Unprecedented Parallel Photosynthetic Losses in a Heterotrophic Orchid Genus

Craig F. Barrett,*[1] Brandon T. Sinn,[1] and Aaron H. Kennedy[2]

[1]Department of Biology, West Virginia University, Morgantown, WV

[2]Mycology and Nematology Genetic Diversity and Biology Laboratory, USDA-APHIS, Beltsville, MD

***Corresponding author:** E-mail: cfb0001@mail.wvu.edu.

Associate editor: Tal Pupko

Abstract

Heterotrophic plants are evolutionary experiments in genomic, morphological, and physiological change. Yet, genomic sampling gaps exist among independently derived heterotrophic lineages, leaving unanswered questions about the process of genome modification. Here, we have sequenced complete plastid genomes for all species of the leafless orchid genus *Hexalectris*, including multiple individuals for most, and leafy relatives *Basiphyllaea* and *Bletia*. Our objectives are to determine the number of independent losses of photosynthesis and to test hypotheses on the process of genome degradation as a result of relaxed selection. We demonstrate four to five independent losses of photosynthesis in *Hexalectris* based on degradation of the photosynthetic apparatus, with all but two species displaying evidence of losses, and variation in gene loss extending below the species level. Degradation in the *atp* complex is advanced in *Hexalectris warnockii*, whereas only minimal degradation (i.e., physical loss) has occurred among some "housekeeping" genes. We find genomic rearrangements, shifts in Inverted Repeat boundaries including complete loss in one accession of *H. arizonica*, and correlations among substitutional and genomic attributes. Our unprecedented finding of multiple, independent transitions to a fully mycoheterotrophic lifestyle in a single genus reveals that the number of such transitions among land plants is likely underestimated. This study underscores the importance of dense taxon sampling, which is highly informative for advancing models of genome evolution in heterotrophs. Mycoheterotrophs such as *Hexalectris* provide forward-genetic opportunities to study the consequences of radical genome evolution beyond what is possible with mutational studies in model organisms alone.

Key words: **heterotroph, phylogenomics, plastome, pseudogene, selection, chloroplast, deletion, evolution.**

Article

Introduction

Heterotrophic plants present prime examples of genomic modification due to relaxed selective constraints on photosynthetic function. They are highly useful as "evolutionary experiments" in that they manage to survive and reproduce under natural conditions despite drastic modifications, in contrast to many induced mutants in model systems which would almost certainly go extinct in the wild. Thus, heterotrophs can increase our knowledge on how whole-organismal systems adjust to such radical changes. Plastid genomes are of particular interest because they are small, nonrecombining (i.e., in the sense of swapping material among homologous chromosomes as in nuclear DNA), and uniparentally inherited; occur in relatively high copy number per cell compared with nuclear and (typically) mitochondrial genomes; and are conserved in gene order, content, and, overall genome structure, with a large and small single copy region bisected by a large Inverted Repeat (IR; typically 20–30 kb) across angiosperms (see Ruhlman and Jansen 2014). The genes and spacers of the IR are among the most conserved regions of the plastome in terms of gene order and substitution rates and contribute to overall structural stability of the chromosome, multimeric structure among plastid chromosomes, and contain the all four of the highly conserved, plastid-encoded ribosomal RNA species (e.g., Zhu et al. 2016). The sequencing of complete plastid genomes (plastomes) has become more feasible because of high-throughput sequencing technology and assembly algorithms, and thus plastomes from throughout the tree of life are being published at a rapid pace (see Gitzendanner et al. 2018).

Plastid genome evolution in heterotrophic plants has received much attention in recent years (e.g., reviewed by Graham et al. [2017], Hadariová et al. [2018], and Wicke and Naumann [2018]), yet only a handful of studies have included taxon sampling of sufficient density to apply phylogenetic, comparative approaches (Wicke et al. 2013, 2016; Barrett et al. 2014, 2018; Feng et al. 2016; Braukmann et al. 2017; Schneider et al. 2018). Many studies have focused on single plastomes, which may fail to capture several key aspects of the process of plastid genome evolution, especially in cases of rapidly accumulated genotypic and phenotypic change due to relaxed selection and increased rates of substitution. Recent studies have begun to incorporate dense sampling of plastomes within clades, for example, at or near the genus and species level (Wicke et al. 2013; Barrett et al. 2014, 2018; Feng et al. 2016; Braukmann et al. 2017; Schneider et al. 2018). This approach is important in that it allows plastome

evolution to be interpreted on a finer scale—and more importantly, from a phylogenetic-comparative perspective—revealing details on the types of changes that occur and their mechanistic causes, not merely the result after several millions of years of evolution on a single plastome.

Species-level sampling in the leafless orchid genus *Corallorhiza* (plastome sequencing) and parasitic vine *Cuscuta* (family Convolvulaceae; DNA probe blotting) has revealed multiple independent shifts to full heterotrophy within single genera, and thus it is likely that the number of hypothesized transitions to complete heterotrophy and loss of photosynthesis have been underestimated in other groups (Revill et al. 2005; Funk et al. 2007; McNeal, Arumugunathan, et al. 2007; McNeal, Kuehl, et al. 2007; Braukmann et al. 2013; Barrett et al. 2014, 2018). Further, it remains to be determined if photosynthetic loss and subsequent plastome degradation can be detected below the species level.

What types of genomic changes occur in heterotrophic plants? Much attention has been given to the evolutionary fates of plastid-encoded gene systems coinciding with or following the evolution of heterotrophy (e.g., dePamphilis and Palmer 1990; Wolfe et al. 1992; Barbrook et al. 2006; Krause 2008, 2011, 2012; Wicke et al. 2011, 2013). Barrett and Davis (2012) and subsequently Barrett et al. (2014) proposed a general conceptual model, in which plastid-encoded gene systems are lost in a more-or-less linear, irreversible pattern (but allowing exceptions), following the general sequence of 1) *ndh* genes, encoding a chloroplast NAD(P)H dehydrogenase; 2) *psa*, *psb*, *rbcL*, *ccsA*, *cemA*, *ycf3*, and *ycf4*, encoding various complexes involved directly in photosynthesis (e.g., photosystems I and II, RuBisCO); 3) *rpo*, a plastid-encoded RNA polymerase; 4) *atp*, encoding a thylakoid ATP synthase; and 5) "housekeeping" genes involved in basic organellar functions such as intron splicing and translation (e.g., *matK*, *rpl*, *rps*, *rrn*, and *trn*). Barrett et al. (2014) combined stages 2 and 3, as the *rpo* complex primarily transcribes photosynthesis-related genes, and thus these two gene classes typically experience concomitant degradation in heterotrophic lineages. Further slight modifications of these conceptual models are proposed by Naumann et al. (2016) and Graham et al. (2017), based on availability of additional heterotrophic plastomes. The consensus among all the aforementioned conceptual models is that stage 2 (i.e., loss of genes involved directly in photosynthesis) represents whole-organismal loss of photosynthetic function, which thus is a major transitory event in both physiology and genome evolution of these plants.

Orchids, with >28,000 accepted species (Plant List 2013), have experienced a greater number of independent transitions to obligate heterotrophy and the resulting losses of photosynthesis than any other major clade, making this the ideal family for testing hypotheses associated with drastic evolutionary changes in the plastid genome (Merckx and Freudenstein 2010). The loss of photosynthesis is hypothesized to have occurred at least 30 times in the orchids alone, representing ~2/3 of all such transitions to fully mycoheterotrophic lifestyles in plants (Freudenstein and Barrett 2010). This form of obligate heterotrophy involves fungus-to-orchid transfer of carbon and essential nutrients. Transitions to full mycoheterotrophy are most likely facilitated by the condition of "initial mycoheterotrophy" in orchids, all of which begin their lives as nonphotosynthetic, mycoparasitic seedlings (protocorms) that eventually become photosynthetic with the development of leaves or other organs capable of photosynthesis (Rasmussen 1995; Bidartondo 2005; Leake 2005). Full mycoheterotrophy can be interpreted as a pedomorphic shift that has occurred many times in separate orchid lineages, and initial mycoheterotrophy can be viewed as a preadaptation to such shifts, for example, with the characteristic mycotrophic coralloid rhizome structure found in heterotrophic species representing a protracted, juvenile protocorm stage (Rasmussen 1995; Freudenstein and Barrett 2010; Merckx, Freudenstein, et al. 2013). These changes are often accompanied by fungal host shifts from "typical" orchid mycorrhizal fungal partners—saprotrophic taxa termed rhizoctonias (teleomorphic families Ceratobasidiaceae, Tulasnellaceae, and Sebacinaceae)—to ectomycorrhizal or even arbuscular mycorrhizal fungi, though some of the former are known to form ectomycorrhizas (Taylor and Bruns 1997; Taylor et al. 2003; Barrett et al. 2010; Kennedy et al. 2011).

Mycoheterotrophs in clades throughout the orchids exhibit several convergent morphoanatomical characteristics during the transition from autotrophy to heterotrophy. These changes have resulted in modifications to several organs, including reduction or loss of roots and leaves, prevalence of cleistogamy (closed-flowers), autonomous self-pollination, reduction or loss of stomata, and reduction in vascular tissues (Leake 1994; Leake et al. 2008; Merckx and Freudenstein 2010).

The leafless orchid genus *Hexalectris* is composed of nine species, all of which have been presumed to be full mycoheterotrophs, distributed throughout Western North America and Mexico (Catling 2004; Kennedy and Watson 2010), with a single species, *H. spicata*, extending into the eastern USA (Catling and Engel 1993). *Hexalectris* is a member of the orchid subtribe Bletiinae (tribe Epidendreae, subfamily Epidendroideae), which contains four genera based on the recent phylogenetic analysis and classification (Chase et al. 2015; Freudenstein and Chase 2015). Bletiinae represents an apt case study in the transition to obligate parastitism, as its four genera encompass different growth forms, including 1) the leafy, green, epiphytic *Chysis*; 2) the leafy, green, terrestrial *Bletia*; 3) the green, terrestrial *Basiphyllaea*, with 1–2 highly reduced or basal leaves; and 4) the leafless, terrestrial, mycoheterotrophic *Hexalectris*, which at least superficially appears to lack visible green tissue (supplementary fig. S1, Supplementary Material online).

Although convergent morphological reduction is apparent in Bletiinae, especially in the leafless *Hexalectris*, it is unclear whether these changes have occurred in parallel with the loss of photosynthesis and resulting plastome degradation. For example, the orchid *Corallorhiza* includes leafless photosynthetic and nonphotosynthetic species, meaning that multiple, independent losses of photosynthesis can occur within a single genus after the loss of leaves (Zimmer et al. 2008; Cameron et al. 2009;

Table 1. Details of Accessions Included in This Study.

Coll.#	Genus	Species	Loc	Length	IR	LSC	SSC	GC	x-cov	#Reads	Mapped	%Plastid Reads
Baco1	*Basiphyllaea*	*corallicola*	FL	158,368	26,910	86,711	17,873	37.3	60.8	13,219,124	139,287	1.05
32	*Bletia*	*roezlii*	JA	158,720	26,860	87,177	17,823	37.2	598.0	14,792,462	940,676	6.36
47	*Hexalectris*	*arizonica*	TX	109,543	4,475	90,507	10,086	35.9	169.8	11,912,444	184,256	1.55
271	*Hexalectris*	*arizonica*	TX	105,192	n/a	n/a	n/a	35.5	485.3	46,290,820	507,602	1.10
377	*Hexalectris*	*arizonica*	AZ	109,461	4,306	90,957	9,892	35.9	415.7	87,738,174	451,735	0.52
36	*Hexalectris*	*brevicaulis*	JA	138,101	31,225	68,147	7,504	37.1	443.0	69,828,048	608,415	0.87
41	*Hexalectris*	*brevicaulis*	JA	137,846	31,263	67,707	7,399	37.1	193.1	26,559,816	170,483	0.64
166	*Hexalectris*	*colemanii*	AZ	117,252	4,865	97,471	10,051	36.2	36.6	6,253,740	37,771	0.60
64	*Hexalectris*	*grandiflora*	TX	152,460	27,090	83,614	14,669	36.8	551.0	49,121,014	834,057	1.70
48	*Hexalectris*	*nitida*	TX	111,720	15,147	71,963	9,063	37.2	72.6	12,620,482	88,569	0.70
79	*Hexalectris*	*nitida*	TX	111,592	15,075	72,400	9,042	37.1	100.4	12,461,696	110,985	0.89
270	*Hexalectris*	*nitida*	TX	111,519	15,148	72,181	9,042	37.1	146.7	13,268,698	161,991	1.22
85	*Hexalectris*	*parviflora*	JA	136,661	27,129	71,757	10,508	37	398.8	46,935,030	540,450	1.15
500	*Hexalectris*	*parviflora*	CH	134,807	26,814	70,342	10,837	37.1	112.1	12,089,618	149,663	1.24
320	*Hexalectris*	*revoluta*	TX	145,345	25,555	83,813	10,422	36.9	206.3	14,764,714	297,143	2.01
44	*Hexalectris*	*spicata*	NC	120,704	23,681	63,566	9,776	37.2	761.3	35,558,544	911,263	2.56
50	*Hexalectris*	*spicata*	TX	120,185	23,582	62,242	10,779	37.2	1137.5	47,730,484	1,348,183	2.83
267	*Hexalectris*	*spicata*	OK	121,646	23,681	63,674	10,610	37.2	932.7	61,605,466	1,125,190	1.83
276	*Hexalectris*	*spicata*	IN	120,027	23,564	62,140	10,759	37.3	45.6	19,701,640	55,164	0.28
49	*Hexalectris*	*warnockii*	TX	119,057	17,332	66,903	17,490	36.9	703.5	38,633,900	829,431	2.15
69	*Hexalectris*	*warnockii*	TX	119,058	17,339	66,938	17,442	36.9	115.3	14,329,294	135,738	0.95

Note.—Coll.#, collection number; Loc, locality (US or Mexican state); length, plastome size in bp; IR, length of the inverted repeat; LSC, large single copy region; SSC, small single copy region; GC, GC content of the plastome excluding one IR copy; x-cov, mean coverage depth of the plastome; #reads, total number of reads generated; mapped, the number of reads mapping to the completed plastome; %plastid reads, percentage of the total read pool mapping to the plastome; and GenBank, NCBI accession number.

