Investigating the Path of Plastid Genome Degradation in an Early-Transitional Clade of Heterotrophic Orchids, and Implications for Heterotrophic Angiosperms

Craig F. Barrett,*[1] John V. Freudenstein,[2] Jeff Li,[1] Dustin R. Mayfield-Jones,[3] Leticia Perez,[1] J. Chris Pires,[3] and Cristian Santos[1]

[1]Department of Biological Sciences, California State University, Los Angeles

[2]Department of Evolution, Ecology, and Organismal Biology and the Museum of Biological Diversity, Ohio State University

[3]Division of Biological Sciences, University of Missouri

***Corresponding author:** E-mail: cbarret5@calstatela.edu.

Associate editor: Naoki Takebayashi

Abstract

Parasitic organisms exemplify morphological and genomic reduction. Some heterotrophic, parasitic plants harbor drastically reduced and degraded plastid genomes resulting from relaxed selective pressure on photosynthetic function. However, few studies have addressed the initial stages of plastome degradation in groups containing both photosynthetic and nonphotosynthetic species. *Corallorhiza* is a genus of leafless, heterotrophic orchids that contains both green, photosynthetic species and nongreen, putatively nonphotosynthetic species, and represents an ideal system in which to assess the beginning of the transition to a "minimal plastome." Complete plastomes were generated for nine taxa of *Corallorhiza* using Illumina paired-end sequencing of genomic DNA to assess the degree of degradation among taxa, and for comparison with a general model of degradation among angiosperms. Quantification of total chlorophyll suggests that nongreen *Corallorhiza* still produce chlorophyll, but at 10-fold lower concentrations than green congeners. Complete plastomes and partial nuclear rDNA cistrons yielded a fully resolved tree for *Corallorhiza*, with at least two independent losses of photosynthesis, evidenced by gene deletions and pseudogenes in *Co. striata* and nongreen *Co. maculata*. All *Corallorhiza* show some evidence of degradation in genes of the NAD(P)H dehydrogenase complex. Among genes with open reading frames, photosynthesis-related genes displayed evidence of neutral evolution in nongreen *Corallorhiza*, whereas genes of the ATP synthase complex displayed some evidence of positive selection in these same groups, though for reasons unknown. *Corallorhiza* spans the early stages of a general model of plastome degradation and has added critical insight for understanding the process of plastome evolution in heterotrophic angiosperms.

Key words: **Orchidaceae, chloroplast, pseudogene, parasite, chloroplast, photosynthesis, chlorophyll.**

Introduction

Parasites often display extreme reduction in morphological and genomic features (e.g., Moran 2002; Lawrence 2005; Morrison et al. 2007). There are many examples of reduced genomes in parasitic animals (e.g., Protasio et al. 2012; Tsai et al. 2013) and fungi (e.g., Katinka et al. 2001; Yoder and Turgeon 2001; Galagan et al. 2003). Heterotrophic plants—sensu lato, those that obtain nutrients either from other plants or from fungi with which they associate (i.e., mycoheterotrophs)—provide unique opportunities to study the types of changes that occur along the transition to parasitism. Some heterotrophic plants retain the ability to carry out photosynthesis (i.e., hemiparasites and partial mycoheterotrophs) whereas others have lost this ability, presumably due to a relaxation of selective constraints associated with photosynthetic function (i.e., holoparasites [Kuijt 1969] and holomycotrophs). Parasitism upon other plants likely evolved a minimum of 12 times among eudicot angiosperms (Westwood et al. 2010; McNeal et al. 2013), whereas mycoheterotrophy has evolved at least 40 times, mainly among monocot angioisperms, with at least 30 independent shifts within the family Orchidaceae alone (Leake 1994; Bidartondo 2005; Freudenstein and Barrett 2010; Leake and Cameron 2010; Merckx and Freudenstein 2010).

