are in longer contigs) ranged from 873 bases to 2,560 bases, averaging 1,506 bases. Following translation with ESTScan (Iseli et al. 1999) and removal of exact duplicates, short assemblies, and perfect subsequences of longer assemblies, a total of 1,438,406 protein sequences was retained with the numbers ranging from 26,320 to 158,408 and averaging 26,320 per transcriptome. The  $N_{50}$  of these translated and cleaned contigs averaged 1,188 bases with the number ranging between 699 and 1,872 bases (details in supplementary table S2, Supplementary Material online).

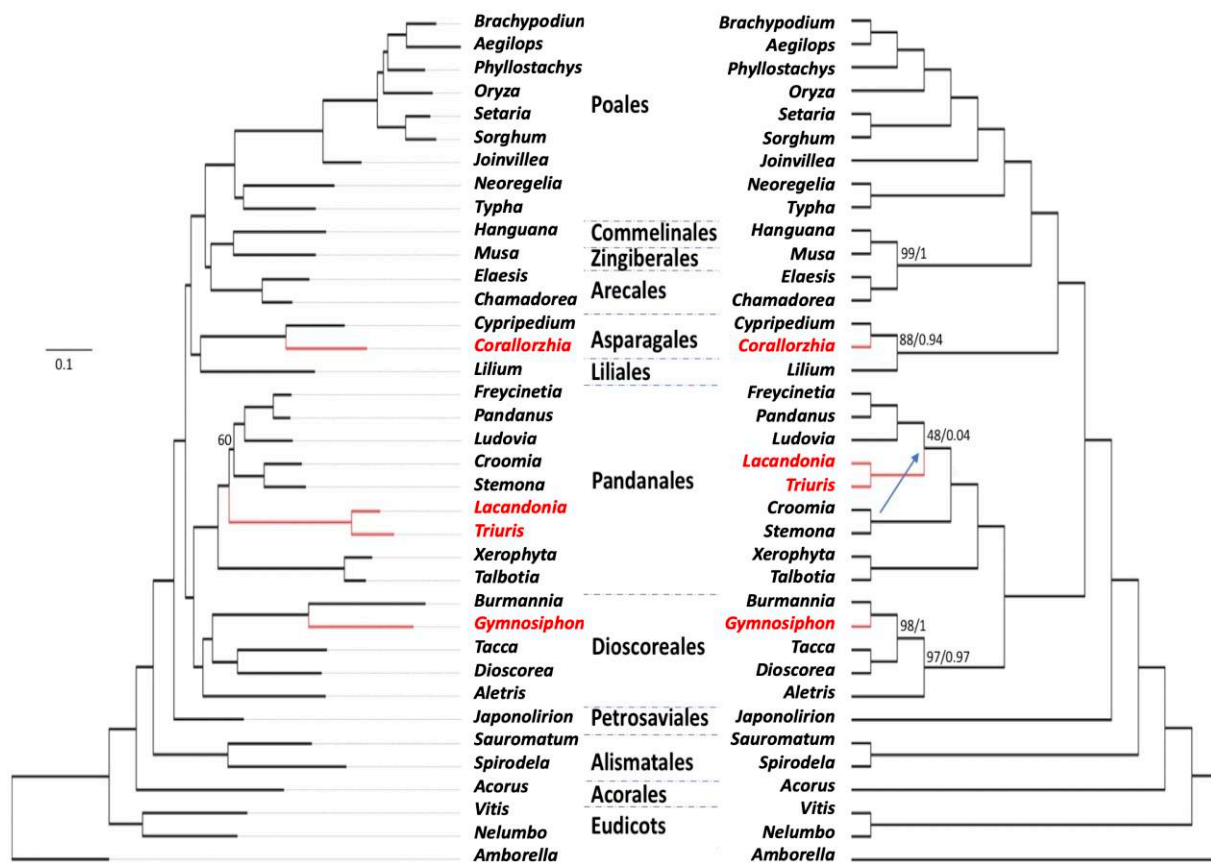
### Single-Copy Gene Representation and Summary

Following global classification of the transcriptomes into the 12-genome orthogroup scaffolds with HMMER (Eddy

1998, 2011) and BLASTp, we determined the occupancies of the transcriptome data in all of the 602 conserved single-copy gene orthogroups (supplementary table S2, Supplementary Material online; supplementary fig. S1, Supplementary Material online). On average, 526 of 602 orthogroups were captured per transcriptome, with the number ranging from 277 (*Triuris*) to 590 (*Sauromatum*). Following removal of potentially duplicated genes and posttranslational cleaning, an average of 394 genes per transcriptome (range 221–491) were retained as high confidence single-copy genes for phylogenomic analysis (supplementary fig. S1, Supplementary Material online). Only two of 24 transcriptomes had fewer than 400 of the single-copy genes. All four full mycoheterotrophs had lower representation of single-copy genes in their transcriptomes relative to their most closely related autotrophic species (supplementary table S2, Supplementary Material online). Single-copy orthogroups captured by each of the mycoheterotroph taxa compared to closest green relatives were: *Corallorhiza* versus *Cypripedium* (501 vs. 572; 12.4% reduction); *Gymnosiphon* versus *Burmannia* (380 vs. 528; 27.5% reduction); *Triuris* (410); and *Lacandonia* (412) versus *Croomia* (587; 29.8% and 30.1% reduction, respectively). Benchmark universally conserved orthologous (BUSCO) assessment of the transcriptomes along with the genomes showed that the number of missing BUSCOs, of 1372 total, ranged from 2 (*Brachypodium distachyon*) to 396 (*Elaeis guineensis*) for the scaffold genomes and from 34 (*Croomia pauciflora*) to 472 (*Joinvillea ascendens*) for the transcriptomes. The reduction pattern for the number of BUSCO genes in the full mycoheterotrophs as compared to their closest green relatives was similar to that seen for single copy genes (above) (supplementary fig. S2, Supplementary Material online). The fully sequenced and annotated genome of the mycoheterotrophic *Gastrodia elata* (Yuan et al. 2018), showed BUSCO representation comparable to transcriptome-based estimates from the other mycoheterotrophs (322 BUSCOs missing from the *Gastrodia* genome, 337 missing from *Corallorhiza* and *Gymnosiphon*, 390 missing from *Lacandonia* and 393 missing from *Triuris*).

### Phylogenetic Tree of the Complete Dataset

We estimated phylogenies based on the 12 scaffold genomes to assess the quality of the gene-family classification database and the utility of single-copy genes for phylogenomic analysis. Phylogenetic trees from the genome scaffolds were constructed from the 602 core single-copy genes using coalescent as well as concatenation methods and are shown in supplementary figure S3, Supplementary Material online. The phylogenetic trees were topologically identical, and all the branches had



**FIG. 2.**—Phylogenetic tree of the 12 genomes and 25 transcriptomes using all 602 putatively single copy genes DNA sequences, partitioned by codon position. Mycoheterotroph lineages have been marked in red. Left: RAxML concatenation tree with inferred branch lengths. Right: ASTRAL coalescent tree. Numbers in black indicate BS (as percentages), followed by local posterior probabilities. All support values are 100% or 1 unless shown. Blue arrow indicates the only branch (Triuridaceae) with a different placement between analyses. Polytomy is rejected (Sayyari and Mirarab 2018) for all nodes but two; all nonzero *P* values are shown in brackets.

100% bootstrap support (BS), regardless of the type of analysis.

The inferred phylogenies from all 37 taxa (fig. 2) for both the RAxML (concatenated) and ASTRAL (coalescent) trees were identical at the level of orders, with just one poorly supported exception at the family level: the position of Triuridaceae within Pandanales. All branches except one had 100% BS in the RAxML tree, and most of the branches had 100% BS and a local posterior probability of 1 in the ASTRAL tree (fig. 2). Both analyses strongly supported placing Pandanales sister to Dioscoreales, which as a group were sister to the clade formed by the commelinids (Arecales, Commelinales, Poales, Zingiberales) and a Liliales + Asparagales clade. Considering monocot relationships as a whole, Commelinales were strongly supported as sister to Zingiberales, and this clade was sister to Arecales, with Poales being the sister group of all of these. As has been seen in other recent analyses based on nuclear gene sequence data (One Thousand Plant Transcriptomes Initiative 2019), Liliales was placed as sister to

Asparagales in the trees based on both coalescence and concatenation. Plastome-based studies place Liliales and Asparagales and successive sister lineages to the commelinids (Givnish et al. 2010, 2018; Soltis et al. 2011; Lam et al. 2018), but we found robust support for the Liliales + Asparagales clade in all of our analyses (fig. 2; supplementary figs. S4–S6, Supplementary Material online; supplementary table S4, Supplementary Material online).

We recovered strongly supported hypotheses for relationships among families of Dioscoreales that were concordant between the two analyses (fig. 2). Burmanniaceae (*Burmannia* and *Gymnosiphon*) were strongly placed as sister to Dioscoreaceae (*Tacca* and *Dioscorea*), and Nartheciaceae (*Aletris*) was the next successive sister to this lineage. Within Pandanales, Pandanaceae (*Freycinetia* and *Pandanus*) was strongly placed with Cyclanthaceae (*Ludovia*) in both analyses, but this clade was sister to Triuridaceae (*Triuris* and *Lacandonia*) in the ASTRAL analysis yet sister to Stemonaceae (*Stemona* and *Croomia*) in the concatenated analysis. The polytomy test for the ASTRAL



tree (Sayyari and Mirarab 2018) showed that the null hypothesis of polytomy was rejected with high confidence ( $P < 0.01$ ) for all nodes except the ones joining Liliales (*Lilium*) to the Asparagales (*Cypripedium* and *Corallorhiza*) ( $P = 0.12$ ) and the branch joining *Aletris* to the other Dioscoreales ( $P = 0.066$ ) (fig. 2). Phylogenetic analysis of 596 genes using amino acid sequences yielded phylogenetic trees that were topologically identical between RAxML and ASTRAL analysis and to the DNA RAxML tree in figure 2 (supplementary fig. S4, Supplementary Material online).

### Assessment of Gene Tree–Species Tree Discordance

A test for the stability of internal branches using PhyloStab (Sheikh et al. 2013) detected no unstable branches in the concatenated or the coalescent tree. Further, no “rogue” taxa were identified in RogueNaRok (Aberer and Stamatakis 2011) analyses of the concatenation or coalescent trees.

Analysis of conflict and concordance showed that there was a moderate to high level of discordance between the individual gene trees and the species tree (fig. 3). In addition to gene tree conflict on the poorly supported ASTRAL resolution for the placement of Triuridaceae (fig. 2), the placement of Asparagales sister to Liliales and the placement of Arecales sister to Commelinales–Zingiberales also exhibited substantial conflict among gene trees, as expected under the coalescence model when there is little time between branching events.

### Expression of Genes for Chlorophyll Biosynthesis and Photosynthetic Reaction Centers

Detection of genes involved in chlorophyll biosynthesis was based on a normalized mapping score, transcripts per million (TPM). Autotrophic taxa consistently showed higher level of expression for genes involved in chlorophyll biosynthesis and photosynthetic reaction centers. Paired *t*-tests of expression levels between a mycoheterotroph and its closest green relative showed significant expression difference between such pairs (one-tailed test, *P*-values for *Lacandonia–Croomia*, *Gymnosiphon–Burmanna*, and *Corallorhiza–Cypripedium* were  $2.4 \times 10^{-8}$ ,  $5.9 \times 10^{-7}$ , and 0.02, respectively). There was no expression difference between these pairs at whole transcriptome levels. Chi-square tests of global expression differences were insignificant except for *Cypripedium–Corallorhiza* pair ( $P = 0.01$ ), where the mycoheterotrophic *Corallorhiza* seemed to have higher gene expression in general compared to the closest green relative *Cypripedium*, even when weighted for number of transcripts detected.