Barrett and Davis 2012; Barrett et al. 2014, 2018). Here, we sampled all species of *Hexalectris* (including multiple accessions for most), plus representatives of the related, leafy *Bletia* and *Basiphyllaea* using high-throughput Illumina sequencing, and assembled complete plastomes. We undertake a phylogenetic, comparative approach to elucidate the broad- and fine-scale processes of plastome modification associated with the transition to full mycoheterotrophy. Specifically, we address the following questions: 1) Do complete plastomes provide resolution and support for relationships among members of Beltiinae, and specifically within *Hexalectris*? 2) What are the major patterns of plastid genome modification across subtribe Bletiinae, particularly within the leafless *Hexalectris*, both among and within species? 3) Is there genomic evidence for loss or retention of photosynthesis in members of the leafless *Hexalectris*, and further, is there evidence for a single loss or multiple losses of photosynthesis in the genus? 4) How do substitution rates and patterns of selection vary among species that have lost photosynthetic apparatus compared with those in which it is retained?

Results

Plastid Genome Data

Plastome Data, Coverage Depth, Percent Plastid DNA

The mean number of paired end reads generated per library was 28,172,786 with a standard deviation of 17,282,194.6 (range = 6,253,740–87,738,174) (table 1). Coverage depth ranged from 36.6 to 1,137.5×, and plastid DNA content per read pool ranged from 0.28% to 6.36%. Percent similarity for all accessions is given in supplementary table S1, Supplementary Material online.

Phylogenetic Analyses

Relationships

We used likelihood, Bayesian, and parsimony to analyze concatenated, filtered locally collinear blocks (LCBs; i.e., the "LCB" matrix, based on all syntenic regions concatenated) and a combined matrix of all protein-coding regions ("CDS" matrix) to resolve and provide support for relationships among *Hexalectris* plastomes. Based on the LCB matrix, we found strong support and posterior probability (pp) for a single topology, with no conflicting relationships among reconstruction methods (fig. 1*A*). *Basiphyllaea* is sister to *Hexalectris*, and within *Hexalectris*, *H. warnockii* is sister to all other accessions. Three accessions of *H. arizonica* are supported as sister to *H. colemanii*, which collectively is sister to two accessions of *H. parviflora*. This entire clade is sister to a single accession of *H. revoluta*. Four accessions of *H. spicata* form a clade, within which accessions follow a topology of ([{276 Indiana, 44 North Carolina}, 267 Oklahoma], 50 Texas). Accessions of *H. spicata* are sister to three accessions of *H. nitida*, all from Texas, USA. The *H. nitida-spicata* clade is sister to a *H. arizonica-colemanii-parviflora-revoluta* clade, and this entire clade is sister to *H. brevicaulis*. All relationships are supported at >99% bootstrap/jackknife (or >0.99 pp), with the exception of *H. spicata* accessions 44 (NC) and 276 (IN). Relationships are nearly identical for the "CDS" matrix based on gene-partitioned analysis of coding regions (fig. 1*B*), except for the placement of *H. revoluta*, which instead is supported as sister to two accessions of *H. parviflora*.

Plastid Genome Structural Evolution

Plastome Size

We compared sizes across representatives of the Bletiinae in order to quantify the overall degree of sequence loss (or gain).

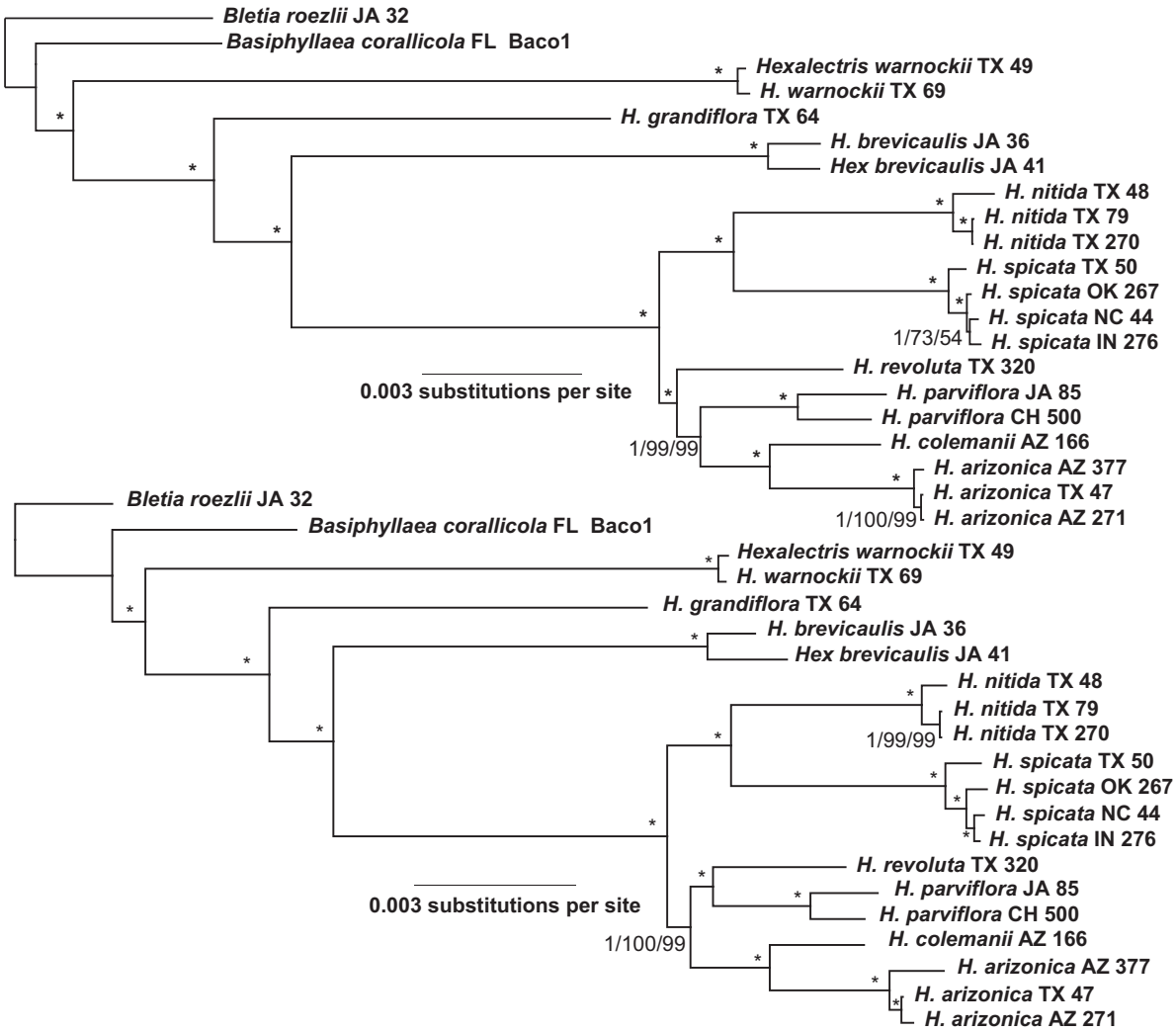

Fig. 1. Phylogenetic trees for LCB (above) and CDS matrices (below). * = pp of 1, ML bootstrap value of 100%, and Parsimony jackknife value of 100%; otherwise, values are indicated respectively. Scale bar and branch lengths = substitutions per site from Bayesian trees.

Plastome length varies little between the leafy *Bletia roezlii* and the reduced-leaved *Basiphyllaea corallicola* at 158,720 and 158,368 bp, respectively. Plastome length varies considerably, however, in the leafless *Hexalectris*, from a maximum of 152,460 bp in *H. grandiflora* to a minimum of 105,192 bp in an accession of *H. arizonica* (accession 271, table 1; fig. 2), representing approximately a 1.5-fold difference in total plastome length.

IR Variation. We compared the size, structure, and gene content in the IR region among members of Bletiinae, to quantify its contribution to overall change in genome evolutionary structural dynamics. The length of the IR in the photosynthetic *Bletia* and *Basiphyllaea* is 26,860 and 26,910 bp, respectively, whereas the IR ranges from 31,225 and 31,263 bp in both accessions of *Hexalectris brevicaulis*, to 4,475 and 4,306 in two of the three accessions of *H. arizonica*. A third accession of *H. arizonica* has no detectable IR region (table 1, fig. 2, and supplementary fig. S2, Supplementary Material online). Evidence for this IR loss comes from analysis of coverage depth: mean coverage depth for accession 271 in the region corresponding to the IR in two other accessions is 587.5× compared with 584.5×—or roughly equivalent—for the rest of the plastome. By contrast, *H. arizonica* accession 47, which has an IR, has a mean coverage depth of 275.8× in the IR and 182.6× outside the IR. Further, the IR junctions and corresponding regions in accession 271 (which lacks an IR) are covered by paired end information, verifying contiguity across regions. The single accession of *H. colemanii* included here has an IR length of 4,865, similar to that in *H. arizonica*. Thus, aside from the loss of the IR in one accession of *H. arizonica*, the net percent change in the size of the IR is +16% in *H. brevicaulis*, and −84% in *H. arizoniza*. On average, the IR regions of *Hexalectris* species are ∼30% smaller than that of the leafy *Bletia* and *Basiphyllaea*. Variation in IR length is predominantly due to boundary shifts relative to insertion/deletion.

Locally Collinear Blocks

We analyzed LCBs in MAUVE in order to characterize the degree of synteny among regions of the plastid genome in *Hexalectris* relative to leafy, photosynthetic relatives. MAUVE

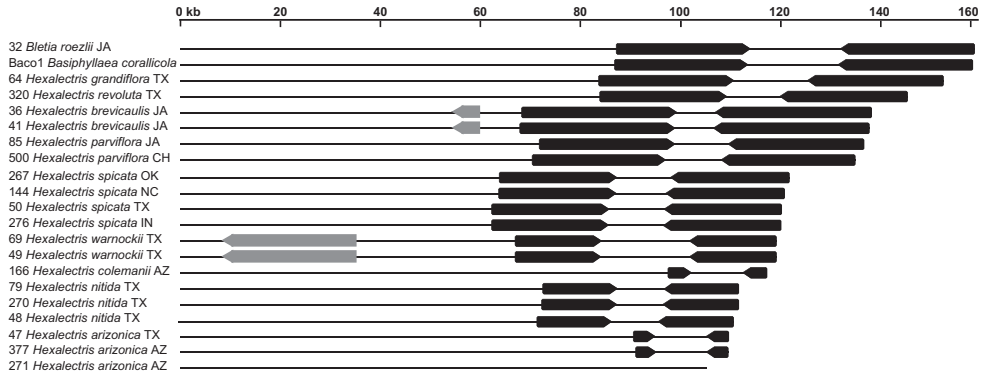

Fig. 2. Scaled representation of plastid genomic regions in *Bletia*, *Basiphyllaea*, and *Hexalectris*, ranked by decreasing total plastome length. Black arrows = IR copies and gray arrows = genomic inversions relative to *Bletia* and other angiosperm plastomes. Thin black lines on the left = LSC region and to the right = small single copy region. Note that *Hexalectris arizonica* accession 271 lacks an IR.

provides evidence for 12 syntenic regions, considering all 21 accessions. Both accessions of *H. warnockii* share a 29-kb inversion in the large single copy (LSC) region (fig. 2 and supplementary fig. S3, Supplementary Material online), containing the following genes: $trnS^{GCU}$, $trnG^{UCC}$, $trnR^{UCU}$, $\psi atpA$, $\psi atpF$, *atpH*, $\psi atpI$, *rps2*, $\psi rpoC2$, $\psi rpoC$, $\psi rpoB$, $trnC^{GCA}$, *petN*, *psbM*, $trnD^{GUC}$, $trnY^{GUA}$, $trnE^{UUC}$, $trnT^{GGU}$, $\psi psbD$, $\psi psbC$, $trnS^{CGA}$, *psbZ*, $trnG^{UUC}$, $trnfM^{CAU}$, *rps14*, $\psi psaB$, $\psi psaA$, $\psi ycf3$, and $trnS^{GGA}$. Both accessions of *H. brevicaulis* share an ~6.7-kb inversion in the LSC as well, containing $\psi psbB$ (5′ partial), *clpP*, *rps12* (5′ exon), *rpl20*, *rps18*, *rpl33*, *psaJ*, $trnP^{UGG}$, $trnW^{CCA}$, *petG*, *petL*, *psbE*, *psbF*, *psbL*, *psbJ*, and *petA*. The double cut and join (DCJ) distance is 1 for both *H. warnockii* and *H. brevicaulis* relative to all other accessions, corresponding to the number of major structural rearrangements of the plastome.

Physical and Functional Gene Losses

We quantified the numbers of putatively functional coding genes, pseudogenes, and wholescale gene losses to summarize the structural evolutionary consequences of relaxed selection pressure on each plastid gene system and genome structure overall. Figure 3 and supplementary figure S4, Supplementary Material online, summarize the number of putatively functional genes, pseudogenes, and physical gene losses among *Bletia*, *Basiphyllaea*, and *Hexalectris*. All plastid genes are putatively functional in *Bletia* and *Basiphyllaea*, as evidenced by intact reading frames. *Hexalectris grandiflora* and *H. revoluta* contain pseudogenes and have lost some members of the *ndh* complex but show no evidence of functional/physical loss of genes essential for photosynthesis. All other accessions of *Hexalectris* show extensive degradation of the *ndh* complex and "photosynthesis-related" genes (e.g., *psa/psb*, *pet*, and *rpo*). *Hexalectris warnockii* further displays evidence of pseudogenes for subunits of the ATP synthase (*atp*) complex, with five of six *atp* genes as pseudogenes (all but *atpH*). No single photosynthesis-related gene is present as an intact reading frame in all species of *Hexalectris*. There is neither evidence of functional or physical loss in any of the protein-coding "housekeeping" genes (e.g., *infA*, *rps/rpl*, *accD*, *matK*, *clpP*, *ycf1*, and *ycf2*) nor evidence of loss-of-function in any ribosomal RNA gene (*rrn*).