Plastid-encoded genes in heterotrophic plants are expected to display evidence of degradation through mutation with little or no selective consequences due to relaxed purifying selection on photosynthesis. The plastid genome has been the target of recent studies in heterotrophic plants, in part brought about by the availability of next-generation sequencing technologies. Examples of heterotrophic taxa for which plastome evolution has been studied include various Orobanchaceae (e.g., *Epifagus*, Wolfe et al. 1992; *Conopholis*, Wimpee et al. 1991; *Hyobanche*/*Harveya*, Wolfe and dePamphilis 1998; Randle and Wolfe 2005; family-wide representative genera, Wicke et al. 2013), Convolvulaceae (*Cuscuta*, Funk et al. 2007; McNeal et al. 2007; Braukmann et al. 2013), monotropoid Ericaceae (Braukmann and Stefanovic 2012; Broe M and Freudenstein J, personal communication), and monocots including Orchidaceae (Delannoy et al. 2011; Logacheva et al. 2011, 2014; Barrett and Davis 2012).

Nearly all of those taxa display some degree of reduction in overall plastome size and gene content, particularly with reduction in photosynthesis-related gene systems, with many "extreme" examples of genomic reduction (e.g., Delannoy et al. 2011; Wicke et al. 2013). In the heterotrophic angiosperm *Rafflesia* and free-living, nonphotosynthetic alga *Polytomella*, failure to recover plastid genomes suggests their complete loss (Molina et al. 2014; Smith and Lee 2014). Although these examples of extreme reduction of the plastid genome are highly interesting, the initial stages in the progression of plastid genome degradation remain poorly understood. In order to address this issue, focus must be oriented toward lower taxa (e.g., see Braukmann et al. 2013) displaying variation in trophic strategy, and thus representing a "snapshot" of the transition from autotrophy to strict heterotrophy.

Building on previous syntheses of plastid genome evolution in heterotrophic plants (e.g., see Krause 2008; Wicke et al. 2011), Barrett and Davis (2012) proposed a model of plastid genome degradation with five "stages": (1) *ndh* → (2) *psa*/*psb*, *pet*, *rbcL*, *ycf3*, *4* → (3) *rpo* → (4) *atp* → (5) *rpl*, *rps*, *rrn*, *trn*, *accD*, *clpP*, *infA*, *matK*, *ycf1*, 2. The rationale behind this model is that in many groups, the NAD(P)H complex (i.e., *ndh* genes, stage 1) is the first plastid-encoded gene system to show evidence of degradation, even in some photosynthetic groups (see Wicke et al. 2011; Barrett and Davis 2012, and references therein). Second, as a result of the transition to a heterotrophic lifestyle, purifying selection is expected to be relaxed for genes involved directly in photosynthesis (stage 2), followed by degradation of the plastid-encoded RNA polymerase complex ("PEP," or rpo-genes; stage 3), due to the role of the latter in transcribing plastid operons encoding many gene products with photosynthesis-related functions. The next complex hypothesized to show evidence of degradation is the ATP synthase complex (*atp*-genes; stage 4). Although this complex is directly involved in synthesizing ATP as part of the photosynthetic process, its function may remain essential in plastids other than chloroplasts (e.g., amyloplasts). The last proposed stage is the degradation of "housekeeping" genes that play roles in basic processes of the plastid itself (e.g., on-site protein synthesis, RNA processing).

An ideal taxon for comparison to the model of Barrett and Davis (2012) is the mycoheterotrophic *Corallorhiza* (the coralroot orchids). *Corallorhiza* is a genus of 12 leafless species (Freudenstein 1992, 1997; Freudenstein and Senyo 2008). The genus contains some species that are partially heterotrophic and have presumably retained the capacity for photosynthesis. These taxa all have some degree of visible green tissue (*Co. trifida*, *Co. odontorhiza*, *Co. wisteriana*, *Co. bulbosa*, *Co. macrantha*, and *Co. maculata* var. *mexicana*; see supplementary fig. S1, Supplementary Material online). Photosynthesis has been demonstrated in one of these species, *Co. trifida*, the "greenest" of the coralroots (Zimmer et al. 2008; Cameron et al. 2009). Other, putatively achlorophyllous coralroot species completely lack visible green tissue (supplementary fig. S1, Supplementary Material online), including the *Co. striata* complex and some members of the *Co. maculata* complex (*Co. mertensiana* and *Co. maculata* vars. *maculata* and *occidentalis*). Preliminary evidence of plastome degradation in *Corallorhiza* comes from heterologous probe experiments (Freudenstein and Doyle 1994a), the presence of RuBisCO Large Subunit (*rbc*L) pseudogenes in some lineages but not others (Barrett and Freudenstein 2008), and sequencing of the complete plastome of *Co. striata* var. *vreelandii* (Barrett and Davis 2012).