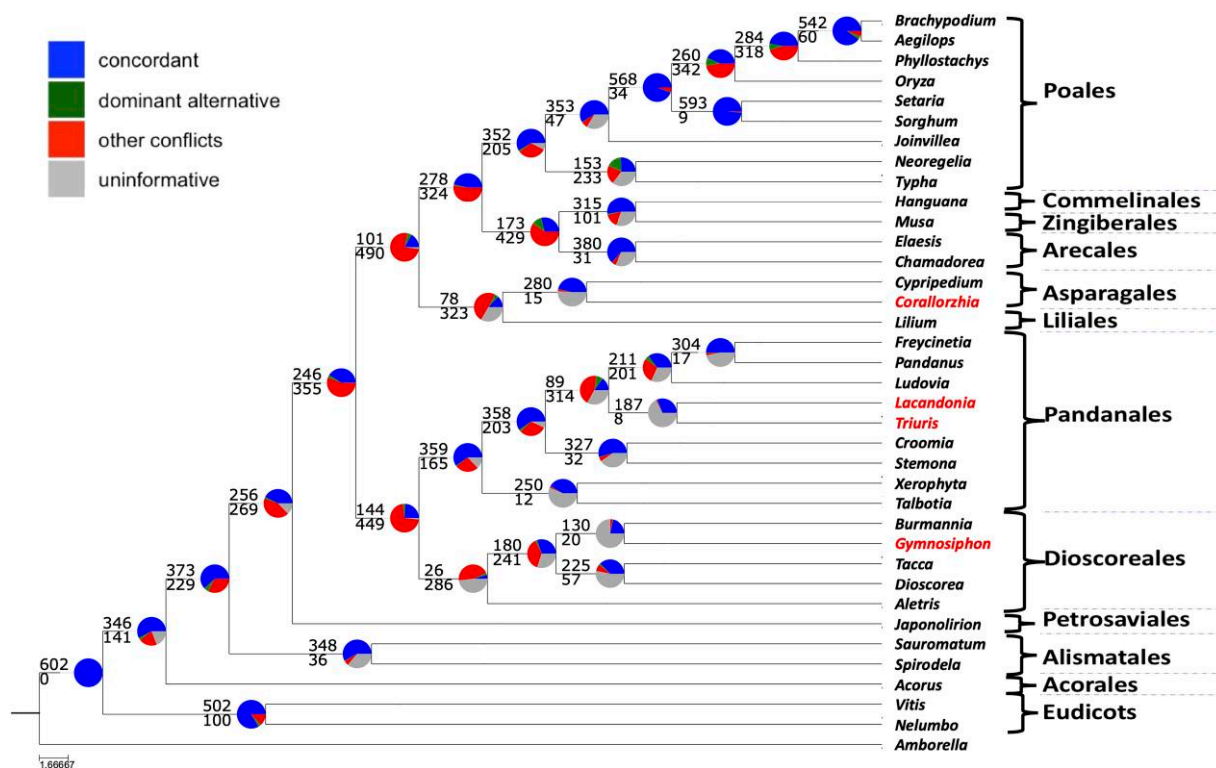
All the genes involved in the common heme and chlorophyll biosynthesis pathway leading up to the formation of protoporphyrin IX were detected in all the transcriptome samples. However, the six enzymes involved in the terminal

(dedicated) pathway for converting protoporphyrin IX to chlorophyll showed a different pattern of expression between mycoheterotrophs and their autotrophic sister lineages (fig. 4 and supplementary table S5, Supplementary Material online). Whereas all six genes were detected from *Corallorhiza*, and all were absent from the *Triuris* and *Lacandonia* transcriptomes, *Gymnosiphon* showed an intermediate level of transcript abundance. Although the two genes involved in the dedicated steps of chlorophyll biosynthesis pathway (magnesium chelatase and magnesium protoporphyrin 9 methyl reductase) were detected along with their primary functional domains, only a small fragment of chlorophyll synthase was detected in the *Gymnosiphon* transcripts, and it contained no functional domain.

The reaction center proteins and light harvesting antennae complex (LHC) genes are involved in binding and stabilizing chlorophyll. In the absence of chlorophyll biosynthesis, it is likely that these genes are no longer essential. We found that the LHC and the Photosystem I and Photosystem II genes showed similar trends (fig. 5A and B and supplementary tables S6–S7, Supplementary Material online). Transcripts of all genes were detected in the photosynthetic “control” plants, 16 of 23 were detected in both *Corallorhiza* and *Gymnosiphon* and one or two in *Lacandonia* (*psb29* only) and *Triuris* (*psb29* and trace levels of *psbQ*).

As a departure from the typical absence or near absence of PSI and PSII transcripts in the mycoheterotrophs, *psb29* was detected in all 7 mycoheterotrophic taxa analyzed. Substantial expression levels were observed for *Lacandonia*, *Gymnosiphon* and *Corallorhiza*. The gene *psbQ* was detected in all mycoheterotrophic taxa except *Lacandonia*. Consistent with these findings, only *psbQ* and *psb29* were reported among these 23 target genes in the annotation of the *Gastrodia* genome (Yuan et al. 2018). We then examined publicly available transcriptome datasets for two other fully mycoheterotrophic plants, the eudicots *Sarcodes sanguinea* and *Monotropa uniflora* (1kp assemblies SERM and LRTN, respectively; Matasci et al. 2014); transcript assemblies for *psb29* were identified from both species. Thus, all seven surveyed mycoheterotrophic plants retain and transcribe *psb29* despite their nonphotosynthetic habit and near or total loss of other PSI and PSII protein genes.

For any LHC, PSI, or PSII gene detected in the mycoheterotrophs, with the exception of *psb29*, the TPM levels were always lower than detected in their green relatives (fig. 5 and supplementary tables S6–S7, Supplementary Material online). Among the mycoheterotrophs, *Gymnosiphon* had trace read levels for five genes (two LHC, one PSI and two PSII) while no reads were detected in *Corallorhiza*. However, whenever *Corallorhiza* and a mycoheterotroph both expressed any of the 21 LHC, PSI, or PSII genes, the



**FIG. 3.**—Summary of concordance and conflict in species tree predicted with ASTRAL III. Numbers above and below each branch indicate the number of homologs concordant and conflicting with that clade, respectively. The colors in pie chart indicate the number of homologs that support the clade, the main alternative for the clade, all other remaining alternatives or that are uninformative. Mycoheterotrophic taxa are indicated in red.

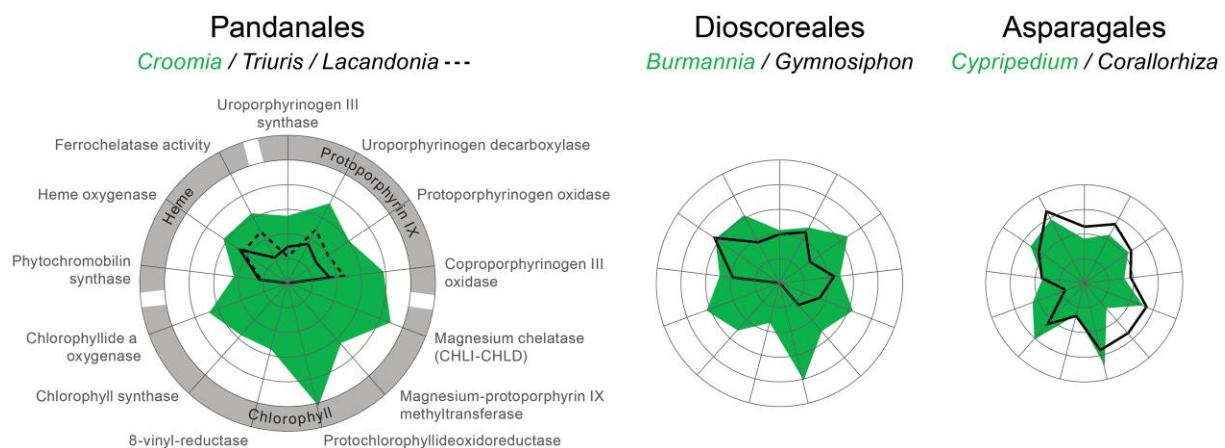
highest levels were always detected in *Corallorhiza*. We also note that *Burmannia biflora* and *Cypripedium acaule* generally showed transcript levels for genes in these pathways that were intermediate between the full autotroph *Croomia* and the trace levels (if any) displayed by full mycoheterotrophs. *Burmannia* transcript levels were consistently higher than *Cypripedium*.

### Comparison of Missing BUSCOs

The four mycoheterotroph transcriptomes used in this study and the *Gastrodia elata* genome were all missing 174 of the total 1372 universal benchmark (BUSCO) genes that were detected in *Arabidopsis* and used for subsequent analyses (fig. 6). We call this set of 174 genes *Missing in Mycoheterotrophs* (MIM) to refer to the set of widely conserved plant genes that are undetected or lost from all of the mycoheterotrophs studied. A further 72 BUSCOs were also shared missing among *Triuris*, *Lacandonia*, *Gymnosiphon*, and *Gastrodia*, but not *Corallorhiza*, while 579 BUSCO genes were missing in one or more of the mycoheterotrophic taxa. Some overlap in gene content among plants is expected due to phylogenetic correlation (*Triuris* and *Lacandonia* in Dioscoreales) because genes lost in a common ancestor are expected to be missing from all descendant lineages. Many BUSCOs (619) were missing

from at least one of five green species sampled in the Pandanales-Dioscoreales clade, but in contrast to the mycoheterotrophs, only 12 were missing from all five photosynthetic taxa (fig. 6). To test this rigorously, we considered the null hypothesis that correlation in the number of missing genes across the five mycoheterotrophs is equal to that expected when accounting for phylogenetic relationships. Using the five autotrophic taxa to estimate the level of shared missing genes due to phylogenetic correlation, we found overwhelming evidence for rejecting our null hypothesis (likelihood ratio test comparing models with or without a taxon set (mycoheterotrophic vs. photosynthetic) to predict the all-missing property, and accounting for repeated measures for each gene:  $X^2 = 179.1$ , degrees of freedom [df] = 1,  $P$ -value  $< 10^{-15}$ ; supplementary table S8, Supplementary Material online). This result was not driven by more genes missing overall in one taxon set versus the other (likelihood ratio test to predict “missing in one or more of the five taxa”:  $X^2 = 2.9$ , df = 1,  $P$ -value = 0.9). This means that the pattern of missing BUSCO genes is far more similar among the mycoheterotrophs than expected by chance even when accounting for phylogenetic correlation.

The now well-established pattern of plastid gene loss and reduction of coding capacity is a dominant theme in



**Fig. 4.**—Transcription of heme and chlorophyll synthesis genes in four mycoheterotrophs and their related autotrophic taxa. Each radar plot shows the same 13 genes from generalized heme synthesis, protoporphyrin IX, and dedicated chlorophyll synthesis. Samples are *Triuris brevistylis* (solid line), *Lacandonia schismatica* (dotted line), *Gymnosiphon panamensis* (solid line), *Corallorhiza striata* (solid line), and related autotrophs *Croomia pauciflora*, *Burmanningia biflora*, and *Cypripedium acaule* (areas shown in green). Expression values are plotted as  $\log_2(\text{TPM} + 1)$ ; concentric inner circles are scaled from inner minimum to maximum for visualization of the full range of transformed expression values on each radar plot; untransformed numerical values are given in [supplementary table S5, Supplementary Material](#) online.

plastid genome evolution in mycoheterotrophic plants (e.g., Graham et al. 2017). In order to understand the likely functions of MIM genes, and gain insights into their convergent disappearance from independent mycoheterotroph lineages, we first extracted the *Arabidopsis* ID numbers for the BUSCOs corresponding to 174 undetected genes that were shared among *Lacandonia*, *Triuris*, *Gymnosiphon*, *Corallorhiza*, and *Gastrodia*, and then performed functional annotation and enrichment analysis with database for annotation, visualization, and integrated discovery (DAVID) (Huang et al. 2009a, 2009b). The functional annotations are presented in [supplementary table S9a, Supplementary Material](#) online and the enrichment analysis of these 174 genes against the background of all 1372 *Arabidopsis* BUSCOs is in [supplementary Table S9b, Supplementary Material](#) online. 145 of 174 (83.3%) genes were significantly enriched for the gene ontology *cellular component* direct term chloroplast ([supplementary table S9b, Supplementary Material](#) online). Other significantly enriched terms (chloroplast thylakoid membrane, chloroplast thylakoid, chloroplast thylakoid lumen, chloroplast stroma) indicate that genes with chloroplast photosynthetic activities predictably represent the largest portion of MIM genes in these nonphotosynthetic plants.

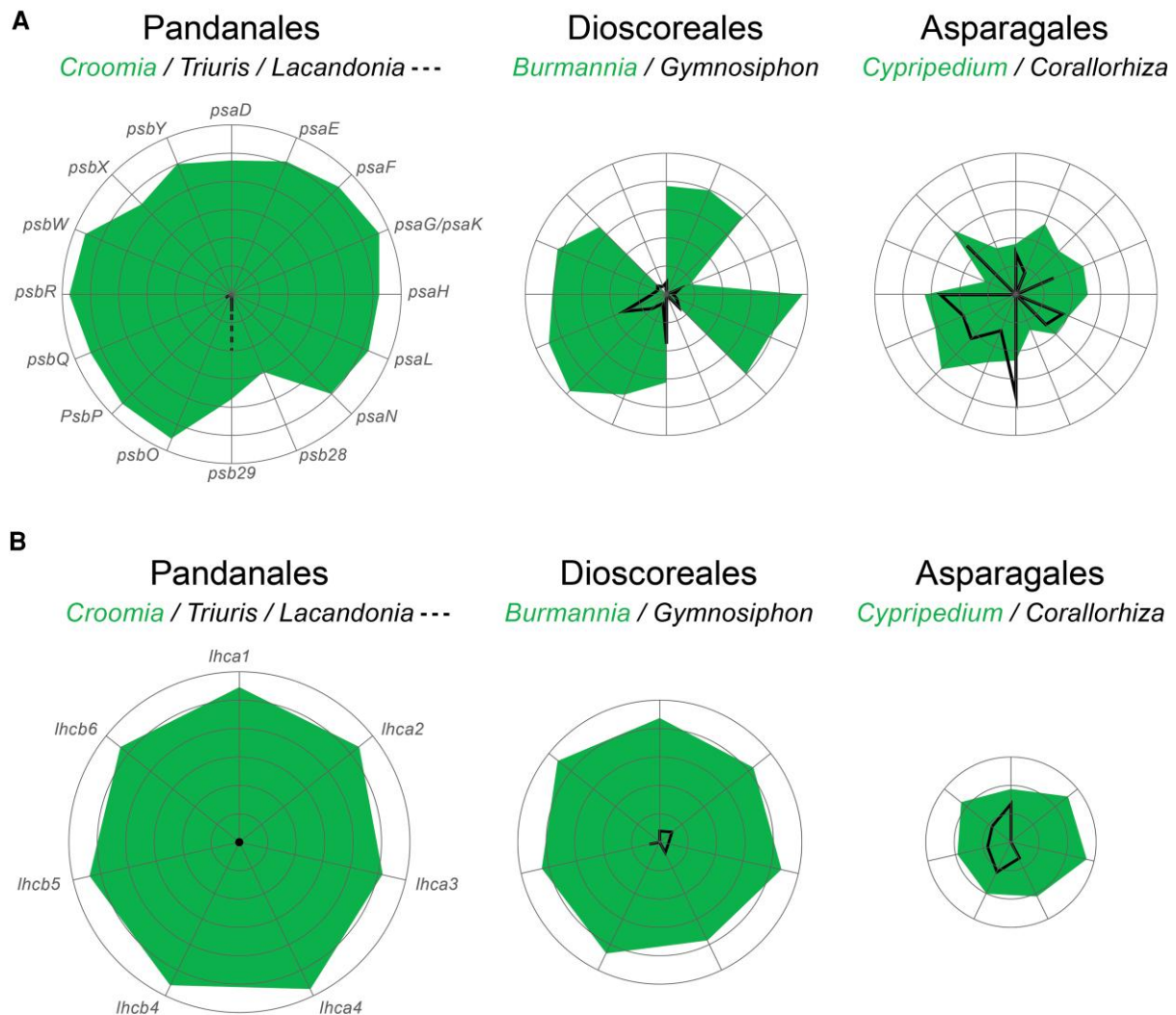
Functional annotation clustering of the 174 MIM genes produced highly significant clusters against the background of all BUSCOs ([supplementary table S9c, Supplementary Material](#) online). Annotation clustering of Uniprot keywords showed that cluster #1, with 86 (49.4%) of the genes was enriched with terms Thylakoid, Chloroplast, and Plastid. Additional smaller clusters include enrichment for uniprot keywords membrane, transmembrane, transmembrane helix, plastoquinone, nicotinamide