We find evidence of intraspecific variation in functional gene content in four of the six species for which multiple individuals were sampled (fig. 3 and supplementary fig. S4, Supplementary Material online). Overall, 8% of all functional and physical gene loss events happen *within* species of *Hexalectris*, based on our sampling. Most gene losses within species occur to small, photosynthesis-related genes, with the exception of three independent functional or physical losses of *rpo* genes (in *H. warnockii* and *H. parviflora*; fig. 3). *rpoB* and *psaJ* have been lost in *H. warnockii* accession 49, whereas they remain as pseudogenes in accession 69. *psbT* and *psbZ* have been lost in accession 41 in *H. brevicaulis*, whereas *psbJ* has become a pseudogene in accession 377 of *H. arizonica* but not in the other two accession of this species. In *H. parviflora*, *psaI* and *psaJ* are pseudogenes in accession 500, whereas *petG* has been lost and *rpoC* has become a pseudogene in accession 85.

Transfer RNA genes are largely conserved, with a few notable exceptions. $trnS^{UGA}$ is missing from both accessions of *H. warnockii*, and $trnT^{GGU}$ is missing from all accessions of *H. arizonica*. $trnL^{CAA}$ presents a more ambiguous case (fig. 3 and supplementary fig. S5, Supplementary Material online), in that there has been a 3-bp deletion relative to all other taxa in both accessions of *H. warnockii*, whereas there has been a 5-bp insertion in a single accession of *H. parviflora* (accession 500 from Chihuahua, Mexico). Predicted secondary structures vary in these accessions, but all length variation is observed in the "variable loop" (V-loop) region.

Loss of Photosynthetic Capacity

Topology Tests for a Single Loss of Photosynthesis

We used Bayes factor calculations based on stepping-stone sampling of the posterior distribution to quantify evidence for a single or multiple losses of photosynthesis in leafless *Hexalectris*. Our analyses provide strong evidence against the topology in which all *Hexalectris* accessions displaying evidence of photosynthetic loss are constrained to be monophyletic, as compared with a negative constraint in which they are nonmonophyletic (mean $\ln L_{\text{monophyly}} = -129{,}194.1$, mean $\ln L_{\text{nonmonophyly}} = -124{,}795.9$, and Bayes

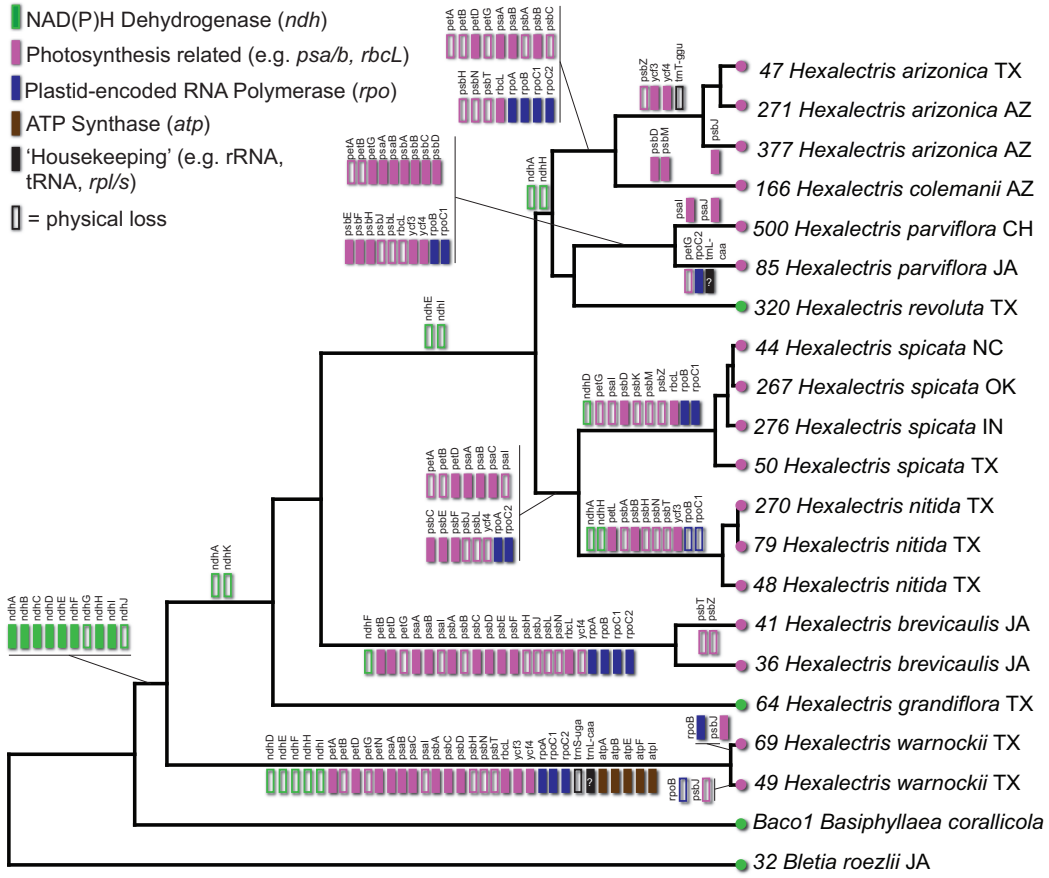

FIG. 3. Functional gene content of *Bletia*, *Basiphyllaea*, and *Hexalectris* based on ancestral state reconstruction of each gene under a loss-only (Dollo) model of transition probabilities, using the "CDS" matrix. Filled rectangles = pseudogenes and open rectangles = physical gene losses. "?" for $trnL^{CAA}$ = evidence of length variation in the "variable loop" region, which may be a precursor to pseudogene formation.

factor = 4,398.2). The same is true for the analysis in which *H. spicata*, *H. nitida*, *H. parviflora*, and *H. revoluta* are constrained to be monophyletic (as in Kennedy and Watson [2010] and Sosa et al. [2016]; mean $\ln L = -135{,}098.1$, Bayes factor = 10,302.2 when compared with negative constraint model).

Ancestral Reconstruction of Photosynthetic Status

We used ancestral state reconstruction to characterize the evolutionary history of photosynthetic loss in leafless *Hexalectris*. Parsimony analysis under a cost matrix weighting maximally against photosynthetic gains reveals four and five independent losses of photosynthesis, based on the LCB and CDS matrices, respectively (supplementary fig. S6, Supplementary Material online): 1) *Hexalectris warnockii*, 2) *H. brevicaulis*, 3) the *H. parviflora-colemanii-arizonica* clade (LCB matrix), 4) the *H. nitida-spicata* clade, and 5) in *H. parviflora* (CDS matrix). An analysis in which all character state changes are equally weighted gives a most parsimonious resolution of three total state changes: two losses (one in *H. warnockii* and the clade containing all taxa excluding *H. warnockii* and *H. grandiflora*); and a gain of photosynthesis in *H. revoluta*. Finally, an analysis in which all changes are constrained to be gains by weighting against losses yields four independent gains of photosynthesis (*Bletia*, *Basiphyllaea*, *H. grandiflora*, and *H. revoluta*).

Maximum likelihood reconstruction of photosynthetic loss under a model restricting photosynthetic gains (i.e., a loss-only model) reveals four to five losses of photosynthesis in *Hexalectris* (fig. 4). Ancestral states for photosynthetic losses in *H. warnockii* and *H. brevicaulis* are unequivocal at >0.99 for a photosynthetic ancestor for both the LCB and CDS matrices. The ancestral state for the *H. spicata-nitida* complex is nonphotosynthetic, though with a probability of 0.63 for the LCB tree and 0.61 for the CDS matrix. For the LCB tree, *H. revoluta* is sister to a clade of (*H. parviflora*, [*H. colemanii*, {*H. arizonica*}]). The ancestral state for the latter clade, excluding *H. revoluta*, is 0.76 in favor of a photosynthetic ancestor, raising the possibility that *H. parviflora* and the *H. colemanii-arizonica* clade each separately lost photosynthesis. Conservatively, reconstruction based on the LCB matrix provides evidence for a minimum of four independent losses in *Hexalectris*. The situation for the CDS matrix is less equivocal, in that the *H. revoluta* is sister to *H. parviflora*, thus providing evidence for at least five independent losses.

Testing among Discrete Trait Models of Photosynthetic Loss

We tested among alternative character state transition models to quantify the number of putative losses and gains of photosynthetic function. The maximum likelihood "equal rates" (ER) and "gain-only" models—in the latter,

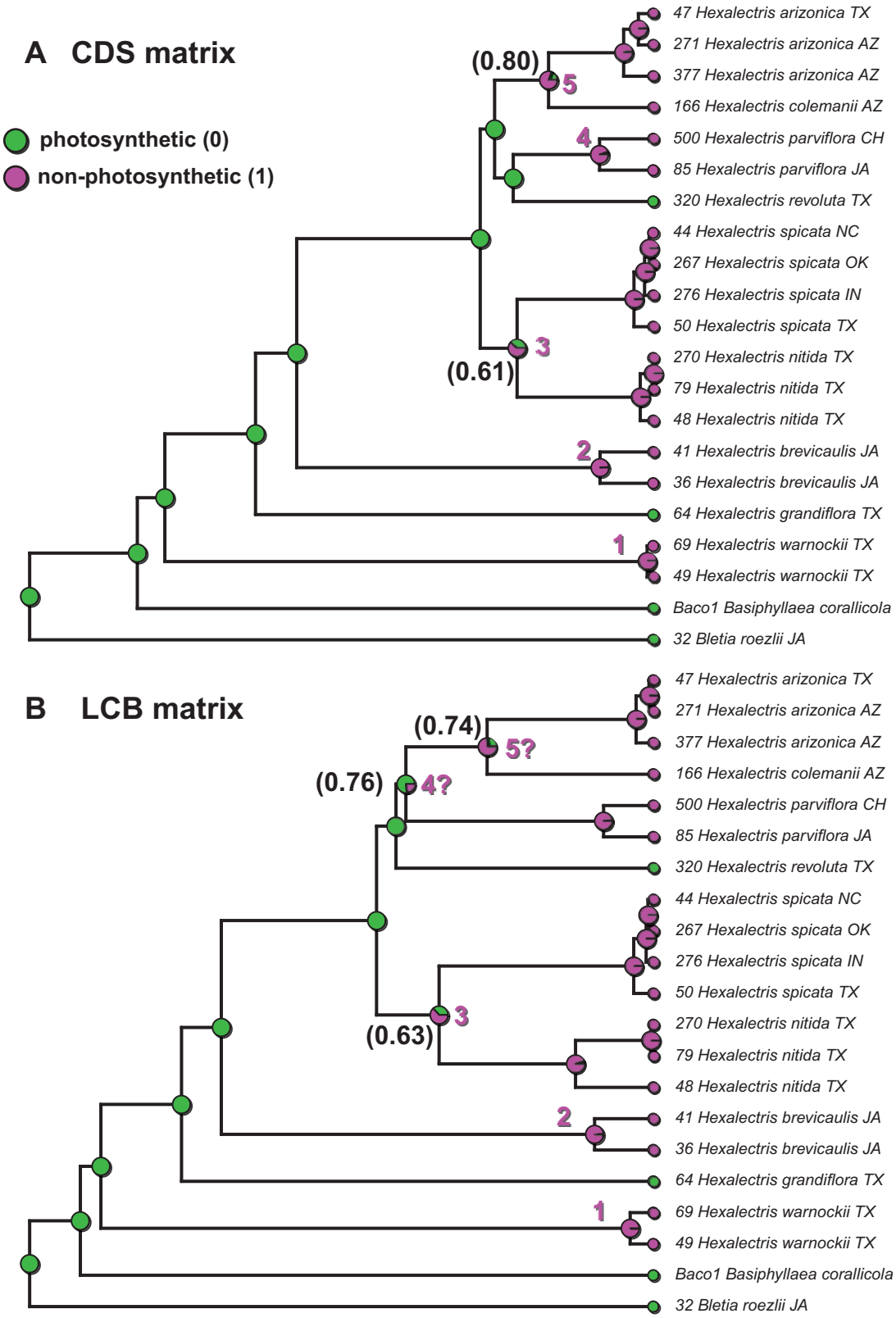

FIG. 4. Trees and ancestral state reconstructions of photosynthetic capability based on (*A*) the "CDS" matrix and (*B*) the "LCB" matrix. Photosynthetic capability is represented as a binary character, under a model allowing only losses of photosynthesis and no gains. Green circles/pie charts represent the probability of a photosynthetic ancestor; red represents a nonphotosynthetic ancestor. Numbers in black next to nodes represent the likelihood fraction of the most likely ancestral state; all other nodes have likelihoods >0.95. Numbers in red indicate the inferred instances of photosynthetic loss. Arrows indicate the alternative position of *Hexalectris revoluta* based on each matrix.

Table 2. Tests of Alternative Models of Transition Probabilities Between Photosynthetic Character States (0 = Photosynthetic and 1 = Nonphotosynthetic).

Model	*q*-Matrix	ln *L*	fp	AICc	AICweight	q_{01}; q_{10}
ER, equal rates	$q_{01} = q_{10}$	−6.78896	1	15.58	0.42	0.71; 0.71
ARD, different rates	$q_{01} \neq q_{10}$	−6.78896	2	17.58	0.15	0.71; 0.72
Gain-only	q_{10} only	−6.82156	1	15.64	0.41	0; 0.68
Loss-only	q_{01} only	−9.70410	1	21.41	0.02	1.14; 0

NOTE.—"*q*-matrix" specifies the specific state transition constraints for each model; "ln *L*" is the likelihood of the model; fp, the number of free model parameters; "AICc" is the corrected Akaike information criterion; "AICweight" is the Akaike weight for a particular model in favor of the alternatives; and "q_{01}; q_{10}" represent the estimated transition probabilities.

photosynthesis can only be gained and not lost (i.e., 1 → 0 only)—had the highest Akaike weights (for the "ER" model, AW = 0.42; for the "gain-only" model AW = 0.41), thus providing the best two explanations among the four competing models of discrete character change (table 2). The "loss-only" model, which allows only photosynthetic losses and no gains (i.e., only 0 → 1) had the lowest weight among the four models compared (AW = 0.02), and the "allowing different rates" model had the second lowest weight (AW = 0.15).