Corallorhiza represents a potentially powerful model clade for answering questions related to the earliest stages of plastome degradation in heterotrophic plants. Furthermore, little is known about the plastid genomes of partial heterotrophs—do they also display evidence of "transitional" plastome degradation? Are all nongreen species of *Corallorhiza* truly achlorophyllous? Which gene complexes are the first to accumulate loss-of-function mutations? Is there a typical progression of pseudogene formation/gene loss within this genus, in accordance with the model of Barrett and Davis (2012), and how do the plastomes of *Corallorhiza* compare with those of other parasitic angiosperms? When comparing plastomes of "green" versus "nongreen" species, are there differences in the numbers of functional genes, overall plastome size, GC content, nonsynonymous versus synonymous substitution rates, and structural rearrangements? In this study, paired-end (PE) Illumina sequencing of genomic DNA was carried out to generate complete plastomes for nine members of *Corallorhiza*, representing all species complexes in the genus, in order to address the following principal hypotheses:

(I) Nongreen members of *Corallorhiza* display plastid genome degradation (pseudogenes, deletions, rearrangements), whereas members with some green tissue are more conserved when both are compared with the plastomes of green, leafy orchid species.

(II) There is a general path of degradation among plastid-encoded gene complexes, corresponding to the model of Barrett and Davis (2012); *Corallorhiza* occupies the relatively early stages in this model.

(III) In addition to pseudogenes and losses, some members of *Corallorhiza* have experienced relaxed purifying selection for loci with intact reading frames, especially in lineages that display evidence of degradation in other regions of the plastome.

Results

Chlorophyll Content

All *Corallorhiza* taxa sampled have detectable levels of chlorophyll (fig. 1). Of the three green taxa included, *Co. trifida* has the highest mean total chlorophyll concentration (31.6 ng/mg), whereas *Co. wisteriana* and *Co. odontorhiza* have slightly less (21.5 and 17.9 ng/mg, respectively); there is relatively wide variation in chlorophyll content among individuals within each green species. All nongreen members of *Corallorhiza* have lower chlorophyll concentrations compared with green *Corallorhiza*, with none exceeding a mean value of 2 ng/mg (fig. 1). Chlorophyll content varies significantly among green versus nongreen taxa after taking phylogeny

into account by way of Phylogenetic Analysis of Variance ($P_{PhyANOVA} = 0.0146$; see fig. 1). Moreover, these values are strikingly consistent across the nongreen species for which chlorophylls were measured, with very little variation within each species.

Genomic Data Sets and Plastomes

Characteristics of the Illumina data sets and resulting finished plastome assemblies are listed in tables 1 and 2. Mean coverage of plastomes for newly sequenced genomic data sets ranges from 81.0 × (*Co. wisteriana*) to 271.3 × (*Co. odontorhiza*). The length of the plastome among the *Corallorhiza* accessions sampled ranges from 151,506 bp in *Co. macrantha* to 137,505 bp in *Co. striata vreelandii* (table 2). On average, plastomes of taxa with at least some visible green tissue are 148.74 kb in length, while those from taxa lacking visible green tissue average 144.98 kb, but the difference in mean total plastome length between these two groupings is nonsignificant after taking relationships into account ($P_{PhyANOVA} = 0.06189$).