adenine dinucleotide phosphate (NADP), and quinone ([supplementary table S9c, Supplementary Material](#) online).

The loss of genes from the plastid genomes of mycoheterotrophic plants (plastid encoded RNA polymerase, *ndh*, and several genes of otherwise unknown function) not otherwise known to be specifically involved with photosynthesis provide important clues that their functions may include processes influencing or directly dependent upon photosynthetic activity. All but 21 of the 174 MIM genes were from BUSCOs contributing to the photosynthesis-thylakoid-membrane common theme ([supplementary table S9b, Supplementary Material](#) online, section D). These 21 other genes participate in a diverse array of functional annotations, including DNA repair, DNA-directed RNA polymerase, pentatricopeptide repeat (organelle RNA editing), transport, and widely varied physiological or biochemical functions. Three of these were annotated as *Domain of Unknown Function* or hypothetical protein.

The progressive loss of plastid genes from mycoheterotrophic and parasitic lineages led us to predict that nuclear gene loss may also be incomplete in younger mycoheterotrophs. Similar enriched categories were observed for the 72 genes that were detected in *Corallorhiza* but not detected in the other four mycoheterotrophs ([supplementary table S10a–S10b, Supplementary Material](#) online). These genes were enriched for the same terms (e.g., chloroplast, plastid, thylakoid), a pattern completely consistent with the theme seen from the 174 MIM genes. However, when the background for the enrichment clustering was changed to the 174 MIM genes, there were no significant clusters, meaning that *Corallorhiza* was likely experiencing a similar



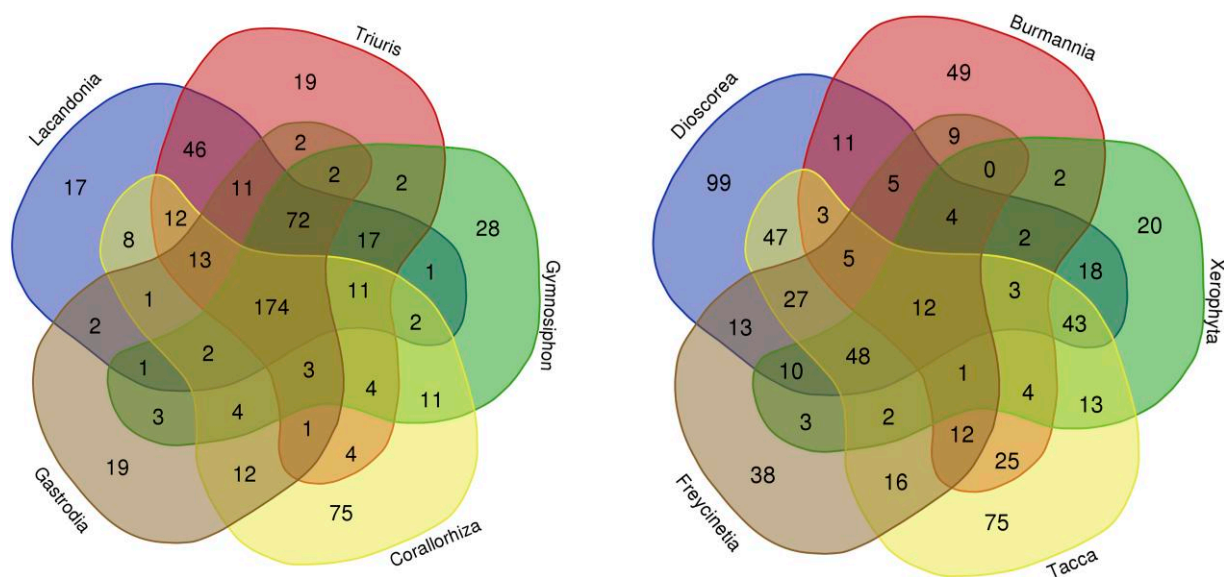


**FIG. 5.**—Transcription of A. Photosystem I and Photosystem II genes and B. light harvesting complex in four mycoheterotrophs *Triuris brevistylis*, *Lacandonia schismatica*, *Gymnosiphon panamensis*, *Corallorhiza striata*, and three related photosynthetic taxa *Croomia pauciflora*, *Burmannia biflora*, and *Cypripedium acaule*. Same color and line patterns as for figure 4. No expression was detected for light harvesting complex genes in *Lacandonia* or *Triuris*; while trace level of *psbQ* was detected in *Triuris*, and *psb29* was detected in both *Triuris* and *Lacandonia*. Expression values are plotted as  $\log_2(\text{TPM} + 1)$ ; untransformed numerical values are given in [supplementary tables S6–S7, Supplementary Material](#) online.

evolutionary trajectory as the other mycoheterotrophs, even though it retained more of the genes lost in the other, older mycoheterotroph lineages. This finding was further corroborated by the 75 BUSCOs that were detected in *Gymnosiphon*, *Lacandonia*, *Triuris*, and *Gastrodia* but not detected in *Corallorhiza* ([supplementary table S11a, Supplementary Material](#) online). This gene set yielded significant clusters when the test was performed against the background of all BUSCOs but no significant clusters were detected when the background was changed to the 174 MIM BUSCOs ([supplementary table S11b\\_Sections A, B, C, Supplementary Material](#) online).

### Signatures of Selection

Most of the genes with significant shifts in purifying selection came from mycoheterotrophs ([Table 1](#) and [supplementary table S12, Supplementary Material](#) online), though we also detected a relaxation of selection in Uroporphyrinogen II synthase and Uroporphyrinogen decarboxylase genes in *Cypripedium*. For genes of the dedicated steps of chlorophyll biosynthesis, relaxation of selection was seen four of six genes in *Corallorhiza*, and both of the retained genes were under relaxed selection in *Gymnosiphon*, whereas their green relatives displayed no relaxation. The rest of the genes in the dedicated



**FIG. 6.**—Comparison of shared and unique missing BUSCOs among the five full mycoheterotrophs (left) and five green Pandanales/Dioscoreales taxa (right) with comparable number of missing BUSCOs as the mycoheterotrophs. One hundred seventy-four MIM genes represent the set of genes lost or undetected from the *Gastrodia* genome and all of four mycoheterotroph transcriptomes.

**Table 1**

Tests of Shifts in Strength of Purifying Selection for Different Chlorophyll Synthesis-related Genes From the Mycoheterotrophs *Gymnosiphon*, *Triuris*, *Lacandonia* and *Corallorhiza*, and Their Green Relatives *Croomia*, *Burmannia*, and *Cypripedium*, Respectively

Stage	Gene Description	<i>Croomia</i>	<i>Triuris</i>	<i>Lacandonia</i>	<i>Burmannia</i>	<i>Gymnosiphon</i>	<i>Cypripedium</i>	<i>Corallorhiza</i>
Heme	Phytochromobilin synthase	No	Yes	Yes	No	No	No	No
	Heme oxygenase	No <sup>a</sup>	Yes	No	No	Yes	No	No
	Ferrochelatase activity	No	Yes	No	No	No	No	No
Protophyrin IX	Uroporphyrinogen II synthase	No	No	No	No	No	Yes	No
	Uroporphyrinogen decarboxylase	No	No	No	No	No	Yes	Yes
	Coproporphyrinogen III oxidase	No	No	No	No	No	No	No
Chlorophyll	Magnesium chelatase	No	N/A	N/A	No	Yes	No	Yes
	Magnesium protoporphyrin 9 methyltransferase	No	N/A	N/A	No	Yes	No	No
	Protochlorophyllide oxygenase	No	N/A	N/A	No	N/A	No	No
	8-vinyl reductase	No	N/A	N/A	No	N/A	No	Yes
	Chlorophyll synthase	No	N/A	N/A	No	N/A	No	Yes
	Chlorophyll a oxygenase	No	N/A	N/A	No	N/A	No	Yes

Yes/no indicate whether relaxation of purifying selection was significant for the branch. One gene in *Croomia* was detected with significantly increased purifying selection (No<sup>a</sup> in table). N/A: test not applicable, transcripts not detected for this gene. Significance test results are given in [supplementary table S12, Supplementary Material online](#).

chlorophyll synthesis pathway were undetected in the mycoheterotrophs. Of all the tests done, there was only one signature of intensification of selection, in the phytochromobilin synthase gene from *Croomia* ([supplementary table S12, Supplementary Material online](#)). Although some of these observations of shifts in the strength of purifying selection are potentially false positives (Fletcher and Yang 2010; Markova-Raina and Petrov 2011), there is an obvious difference in the trends among autotrophs and mycoheterotrophs, with mycoheterotrophs expressing fewer

pathway genes than green plants, and more of the retained genes evolving under relaxed selection than in photosynthetic relatives.

## Discussion

### Phylogenetic Relationships of Monocots and the Independent Origins of Mycoheterotrophic Lineages

We have leveraged genomes and transcriptomes to estimate the relationships among 11 of the 12 monocot

orders recognized by Givnish et al. (2018) and place three independent mycoheterotrophic lineages in Pandanales (*Lacandonia* and *Triuris*), Dioscoreales (*Gymnosiphon*), and Asparagales (*Corallorhiza*). We found that phylogenetic hypotheses based on more than 600 single-copy genes were very robust and independent of analytical method, protein versus DNA sequence, or variation in the number of single-copy genes used in the analysis. Overall, our phylogenomic results are nearly congruent with other broad-scale phylogenetic hypotheses for monocots based on plastomes or few gene datasets, but often with higher support values than from whole plastomes, and the results reflect the signal from hundreds of unlinked genes across a wide range of functional categories. The ordinal level relationship we recovered for Asparagales and Liliales differs from most recent analyses of plastid genes and genomes (Givnish et al. 2010; Givnish et al. 2018; Lam et al. 2018) and 17 genes from all three genomic compartments (Soltis et al. 2011) that resolve Liliales and Asparagales as successive sister lineages to the commelinid clade (Givnish et al. 2010; Givnish et al. 2018; Lam et al. 2018), but similar to what Zeng et al. (2014) and the One Thousand Plant Transcriptomes Initiative (2019) reported using 59 and 410 conserved single-copy nuclear genes, respectively. Although the taxon sampling for Asparagales and Liliales was very sparse in this study, a phylogenomic analysis with deeper sampling (including 13 Liliales and 67 Asparagales taxa) also recovered this sister relationship (Timilsena et al. 2020, 2022; also see McKain et al. 2016; OTPT Initiative 2019; Baker et al. 2022). The level of nuclear gene tree discordance regarding the placement of Liliales and Asparagales was high, with almost equal support for the two alternative relationships, as expected when incomplete lineage sorting and rapid speciation are the sole reason for discordance (Pamilo and Nei 1988; Maddison 1997; Degnan and Rosenberg 2009).

Triuridaceae and Burmanniaceae branches in the phylogenetic trees do appear rather long, but are not vastly accelerated (fig. 2) as is seen in some parasitic and mycoheterotrophic plastid, nuclear rDNA, and mitochondrial gene sequences (Nickrent, Duff, et al. 1997; Nickrent, Yan, et al. 1997; Barkman et al. 2007; Lemaire et al. 2011; Bromham et al. 2013; Lam et al. 2016; Lin et al. 2022). Apparently similar rates of sequence evolution for the nuclear gene markers in the mycoheterotrophs compared to the greatly accelerated rates for plastid genes in some mycoheterotroph lineages (Lam et al. 2016, 2018) need to be examined with a rigorous analysis of evolutionary rates and constraints for the 602 underlying gene trees, but the apparent lack of extreme heterogeneity in phylogenetic placements consistent with plastid and mitochondrial estimates does suggest additional confidence in nuclear genes as a tool for phylogenetic reconstruction.