Evolutionary Rate Variation and Photosynthetic Loss

Photosynthetic Capacity and Substitution Rate Associations

We used TraitRateProp and RELAX to explore evidence of different substitution rates and relaxed selection associated with photosynthetic loss. We found evidence of different substitution rates in putatively nonphotosynthetic taxa among four genes encoding the ATP synthase complex (*atpA*, *atpE*, *atpF*, and *atpI*; here excluding *H. warnockii* which contains *atp* pseudogenes), three ribosomal protein genes (*rps3*, *rps4*, and *rps8*), the ATP-dependent Clp protease proteolytic subunit (*clpP*), *ycf1*, and *ycf2* (supplementary table S2, Supplementary Material online). All other genes showed no significant substitution rate changes in association with photosynthetic loss. Only two loci displayed evidence of relaxed selective constraints associated with the loss of photosynthesis: *atpA* (*H. warnockii* excluded from analysis) and *rpl14* (supplementary table S2, Supplementary Material online). Analyses of genes concatenated by functional class (*atp*, *accD-clpP-matK*, *rpl*, and *rps*) yielded increased substitution rates associated with photosynthetic loss for the *atp* complex (excluding *H. warnockii*), *accD-clpP-matK* as a concatenated alignment, and all *rps* genes, encoding small-subunit ribosomal proteins (supplementary table S3, Supplementary Material online). RELAX analyses of the same concatenated matrices reveal only evidence for intensified selection in the *atp* complex, again, excluding *H. warnockii*.

Correlations among Genomic and Substitutional Features

dS and d*N* are significantly correlated with one another, and with total plastome length, the aggregate length of protein-coding genes, the number of functional genes, the number of tandem repeats, and indels (summarized in supplementary table S3, Supplementary Material online). dS is further correlated with DCJ distance (i.e., rearrangements), whereas d*N* is further correlated with IR length. GC content is not significantly correlated with any other trait. Total plastome length is positively correlated with the length of coding DNA, the number of functional genes, tandem repeats, and indels. A nearly identical set of correlations is found for the cumulative length of coding DNA as for total plastome length. The number of functional genes is positively correlated with tandem repeat content and negatively correlated with indel content. DCJ distance is correlated with both tandem (negative) and dispersed repeats (positive), as well as indel content (positive). Lastly, IR length is positively correlated with both tandem and dispersed repeat content.

Discussion

We present compelling evidence of four to five independent losses of photosynthesis in a single genus of mycoheterotrophic orchid, which is unprecedented among studies published thus far on plastid genome evolution in heterotrophic plants. This is particularly striking because it indicates that independent loss of photosynthesis has occurred within approximately half of the species in this genus. Our findings therefore indicate that the number of independent transitions to an obligately mycoheterotrophic, nonphotosynthetic lifestyle has likely been underestimated in orchids, and also across land plants. Our conclusions are based on evidence from 1) plastome size reduction, photosynthetic gene loss, and pseudogenization in several species of *Hexalectris*, but not in others; 2) a strongly supported phylogenetic hypothesis based on complete plastid genomes; and 3) hypothesis tests of photosynthetic loss. We further find evidence of accelerated substitution rates, inter- and intra-specific variation in IR dynamics, and correlations among substitutional parameters and genomic structural features.

Phylogenetic Analyses

We recover phylogenetic relationships similar to those in previous studies of Bletiinae based on Sanger sequencing (nuclear ITS and four plastid DNA) that include representatives of *Hexalectris* (Sosa 2007; Kennedy and Watson 2010; Sosa et al. 2016). Our results also support the revised species circumscriptions within the *H. spicata* complex proposed by Kennedy and Watson (2010). The main difference in topology found here is in the placement of the *H. arizonica-colemanii* clade; we find strong evidence for the former as a member of a clade with *H. parviflora* and *H. revoluta*, whereas relationships recovered in the most recent study placed *arizonica-colemanii* as sister to *revoluta-parviflora-nitida-spicata* (Sosa et al. 2016). Bayesian topology tests based on whole plastome data provide strong evidence against the latter topology. Only the placement of *H. revoluta* differs between our analyses of the LCB and CDS matrices, with strong support in both cases, as sister of *H. parviflora* (CDS matrix) or sister to a clade of *H. parviflora-colemanii-arizonica* (LCB matrix).

Plastid Genome Structural Evolution

Plastome size evolution in *Hexalectris* follows a similar pattern observed in recent studies on heterotrophic angiosperms, including: 1) the holoparasitic Orobanchaceae (Wicke et al. 2013, 2016), 2) fully mycoheterotrophic members of the orchid genus *Corallorhiza* (Barrett and Davis 2012; Barrett et al. 2014, 2018), 3) fully mycoheterotrophic members of the orchid tribe Neottieae (Feng et al. 2016), and 4) the monotropoid Ericaceae (Braukmann et al. 2017). Plastomes are more highly reduced in *Hexalectris* than in *Corallorhiza*, but not to the extent observed in Orobanchaceae, Neottieae, or monotropoid Ericaceae. *Hexalectris* occupies a somewhat intermediate position on the continuum of plastome degradation relative to other groups that have been sufficiently sampled. *Hexalectris* is estimated to have diverged from its closest leafy ancestor (most likely *Basiphyllaea*) about 33 Ma, in the late Oligocene (Sosa et al. 2016). *Hexalectris warnockii*—representing the earliest divergence in the genus—is estimated to have split from the ancestor of the remaining *Hexalectris* species about 24 Ma (Sosa et al. 2016). For comparison, the crown age of *Corallorhiza* is estimated to be ~9 Ma, having lost ~20% of its plastome in the most highly reduced species, *C. bentleyi* and *C. involuta*. *Hexalectris arizonica*, representing the smallest plastome in the genus, has been reduced by 34%, compared with the leafy *B. roezlii* (table 1, fig. 2). In contrast, the stem age and minimum plastome sizes among other clades containing heterotrophs are estimated to be 40–50 Ma and 68 kb in *Conopholis*, representing holoparasitic Orobanchaceae (Wicke et al. 2016); ~90 Ma and 27 kb in *Hydnora visseri* representing holoparasitic Hydnoraceae (Naumann et al. 2013, 2016); ~80 Ma and 11 kb in *Pilostyles aethiopica* representing holoparasitic Apodanthaceae (Bellot and Renner 2014, 2015); and ~100 Ma and 0 kb in *Rafflesia lagascae* representing Rafflesiaceae (Molina et al. 2014). Thus, time since divergence is likely a key factor in determining the degree of plastome degradation in heterotrophic lineages, in addition to the degree of weakening of negative selection on photosynthetic function and associated acceleration of mutation rates associated with a purely heterotrophic lifestyle (Wicke and Naumann 2018).

Loss of Photosynthetic Capacity

We provide strong evidence demonstrating that several independent losses of photosynthesis have occurred in parallel in *Hexalectris*. Bayes factor comparisons constraining all taxa with genomic evidence of photosynthetic loss to be monophyletic versus nonmonophyletic allow us to conclude that a single loss of photosynthesis is highly unlikely. Ancestral state reconstructions of each plastid gene's functional capacity and of photosynthetic capability as a binary character under a Dollo model (i.e., no reversals back to photosynthesis once it is lost) suggest that photosynthesis has been lost multiple times in the genus (figs. 3 and 4). Ironically, such a scenario is least parsimonious and has the lowest likelihood among various alternative models allowing photosynthetic gain, but these are not biologically plausible.

ndh Genes

We detect pseudogenes and losses of NAD(P)H dehydrogenase subunit genes in all members of *Hexalectris* (i.e., coinciding with the loss of leaves) but not in leafy relatives *Bletia* and *Basiphyllaea*. Loss of *ndh* genes corresponds to the first stage of degradation in many recent conceptual models of plastome degradation in heterotrophic plants (Wicke et al. 2011, 2016; Barrett and Davis 2012; Barrett et al. 2014; Naumann et al. 2016; Graham et al. 2017; Wicke and Naumann 2018). Our findings are unsurprising in that loss of the *ndh* complex is well documented in other clades containing hemiparasites and partial mycoheterotrophs (e.g., Barrett et al. 2014; Feng et al. 2016), as well as in leafy, otherwise photosynthetic orchids (Lin et al. 2017). Losses of *ndh* in the latter may be associated with the tendency of many orchids to grow in low-light, canopy habitats as epiphytes, or in dark, shaded woods as terrestrials (e.g., Wicke et al. 2011; Lin et al. 2017). The *ndh* complex largely functions as an electron recycling system, allowing the fine tuning of photosynthesis (Peltier et al. 2016). Thus, in epiphytes or terrestrial orchids growing in relatively dark, forested habitats, this complex may be nonessential; this situation may be augmented by dependence upon host-derived carbon and decreasing dependence upon photosynthetic carbon in parasites, explaining why *ndh* genes are lost in nearly every parasite examined to date.

The relative degree of reliance on photosynthesis in *Hexalectris* and close leafy relatives *Bletia* (with large, plicate leaves) and *Basiphyllaea* (with 1–2 reduced, conduplicate leaves) remains to be explored from a physiological perspective. Of particular interest would be to assess photosynthetic capacity in all species of *Hexalectris*, specifically to determine the degree to which *H. grandiflora* and *H. revoluta*, which retain intact reading frames for all plastid genes except *ndh*, rely more heavily on photosynthesis than their congeners which have physically or functionally lost most of these genes. Evidence from the visibly green, leafless *Corallorhiza trifida* suggests that despite the loss of several *ndh* genes, this species retains photosynthetic capacity but is photosynthetically inefficient and thus relies almost entirely on mycotrophic carbon for nutrition (Cameron et al. 2009). The situation is not as clear in *Hexalectris* as it is in *Corallorhiza* (Barrett et al. 2014); further study is needed to determine chlorophyll content among species of *Hexalectris* and relatives and to quantify the relative importance of photosynthetic versus mycotrophic carbon among species. Specifically, what if any photosynthetic consequences are associated with the loss of the *ndh* complex in species otherwise capable of photosynthesis?

Photosynthesis-Related Genes

Photosynthesis-related genes (e.g., *ccsA*, *cemA*, *pet*, *psa*/*psb*, *rbcL*, *ycf3*, and *ycf4*; supplementary fig. S5, Supplementary Material online) and genes of the plastid-encoded RNA polymerase that transcribes many of the former (*rpoA*, *B*, *C1*, and *C2*) show extensive degradation and loss among several species of *Hexalectris*, in parallel across multiple lineages (fig. 3). Each of these independent transitions represents

the second phase in recent models of plastome degradation, following *ndh* loss (see references above). This phase is largely hypothesized to represent photosynthetic loss.

We find evidence for functional and/or physical gene loss below the species level in four of six species in which >1 individual was sampled. This suggests a continuum of plastome degradation and functional gene capacity that extends to the population level, and that such patterns of variation may be pervasive among heterotrophs. Our findings in *Hexalectris* corroborate those in the *Corallorhiza striata* complex, in that plastome degradation at or below the species level might be very common, arguing for more intensive sampling of plastomes within species (Barrett et al. 2018). Among the most potentially informative are those species which display evidence of being "on the edge" of photosynthetic loss. Certainly *H. grandiflora* and *H. revoluta* would qualify, as well as some putative partially mycoheterotrophic species of *Corallorhiza*, which display evidence of the beginnings of plastome degradation (e.g., *ψpsbI*, *ψpsbM*, and *ψcemA*; Barrett et al. 2014).

What drives such changes at the population level? Relaxed selection and genetic drift in small populations would seem to be an appropriate explanation in most full mycoheterotrophs, which typically have low population densities, patchy distributions, and rely on self-pollination or vegetative propagation to reproduce (Merckx, Smets, et al. 2013). However, an alternative hypothesis would be that site specific or regional environmental conditions modulate selection pressures. Or, the functioning of many pathways has degraded over time such that each can be disrupted idiosyncratically among populations; that is, these pathways may be "nearly broken," and the pattern we observe is the result of chance. To our knowledge, the relative importance of these scenarios has not been addressed in mycoheterotrophs.

atp Genes

This naturally leads to the question of *how* plastome degradation proceeds beyond loss of the *ndh* complex (stage 1) and photosynthesis-related genes (stage 2), as this seems to be a "sticking point" or stable equilibrium phase for many taxa that have recently lost photosynthesis (sensu Barrett and Davis 2012; Barrett et al. 2014; Naumann et al. 2016; Wicke et al. 2016). *Hexalectris warnockii* is the only species that displays evidence of advanced degradation in the *atp* complex (fig. 3 and supplementary fig. S4, Supplementary Material online; Barrett and Kennedy 2018). Thus, *H. warnockii* may have entered a new state of cellular equilibrium (sensu Wicke and Naumann 2018) wherein the function of a plastid ATP synthase is no longer needed. It has been proposed that ATP synthase is retained in heterotrophic plants beyond photosynthetic gene loss to maintain a proton gradient across the thylakoid membrane, which in turn is essential for the proper functioning of the twin-arginine translocator (*Tat*) complex responsible for moving folded proteins across the thylakoid membrane (Kamikawa et al. 2015). In fact, the pH gradient across the thylakoid is the primary mechanism of *Tat* activity and thus folded protein transport (New et al. 2018). This "*atp-Tat*" coevolutionary hypothesis is based on observations of correlated retention of both the plastid-encoded ATP synthase and *Tat* genes independently across lineages of nonphotosynthetic plants and diatoms. This may explain retention of *atp* genes despite loss of photosynthetic genes (e.g., *psa*, *psb*, and *rbcL*) in genera such as *Hexalectris* (excepting *H. warnockii*) and *Corallorhiza*, in which many recent losses of photosynthesis have occurred.