Relationships among families within Pandanales and Dioscoreales were congruent between the coalescent and concatenation analyses and were strongly supported in both cases. The placement of Triuridaceae in the ASTRAL analysis alone was poorly supported, but is concordant with recent phylogenetic analyses using plastomes (Lam et al. 2015, 2018; Givnish et al. 2018; Soto Gomez et al. 2020) and mitochondrial sequences (Soto Gomez et al. 2020; Lin et al. 2022).

We observed a difference in the position of the branches for *Triuris* and *Lacandonia* and *Croomia* and *Stemona* between coalescent and concatenated ML analyses. Assessment of the individual gene trees revealed two alternative topologies in very similar proportions, so much so that one alternative resolution was supported by almost the same proportion of genes (39%) as the most frequent resolution (42%). This result is not consistent with coalescence model expectation of incomplete lineage sorting as the sole reason for discordance. This may instead indicate biases in gene tree estimation due to misspecification of the substitution process (e.g., the long-branch issues for Triuridaceae noted by Mennes et al. 2013), or possibly gene flow between an ancestor of the Pandanaceae and Cyclanthaceae clade and an ancestor of lineages leading to Triuridaceae or Stemonaceae. If gene tree discordance was solely due to incomplete lineage sorting in the face of rapid diversification, we would expect nearly equal frequencies of gene trees with Triuridaceae and Stemonaceae as sister lineages, and the observed gene trees with Triuridaceae or Stemonaceae as sister to the Pandanaceae + Cyclanthaceae (Degnan and Rosenberg 2009; Liu, Yu, Kubatko, et al. 2009; Liu, Yu, Pearl, et al. 2009).

### Molecular Evolution of Mycoheterotrophs

When a plant transitions to an increasingly heterotrophic lifestyle, selective constraints on photosynthesis genes are relaxed, and can be completely lost in nonphotosynthetic lineages. Such radical shifts in functional constraints impacting large suites of genes have long made nonphotosynthetic parasitic plants and mycoheterotrophs valuable “evolutionary genetic mutants” for understanding plastid genome evolution and gene function. Pseudogenization and large-scale losses of plastid-encoded photosynthesis genes have been reported by numerous studies of both nonphotosynthetic and minimally photosynthetic plants and algae that have secondarily lost photosynthetic capability (e.g., dePamphilis and Palmer 1990; Wolfe et al. 1992; Delannoy et al. 2011; Wickett et al. 2011; Barrett and Davis 2012; Lam et al. 2015, 2018; Wicke et al. 2016; Graham et al. 2017; Hadariová et al. 2018; Dorrell et al. 2019), showing that the normally slowly-evolving plastid genomes display highly specific convergent gene losses



when selective constraints are lost, leaving behind a collection of genes with nonphotosynthetic function.

More recently, it has become clear that widespread gene loss extends also to large numbers of genes from the nuclear genomes of a mycoheterotrophic orchid (Yuan et al. 2018), and four parasitic plants (Sun et al. 2018; Vogel et al. 2018; Yoshida et al. 2019; Cai et al. 2021), where not only photosynthesis genes but also genes involved in signaling and defense (Cai et al. 2021) and plastid genome repair (Schelkunov et al. 2021) are greatly reduced in number. Evidence in this paper expands our understanding of reductive evolution in nonphotosynthetic plants by focusing on expression shifts, changes in evolutionary constraints, and probable losses of chlorophyll and photosynthetic pathway genes as well as benchmark single copy genes in four different mycoheterotrophic plants of widely different ages (i.e., time since loss of photosynthesis).

However, functional *retention* of genes typically involved in photosynthesis and related pathways, when most other genes in this category have been lost, suggests potential additional or alternative functions for these genes. We observed several such examples of retained nuclear genes with direct photosynthetic roles. The Photosystem II gene *psb29* is retained and transcribed in all seven mycoheterotrophic plants we examined. This gene encodes a protein that is localized in both the outer plastid membrane and stroma and is involved in Photosystem II repair (Bečková et al. 2017). However, the *psb29*-encoded protein also interacts with G protein alpha 1 (GPA1) in *Arabidopsis* for regulating sugar signaling between the plastid and plastid membrane (Huang et al. 2006). Considering the mycoheterotrophic lineages as multiple independent evolutionary genomic experiments, it is highly likely that *psb2* is retained for a nonphotosynthetic function such as the continued need for plastid sugar signaling. Ifuku et al. (2005) found that *psbQ* is not essential for regularization and stabilization of Photosystem II in higher plants, so its retention in mycoheterotrophic lineages is consistent with it retaining a nonphotosynthetic function in these plants as well as in photosynthetic plants.

Wickett et al. (2011) found that the chlorophyll biosynthetic genes were retained and expressed in the holoparasitic parasitic plant *Phelipanche aegyptiaca*, Orobanchaceae. Other nonphotosynthetic plants have been reported to frequently have very low but detectable amounts of chlorophyll in absence of photosynthetic function (Cummings and Welschmeyer 1998) and may therefore have secondary function(s) in at least these plants. In mycoheterotrophic species, the bulk of plastid-encoded photosynthesis genes are lost first, while photosynthesis genes with secondary functions are commonly retained (reviewed in Graham et al. 2017). In this analysis, we found that nuclear genes involved in heme synthesis via the common protoporphyrin pathway were expressed in the

mycoheterotrophs in addition to all the green plants. Retention of generalized heme synthesis is not surprising because hemes are cofactors for various enzymes including catalases, peroxidases, cytochromes and also parts of sensory molecules and components in non-photosynthetic electron transport system for energy recovery (Heinemann et al. 2008; Wickett et al. 2011). These molecules are essential for survival of both autotrophic (green) and heterotrophic plants. Consistent with this, the plastid-encoded *tnE* gene, which has a secondary role in heme biosynthesis, is thought to be one of the last plastid encoded genes to be lost in heterotrophic plants in general (e.g., Graham et al. 2017).

The magnesium insertion step catalyzed by magnesium chelatase is the first committed step for chlorophyll synthesis that separates the generalized heme and protoporphyrin synthesis from dedicated chlorophyll synthesis pathway (Walker and Weinstein 1994; Bollivar 2006). The expression of genes involved in the pathway after this step follows a different pattern. All of the six genes from *Corallorhiza striata* are detected in this analysis, but unlike the holoparasite *Phelipanche aegyptiaca* (Wickett et al. 2011), the other full mycoheterotrophs either do not contain or did not express these genes at detectable levels in our study. This is consistent with the prior findings that *Corallorhiza striata* is a recent mycoheterotroph with its plastid genome in an early stage of degradation (Barrett and Davis 2012; Barrett et al. 2018). In fact, trace amounts of chlorophyll are detectable in *Corallorhiza* using fluorometry, even in non-green species such as *C. striata* and *C. maculata* (Barrett et al. 2014). The pattern and level of gene loss also seems to correspond to the age of the mycoheterotrophic lineage. Triuridaceae, including both *Lacandonia* and *Triuris*, is an ancient family with a crown age of around 75 million years ago (MYA) (Mennes et al. 2013, Givnish et al. 2018).

*Burmannia* and *Gymnosiphon* are younger, with an estimated crown age between 40 and 55 MYA (Merckx et al. 2008). The orchid *Corallorhiza striata* is a relatively young mycoheterotroph, in a non-photosynthetic crown lineage for an estimated 7.3 million years (MY), separated from the leafy green relative *Oreorchis* for 10–15 MY (Barrett et al. 2018) and its plastid genome is clearly still in an earlier transitional stage of degradation relative to the “older” mycoheterotroph lineages included in this study (Barrett and Davis 2012). Thus, it is to be expected that some of the nuclear-encoded photosynthesis genes would still be retained and potentially expressed in this species, albeit under relaxed selective constraint if there was no evolutionary pressure to retain the pathway (Sen and Ghosh 2013).

We demonstrated a reduction in transcript levels in genes for chlorophyll synthesis and photosynthesis in *Cypripedium acaule* and *Burmannia biflora* relative to *Croomia* (figs. 4 and 5) and a relaxation of selective constraint in chlorophyll biosynthesis genes (Uroporphyrinogen II synthase and

Uroporphyrinogen decarboxylase) in *Cypripedium acaule*. Carbon ratio analyses of different species within *Cypripedium* and *Burmannia* showed that some members of these genera are partial mycoheterotrophs (Bushman 2006; Gebauer et al. 2016; Bolin et al. 2017; Suetsugu et al. 2021), but the species we sequenced were not tested. Although they are both leafy and green, our results suggest that the species included in this study may in fact be partial mycoheterotrophs at maturity rather than fully autotrophic species, as we had assumed at the outset of our research. These findings also introduce new lines of evidence coming from analyses of transcriptome data that may help identify partial mycoheterotrophy and early genomic shifts associated with this nutritional state. Further analyses, using isotope analyses (Preiss and Gebauer 2008; Hynson et al. 2016), are needed to confirm these conclusions.

Although retained and transcribed (albeit at lower levels than the putative partial mycoheterotroph *Cypripedium* or the fully autotrophic *Croomia*), we detected significant relaxation of purifying selection in four of the six *Corallorhiza* genes from the dedicated chlorophyll synthesis pathway. Similar relaxation of selection in both such genes detected from *Gymnosiphon* indicates that selective constraints are being released from large sets of nuclear genes in these plants. Chlorophyll synthesis may occur in recent full mycoheterotrophs, presumably at a lower level, as trace levels of chlorophyll in these plants can protect them from excess light exposure (Sakuraba et al. 2010; Wickett et al. 2011; Ballottari et al. 2012) which may also help to explain their frequent occurrence in low light forested conditions (Merckx, Smets, et al. 2013). The fact that most of the detected genes in the dedicated steps of chlorophyll synthesis showed relaxation of selection in mycoheterotrophs, and that most of the antennae protein and light harvesting genes were not detected in those taxa, is an indication that the transcribed genes are following in an evolutionary degradation process in the nuclear genome that parallels the well-established pattern for plastid genomes (Wicke et al. 2016; Graham et al. 2017).

A total of 174 MIM genes were found to be absent from all four mycoheterotroph transcriptomes and the *Gastrodea* genome. This shared set of missing genes, like the other lost chlorophyll and photosynthetic nuclear genes documented above, represents a predictable, but dramatically convergent set of gene losses, similar in kind but far larger in number than the convergent deletion of photosynthesis genes from the plastid genomes (Graham et al. 2017; Hadariová et al. 2018; Wicke and Naumann 2018) of mycoheterotrophs and parasitic plants. The 72 additional genes detected in the recent mycoheterotroph *Corallorhiza striata* that were missing in the other four mycoheterotrophs have a functional profile indistinguishable from the 174 MIM genes. This suggests that, given enough time, a minimum of 246 (=174+72) otherwise highly

conserved BUSCO genes are likely among the gene set that is dispensable in secondarily nonphotosynthetic plants. *Corallorhiza*, therefore, gives important clues to the progression of gene silencing and loss in the nuclear genome, as it has for early stages of plastid genome reduction (Barrett and Davis 2012). Careful phylogenomic studies across numerous such lineages should further expand the number of convergently lost and unexpectedly retained genes to include genes in larger gene families with complex birth and death processes.

Although most MIM genes are predictably tied to photosynthesis, not all MIM genes are known to be associated with photosynthetic activity in model plants, or even to a plastid function. For example, BUSCO gene comparisons identify an intriguing set of 21 shared MIM genes that lack annotations. Their convergent loss in independent mycoheterotrophic lineages suggests that the functions of these genes, though unknown, could be related to the loss of photosynthesis. An exciting alternative possibility is that some of these convergently lost genes could be marking genetic changes associated with other common traits of mycoheterotrophs, such as the loss or major reduction of leaves and roots (fig. 1). More detailed transcriptomic evidence, and experimental studies are needed to distinguish which of these explanations may apply to particular MIM genes.

In conclusion, the loss of chlorophyll synthetic and photosynthesis genes in mycoheterotrophs reveal dramatic changes correlated with the reduction and loss of photosynthetic dependence; these nuclear gene losses appear to be both progressive and strongly convergent in independent lineages. This process begins prior to the loss of photosynthesis, and closely parallels major stages of plastid gene evolution in mycoheterotrophic and parasitic plants (Wicke et al. 2016; Graham et al. 2017; Wicke and Naumann 2018), but it also involves major shifts in gene expression that are detectable through transcriptomic analysis: 1) Green, putatively autotrophic plants *Burmannia* and *Cypripedium* showed a reduction of expression or a relaxation of selection in some genes, suggesting that these could reflect initial genomic changes seen in partial mycoheterotrophs; 2) the “recent” full mycoheterotroph *Corallorhiza* has apparently lost 174 MIM genes in common with other mycoheterotroph lineages, but it expresses all of the genes for the dedicated chlorophyll synthesis pathway and some for photosynthetic processes, with relaxed selective constraints and higher expression levels for retained photosynthesis genes than any detected from the other mycoheterotrophs; 3) *Gymnosiphon*, which is an intermediate-aged mycoheterotrophic lineage, expressed a similar gene complement as *Corallorhiza*, and has experienced relaxed selection of chlorophyll biosynthetic pathway genes, magnesium chelatase and magnesium protoporphyrin 9 methyl transferase with decreased expression

levels; and 4) the “ancient mycoheterotrophs” *Triuris* and *Lacandonia* completely lacked transcripts for most genes encoding photosynthetic proteins, consistent with mycoheterotrophs undergoing major genetic modifications following the transition from autotrophic to mycoheterotrophic lifestyle. The retention of the heme biosynthesis pathway follows a similar pattern detected for holoparasitic plant lineages like *Orobancha*, *Phelipanche* and *Cuscuta* (McNeal et al. 2007; Wickett et al. 2011), except that older mycoheterotrophs have lost expression of all genes in the dedicated chlorophyll synthesis pathway, possibly indicative of the young age of Orobanchaceae compared to the ancient mycoheterotrophs in this study. The estimated stem age for Orobanchaceae is 35.9 MYA (Magallón et al. 2015), while multiple individual holoparasite lineages within the family such as *Epifagus* are younger still (Naumann et al. 2013). A similar progressive loss of photosynthesis genes has been extensively documented in the plastomes of parasitic and mycoheterotrophic plants. In the plastome, different levels of degradation, ranging from minimal loss to extreme reduction in gene number, has been observed following the transition to parasitic or mycoheterotrophic lifestyles. The relaxation of selection in genes from the dedicated chlorophyll synthesis pathway in *Corallorhiza* and *Gymnosiphon* follows a similar fate proposed for the removal of selective pressure to retain corresponding photosynthesis genes in the plastid genome (Barrett and Davis 2012; Barrett et al. 2014; Graham et al. 2017).

This study also shows that transcriptome datasets can be used effectively for phylogenetic reconstruction of plant relationships, regardless of their dependence on an external carbon source. Although very high rates of sequence evolution have often been documented for mycoheterotroph plastid and nuclear rDNA genes, challenging phylogenetic analysis, branch lengths for nuclear single copy genes fall within the range of variation in overall rates seen in monocots as a whole. Our application of single copy nuclear genes that are widely conserved across monocots and other species and are generally indispensable housekeeping genes (Duarte et al. 2010; De Smet et al. 2013) ameliorates the risk of relying on paralogs that could have led to bias in the inference of species trees, and creates a robust platform for investigating the evolution of gene and expression reduction in mycoheterotrophic plants. Transcriptomic analysis thus has great utility for understanding molecular evolutionary aspects of the origins of both mycoheterotrophy and parasitism in green plant lineages.

As larger numbers of nonphotosynthetic plants have their genomes and transcriptomes sequenced and analyzed, it will be possible to determine how many of the conclusions reached in this and other recent genome studies apply more generally across nonphotosynthetic organisms with photosynthetic ancestry, including the numerous

lineages of nonphotosynthetic parasitic plants (Barkman et al. 2007; Nickrent 2020), algae and other protists (Hadariová et al. 2018; Dorrell et al. 2019), as well as the many mycoheterotrophic lineages (Merckx, Mennes, et al. 2013) not yet exposed to genomic or transcriptomic analysis. This approach will also enable a comparative analysis of adaptive changes associated with the gain of heterotrophic capabilities (Yang et al. 2015; Sun et al. 2018; Yoshida et al. 2019) and the extent to which gene loss drives adaptive evolution (McNeal et al. 2009; Albalat and Canestro 2016; Sun et al. 2018; Helsen et al. 2021) in heterotrophic plants. With sufficient numbers of independent “evolutionary experiments” representing many time points up to and following loss of photosynthesis, the common and exceptional patterns across lineages, as well as the dynamics and mechanisms of radical genomic change, can be more clearly defined and understood.

## Materials and Methods

### Taxon Sampling Data Collection and Sequencing

A total of 37 taxa were represented in the study (supplementary tables S1–S3, Supplementary Material online), including 20 newly sequenced taxa. Five transcriptomes were obtained from 1KP (Matasci et al. 2014), and 12 genome datasets were obtained from biological repositories including Phytozome (Goodstein et al. 2012), Ensembl Plants (Bolser et al. 2017), NCBI (Agarwala et al. 2017) and others. The final dataset represents 11 of 12 monocot orders (Givnish et al. 2018), and every family recognized by Angiosperm Phylogeny Group (2016) for the two focal orders, Pandanales and Dioscoreales.

Mycoheterotrophy was represented by four fully mycoheterotrophic species in three families: *Corallorhiza striata*, (Orchidaceae, Asparagales), *Gymnosiphon panamensis*, (*Burmanniaceae*, *Dioscoreales*), and *Lacandonia schismatica* and *Triuris brevistylis* (Triuridaceae, Pandanales). We also designate these four species as *young*, *intermediate*, and *ancient* mycoheterotrophic lineages, in a relative sense, based on previously estimated times of divergence of each lineage from their nearest photosynthetic relatives. *Corallorhiza striata*, which is estimated to have diverged ca. 10–15 MYA from the leafy green *Oreorchis* (Barrett et al. 2018) represents the “young” category. In the “intermediate” category is *Gymnosiphon*, estimated to have diverged from the green relative *Burmannia bicolor* ca. 56 MYA (Givnish et al. 2018). Finally, in the “ancient” category is the family Triuridaceae, represented here by both *Lacandonia* and *Triuris*, with a crown age of ca. 75 MY (Mennes et al. 2013; Givnish et al. 2018; Lam et al. 2018).

Photosynthetic, leafy relatives were included for close comparisons to each mycoheterotrophic lineage: *Croomia* (Stemonaceae, Pandanales), *Burmannia biflora*



(Burmanniaceae), scored as autotrophic in Merckx et al. 2006 and *Cypripedium acaule* (Orchidaceae). The genera *Burmannia* and *Cypripedium* have both been found to include one or more partial mycoheterotroph species based on isotopic data (Gebauer et al. 2016; Bolin et al. 2017), but these species have not been subject to isotopic analysis. Full details for these taxa are shown in [supplementary table S2, Supplementary Material](#) online.

### Libraries and Sequencing

Fresh aboveground tissues from individual plants were collected and flash frozen in liquid nitrogen. RNA extraction and purification was performed as described by Johnson et al. (2012). Illumina libraries were prepared with the NEBNext Ultra Directional RNA Library Prep Kit for Illumina (New England Biolabs, Inc.) using the manufacturer's protocol and sequenced on the Illumina Hi-Seq or Next-Seq platforms at the Cold Spring Harbor Laboratories genomics core facilities. Depending on the taxon and sample, paired end reads were generated that varied from 70 base pairs to 150 base pairs in length, with a mean of 33.1 million PE reads (range: 8.7 million to 89.2 million bases per taxon; [supplementary table S2, Supplementary Material](#) online).

### Data Cleaning and Assembly, Post-processing, and Translation

Raw reads were filtered and trimmed to remove low-quality bases and short read fragments as well as contaminating adapter sequences with Trimmomatic (Bolger et al. 2014) using the following parameters: ILLUMINACLIP:2:30:10 LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:36. FastQC v0.11.5 (Andrews 2010) was used to assess the overall sequence quality before and after trimming. Cleaned reads were de novo assembled using Trinity (Grabherr et al. 2011; Haas et al. 2013) with the default parameters. The resulting transcriptome assemblies were post-processed with the PlantTribes tool (Wafula et al. 2022) *AssemblyPostProcessor* (<https://github.com/dePamphilis/PlantTribes>) using *Oryza sativa* ESTScan score matrices (Iseli et al. 1999) to produce non-redundant sets of predicted coding sequences and their corresponding translations ([supplementary table S2, Supplementary Material](#) online). Additional filtering was performed on post-processed assemblies to remove non-plant contaminant sequences by screening transcripts against NCBI non-redundant protein database using BLASTp (Altschul et al. 1990) (E-value =  $1e-5$ ) as shown in [supplementary tables S13a and S13b, Supplementary Material](#) online.

### Gene Family Circumscription and Transcriptome Sorting

Proteomes of the 12 plant genomes were circumscribed into narrowly defined gene lineages (i.e., orthogroups)

using the OrthoMCL version 2.0.9 algorithm (Enright et al. 2002; Li et al. 2003) with BLASTp E-value cutoff of  $1e-05$  and inflation value of 1.2. We selected nine species to represent lineages of monocots for which genome sequence data are available, including six from Poales (*Aegilops tauschii*, *Brachypodium distachyon*, *Phyllostachys heterocycla*, *Oryza sativa*, *Sorghum bicolor*, and *Setaria italica*), one from Arecales (*Elaeis guineensis*), one from Zingiberales (*Musa acuminata*), and one from Alismatales (*Spirodela polyrhiza*). Additionally, we selected three outgroup genomes, *Vitis vinifera*, *Nelumbo nucifera*, and *Amborella trichopoda* to enable rooting of monocot gene and species relationships. A total of 24,873 orthogroups containing at least two genes were circumscribed, 22,571 of which included at least one monocot taxon. We performed a second round of circumscription to connect distant but potentially related orthogroups into larger hierarchical gene families (i.e., super-orthogroups) to represent traditional families characterized by functional domains as described by Wall et al. (2008). To provide a flexible resource enabling the identification of widely divergent gene family members, an average of 3,074 super-orthogroups were circumscribed for 10 MCL stringencies with inflation values ranging from 1.2 to 5.0 ([supplementary tables S14–S15, Supplementary Material](#) online). Orthogroups were then annotated with functional metadata from protein functional databases including Swiss-Prot (Bairoch and Apweiler 2000), TrEMBL, TAIR (Rhee et al. 2003), Pfam (Finn et al. 2014), InterProScan (McDowall and Hunter 2011), and Gene Ontology (Ashburner et al. 2000; The Gene Ontology Consortium 2019). Finally, post-processed transcriptome assemblies were sorted into the 12-plant genome OrthoMCL orthogroup scaffold using the PlantTribes tool *GeneFamilyClassifier* (<https://github.com/dePamphilis/PlantTribes>) and orthogroup fasta files (CDS and inferred protein sequences) were compiled with the PlantTribes tool *GeneFamilyIntegrator* ([supplementary table S16, Supplementary Material](#) online).

### Species Tree Inference

We selected 602 single-copy orthogroups for gene and species tree analyses; these orthogroups included one and only one homolog in each of the 12 genomes used for orthogroup circumscription. For each inferred single-copy orthogroup, protein multiple sequence alignments were performed using the L-INS-I algorithm in MAFFT (Katoh et al. 2009; Katoh and Standley 2013, 2014) with 1000 iterative refinements and then converted to codon alignments (<https://github.com/dePamphilis/PlantTribes>). If more than one transcript assembly for a particular species was sorted into a given orthogroup, they were collapsed into a consensus sequence if there was less than 5%

conflict between the sequences in the region of overlap (Wickett et al. 2014). Otherwise, the orthogroup was treated as missing for species with multiple, more divergent transcript assemblies. Amino-acid translations for single-copy nucleotide consensus transcripts were predicted using PlantTribes as previously described and iteratively added to the orthogroup protein backbone alignments using MAFFT. Only transcripts with greater than 70% amino acid coverage were retained in the single-copy orthogroup alignments. Protein alignments were converted to codon alignments and trimmed using trimAl (Capella-Gutierrez et al. 2009) to remove sites with gaps in more than 70% of the orthogroup taxa. Super-matrix phylogenetic trees were estimated using RAxML version 8.2 (Stamatakis 2014) for protein and codon single-copy orthogroup alignments that were concatenated and partitioned using FASconCat (Kuck and Meusemann 2010). Additionally, protein and codon single-copy gene trees were inferred with RAxML and used to estimate coalescent-based species trees with ASTRAL-III (Mirarab and Warnow 2015; Zhang et al. 2018).

Concatenated trees and ASTRAL III species trees were estimated from both amino-acid and DNA alignments using all the single copy genes as well as different subsets of single copy genes based on the number of taxa that contain the gene (supplementary table S4, Supplementary Material online). Partitioning was based on codon position and by gene for the DNA alignments and based on genes for the protein alignments. Support values were calculated using standard bootstrap percentages and local posterior probabilities. Additionally, we also evaluated all the taxa in the dataset with BUSCO (Benchmarking Universal Single Copy Orthologs) software (Simao et al. 2015; Waterhouse et al. 2017) using orthoDB version 10 for embryophytes to estimate the completeness of the transcriptomes for universally conserved low copy genes, and also testing the BUSCO orthogroups as a distinct alternative dataset for phylogeny estimation. Complete and single copy orthogroups were extracted from each of the transcriptomes and genomes, and gene trees were estimated as in the single copy gene analyses. The full list and details of the analyses are presented in supplementary table S16, Supplementary Material online.