Only *atpH* remains putatively functional and is under negative selection in *H. warnockii*, the species which perhaps represents the earliest loss of photosynthesis in the genus (fig. 3; Barrett and Kennedy 2018). *atpH*, in part, encodes a CF_0 membrane C_{14} ring and central "stalk" of the ATP synthase complex embedded within the thylakoid membrane (Allen et al. 2011; Hahn et al. 2018). When protonated, the C_{14} ring (encoded by *atpH*) rotates, and distributes protons to the proton exit channel protein encoded by plastid *atpI*, allowing protons to escape across the thylakoid membrane along their pH gradient, and driving the primary mechanism of ATP production. The extramembranous catalytic core of ATP synthase, where ATP synthesis occurs, is encoded by the plastid *atpA*, *atpB*, and *atpE*, and nuclear *atpC* (Allen et al. 2011). These genes have all been lost from *H. warnockii* along with *atpF*, the latter comprising part of the "peripheral stalk" that acts to stabilize the rotating mechanism of the central stalk and intramembranous core (Hahn et al. 2018). Even if the product of *atpH* is somehow functional in *H. warnockii*, the proton channel encoded by *atpI* is not, having been functionally lost, and thus this species may have lost the primary mechanism of proton gradient maintenance across the thylakoid membrane.

It can be assumed that the thylakoid ATP-ase converting ADP to ATP in *H. warnockii* is nonfunctional, and it remains to be discovered what if any function the membrane-embedded *atpH* retains, as this is the only member of the complex that is putatively functional. The ability of a chloroplast to translocate proteins into the thylakoid membrane is likely to be integral to the functioning of the plastid overall (Mori et al. 2001; Cline and Theg 2007; New et al. 2018), and thus degradation of the *atp* complex, given its putative importance in maintaining the conditions necessary for basic organellar function, likely represents a major transitory event in the basic biology of these plants, and others that have lost ATP-ase function independently. It remains to be explored whether loss of the *atp* complex is associated with a whole-scale loss of chloroplast function, a decreased rate of chloroplast synthesis, or some other change in cellular-organellar functional dynamics.

The situation of *atp* gene loss in *H. warnockii* lies in contrast to that in all other *Hexalectris* species, which taken together display some evidence of intensified selection. In the RELAX alternative model, small values of dN/dS decrease in the test lineages relative to the photosynthetic reference lineages, whereas those sites with values dN/dS increase relative to reference lineages (table 3; $K = 1.96$, $P = 0.031$). The interpretation is that sites under negative selection and positive selection have both experienced intensified selection pressures. Given the importance of the ATP synthase complex in maintaining the proton gradient across the thylakoid

Table 3. Results of RELAX Analyses Based on Genes Concatenated By Functional Class (*atp*, *rps*, and *rpl*; except in the case of *accD*, *clpP*, and *matK* Which Do Not Share Similar Functions).

Model	log *L*	*K*, *P*-val	Selection	Branch	ω_1	ω_2	ω_3
All *atp*							
Alternative	**−8,293.4**	**1.96, 0.031**	**Intensification**	**Test**	**0.01 (57.93%)**	**0.06 (37.31%)**	**10.17 (4.76%)**
				Reference	**0.07 (57.93%)**	**0.25 (37.31%)**	**3.26 (4.76%)**
Null	**−8,295.7**			**Test**	**0.00 (58.21%)**	**0.02 (35.58%)**	**6.06 (6.21%)**
				Reference	**0.00 (58.21%)**	**0.02 (35.58%)**	**6.06 (6.21%)**
accD*, *clpP*, *matK							
Alternative	**−6,861.3**	**0.85, 0.180**	**Relaxation**	**Test**	**0.00 (3.08%)**	**0.43 (96.78%)**	**90.82 (0.14%)**
				Reference	**0.00 (3.08%)**	**0.37 (96.78%)**	**206.53 (0.14%)**
Null	**−6,862.2**			**Test**	**0.00 (1.66%)**	**0.40 (98.23%)**	**172.60 (0.10%)**
				Reference	**0.00 (1.66%)**	**0.40 (98.23%)**	**172.60 (0.10%)**
All *rps*							
Alternative	**−8,319.9**	**2.54, 0.097**	**Intensification**	**Test**	**0.00 (72.92%)**	**0.22 (12.65%)**	**2.74 (14.43%)**
				Reference	**0.01 (72.92%)**	**0.55 (12.65%)**	**1.49 (14.43%)**
Null	**−8,321.2**			**Test**	**0.00 (72.30%)**	**0.61 (11.21%)**	**1.78 (16.49%)**
				Reference	**0.00 (72.30%)**	**0.61 (11.21%)**	**1.78 (16.49%)**
All *rpl*							
Alternative	**−5,321.3**	**1.13, 0.340**	**Intensification**	**Test**	**0.00 (17.27%)**	**0.00 (79.02%)**	**23.74 (3.71%)**
				Reference	**0.00 (17.27%)**	**0.00 (79.02%)**	**16.41 (3.71%)**
Null	**−5,321.8**			**Test**	**0.01 (80.86%)**	**0.04 (15.67%)**	**21.22 (3.47%)**
				Reference	**0.01 (80.86%)**	**0.04 (15.67%)**	**21.22 (3.47%)**

NOTE.—"Model" refers to the null versus alternative models; "log *L*" is the log-likelihood for a particular model; K = the value of the "*K*" parameter, where $K > 1$ indicates intensified selection and $K < 1$ indicates relaxed selection, and *P*-val = the significance of *K*; "Selection" indicates the predominant mode of selection for the test branches; "Branch" distinguishes parameter estimates for test versus reference branches; ω_1–ω_3 are the three estimates of dN/dS for the three site classes in the model, for test and reference branches, with the proportion of sites under each value given in parentheses. Analyses of single genes can be found in supplementary table S3, Supplementary Material online.

membrane, it is intuitive that selection might be intensified following the loss of photosynthesis in these lineages, either for sequence conservation (negative selection), putatively novel or modified function (positive selection), or both. A similar situation has been observed among lineages of holoparasitic Orobanchaceae, in which many "housekeeping" genes experience intensified selection following periods of relaxed selection of functional/physical gene loss in other complexes (i.e., photosynthesis-related genes; Wicke et al. 2016; Wicke and Naumann 2018).

Housekeeping Genes

Housekeeping genes are largely intact in *Hexalectris*, with three exceptions. The transfer RNA gene $trnT^{GGU}$ is missing from *H. arizonica*, and $trnS^{UGA}$ is missing from *H. warnockii*. Further, indel variation in the "variable loop" region of $trnL^{CAA}$ in one accession of *H. parviflora* and both accessions of *H. warnockii* (supplementary fig. S6, Supplementary Material online) suggests a modification or possibly loss-of-function in this tRNA gene, as length of the variable loop is important in recognition of aminoacyl tRNA synthetase and helps in tRNA molecular stability (Sun and Caetano-Anollés 2009). These findings, taken with the evidence of relaxed selection in a single housekeeping gene (*rpl14*; supplementary table S3, Supplementary Material online), open the possibility that some species of *Hexalectris* have begun to shift beyond "phase 2" of recent models of plastome evolution in parasites (see citations above), defined by the loss of housekeeping genes.

We find similarities between *Hexalectris* and *Corallorhiza*, in which some members of the *Corallorhiza striata* complex have lost $trnT^{GGU}$ just as is the case in *Hexalectris* (Barrett et al. 2014). It should be noted that the tRNAs lost in *Hexalectris* involve some level of redundancy among amino acids: each loss or modification involves a plastid-encoded tRNA that specifies at least two anticodons for a particular amino acid (i.e., *trnT* [$n = 2$], *trnS* [$n = 3$], and *trnL* [$n = 3$] if we assume the function of the latter is somehow compromised; fig. 3 and supplementary fig. S6, Supplementary Material online). It is tempting to suggest, based on these observations, that tRNA genes with some level of amino acid redundancy may be lost more easily. $trnT^{UGU}$ and $trnT^{GGU}$ are lost at nearly equal frequency among holoparasites and full mycoheterotrophs, but *trnT* is by no means the most frequently lost among all tRNAs (see supplementary information in Graham et al. [2017]), and $trnS^{UGA}$ is the least frequently lost among the three tRNAs with serine anticodons.

Photosynthetic Loss

Loss of photosynthesis has been an ongoing process in *Hexalectris* over the past ~33 My (range = 38.6–14.6 Ma for the stem lineage of *Hexalectris* based on divergence time analysis of Sosa et al. [2016]), beginning with the shared, leafless condition and loss-of-function in *ndh* genes. Loss of photosynthesis genes and the *atp* complex in *H. warnockii* may have begun as early as 30 Ma (Sosa et al. 2016; Barrett and Kennedy 2018), but it is impossible to discern the order in which members of these gene functional classes were lost in this lineage. Based on the phylogenetic position of the putatively photosynthetic *H. grandiflora*, four of the five putative losses of photosynthesis must have occurred within the last ~20 My in *Hexalectris* excluding *H. warnockii*. Further, due to

the nested placement of *H. revoluta*, photosynthetic losses among *H. arizonica*, *colemanii*, and *parviflora* must have occurred within the last ~10 My. Thus, many of the independent losses of photosynthesis in *Hexalectris* have occurred relatively recently.

The possibility remains that missing or pseudogenized copies of plastid genes have been duplicated or transferred to the nuclear genome, and that these genes may remain functional. Indeed, several plastid genes and gene fragments were found in the nuclear and mitochondrial genomes of holoparasitic Orobanchaceae, but it is doubtful that any involved in photosynthesis remain functional (Cusimano and Wicke 2016). We see no reason why retention of functional copies of plastid-photosynthetic genes in the nuclear genome might be adaptive, given the heterotrophic habit of taxa such as *Hexalectris*, and indeed there exist no compelling examples from other studies that this is the case. For example, *ndh* genes lost from the plastome were not detected as functional copies in the nuclear genomes of several orchid taxa (Lin et al. 2017). The only plausible alternative hypothesis for the patterns observed here may be photosynthetic "rescue" via hybrid introgression. Such a scenario would require pollen to be transferred from a nonphotosynthetic species of *Hexalectris* to a putatively photosynthetic (and possibly now extinct) member of this clade, such that the offspring would inherit a plastid genome with photosynthetic functionality. The only known example even remotely approaching such a scenario is in the orchid genus *Cymbidium*, between the leafless, partial mycoheterotroph *Cymbidium macrorhizon* and the leafy relative *Cymbidium ensifolium* (Ogura-Tsujita et al. 2014). However, in that case, the hybrid was the result of artificial pollination, and the plastid-photosynthetic machinery is still largely intact in the leafless *Cymbidium macrorhizon* (Motomura et al. 2010; Kim and Chase 2017; Kim et al. 2017; Suetsugu et al. 2018).

Under a scenario of gene flow between a nonphotosynthetic pollen donor and a photosynthetic recipient, we would expect maternal inheritance of a "functionally photosynthetic" plastome, and also plastid-nuclear gene tree discordance. In fact, based on nuclear ITS data from a previous study (Kennedy and Watson 2010), putatively photosynthetic *H. grandiflora* and *H. revoluta* occupy very similar phylogenetic positions, nested among nonphotosynthetic species of *Hexalectris*, suggesting that this scenario is unlikely. If such a scenario was to have occurred in *Hexalectris*, it is likely that Müller-Dobzhansky incompatibilities would arise (Dobzhansky 1936; sensu Orr 1995). In other words, a putatively photosynthetic plastome would not likely be able to carry out photosynthetic function within a nuclear-genomic, cellular "environment" of photosynthetic loss-of-function, or even form a viable zygote for that matter (see discussion in Barrett et al. [2018]). It seems unlikely that photosynthesis could be "rescued" in such a way through hybridization, assuming the extensive degradation of the plastid-encoded photosynthetic machinery is paralleled in the nuclear genome (e.g., *Hypopitys*, Ravin et al. 2016; *Gastrodia*, Yuan et al. 2018). This hypothesis could potentially be tested by experimental crosses in vivo between a nonphotosynthetic pollen donor (e.g., *H. spicata*) and a photosynthetic pollen recipient (e.g., *H. grandiflora*) where they are known to co-occur locally (e.g., Texas, USA).

Evolutionary Rate Variation

Functional and physical losses of plastid genes are just one dimension of the overall process of plastome degradation. Although protein and RNA coding capacity are obviously of utmost importance to photosynthetic function, these changes often occur in parallel with other important genomic changes, including rearrangements, modifications of IR boundaries, changes in base composition (e.g., GC content), proliferation of tandem and dispersed repeats, increased overall substitution rates, and relaxed selective pressure on housekeeping genes with basic organellar functions (Wicke et al. 2013, 2016; Barrett et al. 2018). We find evidence for increased overall substitution rates associated with photosynthetic loss for ten genes in *Hexalectris* (here not considering those that have already become pseudogenes), five of which are *atp* genes (excluding *H. warnockii*, which has several *atp* pseudogenes), and five of which are housekeeping genes (supplementary table S2, Supplementary Material online). Only two genes showed evidence of significantly relaxed selective pressure (i.e., elevated d*N*/d*S* ratios), as indicated by our RELAX analysis (*atpA* and *rpl14*; supplementary table S2, Supplementary Material online). Increased overall rates of substitution and indels may be coincident with relaxed selective constraints on plastid genes, and further may obscure detection of relaxed selection via elevated d*N*/d*S* ratios using conventional methods, as both the numerator and denominator are elevated. Further, we interpret these findings with caution based on the often low information content in single gene analyses, in which there are few variable positions.