In addition to bootstrap percentages and local posterior probabilities, the robustness of phylogenetic inferences was assessed using multiple measures of phylogenetic stability. Each branch of the estimated ASTRAL species tree was tested for polytomy (Sayyari and Mirarab 2018). Phylostab (Sheikh et al. 2013) measures the stability of trees, subtrees, and individual taxa with respect to changes in the input data (sequences). Leaf stability index as implemented in RogueNaRok (Goloboff and Szumik 2015) measures the stability of trees based on quartet frequencies. RogueNaRok was also used to identify rogue taxa or taxa

that cannot be accurately placed anywhere in the phylogenetic tree. In addition, we used phyutility (Smith and Dunn 2008) to identify possible alternative locations for certain clades or taxa, using the “lineage movement” function. This analysis attempts to identify if and where a lineage moves, given a set of trees and a consensus tree and a target clade. Finally, to evaluate the level of discordance between gene trees and species trees, we mapped the individual gene trees to the ASTRAL species tree using PhyParts (Smith et al. 2015).

### BUSCO Assessment and Comparison With *Gastrodia elata* Genome

In order to better interpret the transcriptome-wide pattern of gene losses (or non-expression) between the different green and mycoheterotrophic taxa and to compare the pattern with the genome of *Gastrodia elata*, an obligately mycoheterotrophic orchid (Yuan et al. 2018), we ran BUSCO analysis of the inferred protein sequences for the *Gastrodia elata* genome. We then looked at all the unique and shared genes between the five mycoheterotrophic taxa as well as another set of five green taxa from the Pandanales or Dioscoreales clades with a comparable level of missing BUSCOs, to detect any differences in the pattern of gene losses in the two groups.

To assess the pattern of shared gene loss, we used mixed logistic regression models to explain the binary response variable “missing in all 5 taxa” versus “not missing in all 5 taxa”, using the taxon set as a predictor variable: mycoheterotrophic versus photosynthetic. To account for pairing by gene, that is, repeated measurement for each gene across the two taxon sets, we used the gene ID as a random effect. We then conducted a likelihood ratio test to compare the model without a taxon set predictor (our null hypothesis that shared missing is equal across the two taxon sets) to a model with taxon set as a predictor.

The analysis above should be conducted on taxon sets that have similar proportions of missing genes, because an increase in missingness is expected to lead to an increase in shared missingness. To check that the set of five green taxa was chosen appropriately to meet this requirement, we repeated the analysis above with the binary response “missing in 1 or more of the 5 taxa” versus “missing in none of the 5 taxa”. Overall gene missingness was considered similar in both sets if the likelihood ratio test provided no evidence of a difference between the mycoheterotrophic and photosynthetic taxa. For fitting these mixed logistic models, we used function glmer from the R package lme4 (Bates et al. 2015) in the R environment v4.1.0 (R Core Team 2021).

The BUSCOs missing from all five mycoheterotrophic taxa were further examined as follows: TAIR IDs (Rhee et al. 2003) of corresponding *Arabidopsis thaliana* genes

from each of the common missing BUSCOs were examined to identify the putative functions of the missing genes. DAVID (Huang et al. 2009a, 2009b) version 6.8 was used for functional annotation of the *Arabidopsis* sequences and to identify functional groups that were over- or under-represented. Enrichment analysis of the 174 shared missing genes was performed against the background of all *Arabidopsis thaliana* genes, as well as against the background of all BUSCO *Arabidopsis* genes. Similar tests were also repeated for gene sets missing from other subsets of mycoheterotrophic taxa.

### Evolutionary Fate of Photosynthetic and Chlorophyll Biosynthetic Genes

The *Arabidopsis thaliana* proteome was obtained from Phytozome 12, and genes were sorted into the 12-genome OrthoMCL gene family circumscriptions using *GeneFamilyClassifier* in Plant Tribes (Wafula et al. 2022). Genes representing the heme and chlorophyll biosynthetic pathway, and light harvesting and photosynthetic pathways were selected following Wickett et al. (2011), and *Arabidopsis* gene annotations were associated with corresponding orthogroups. Transcripts for *Croomia*, *Gymnosiphon*, *Triuris*, *Lacandonia*, *Cypripedium* and *Corallorhiza* were also annotated independently of their orthogroup assignment to further validate the annotations. Focal genes were screened for contaminants and only green plant sequences with specific functional domains were retained. Expression was quantified from each transcriptome as TPM using Salmon (Patro et al. 2017). Because multiple isoforms could be observed for a given gene, the level of expression was calculated as the sum of TPM of reads assigned to any isoform of the gene of interest in each taxon (supplementary tables S5–S7, Supplementary Material online). In addition to Hidden Markov Model (HMM) searching with HMMER 3.1b2 (Eddy 1998, 2011), we searched for highly divergent genes that may have been missed using low stringency blast (tBLASTx, cutoff e-5) of the collection of available genes against the transcriptome assemblies with a missing gene. In cases where only small fragments without functional domains were detected, we attempted to recover the full-length sequences using targeted assembly with the *AssemblyPostProcessor* in PlantTribes (<https://github.com/dePamphilis/PlantTribes>). These sequences were further verified using protein BLAST and integrated into the alignment scaffolds.

For each of the focal gene families, protein alignments were first created using the MAFFT L-INS-I algorithm. Short fragments of poor alignment homology were removed, and the remaining sequences were realigned. The DNA sequences were then mapped onto their corresponding protein alignments. Gene trees were estimated from DNA alignments with 100 rapid bootstrap replicates in

RAxML (GTRGAMMA model, partitioned by codon positions). Relaxation or intensification of selective constraints on branches of interest (i.e., mycoheterotrophs) was detected with RELAX (Wertheim et al. 2015) using the HyPhy package (Pond et al. 2005) in the Datamonkey adaptive evolution server (Pond and Frost 2005; Delport et al. 2010; Weaver et al. 2018). RELAX tests for the relaxation or intensification of selection in a given set of test branches with respect to reference branches by using a selection intensity parameter, K, followed by likelihood ratio test. Because of the incompleteness of some of the sequences recovered from RNASeq, we were unable to find enough sites to test the selection for all of the genes.

### Supplementary material

Supplementary data are available at *Genome Biology and Evolution* online (<http://www.gbe.oxfordjournals.org/>).

### Acknowledgements

Data generation was performed under the Monocot Tree of Life project (MonAToL), DEB-0829868, at Cold Spring Harbor Laboratories, University of Georgia, Global Biologics, LLC (Sean Blake) and Penn State University. We thank Chang Liu, Charlotte Quigley, and Lisa DeGironimo for assisting with RNA isolations; Naomi Altman for advice on statistical analyses; and Heather Hines, Kateryna Makova, John Carlson, Paula Ralph, Huiting Zhang, Marcos Caraballo, Elizabeth Kelly, and two reviewers for helpful comments. P.R.T. and E.K.W. and computer resources were supported in part by DEB-0829868 and IOS-1238057 and by the departments of Biology and Plant Science at Penn State University. E.R.A.-B.'s research on *Lacandonia* was supported by grants UC-MEXUS-CONACYT CN12-571, ECO-IE271, CONACYT 0435/B-1, 90565, 167705, and DGAPA-PAPIIT: IN221406, IN226510-3. C.F.B.'s research on *Corallorhiza* was supported by the California State University Program for Research and Education in Biotechnology, and the West Virginia University Program to Stimulate Competitive Research. We thank Evgenia Kriventseva for access to OrthoDB version 10 prior to its public release, and Sarah Friedrich for radar graphs for figures 4 and 5.

### Data Availability

Twenty newly sequenced transcriptome datasets have been deposited at Genbank under Bioproject PRJNA313089; all other transcriptomes and genomes used in this study were obtained from Genbank (<https://www.ncbi.nlm.nih.gov/genbank/>), OneKP (Matasci et al. 2014), and Phytozome (<https://phytozome-next.jgi.doe.gov/>) as listed in supplementary table S2, Supplementary Material online. The data analysis pipeline is available at [github.com/dePamphilis/PlantTribes](https://github.com/dePamphilis/PlantTribes). The 12-genome gene family

classification and orthogroup annotation is available at [bigdata.bx.psu.edu/PlantTribes\\_scaffolds/data/12Gv1.0.tar.bz2](http://bigdata.bx.psu.edu/PlantTribes_scaffolds/data/12Gv1.0.tar.bz2).

## Literature Cited

- The Angiosperm Phylogeny Group, et al. 2016. An update of the angiosperm phylogeny group classification for the orders and families of flowering plants: aPG IV. *Botanical J Lin Soc.* 181:1–20.
- The Gene Ontology Consortium. 2019. The gene ontology resource: 20 years and still GOing strong. *Nucleic Acids Res.* 47:D330–D338.
- One thousand plant transcriptomes initiative (OTPT Initiative). 2019. One thousand plant transcriptomes and the phylogenomics of plants. *Nature* 574:679–685.
- Aberer AJ, Stamatakis A. 2011. A simple and accurate method for rogue taxon identification. 2011 IEEE International Conference on Bioinformatics and Biomedicine, Atlanta, GA, 2011:118–122.
- Agarwala R, et al. 2017. Database resources of the national center for biotechnology. *Nucleic Acids Res.* 45:D12–D17.
- Albalat R, Canestro C. 2016. Evolution by gene loss. *Nat Rev Genet.* 17:379–391.
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J Mol Biol.* 215:403–410.
- Andrews A. 2010. FastQC: a quality control tool for high throughput sequence data. <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>; last accessed Aug 29, 2021.
- Ashburner M, et al. 2000. Gene ontology: tool for the unification of biology. *The gene ontology consortium.* *Nat Genet.* 25:25–29.
- Atul-Nayyar A, Hamel C, Hanson K, Germida J. 2009. The arbuscular mycorrhizal symbiosis links N mineralization to plant demand. *Mycorrhiza* 19:239–246.
- Bairoch A, Apweiler R. 2000. The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. *Nucleic Acids Res.* 28: 45–48.
- Baker WJ, et al. 2022. A comprehensive phylogenomic platform for exploring the angiosperm tree of life. *Syst Biol.* 71:301–319.
- Ballottari M, Girardon J, Dall’Osto L, Bassi, R. 2012. Evolution and functional properties of Photosystem II light harvesting complexes in eukaryotes. *Bioch et Biophys Acta (BBA) - Bioenergetics.* 1817: 143–157.
- Barkman TJ, et al. 2007. Mitochondrial DNA suggests at least 11 origins of parasitism in angiosperms and reveals genomic chimerism in parasitic plants. *BMC Evol Biol.* 7:248.
- Barrett CF. 2010. Mycoheterotrophy and diversity in Orchidaceae. In: Merckx V, editors. *Mycoheterotrophy.* New York (NY): Springer. p. 25–37.
- Barrett CF, et al. 2014. Investigating the path of plastid genome degradation in an early-transitional clade of heterotrophic orchids, and implications for heterotrophic angiosperms. *Mol Biol Evol.* 31: 3095–3112.
- Barrett CF, Davis JI. 2012. The plastid genome of the mycoheterotrophic *Corallorhiza striata* (Orchidaceae) is in the relatively early stages of degradation. *Am J Bot.* 99:1513–1523.
- Barrett CF, Kennedy AH. 2018. Plastid genome degradation in the endangered, mycoheterotrophic, north American orchid *Hexalectris warnockii*. *Genome Biol Evol.* 10: 1657–1662.
- Barrett CF, Sinn BT, Kennedy AH. 2019. Unprecedented parallel photosynthetic losses in a heterotrophic orchid genus. *Mol Biol Evol.* 36: 1884–1901.
- Barrett CF, Wicke S, Sass C. 2018. Dense infraspecific sampling reveals rapid and independent trajectories of plastome degradation in a heterotrophic orchid complex. *New Phytologist.* 218:1192–1204.
- Barrett CF, et al. 2016. Plastid genomes reveal support for deep phylogenetic relationships and extensive rate variation among palms and other commelinid monocots. *New Phytol.* 209:855–870.
- Bates D, Maechler M, Bolker B, Walker S. 2015. Fitting linear mixed-effects models using lme4. *J Stat Softw.* 67:1–48.
- Bečková M, et al. 2017. Structure of Psb29/Thf1 and its association with the FtsH protease complex involved in photosystem II repair in cyanobacteria. *Philos Trans R Soc Lond B Biol Sci.* 372: 20160394.
- Bidartondo MI. 2005. The evolutionary ecology of myco-heterotrophy. *New Phytol.* 167:335–352.
- Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for illumina sequence data. *Bioinformatics* 30:2114–2120.
- Bolin JF, Tennakoon KU, Majid MBA, Cameron DD. 2017. Partial mycoheterotrophy in *Burmannia coelestis*. *Plant Species Biol.* 32: 74–80.
- Bollivar DW. 2006. Recent advances in chlorophyll biosynthesis. *Photosynth Res.* 90:173–194.
- Bolser DM, Staines DM, Perry E, Kersey PJ. 2017. Ensembl plants: integrating tools for visualizing, mining, and analyzing plant genomic data. *Methods Mol Biol.* 1533:1–31.
- Borner T, Zhelyazkova P, Legen J, Schmitz-Linneweber C. 2014. Chloroplast gene expression–RNA synthesis and processing. In: Theg SM and Wollman FA, editors. *Plastid biology.* New York (NY): Springer. p. 3–48.
- Bromham L, Cowman PF, Lanfear R. 2013. Parasitic plants have increased rates of molecular evolution across all three genomes. *BMC Evol Biol.* 13:126.
- Bronstein JL. 1994. Conditional outcomes in mutualistic interactions. *Trends Ecol Evolut.* 9:214–217.
- Bushman M. 2006. Two records of achlorophyllous *Cypripedium* from Wisconsin. 2006. *Mich Bot.* 45:193–196.
- Cai L, et al. 2021. Deeply altered genome architecture in endoparasitic flowering plant *Sapria himalayana* griff. (Rafflesiaceae). *Curr Biol.* 31:1002–1011.
- Capella-Gutierrez S, Silla-Martinez JM, Gabaldon T. 2009. Trimal: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25:1972–1973.
- Cummings MP, Welschmeyer NA. 1998. Pigment composition of putatively achlorophyllous angiosperms. *Plant Syst Evol.* 210: 105–111.
- Degnan JH, Rosenberg NA. 2009. Gene tree discordance, phylogenetic inference and the multispecies coalescent. *Trends Ecol Evol.* 24: 332–340.
- Delannoy E, Fujii S, Colas des Francs-Small C, Brundrett M, Small I. 2011. Rampant gene loss in the underground orchid *Rhizanthella gardneri* highlights evolutionary constraints on plastid genomes. *Mol Biol Evol.* 28:2077–2086.
- Delavault PM, Russo NM, Lusson NA, Patrick A, Thalouarn PA. 1996. Organization of the reduced plastid genome of *Lathraea clandestina*, an achlorophyllous parasitic plant. *Phys Plantarum.* 96: 674–682.
- Delpont W, Poon AF, Frost SD, Kosakovsky Pond SL. 2010. Datamonkey 2010: a suite of phylogenetic analysis tools for evolutionary biology. *Bioinformatics* 26:2455–2457.
- dePamphilis CW, Palmer JD. 1990. Loss of photosynthetic and chlororespiratory genes from the plastid genome of a nonphotosynthetic plant. *Nature* 348:337–339.
- De Smet R, et al. 2013. Convergent gene loss following gene and genome duplications creates single-copy families in flowering plants. *Proc Natl Acad Sci U S.* 110:2898–2903.
- Dorrell RG, et al. 2019. Principles of plastid reductive evolution illuminated by nonphotosynthetic chrysophytes. *Proc Natl Acad Sci.* 116: 6914–6923.