What causes increases in substitution rates in heterotrophic plants? A number of possibilities exist, including relaxed negative selection, positive selection, low effective population sizes, growth form differences (thus related to generation time and the inheritance of mutations), and genome-wide selection for increased overall substitution rates to help evade host defenses in coevolutionary arms races (Bromham et al. 2013). First, increased substitution rates may be a result of relaxed selective pressures on genes being analyzed, even housekeeping genes. However, although this is obvious for photosynthesis-related genes, it does not fully explain why both synonymous and nonsynonymous rates tend to increase overall and across all three plant genomes, nor does it account for increased rates in noncoding regions (e.g., Nickrent and Starr 1994; Bromham et al. 2013). In the present study, d*N*/d*S* ratios among protein-coding genes are elevated for only a few genes, whereas a substantially greater number of genes display evidence of overall rate increases associated with the putative loss of photosynthesis (supplementary table S2, Supplementary Material online).

One possibility is relaxed selective pressure on the functions of nuclear-encoded genes with products involved in DNA replication, recombination, and repair (Day and Madesis 2007; Barnard-Kubow et al. 2014; Zhang et al. 2015, 2016). These genes include, for example, "proofreading

enzymes" that repair DNA breaks, misincorporated bases, or illegitimate recombinants (e.g., *WHIRLY1* functions in plastome stability; Maréchal et al. 2009). Relaxed selection in some of these genes, and presumed decreases in their overall functional performance, shows a positive correlation with the degree of plastome rearrangement (e.g., in Geraniaceae; Zhang et al. 2016). However, the reasons as to why these genes with genome-stabilizing functions would experience relaxed selection in heterotrophs are unclear, though theoretical models predict that increased mutation rates overall may be beneficial to parasites in situations of coevolutionary arms races, despite the "cost" associated with high mutational loads (e.g., Haraguchi and Sasaki 1996).

An alternative explanation is that increased substitution rates are directly linked to nutritional mode, that is, the loss of photosynthesis leading to a completely heterotrophic lifestyle itself causes the observed increase. A plausible, but as-of-yet untested hypothesis is that in order to compensate for carbon deficits resulting from the loss of photosynthesis, fully mycoheterotrophic plants need to "ramp up" their metabolic rates. This may cause an increase in reactive free radical species, and associated oxidative stress, which is well known to be linked to increased DNA damage. For example, genes involved in oxidative stress response in albino *Epipactis helleborine* are upregulated, possibly in association with increased reliance on fungal carbon (Suetsugu et al. 2017). Although many ecophysiological studies in partially and fully heterotrophic plants have focused on photosynthetic loss per se, few have focused on relative respiratory metabolic rates among lineages of leafy, partially heterotrophic, and fully heterotrophic plants. Such a study would be particularly powerful in a phylocomparative framework, to tease apart the association of metabolic rate, substitution rate, and phylogenetic relationships. One would expect to observe increased respiration rates and also increased expression of respiratory genes and genes involved in oxidative stress responses in fully heterotrophic taxa relative to partial heterotrophs and especially to leafy relatives, under identical conditions.

Conclusions

We demonstrate a minimum of four losses of photosynthesis within a single orchid genus of only nine species, which is hitherto unprecedented among heterotrophic plants. Thus, a previous estimate of 30 transitions to full mycoheterotrophy in orchids is likely an underestimate. A conservative estimate based on the current study and a previous study in *Corallorhiza* ($\geq$2 transitions; Barrett et al. 2018) increases this number to a minimum of 35 independent transitions within the orchids, underscoring the importance of dense sampling within clades containing partial and full heterotrophs. Future studies should include additional sampling of plastomes of several *Hexalectris* species at the population level to further reveal details of the patterns by which photosynthesis is lost. Data from nuclear genomes, transcriptomes, proteomes, physiology (carbon and nitrogen isotopic data, chlorophyll content, etc.), basic cellular data (relative populations of mitochondria, chloroplasts and other plastids), and ecological niche parameters should also be investigated to provide much needed biological context for evolutionary mechanisms that drive radical genomic, morphological, and physiological changes observed in *Hexalectris* and relatives. Heterotrophic plants such as *Hexalectris* afford the ability to examine systems-level alterations to entire pathways as a result of loss- or gain-of-function in others, and provide forward-genetic opportunities beyond mutational studies in model systems.

Materials and Methods

Plastid Genomes

Material, DNA Extraction, and Sequencing

We collected 21 accessions across sites in North America, including 19 accessions of *Hexalectris* (9 species), and 1 representative each for the green, leaf-bearing *Ba. corallicola* and *B. roezlii* (table 1). Six of the nine species of *Hexalectris* are represented by more than one individual. We dried floral and bract material in silica gel and deposited voucher specimens in the Willard Sherman Turrell Herbarium at Miami University (Oxford, OH) or the West Virginia University Herbarium (Morgantown, WV; table 1). We extracted total genomic DNAs using a modified CTAB protocol (Doyle and Doyle 1987) and checked nucleic acid quality with a NanoDrop Spectrophotometer (Thermo Fisher Scientific, Waltham, MA). We quantified double stranded DNA using nucleic acid staining via Qubit fluorometry (Thermo Fisher Scientific).

We sheared total gDNAs to ~450 bp with a Covaris E220 Focused-ultrasonicator (Covaris, Inc., Woburn, MA) and prepared/indexed libraries for shotgun sequencing following Glenn et al. (2016) with slight modifications as in Barrett and Kennedy (2018). We quantified indexed libraries using nucleic acid staining and pooled at equimolar ratios, and then assessed library size distribution on an Agilent Bioanalyzer (Agilent Technologies, Santa Clara, CA) prior to sequencing on two lanes of an Illumina HiSeq1500 (Illumina, Inc., San Diego, CA) at the West Virginia University-Marshall University Shared Sequencing Facility (100-bp paired ends), with ten libraries from another project.

Data Processing, Plastome Assembly

We conducted initial assemblies using Novoplasty v. 2.7.2, a PERL-based assembler that begins with a plastid sequence as a starting seed, and generates a complete, circularized plastome, given sufficient coverage (Dierckxsens et al. 2017). We used the following parameters: target genome size 50–200 kb, kmer size = 45, and insert size range = 1.8×. We also used VELVET v.1.2.1 for de novo assembly (Zerbino and Birney 2008), with the following settings: kmer length = 71, insert size = 450 bp, scaffolding = off, minimum contig length = 300, and minimum coverage = 10×. We then aligned VELVET contigs to circularized draft plastomes from NOVOPlasty using the native Geneious v.10 "map to reference" function. Further, we mapped original read pools to each draft reference to check for misassemblies or coverage gaps. Accessions with very high coverage were downsampled to ~200× coverage depth and reassembled de novo as

above, to avoid errors associated with misinterpretation of sequencing errors as polymorphisms by the assembler. We visually verified contiguous coverage and spanning of the IR boundaries using paired end data, and then used the self-dotplot function in Geneious to verify the IR. We extracted and aligned the two IR copies for each accession using the MAFFT v.7 Geneious plugin (Katoh and Standley 2013) to verify identical sequence. We completed annotations for all accessions in DOGMA (Wyman et al. 2004) and submitted them to GenBank under NCBI accessions MK726066–MK726085.

Data Set Construction, Alignment

We initially aligned the annotated plastomes, excluding one copy of the IR, using the ProgressiveMAUVE algorithm in Geneious (Darling et al. 2010; parameters = autocalculate seed weight, minimum LCB score), which identifies locally collinear, syntenic blocks (LCB), and allows visualization of genomic structural rearrangements. We scored the degree of rearrangement as the DCJ distance from each accession to *B. roezlii*, which shows no evidence of genomic rearrangement relative to the highly conserved plastome configuration characteristic of monocot angiosperms (as assessed by a previous alignment in MAUVE as above to *Typha latifolia* [NCBI accession GU195652; Poales, Typhaceae] and *Phoenix dactylifera* NCBI accession GU811709 [Arecales, Arecaceae]). We realigned each collinear block identified by MAUVE using the MAFFT plugin for Geneious (default parameters). We then concatenated the blocks, and filtered sites with missing data or gaps; hereafter, we refer to this as the "LCB" matrix, which includes both coding and noncoding sequence. We also extracted each protein-coding DNA sequence (CDS; hereafter the "CDS" matrix) from the data set using a custom UNIX command and aligned each with MACSE v.2 (Ranwez et al. 2018), which preserves reading frame and allows incorporation of sequencing errors or sequences with frameshifts (default parameters). We kept genes for which all taxa had intact reading frames, and then concatenated these into a single plastid supermatrix using TriFusion (http://odiogosilva.github.io/TriFusion/; last accessed December 10, 2018).

Phylogenetic Analysis

Estimation of Relationships

We subjected the two data matrices above to parsimony, likelihood, and Bayesian analyses. We conducted parsimony searches in TNT v.1.1 (Goloboff et al. 2008), using *Bletia* as the outgroup (as for all subsequent phylogenetic analyses). We ran 1,000 random addition sequences and Tree Bisection-Reconnection branch swapping, keeping up to 1,000 trees. We assessed parsimony branch support via Jackknife resampling, with 37% character removal probability. We chose the best-fit substitution model for each data set in Mega v.7 (Kumar et al. 2016), under the Bayesian Information Criterion. We generated maximum likelihood trees in RAxML v.8.2 (Stamatakis 2014) under an unpartitioned and gene-partitioned GTR + Γ model, respectively, with the default number of rate categories across sites (25). We initiated 10 independent searches from random starting seeds and assessed support via nonparametric "standard" bootstrapping, with 1,000 pseudoreplicates. We generated Bayesian estimates of plastid relationships in MrBayes v.3.2.6 (Ronquist et al. 2012), under a GTR + Γ model with four rate categories, using both unpartitioned and partitioned models, as above. We ran four independent chains (three heated, one cold) for 10^7 generations, discarding the first 20% as burn-in. We confirmed chain stationarity and convergence by standard deviation of split frequencies <0.01, and effective sample size >500 for each parameter in Tracer v.1.7.1 (http://tree.bio.ed.ac.uk/software/tracer/; last accessed December 10, 2018).

Plastid Genome Structural Evolution

Physical and Functional Gene Losses

We scored each gene present in the "typical" monocot angiosperm gene set as putatively functional, physically lost (i.e., completely deleted), or functionally lost as a pseudogene among the accessions. Here, we denoted pseudogenes as "ψ," and defined these conservatively by at least 10% of the open reading frame being disrupted by internal stop codons, nontriplet frameshifts, or large deletions. Accessions displaying evidence of functional or physical loss in $\geq$10 photosynthesis-related genes were concluded to have lost the ability to carry out photosynthesis. We chose this conservative definition to avoid assigning a loss of photosynthesis to accessions with only limited evidence of functional gene loss. We used tRNAscan-SE (Lowe and Chan 2016) and ARAGORN (Laslett and Canback 2004) to detect putative transfer RNA pseudogenes, secondary structural rearrangements, or substitutions affecting anticodon sites. For each, we chose default search parameters and submitted full sets of tRNAs (excluding duplicates from the IR) as unaligned FASTA files. We then aligned each homologous tRNA across all accessions in MAFFT and annotated the anticodon identified by ARAGORN in Geneious to check for substitutions or indels.

Evolution of Photosynthetic Capacity

Topology Tests for a Single Loss of Photosynthesis

We constrained all accessions with evidence of degradation in photosynthetic machinery (see above) to be monophyletic (i.e., all accessions excluding *Bletia*, *Basiphyllaea*, *H. grandiflora*, and *H. revoluta*). We set an alternative, negative constraint specifying nonmonophyly of these accessions. Further, to test a recent phylogenetic hypothesis of relationships in *Hexalectris* (Sosa et al. 2016), we specified a monophyly constraint including all accessions of *H. spicata*, *H. nitida*, *H. parviflora*, and *H. revoluta*. We used stepping-stone sampling in MrBayes v.3.2.6, specifying a GTR + Γ+I model with four rate categories on the filtered data set of concatenated LCBs, and using 10^7 steps of the Markov Chain Monte Carlo sampler with the first 20% as burn-in, and sampling every 5,000 steps. We checked convergence by verifying that the standard deviation of split frequency for each run was <0.01. We then used the average estimates of the marginal likelihoods in a Bayes factor comparison, given by BF = the average of the

marginal likelihood of full mycoheterotroph monophyly divided by the average of the marginal likelihood of nonmonophyly of full mycoheterotrophs. This formula simplifies to 2× the difference in log-likelihoods between the two marginal likelihood averages. A Bayes factor $>|100|$ indicates very strong evidence in favor of one model over another.

Ancestral Reconstruction of Photosynthetic Status

We reconstructed photosynthetic status at the ancestral nodes for the plastid tree as a discrete character with three states for each gene: functional (0; as this is the ancestral state among orchids and within Bletiinae), pseudogene (1), or physically lost (2). We specified a transition matrix allowing changes from $0 \rightarrow 1$, $1 \rightarrow 2$, and $0 \rightarrow 2$ with equal probabilities, and restricting all reversals. We use the R packages "ape," "geiger," and "phytools" for all analyses (Harmon et al. 2008; Revell 2012; Paradis and Schliep 2019).

Testing Discrete Trait Models of Photosynthetic Loss

We then coded photosynthetic status as a binary character, based on evidence from extensive degradation of photosynthesis-related genes (0 = photosynthetic and 1 = nonphotosynthetic). First, we used a parsimony criterion and cost matrices to constrain character state changes in terms of loss and gain of photosynthesis. We used equal weights for character state changes in the first analysis (all costs = 1 step), and two cost constraints including weights against loss and gain of photosynthesis. Constraints on state changes were conducted in TNT by increasing the cost to the maximum allowable (cost = 99 steps) for the state change of interest. Searches for most parsimonious trees were conducted as above, and character state changes under each cost matrix were optimized on the most parsimonious tree.

We fit four Markov (mk model, sensu Lewis 2001) models of discrete character evolution under a maximum likelihood framework to the LCB tree, using modifications of the character state rate matrix, that is, the "*q*-matrix." We fit the following alternative models of trait change: 1) "ER" or equal rates ($q_{01} = q_{10}$), 2) "ARD" or allowing rate differences ($q_{01} \neq q_{10}$), 3) a "loss-only" model in which only photosynthetic loss can occur (q_{01} only), and 4) a "gain-only" model in which only reversals to photosynthesis from a nonphotosynthetic state are allowed (q_{10} only). We evaluated the fit of each model using the corrected Akaike information criterion (AICc; Burnham and Anderson 2002) and Akaike weights in the R package qpcRv.1.4-1 (Spiess 2018).