- Duarte JM, et al. 2010. Identification of shared single copy nuclear genes in *Arabidopsis*, *Populus*, *Vitis* and *Oryza* and their phylogenetic utility across various taxonomic levels. *BMC Evol Biol.* 10:61.
- Eddy S. 1998. Profile hidden markov models. In: *Bioinformatics*. England: Oxford. Vol 14, p. 755–763.
- Eddy S. 2011. Accelerated profile HMM searches. *PLoS Comput Biol.* 7: e1002195.
- Enright AJ, Van Dongen S, Ouzounis CA. 2002. An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Res.* 30: 1575–1584.
- Feng YL, et al. 2016. Lineage-specific reductions of plastid genomes in an orchid tribe with partially and fully mycoheterotrophic species. *Genome Biol Evol.* 8:2164–2175.
- Finn RD, et al. 2014. Pfam: the protein families database. *Nucleic Acids Res.* 42:D222–D230.
- Fletcher W, Yang Z. 2010. The effect of insertions, deletions and alignment errors on the branch-site test of positive selection. *Mol Biol Evol.* 27(10):2257–2267.
- Freudenstein JV, Barrett CF. 2010. Mycoheterotrophy and diversity in Orchidaceae. In: Seberg O, Petersen G, Barfod AS, Davis JJ, editors. *Diversity, phylogeny, and evolution in the Monocotyledons*. Aarhus: Aarhus University Press. p. 25–37.
- Funk HT, Berg S, Krupinska K, Maier UG, Krause K. 2007. Complete DNA sequences of the plastid genomes of two parasitic flowering plant species, *Cuscuta reflexa* and *Cuscuta gronovii*. *BMC Plant Biol.* 2007:7.
- Gebauer G, Preiss K, Gebauer AC. 2016. Partial mycoheterotrophy is more widespread among orchids than previously assumed. *New Phytol.* 211:11–15.
- Givnish TJ, et al. 2010. Assembling the tree of the monocotyledons: plastome sequence phylogeny and evolution of Poales. *Ann Missouri Bot Gard.* 97:584–616.
- Givnish TJ, et al. 2018. Monocot plastid phylogenomics, timeline, net rates of species diversification, the power of multi-gene analyses, and a functional model for the origin of monocots. *Am J Bot.* 105:1888–1910.
- Goloboff PA, Szumik CA. 2015. Identifying unstable taxa: efficient implementation of triplet-based measures of stability, and comparison with phyutility and RogueNaRok. *Mol Phylogenet Evol.* 88: 93–104.
- Goodstein DM, et al. 2012. Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Res.* 40:D1178–D1186.
- Grabherr MG, et al. 2011. Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. *Nat Biotechnol.* 29:644–652.
- Graham SW, Lam VK, Merckx VSFT. 2017. Plastomes on the edge: the evolutionary breakdown of mycoheterotroph plastid genomes. *New Phytol.* 214:48–55.
- Haas BJ, et al. 2013. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat Protoc.* 8:1494–1512.
- Hadariová L, Vesteg M, Hampl V, Krajčovič J. 2018. Reductive evolution of chloroplasts in non-photosynthetic plants, algae and protists. *Curr Genet.* 64:365–387.
- Hahn A, Vonck J, Mills DJ, Meier T, Kuhlbrandt W. 2018. Structure, mechanism, and regulation of the chloroplast ATP synthase. *Science* 360:620–628.
- Hajdukiewicz PTJ, Allison LA, Maliga P. 1997. The two RNA polymerases encoded by the nuclear and the plastid compartments transcribe distinct groups of genes in tobacco plastids. *EMBO J.* 16:4041–4048.
- Heinemann IU, Jahn M, Jahn D. 2008. The biochemistry of heme biosynthesis. *Arch Biochem Biophys.* 474:238–251.
- Helsen J, et al. 2021. Gene loss predictably drives evolutionary adaptation. *Mol Biol Evol.* 37:2989–3002.
- Hoffmann D, Vierheilig H, Riegler P, Schausberger P. 2009. Arbuscular mycorrhizal symbiosis increases host plant acceptance and population growth rates of the two-spotted spider mite *Tetranychus urticae*. *Oecologia* 158:663–671.
- Huang J, et al. 2006. The plastid protein THYLAKOID FORMATION1 and the plasma membrane G-protein GPA1 interact in a novel sugar-signaling mechanism in *Arabidopsis*. *Plant Cell.* 18(5): 1226–1238.
- Huang DW, Sherman BT, Lempicki RA. 2009a. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* 37:1–13.
- Huang DW, Sherman BT, Lempicki RA. 2009b. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc.* 4:44–57.
- Hynson NA, Schiebold JMI, Gebauer G. 2016. Plant family identity distinguishes patterns of carbon and nitrogen stable isotope abundance and nitrogen concentration in mycoheterotrophic plants associated with ectomycorrhizal fungi. *Ann Bot.* 118:467–479.
- Ifuku K, Yamamoto Y, Ono TA, Ishihara S, Sato F. 2005. PsbP protein, but not PsbQ protein, is essential for the regulation and stabilization of photosystem II in higher plants. *Plant Physiol.* 2139: 1175–1184.
- Imhof S. 2010. Are monocots particularly suited to develop mycoheterotrophy? In: Seberg O Petersen G Barfod A and Davis JJ, editors. *Diversity, phylogeny and evolution in the monocotyledons*. Copenhagen: Aarhus University Press. p. 11–23.
- Iseli C, Jongeneel CV, Bucher P. 1999. ESTScan: a program for detecting, evaluating, and reconstructing potential coding regions in EST sequences. *Proc Int Conf Intell Syst Mol Biol.* 138–148.
- Ishizaki Y, et al. 2005. A nuclear-encoded sigma factor, *Arabidopsis* SIG6, recognizes sigma-70 type chloroplast promoters and regulates early chloroplast development in cotyledons. *Plant J.* 42: 133–144.
- Jacquemyn H, Merckx VSFT, Shefferson R. 2019. Mycorrhizal symbioses and the evolution of trophic modes in plants. *J Ecol.* 107: 1567–1581.
- Jakalski M, Minasiewicz J, Caius J, Sellose MA, Delannoy E. 2021. The genomic impact of mycoheterotrophy in orchids. *Front Plant Sci.* 12:9. Jun 2021.
- Johnson MTJ, et al. 2012. Evaluating methods for isolating total RNA and predicting the success of sequencing phylogenetically diverse plant transcriptomes. *PLoS One* 7:e50226.
- Joyce EM, et al. 2018. Evolution of *Geosiris* (Iridaceae): historical biogeography and plastid-genome evolution in a genus of non-photosynthetic tropical rainforest herbs disjunct across the Indian ocean. *Aus Syst Bot.* 31:504–522.
- Kamikawa R, et al. 2015. Proposal of a twin arginine translocator system-mediated constraint against loss of ATP synthase genes from non-photosynthetic plastid genomes. *Mol Biol Evol.* 32: 2598–2604.
- Katoh K, Asimenos G, Toh H. 2009. Multiple alignment of DNA sequences with MAFFT. *Methods Mol Biol.* 537:39–64.
- Katoh K, Standley DM. 2013. MAFFT Multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol.* 30:772–780.
- Katoh K, Standley DM. 2014. MAFFT: iterative refinement and additional methods. *Methods Mol Biol.* 1079:131–146.
- Klimpert NJ, et al. 2022. Phylogenomics and plastome evolution of a Brazilian mycoheterotrophic orchid, *Pogoniopsis schenckii*. *Am J Bot.* 2022:1–22.
- Kuck P, Meusemann K. 2010. FASconCAT: convenient handling of data matrices. *Mol Phylogenet Evol.* 56:1115–1118.