Evolutionary Rate Variation and Photosynthetic Loss

Photosynthetic Capacity and Substitution Rate Associations

We used TraitRateProp (Mayrose and Otto 2011; Karin, Ashkenazy, et al. 2017; Karin, Wicke et al. 2017) to test associations between photosynthetic status as a binary trait (0 = photosynthetic and 1 = nonphotosynthetic) and substitution rates for protein-coding genes with intact reading frames. Genes with extensive loss or evidence of pseudogenization in >2 taxa were excluded (e.g., *psa*/*psb*, *pet*, and *rpo*). TraitRateProp uses an ultrametric tree (here via penalized likelihood via the "APE" package in R), multiple sequence alignment, a binary trait, and a joint maximum likelihood model of sequence and trait evolution to test the hypothesis of trait/rate association versus a null model of no association via stochastic trait mapping. We additionally included *atp* genes, some of which are degraded in *H. warnockii* (Barrett and Kennedy 2018; see below), by pruning this species from the tree in R via the "drop.tip" function ("APE") and excluding this species from the sequence alignment and trait data.

Analysis of Relaxed Selective Constraints across Taxa

We used the RELAX model in HyPhy v.2.4.4 (Kosakovsky Pond et al. 2005; Wertheim et al. 2015) to test the hypothesis of relaxed versus intensified selective constraints based on nonsynonymous/synonymous substitution ratios ($\omega = dN/dS$, i.e., the proportion of nonsynonymous and substitutions per respective site) in nonphotosynthetic accessions versus photosynthetic reference accessions. We used *Bletia*, *Basiphyllaea*, *H. grandiflora*, and *H. revoluta* as reference photosynthetic branches, and others as "test" nonphotosynthetic branches. Analyses were run for all loci used in TraitRate analyses (above). Specifically, RELAX allows testing of a null model of no relaxed selection against an alternative model in which test branches are allowed to have differing rates of ω, indicated by the selection intensity parameter, K, where $K > 1$ indicates intensified selection and $K < 1$ indicates relaxed selection. These are evaluated via a log-likelihood ratio test. All analyses were run via commandline on a local server, and resulting .json result files were parsed with a custom UNIX script.

Correlations among Genomic and Substitutional Features

We used the program CoEvol v.1.5 (Lartillot and Poujol 2011) to test correlations among genomic and substitutional features. Specifically, we analyzed dS, dN, GC content, total genome length, the number of functional genes (i.e., those with no evidence of functional or physical loss), the degree of rearrangement (DCJ distance, see below), length of the IR, insertions/deletions, tandem repeats, and dispersed repeats. CoEvol uses a Bayesian model to allow GC frequencies and substitution rate parameters to vary across branches of a user-defined tree (here, the CDS matrix topology), and models various continuous traits (phenotype, life-history, genome size, etc.) with the objective of testing for correlations among these traits and substitution parameters while controlling for others. Here, we treat genomic features (repeat content, rearrangements, etc.), as the "phenotype" of the plastome.

We used REPuter (Kurtz et al. 2001) to find genomic repeats >20 bp in length, with a Hamming distance of 3, and maximum *e*-value of 10^{-3}. Specifically, we searched for forward–forward, forward–reverse, forward–compliment, and palindromic repeats. We transformed data as follows, following the recommendations of Lartillot and Poujol (2011): total genome length ($e^{x/10,000}$), length of coding DNA ($e^{x/10,000}$), # functional genes ($e^{x/10}$), DCJ distance (e^{x}), IR length ($e^{x/10,000}$), repeats classes ($e^{x/10}$), and indels ($e^{x/1-x}$) to avoid zeroes. We ran two independent chains of CoEvol for

10^5 generations of the Markov Chain Monte Carlo, discarding the first 10% as burn-in and sampling every five steps. We used minimum effective sample sizes of 500 and discrepancies between chains of <0.1 for each parameter to assess convergence, using the executable "tracecomp" which is part of the CoEvol suite. Once the chains converged, we summarized the results with "readcoevol." We considered pp's of >0.9 for positively correlated attributes and <0.1 for negatively correlated attributes (1 − pp) to be significant.

Supplementary Material

Supplementary data are available at *Molecular Biology and Evolution* online.

Acknowledgments

This work was supported by a WVU-PSCoR grant to C.F.B. We thank Troy Meyers (Meyers Conservatory) for providing material of *Ba. corallicola*. We thank Big Bend National Park, Cedar Ridge Preserve and Dogwood Canyon Audubon Center (Audubon Society; Dallas, TX), Coronado National Forest, Briarwood Nature Preserve (Natchitoches Parish, LA), Indiana Department of Natural Resources and Harrison-Crawford State Forest, North Carolina Department of Environment and Natural Resources and Stone Mountain State Park, the Nature Conservancy of Texas and Davis Mountains Preserve, and the University of Science and Arts of Oklahoma for collecting permission and access to land. We thank Paul Martin Brown, Marcy Brown-Marsden, Pablo Carrillo-Reyes, Ronald Coleman, Frank Farruggia, John Karges, Kellie Kennedy, Allison Leavitt, Larry Magrath, Aaron Rodríguez, Al Schotz, Joe Sirotnak, Victoria Sosa, Bonnie Stolp, and Michael Vincent for assistance with permits and collections. Thank you to the Miami University's Willard Sherman Turrell Herbarium (MU) and the West Virginia University Herbarium (WV) for collections support and specimen curation. We thank Donald Pimerano and Jun Fan at the Marshall University Genomics Core Facility for sequencing assistance, supported in part by a National Institutes of Health grant to the WV-IDeA Network of Biomedical Research Excellence (WV-INBRE) program (2P20 GM103434).

References

Allen JF, de Paula WBM, Puthiyaveetil S, Nield J. 2011. A structural phylogenetic map for chloroplast photosynthesis. *Trends Plant Sci*. 16(12):645–655.

Barbrook AC, Howe CJ, Purton S. 2006. Why are plastid genomes retained in non-photosynthetic organisms? *Trends Plant Sci*. 11(2):101–108.

Barnard-Kubow KB, Sloan DB, Galloway LF. 2014. Correlation between sequence divergence and polymorphism reveals similar evolutionary mechanisms acting across multiple timescales in a rapidly evolving plastid genome. *BMC Evol Biol*. 14:268.

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, Freudenstein JV, Li J, Mayfield-Jones DR, Perez L, Pires JC, Santos C. 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(12):3095–3112.

Barrett CF, Freudenstein JV, Taylor DL, Koljalg U. 2010. Rangewide analysis of fungal associations in the fully mycoheterotrophic *Corallorhiza striata* complex (Orchidaceae) reveals extreme specificity on ectomycorrhizal *Tomentella* (Thelephoraceae) across North America. *Am J Bot*. 97(4):628–643.

Barrett CF, Kennedy AH. 2018. Plastid genome degradation in the endangered, mycoheterotrophic, North American orchid *Hexalectris warnockii*. *Genome Biol Evol*. 10(7):1657–1662.

Barrett CF, Wicke S, Sass C. 2018. Dense infraspecific sampling reveals rapid and independent trajectories of plastome degradation in a heterotrophic orchid complex. *New Phytol*. 218(3):1192–1204.

Bellot S, Renner SS. 2014. Exploring new dating approaches for parasites: the worldwide Apodanthaceae (Cucurbitales) as an example. *Mol Phylogenet Evol*. 80:1–10.

Bellot S, Renner SS. 2015. The plastomes of two species in the endoparasite genus *Pilostyles* (Apodanthaceae) each retain just five or six possibly functional genes. *Genome Biol Evol*. 8(1):189–201.

Bidartondo MI. 2005. The evolutionary ecology of myco-heterotrophy. *New Phytol*. 167(2):335–352.

Braukmann T, Kuzmina M, Stefanovic S. 2013. Plastid genome evolution across the genus *Cuscuta* (Convolvulaceae): two clades within subgenus *Grammica* exhibit extensive gene loss. *J Exp Bot*. 64(4):977–989.

Braukmann TWA, Broe MB, Stefanovic S, Freudenstein JV. 2017. On the brink: the highly reduced plastomes of nonphotosynthetic Ericaceae. *New Phytol*. 216(1):254–266.

Bromham L, Cowman PF, Lanfear R. 2013. Parasitic plants have increased rates of molecular evolution across all three genomes. *BMC Evol Biol*. 13(1):126.

Burnham KP, Anderson DR. 2002. Model selection and multimodel inference: a practical information-theoretic approach. New York: Springer.

Cameron DD, Preiss K, Gebauer G, Read DJ. 2009. The chlorophyll-containing orchid *Corallorhiza trifida* derives little carbon through photosynthesis. *New Phytol*. 183(2):358–364.

Catling PM. 2004. A synopsis of the genus *Hexalectris* in the United States and a new variety of *Hexalectris revoluta*. *Native Orchid Conf J*. 1:5–25.

Catling PM, Engel VS. 1993. Systematics and distribution of *Hexalectris spicata* var. *arizonica* (Orchidaceae). *Lindleyana* 8:119–125.

Chase MW, Cameron KM, Freudenstein JV, Pridgeon AM, Salazar G, Van den Berg C, Schuiteman A. 2015. An updated classification of Orchidaceae. *Bot J Linn Soc*. 177(2):151–174.

Cline K, Theg SM. 2007. The *Sec* and *Tat* protein translocation pathways in chloroplasts. In: Dalbey RE, Koehler CM, Tamanoi F, editors. The enzymes: molecular machines involved in protein transport across cellular membranes. Vol. 25. London: Elsevier. p. 463–492.

Cusimano N, Wicke S. 2016. Massive intracellular gene transfer during plastid genome reduction in non-green Orobanchaceae. *New Phytol*. 210(2):680–693.

Darling AE, Mau B, Perna NT. 2010. ProgressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. *PLoS One* 5(6):e11147.

Day A, Madesis P. 2007. DNA replication, recombination, and repair in plastids. In: Bock R, editor. Cell and molecular biology of plastids. Topics in current genetics. Vol. 19. Berlin/Heidelberg (Germany): Springer. p. 65–119.

dePamphilis CW, Palmer JD. 1990. Loss of photosynthetic and chlororespiratory genes from the plastid genome of a parasitic flowering plant. *Nature* 348(6299):337–339.

Dierckxsens N, Mardulyn P, Smits G. 2017. NOVOPlasty: de novo assembly of organelle genomes from whole genome data. *Nucleic Acids Res*. 45(4):e18.

Dobzhansky T. 1936. Studies on hybrid sterility. II. Localization of sterility factors in *Drosophila pseudoobscura* hybrids. *Genetics* 21(2):113–135.

Doyle JJ, Doyle JL. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem Bull*. 19:11–15.

Feng YL, Wicke S, Li JW, Han Y, Lin CS, Li DZ, Zhou TT, Huang WC, Huang LQ, Jin XH. 2016. Lineage-specific reductions of plastid genomes in an orchid tribe with partially and fully mycoheterotrophic species. *Genome Biol Evol*. 8(7):2164–2175.

Freudenstein JV, Barrett CF. 2010. Mycoheterotrophy and diversity in Orchidaceae. In: Seberg O, Petersen G, Barfod A, Davis JI, editors. Diversity, phylogeny and evolution in the monocotyledons. Aarhus (Denmark): Aarhus University Press. p. 25–37.

Freudenstein JV, Chase MW. 2015. Phylogenetic relationships in Epidendroideae (Orchidaceae), one of the great flowering plant radiations: progressive specialization and diversification. *Ann Bot*. 115(4):665–681.

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*. 7(1):45.

Gitzendanner MA, Soltis PS, Yi T-S, Li D-Z, Soltis DE. 2018. Chapter ten—plastome phylogenetics: 30 years of inferences into plant evolution. In: Chaw S-M, Jansen RK, editors. Advances in botanical research. Cambridge, Massachusetts, USA:Academic Press. p. 293–313.

Glenn TC, Nilsen R, Kieran TJ, Finger JW, Pierson TW, Bentley KE, Hoffberg S, Louha S, Garcia-De-Leon FJ, Angel del Rio Portilla M. 2016. Adapterama I: universal stubs and primers for thousands of dual-indexed Illumina libraries (iTru & iNext). bioRxiv. https://doi.org/10.1101/049114 (last accessed 15 December, 2018).

Goloboff PA, Farris JS, Nixon KC. 2008. TNT, a free program for phylogenetic analysis. *Cladistics* 24(5):774–786.

Graham SW, Lam VKY, Merckx V. 2017. Plastomes on the edge: the evolutionary breakdown of mycoheterotroph plastid genomes. *New Phytol*. 214(1):48–55.

Hadariová L, Vesteg M, Hampl V, Krajcovic J. 2018. Reductive evolution of chloroplasts in non-photosynthetic plants, algae and protists. *Curr Genet*. 64(2):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.

Haraguchi Y, Sasaki A. 1996. Host-parasite arms race in mutation modifications: indefinite escalation despite a heavy load? *J Theor Biol*. 183(2):121–137.

Harmon LJ, Weir JT, Brock CD, Glor RE, Challenger W. 2008. GEIGER: investigating evolutionary radiations. *Bioinformatics* 24(1):129–131.

Kamikawa R, Tanifuji G, Ishikawa SA, Ishii KI, Matsuno Y, Onodera NT, Ishida KI, Hashimoto T, Miyashita H, Mayama S, 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(10):2598–2604.

Karin EL, Ashkenazy H, Wicke S, Pupko T, Mayrose I. 2017. TraitRateProp: a web server for the detection of trait-dependent evolutionary rate shifts in sequence sites. *Nucleic Acids Res*. 45:W260–W264.

Karin EL, Wicke S, Pupko T, Mayrose I. 2017. An integrated model of phenotypic trait changes and site-specific sequence evolution. *Syst Biol*. 66:917–933.

Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol*. 30(4):772–780.

Kennedy AH, Taylor DL, Watson LE. 2011. Mycorrhizal specificity in the fully mycoheterotrophic *Hexalectris* Raf. (Orchidaceae: Epidendroideae). *Mol Ecol*. 20(6):1303–1316.

Kennedy AH, Watson LE. 2010. Species delimitations and phylogenetic relationships within the fully myco-heterotrophic *Hexalectris* (Orchidaceae). *Syst Bot*. 35(1):64–76.

Kim HT, Chase MW. 2017. Independent degradation in genes of the plastid *ndh* gene family in species of the orchid genus *Cymbidium* (Orchidaceae; Epidendroideae). *PLoS One* 12(11):e0187318.

Kim YK, Kwak MH, Chung MG, Kim HW, Jo S, Sohn JY, Cheon SH, Kim KJ. 2017. The complete plastome sequence of the endangered orchid *Cymbidium macrorhizon* (Orchidaceae). *Mitochondrial DNA Part B Resour*. 2(2):725–727.

Kosakovsky Pond SLK, Frost SDW, Muse SV. 2005. HyPhy: hypothesis testing using phylogenies. *Bioinformatics* 21(5):676–679.

Krause K. 2008. From chloroplasts to "cryptic" plastids: evolution of plastid genomes in parasitic plants. *Curr Genet*. 54(3):111–121.

Krause K. 2011. Piecing together the puzzle of parasitic plant plastome evolution. *Planta* 234(4):647–656.

Krause K. 2012. Plastid genomes of parasitic plants: a trail of reductions and losses. In Bullerwell C, editor. Organelle genetics: evolution of organelle genomes and gene expression. Berlin (Germany): Springer. p. 79–103.

Kumar S, Stecher G, Tamura K. 2016. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. *Mol Biol Evol*. 33(7):1870–1874.

Kurtz S, Choudhuri JV, Ohlebusch E, Schleiermacher C, Stoye J, Giegerich R. 2001. REPuter: the manifold applications of repeat analysis on a genomic scale. *Nucleic Acids Res*. 29(22):4633–4642.

Lartillot N, Poujol R. 2011. A phylogenetic model for investigating correlated evolution of substitution rates and continuous phenotypic characters. *Mol Biol Evol*. 28(1):729–744.

Laslett D, Canback B. 2004. ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. *Nucleic Acids Res*. 32(1):11–16.

Leake JR. 1994. The biology of myco-heterotrophic (saprophytic) plants. *New Phytol*. 127(2):171–216.

Leake JR. 2005. Plants parasitic on fungi- unearthing the fungi in myco-heterotrophs and debunking the 'saprophytic' plant myth. *Mycologist* 19:113–122.

Leake JR, Cameron DD, Beerling DJ. 2008. Fungal fidelity in the myco-heterotroph-to-autotroph life cycle of Lycopodiaceae: a case of parental nurture? *New Phytol*. 177(3):572–576.

Lewis PO. 2001. A likelihood approach to estimating phylogeny from discrete morphological character data. *Syst Biol*. 50(6):913–925.

Lin C-S, Chen JJW, Chiu C-C, Hsiao HCW, Yang C-J, Jin X-H, Leebens-Mack J, de Pamphilis CW, Huang Y-T, Yang L-H, et al. 2017. Concomitant loss of NDH complex-related genes within chloroplast and nuclear genomes in some orchids. *Plant J*. 90(5):994–1006.

Lowe TM, Chan PP. 2016. tRNAscan-SE on-line: integrating search and context for analysis of transfer RNA genes. *Nucleic Acids Res*. 44(W1):W54–W57.

Maréchal A, Parent JS, Veronneau-Lafortune F, Joyeux A, Lang BF, Brisson N. 2009. *WHIRLY* proteins maintain plastid genome stability in *Arabidopsis*. *Proc Natl Acad Sci U S A*. 106(34):14693–14698.

Mayrose I, Otto SP. 2011. A Likelihood method for detecting trait-dependent shifts in the rate of molecular evolution. *Mol Biol Evol*. 28(1):759–770.

McNeal JR, Arumugunathan K, Kuehl JV, Boore JL, Depamphilis CW. 2007. Systematics and plastid genome evolution of the cryptically photosynthetic parasitic plant genus *Cuscuta* (Convolvulaceae). *BMC Biol*. 5:55.

McNeal JR, Kuehl JV, Boore JL, de Pamphilis CW. 2007. Complete plastid genome sequences suggest strong selection for retention of photosynthetic genes in the parasitic plant genus *Cuscuta*. *BMC Plant Biol*. 7(1):57.

Merckx V, Freudenstein JV. 2010. Evolution of mycoheterotrophy in plants: a phylogenetic perspective. *New Phytol*. 185(3):605–609.

Merckx VSFT, Freudenstein JV, Kissling J, Christenhusz JM, Stotler RE, Crandall-Stotler B, Wickett NJ, Rudall PJ, de Kamer HM, Maas P. 2013. Taxonomy and classification. In: Merckx VSFT, editor. Mycoheterotrophy: the biology of plants living on fungi. New York: Springer. p. 19–101.

Merckx VSFT, Smets EF, Specht CD. 2013. Biogeography and conservation. In: Merckx VSFT, editor. Mycoheterotrophy: the biology of plants living on fungi. New York: Springer. p. 103–156.

Molina J, Hazzouri KM, Nickrent D, Geisler M, Meyer RS, Pentony MM, Flowers JM, Pelser P, Barcelona J, Inovejas SA, et al. 2014. Possible loss of the chloroplast genome in the parasitic flowering plant *Rafflesia lagascae* (Rafflesiaceae). *Mol Biol Evol*. 31(4):793–803.

Mori H, Summer EJ, Cline K. 2001. Chloroplast *TatC* plays a direct role in thylakoid Delta pH-dependent protein transport. *FEBS Lett.* 501(1):65–68.
Motomura H, Selosse MA, Martos F, Kagawa A, Yukawa T. 2010. Mycoheterotrophy evolved from mixotrophic ancestors: evidence in *Cymbidium* (Orchidaceae). *Ann Bot.* 106(4):573–581.
Naumann J, Der JP, Wafula EK, Jones SS, Wagner ST, Honaas LA, Ralph PE, Bolin JF, Maass E, Neinhuis C, et al. 2016. Detecting and characterizing the highly divergent plastid genome of the nonphotosynthetic parasitic plant *Hydnora visseri* (Hydnoraceae). *Genome Biol Evol.* 8(2):345–363.
Naumann J, Salomo K, Der JP, Wafula EK, Bolin JF, Maass E, Frenzke L, Samain M-S, Neinhuis C, dePamphilis CW, 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(11):e79204.
New CP, Ma QQ, Dabney-Smith C. 2018. Routing of thylakoid lumen proteins by the chloroplast twin arginine transport pathway. *Photosynth Res.* 138(3):289–301.
Nickrent DL, Starr EM. 1994. High-rates of nucleotide substitution in nuclear small-subunit (18s) rDNA from holoparasitic flowering plants. *J Mol Evol.* 39(1):62–70.
Ogura-Tsujita Y, Miyoshi K, Tsutsumi C, Yukawa T. 2014. First flowering hybrid between autotrophic and mycoheterotrophic plant species: breakthrough in molecular biology of mycoheterotrophy. *J Plant Res.* 127(2):299–305.
Orr HA. 1995. The population-genetics of speciation—the evolution of hybrid incompatibilities. *Genetics* 139(4):1805–1813.
Paradis E, Schliep K. 2019. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* 35(3): 526–528.
Peltier G, Aro EM, Shikanai T. 2016. NDH-1 and NDH-2 plastoquinone reductases in oxygenic photosynthesis. *Annu Rev Plant Biol.* 67:55–80.
Plant List. 2013. Version 1.1. Published on the internet. Available from: http://www.theplantlist.org (last accessed November 16, 2018).
Ranwez V, Douzery EJP, Cambon C, Chantret N, Delsuc F. 2018. MACSE v2: toolkit for the alignment of coding sequences accounting for frameshifts and stop codons. *Mol Biol Evol.* 35(10):2582–2584.
Rasmussen H. 1995. Terrestrial orchids: from seed to mycotrophic plant. Cambridge: Cambridge University Press.
Ravin NV, Gruzdev EV, Beletsky AV, Mazur AM, Prokhortchouk EB, Filyushin MA, Kochieva EZ, Kadnikov VV, Mardanov AV, Skryabin KG. 2016. The loss of photosynthetic pathways in the plastid and nuclear genomes of the non-photosynthetic mycoheterotrophic eudicot *Monotropa hypopitys*. *BMC Plant Biol.* 16:153–161.
Revell LJ. 2012. PHYTOOLS: an R package for phylogenetic comparative biology (and other things). *Methods Ecol Evol.* 3(2):217–223.
Revill MJW, Stanley S, Hibberd JM. 2005. Plastid genome structure and loss of photosynthetic ability in the parasitic genus *Cuscuta*. *J Exp Bot.* 56(419):2477–2486.
Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP. 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst Biol.* 61(3):539–542.
Ruhlman TA, Jansen RK. 2014. The plastid genomes of flowering plants. *Methods Mol Biol.* 1132:3–38.
Schneider AC, Chun H, Stefanovic S, Baldwin BG. 2018. Punctuated plastome reduction and host-parasite horizontal gene transfer in the holoparasitic plant genus *Aphyllon*. *Proc R Soc B* 285(1887):20181535.
Sosa V. 2007. A molecular and morphological phylogenetic study of subtribe Bletiinae (Epidendreae, Orchidaceae). *Syst Bot.* 32(1):34–42.
Sosa V, Cameron KM, Angulo DF, Hernandez-Hernandez T. 2016. Life form evolution in epidendroid orchids: ecological consequences of the shift from epiphytism to terrestrial habit in *Hexalectris*. *Taxon* 65(2):235–248.
Spiess AN. 2018. qpcR: modelling and analysis of real-time PCR data. R package version 1.4-1. Available from: https://CRAN.R-project.org/package=qpcR (last accessed December 15, 2018).
Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30(9):1312–1313.
Suetsugu K, Ohta T, Tayasu I. 2018. Partial mycoheterotrophy in the leafless orchid *Cymbidium macrorhizon*. *Am J Bot.* 105(9):1595–1600.
Suetsugu K, Yamato M, Miura C, Yamaguchi K, Takahashi K, Ida Y, Shigenobu S, Kaminaka H. 2017. Comparison of green and albino individuals of the partially mycoheterotrophic orchid *Epipactis helleborine* on molecular identities of mycorrhizal fungi, nutritional modes and gene expression in mycorrhizal roots. *Mol Ecol.* 26(6):1652–1669.
Sun FJ, Caetano-Anollés G. 2009. The evolutionary significance of the long variable arm in transfer RNA. *Complexity* 14(5):26–39.
Taylor DL, Bruns TD. 1997. Independent, specialized invasions of ectomycorrhizal mutualism by two nonphotosynthetic orchids. *Proc Natl Acad Sci U S A.* 94(9):4510–4515.
Taylor DL, Bruns TD, Szaro TM, Hodges SA. 2003. Divergence in mycorrhizal specialization within *Hexalectris spicata* (Orchidaceae), a nonphotosynthetic desert orchid. *Am J Bot.* 90(8):1168–1179.
Wertheim JO, Murrell B, Smith MD, Pond SLK, Scheffler K. 2015. RELAX: detecting relaxed selection in a phylogenetic framework. *Mol Biol Evol.* 32(3):820–832.
Wicke S, Muller KF, dePamphilis CW, Quandt D, Bellot S, Schneeweiss GM. 2016. Mechanistic model of evolutionary rate variation en route to a nonphotosynthetic lifestyle in plants. *Proc Natl Acad Sci U S A.* 113(32):9045–9050.
Wicke S, Muller KF, de Pamphilis CW, Quandt D, Wickett NJ, Zhang Y, Renner SS, Schneeweiss GM. 2013. Mechanisms of functional and physical genome reduction in photosynthetic and nonphotosynthetic parasitic plants of the broomrape family. *Plant Cell* 25(10):3711–3725.
Wicke S, Naumann J. 2018. Molecular evolution of plastid genomes in parasitic flowering plants. In: Chaw S-M, Jansen RK, editors. Advances in botanical research. Cambridge, Massachusetts, USA:Academic Press. p. 315–347.
Wicke S, Schneeweiss GM, dePamphilis CW, Muller KF, Quandt D. 2011. The evolution of the plastid chromosome in land plants: gene content, gene order, gene function. *Plant Mol Biol.* 76(3-5):273–297.
Wolfe KH, Morden CW, Palmer JD. 1992. Function and evolution of a minimal plastid genome from a non-photosynthetic parasitic plant. *Proc Natl. Acad Sci U S A.* 89(22):10648–10652.
Wyman SK, Jansen RK, Boore JL. 2004. Automatic annotation of organellar genomes with DOGMA. *Bioinformatics* 20(17):3252–3255.
Yuan Y, Jin XH, Liu J, Zhao X, Zhou JH, Wang X, Wang DY, Lai CJS, Xu W, Huang JW, et al. 2018. The *Gastrodia elata* genome provides insights into plant adaptation to heterotrophy. *Nat Commun.* 9(1):1615.
Zerbino DR, Birney E. 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res.* 18(5):821–829.
Zhang J, Ruhlman TA, Sabir J, Blazier JC, Jansen RK. 2015. Coordinated rates of evolution between interacting plastid and nuclear genes in Geraniaceae. *Plant Cell* 27(3):563–573.
Zhang J, Ruhlman TA, Sabir JSM, Blazier JC, Weng ML, Park S, Jansen RK. 2016. Coevolution between nuclear-encoded DNA replication, recombination, and repair genes and plastid genome complexity. *Genome Biol Evol.* 8(3):622–634.
Zhu A, Guo W, Gupta S, Fan W, Mower JP. 2016. Evolutionary dynamics of the plastid inverted repeat: the effects of expansion, contraction, and loss on substitution rates. *New Phytol.* 209(4):1747–1756.
Zimmer K, Meyer C, Gebauer G. 2008. The ectomycorrhizal specialist orchid *Corallorhiza trifida* is a partial myco-heterotroph. *New Phytol.* 178(2):395–400.