- Lam VKY, et al. 2018. Phylogenomic inference in extremis: a case study with mycoheterotroph plastomes. *Am J Bot.* 105:480–494.
- Lam VKY, Merckx VSFT, Graham SW. 2016. A few-gene plastid phylogenetic framework for mycoheterotrophic monocots. *Am J Bot.* 103:692–708.
- Lam VK, Soto Gomez M, Graham SW. 2015. The highly reduced plastome of mycoheterotrophic *Sciaphila* (Triuridaceae) is colinear with its green relatives and is under strong purifying selection. *Genome Biol Evol.* 7:2220–2236.
- Leake JR. 1994. The biology of myco-heterotrophic ('Saprophytic') plants. *New Phytol.* 127:171–216.
- Leebens-Mack JL, dePamphilis CW. 2002. Power analysis of tests for loss of selective constraint in cave crayfish and nonphotosynthetic plant lineages. *Mol Biol Evol.* 19:1292–1302.
- Lemaire B, Huysmans S, Smets E, Merckx V. 2011. Rate accelerations in nuclear 18S rDNA of mycoheterotrophic and parasitic angiosperms. *J Plant Res.* 124:561–576.
- Li L, Stoeckert CJ Jr, Roos DS. 2003. OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res.* 13:2178–2189.
- Lim GS, Barrett CF, Pang CC, Davis JI. 2016. Drastic reduction of plastome size in the mycoheterotrophic *Thismia tentaculata* relative to that of its autotrophic relative *Tacca chantrieri*. *Am J Bot.* 103:1129–1137.
- Lin CS, et al. 2017. Concomitant loss of NDH complex-related genes within chloroplast and nuclear genomes in some orchids. *Plant J.* 90:994–1006.
- Lin Q, et al. 2022. Mitochondrial genomic data are effective at placing mycoheterotrophic lineages in plant phylogeny. *New Phytol.* 236:1908–1921.
- Lin Q, Ané C, Givnish TJ, Graham SW. 2021. A new carnivorous plant lineage (*Triantha*) with a unique sticky-inflouescence trap. *Proc Natl Acad Sci.* 118:e2022724118.
- Liu L, Yu L, Kubatko L, Pearl DK, Edwards SV. 2009. Coalescent methods for estimating phylogenetic trees. *Mol Phylogenet Evol.* 53:320–328.
- Liu L, Yu L, Pearl DK, Edwards SV. 2009. Estimating species phylogenies using coalescence times among sequences. *Syst Biol.* 58:468–477.
- Logacheva MD, Schelkunov MI, Nuraliev MS, Samigullin TH, Penin AA. 2014. The plastid genome of mycoheterotrophic monocot *Petrosavia stellaris* exhibits both gene losses and multiple rearrangements. *Genome Biol Evol.* 6:238–246.
- Maddison WP. 1997. Gene trees in species trees. *Syst Biol.* 46:523–536.
- Magallón S, Gómez-Acevedo S, Sánchez-Reyes LL, Hernández-Hernández T. 2015. A metacalibrated time-tree documents the early rise of flowering plant phylogenetic diversity. *New Phytol.* 207:437–453.
- Markova-Raina P, Petrov D. 2011. High sensitivity to aligner and false positives in the estimates of positive selection in the 12 *Drosophila* genomes. *Genome Res.* 21:863–874.
- Matasci N, et al. 2014. Data access for the 1,000 plants (1KP) project. *Gigascience* 3:17.
- McDowall J, Hunter S. 2011. Interpro protein classification. *Methods Mol Biol.* 694:37–47.
- McKain MR, et al. 2016. A phylogenomic assessment of ancient polyploidy and genome evolution across the Poales. *Genome Biol Evol.* 8:1150–1164.
- McKendrick SL, Leake JR, Taylor DL, Read DJ. 2000. Symbiotic germination and development of myco-heterotrophic plants in nature: ontogeny of *Corallorhiza trifida* and characterization of its mycorrhizal fungi. *New Phytol.* 145:15.
- McNeal JR, Kuehl JV, Boore JL, dePamphilis CW. 2007. Complete plastid genome sequences suggest strong selection for retention of photosynthetic genes in the parasitic plant genus *Cuscuta*. *BMC Plant Biol.* 7:57.
- McNeal JR, Kuehl JV, Boore JL, Leebens-Mack J, dePamphilis CW. 2009. Parallel loss of plastid introns and their maturase in the genus *Cuscuta*. *PLoS One* 4:e5982.
- Mennes CB, et al. 2015. Ancient gondwana break-up explains the distribution of the mycoheterotrophic family Corsiaceae (Liliales). *J Biogeography* 42:1123–1136.
- Mennes CB, Smets EF, Moses SN, Merckx VSFT. 2013. New insights in the long-debated evolutionary history of Triuridaceae (Pandanales). *Mol Phylogenet Evol.* 69:994–1004.
- Merckx V, et al. 2006. Phylogeny and evolution of Burmanniaceae (Dioscoreales) based on nuclear and mitochondrial data. *Am J Bot.* 93:1684–1698.
- Merckx VSFT, et al. 2013. Taxonomy and classification. In: Merckx V, editors. *Mycoheterotrophy*. New York (NY): Springer. p. 19–102.
- Merckx V, Bakker FT, Huysmans S, Smets E. 2009. Bias and conflict in phylogenetic inference of myco-heterotrophic plants: a case study in thismiaceae. *Cladistics* 25:64–77.
- Merckx V, Bidartondo MI, Hynson NA. 2009. Myco-heterotrophy: when fungi host plants. *Ann Bot* 104:1255–1261.
- Merckx VSFT, Mennes CB, Peay KG, Gemel J. 2013. Evolution and diversification. In: Merckx V, editors. *Mycoheterotrophy*. New York (NY): Springer. p. 215–244.
- Merckx VSFT, Smets EF, Specht CD. 2013. Biogeography and distribution. In: Merckx V, editors. *Mycoheterotrophy*. New York (NY): Springer. p. 103–156.
- Merckx V, et al. 2008. Diversification of myco-heterotrophic angiosperms: evidence from Burmanniaceae. *BMC Evol Biol.* 8:178.
- Mirarab S, Warnow T. 2015. ASTRAL-II: coalescent-based species tree estimation with many hundreds of taxa and thousands of genes. *Bioinformatics* 31:i44–i52.
- Molina J, et al. 2014. Possible loss of the chloroplast genome in the parasitic flowering plant *Rafflesia lagascae* (Rafflesiaceae). *Mol Biol Evol.* 31:793–803.
- Naumann J, et al. 2013. Single-copy nuclear genes place haustorial Hydnoraceae within Piperales and reveal a cretaceous origin of multiple parasitic angiosperm lineages. *PLoS One* 8:e79204.
- Nevill PG, et al. 2019. Plastome-wide rearrangements and gene losses in carnivorous Droseraceae. *Genome Biol Evol.* 11:472–485.
- Nickrent DL. 2020. Parasitic angiosperms: how often and how many? *Taxon* 69:5–27.
- Nickrent DL, Duff RJ, Konings DAM. 1997. Structural analyses of plastid-derived 16S rRNAs in holoparasitic angiosperms. *Plant Mol Biol.* 34:731–743.
- Nickrent DL, Yan OY, Duff RJ, dePamphilis CW. 1997. Do nonasterid holoparasitic flowering plants have plastid genomes? *Plant Mol Biol.* 34:717–729.
- Pamilo P, Nei M. 1988. Relationships between gene trees and species trees. *Mol Biol Evol.* 5:568–583.
- Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. 2017. Salmon provides fast and bias-aware quantification of transcript expression. *Nat Methods.* 14:417–419.
- Petersen G, Zervas A, Pedersen HA, Seberg O. 2018. Genome reports: contracted genes and dwarfed plastome in mycoheterotrophic *Sciaphila thaidanica* (Triuridaceae, Pandanales). *Genome Biol Evol.* 10:976–981.
- Pond SL, Frost SD. 2005. Datamonkey: rapid detection of selective pressure on individual sites of codon alignments. *Bioinformatics* 21:2531–2533.
- Pond SLK, Frost SDW, Muse SV. 2005. Hyphy: hypothesis testing using phylogenies. *Bioinformatics* 21:676–679.

- Preiss K, Gebauer G. 2008. A methodological approach to improve estimates of nutrient gains by partially myco-heterotrophic plants. *Isot Environ Health Stud.* 44:393–401.
- R Core Team. 2021. R: A language and environment for statistical computing. R foundation for statistical computing, Vienna, Austria. <https://www.R-project.org/>
- Rasmussen HN. 1995. *Terrestrial orchids: from seed to mycotrophic plant.* New York: Cambridge University Press.
- Rhee SY, et al. 2003. The Arabidopsis information resource (TAIR): a model organism database providing a centralized, curated gateway to *Arabidopsis* biology, research materials and community. *Nucleic Acids Res.* 31:224–228.
- Sakuraba Y, Yokono M, Akimoto S, Tanaka R, Tanaka A. 2010. Deregulated chlorophyll b synthesis reduces the energy transfer rate between photosynthetic pigments and induces photodamage in *Arabidopsis thaliana*. *Plant Cell Physiol.* 51:1055–1065.
- Sayyari E, Mirarab S. 2018. Testing for polytomies in phylogenetic species trees using quartet frequencies. *Genes (Basel).* 9:132.
- Schelkunov MI, Nuraliev MS, Logacheva MD. 2021. Genomic comparison of non-photosynthetic plants from family Balanophoraceae with their photosynthetic relatives. *Peer J.* 9:e12106.
- Schwender J, Goffman F, Ohlrogge JB, Shachar-Hill Y. 2004. Rubisco without the calvin cycle improves the carbon efficiency of developing green seeds. *Nature* 432:779–782.
- Sen K, Ghosh TC. 2013. Pseudogenes and their composers: delving in the 'debris' of human genome. *Brief Funct Genomics.* 12:536–547.
- Sheikh SI, Kahveci T, Ranka S, Burleigh JG. 2013. Stability analysis of phylogenetic trees. *Bioinformatics* 29:166–174.
- Silva SR, et al. 2016. The chloroplast genome of *Utricularia reniformis* sheds light on the evolution of the *ndh* gene complex of terrestrial carnivorous plants from the Lentibulariaceae family. *PLoS One* 11: e0165176.
- Simao FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31: 3210–3212.
- Smith SA, Dunn CW. 2008. Phyutility: a phyloinformatics tool for trees, alignments and molecular data. *Bioinformatics* 24:715–716.
- Smith SA, Moore MJ, Brown JW, Yang Y. 2015. Analysis of phylogenomic datasets reveals conflict, concordance, and gene duplications with examples from animals and plants. *BMC Evol Biol.* 15: 150.
- Smith SE, Read DJ. 2010. *Mycorrhizal symbiosis.* London: Academic Press.
- Soltis DE, et al. 2011. Angiosperm phylogeny: 17 genes, 640 taxa. *Am J Bot.* 98:704–730.
- Soto Gomez MS, et al. 2020. A bi-organellar phylogenomic study of Pandanales: inference of higher-order relationships and unusual rate-variation patterns. *Cladistics* 36:481–504.
- Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30: 1312–1313.
- Su HJ, et al. 2019. Novel genetic code and record-setting AT-richness in the highly reduced plastid genome of the holoparasitic plant *Balanophora*. *Proc Natl Acad Sci.* 116:934–943.
- Suetsugu K, Yamato M, Matsubayashi J, Tayasu I. 2021. Partial and full mycoheterotrophy in green and albino phenotypes of the slipper orchid *Cypripedium debile*. *Mycorrhiza* 31:301–312.
- Sun G, et al. 2018. Large scale gene losses underlie the genome evolution of the parasitic plant *Cuscuta australis*. *Nature Comm.* 9: 2683.
- Taylor DL, et al. 2013. Progress and prospects for the ecological genetics of mycoheterotrophs. In: Merckx V, editors. *Mycoheterotrophy.* New York (NY): Springer. p. 267–296.
- Taylor DL, Bruns TD, Leake JR, Read DJ. 2002. Mycorrhizal specificity and function in myco-heterotrophic plants. In: van der Heijden MGA and Sanders IR, editors. *Mycorrhizal ecology. Ecological studies (analysis and synthesis).* Vol 157. Berlin (Heidelberg): Springer. p. 375–413.
- Timilsena PR. 2020. *Phylogenomics and comparative transcriptomic analyses of the monocots [Ph.D. Dissertation].* [University Park (PA)]: The Pennsylvania State University.
- Timilsena PR, et al. 2022. Phylogenomic resolution of order- and family-level monocot relationship using 602 nuclear genes and 1375 BUSCO genes. *Front Plant Sci.* 13:876779.
- Vogel A, et al. 2018. Footprints of parasitism in the genome of the parasitic flowering plant *Cuscuta campestris*. *Nat Commun.* 9: 2515.
- Wafula EK, et al. 2022. Planttribes 2: tools for comparative gene family analysis in plant genomics. *Front Plant Sci.* doi: 10.3389/fpls.2022.1011199.
- Walker CJ, Weinstein JD. 1994. The magnesium-insertion step of chlorophyll biosynthesis is a two-stage reaction. *Biochem J.* 299: 277–284.
- Wall PK, et al. 2008. PlantTribes: a gene and gene family resource for comparative genomics in plants. *Nucleic Acids Res.* 36: D970–D976.
- Waterhouse RM, et al. 2017. BUSCO Applications from quality assessments to gene prediction and phylogenomics. *Mol Biol Evol.* 35: 543–548.
- Waterman RJ, Bidartondo MI. 2008. Deception above, deception below: linking pollination and mycorrhizal biology of orchids. *J Exp Bot.* 59:1085–1096.
- Weaver S, et al. 2018. Datamonkey 2.0: a modern web application for characterizing selective and other evolutionary processes. *Mol Biol Evol.* 35:773–777.
- Wertheim JO, Murrell B, Smith MD, Pond SLK, Scheffler K. 2015. RELAX: detecting relaxed selection in a phylogenetic framework. *Mol Biol Evol.* 32:820–832.
- Wicke S, et al. 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, et al. 2016. Mechanistic model of evolutionary rate variation en route to a nonphotosynthetic lifestyle in plants. *Proc Natl Acad Sci.* 113:9045–9050.
- Wicke S, Naumann J. 2018. Molecular evolution of plastid genomes in parasitic flowering plants. *Adv Bot Res.* 85:315–347.
- Wicke S, Schäferhoff B, dePamphilis CW, Müller KF. 2014. Disproportional plastome-wide increase of substitution rates and relaxed purifying selection in genes of carnivorous Lentibulariaceae. *Mol Biol Evol.* 31:529–545.
- Wicke S, Schneeweiss GM, dePamphilis CW, Müller KF, Quandt D. 2011. The evolution of the plastid chromosome in land plants: gene content, gene order, gene function. *Plant Mol Biol.* 76: 273–297.
- Wickett NJ, et al. 2007. Functional gene losses occur with minimal size reduction in the plastid genome of the parasitic liverwort *Aneura mirabilis*. *Mol Biol Evol.* 25:393–401.
- Wickett NJ, et al. 2011. Transcriptomes of the parasitic plant family Orobanchaceae reveal surprising conservation of chlorophyll synthesis. *Curr Biol.* 21:2098–2104.
- Wickett NJ, et al. 2014. Phylotranscriptomic analysis of the origin and early diversification of land plants. *Proc Natl Acad Sci.* 1114: E4859–E4868.
- Wolfe AD, dePamphilis CW. 1998. The effect of relaxed functional constraints on the photosynthetic gene *rbL* in photosynthetic and nonphotosynthetic parasitic plants. *Mol Biol Evol.* 15: 1243–1258.

- Wolfe KH, Morden CW, Palmer JD. 1992. Function and evolution of a minimal plastid genome from a nonphotosynthetic plant. *Proc Natl Acad Sci.* 89:10648–10652.
- Yang Z, et al. 2015. Comparative transcriptome analyses reveal core parasitism genes and suggest gene duplication and repurposing as sources of structural novelty. *Mol Biol Evol.* 32:767–790.
- Yoshida S, et al. 2019. Genome sequence of *Striga asiatica* provides insight into the evolution of plant parasitism. *Curr Biol.* 29:1–12.
- Yu QB, Huang C, Yang ZN. 2014. Nuclear-encoded factors associated with the chloroplast transcription machinery of higher plants. *Front Plant Sci.* 5:316.
- Yuan Y, et al. 2018. The *Gastrodia elata* genome provides insights into plant adaptation to heterotrophy. *Nat Commun.* 9:1615.
- Zeng L, et al. 2014. Resolution of deep angiosperm phylogeny using conserved nuclear genes and estimates of early divergence times. *Nat Commun.* 5:4956–4956.
- Zhang C, Rabiee M, Sayyari E, Mirarab S. 2018. ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinformatics* 19:153.
- Zimmer EA, Wen J. 2012. Using nuclear gene data for plant phylogenetics: progress and prospects. *Mol Phylogenet Evol.* 65:774–785.

**Associate editor:** Dr. Itay Mayrose