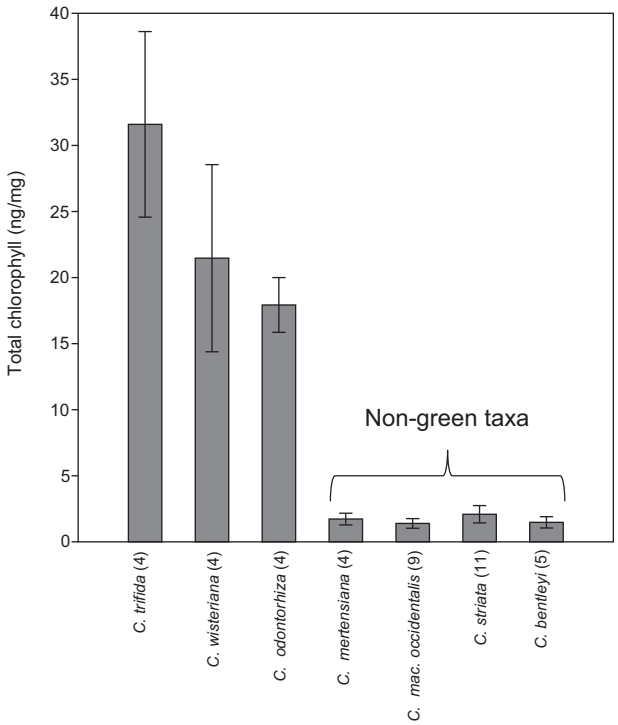

Fig. 1. Total chlorophyll concentration (chlorophylls a and b combined, as measured by UV-Vis spectroscopy), averaged across multiple individuals in each taxon. Numbers adjacent to taxon names are samples sizes (# individuals). Error bars represent standard deviations.

The length of the inverted repeat (IR) ranges from 25,775 bp in *Co. odontorhiza* to 27,243 bp in *Co. maculata* var. *mexicana*, with no major variation in gene content, nor evidence of any gene being incorporated into or lost from the IR. *Corallorhiza striata* var. *vreelandii* has by far the shortest large single copy (LSC, 72,152 bp). This is followed by *Co. maculata* var. *maculata* (80,401 bp), *Co. mertensiana* (81,146 bp), and *Co. maculata* var. *occidentalis* (81,363 bp); *Co. macrantha* (84,253 bp) and *Co. maculata* var. *mexicana* (84,349 bp) have the longest LSC regions. Thus, the range of length variation among members of the *Co. maculata* complex for the LSC region (3,852 bp) is substantial. Nongreen taxa have significantly shorter LSC regions after accounting for phylogenetic relationships ($P_{PhyANOVA} = 0.0148$). Unique to *Co. maculata* var. *maculata* is an approximately 16-kb inversion in the LSC region, with breakpoints occurring near *ycf4* and within the *petD* intron (supplementary fig. S2, Supplementary Material online).

The length of the small single copy (SSC) also varies substantially, ranging from 11,743 bp in *Co. wisteriana* to 14,427 in *Co. trifida*, but there is no significant difference in length between the SSC regions of the plastome for green versus nongreen taxa ($P_{PhyANOVA} = 0.7382$), nor is there any correlation between SSC length and chlorophyll content ($P_{IC} = 0.4093$). GC content is lowest in *Co. striata vreelandii* (34.14%), and highest in *Co. trifida* (35.74%), but does not differ significantly among green versus nongreen taxa nor

Table 1. Voucher Information, GenBank Accession Numbers, the Number of PE Illumina Reads (#reads), and Mean Coverage of Each Plastome.

Species	Voucher	GenBank[a]	#reads	×-Cov[b]
Corallorhiza bulbosa A. Rich. & Galeotti	CFB 238[c] (OS, MEXU)[d]	KM390013	10,757,100	243.8
Co. macrantha Schltr.	Salazar 8191[c] (OS, MEXU)	KM390017	9,222,076	136.4
Co. maculata (Raf.) Raf. var. *maculata*	JVF 2919 (OS)	KM390014	11,498,452	109.1
Co. maculata var. *mexicana* (Lindl.) Freudenstein	CFB 232[c] (OS, MEXU)	KM390015	7,828,614	118.2
Co. maculata var. *occidentalis* (Lindl.) Ames	JVF 2095 (OS)	KM390016	13,204,554	225.1
Co. mertensiana Bong.	JVF 1999 (OS)	KM390018	14,540,670	212.0
Co. odontorhiza (Willd.) Poir.	JVF 2778 (OS)	KM390021	21,861,728	271.3
Co. striata L. var. *vreelandii*[e] (Rydb.) L.O. Williams	Taylor 341 (UNM)	JX087681	3,204,074	43.9
Co. trifida Châtel.	JVF 2676 (OS)	KM390019	13,617,730	236.9
Co. wisteriana Conrad	JVF 2462 (OS)	KM390020	9,992,438	81.0

[a]GenBank accession numbers will be included contingent upon a favorable review.
[b]Mean depth of coverage of each plastome (×-cov), based on remapping the original read pool to the finished plastome.
[c]Duplicate vouchers at Universidad Nacional Autónoma de México (UNAM).
[d]Herbarium codes: (OS), Ohio State University Herbarium; (UNM), University of New Mexico Herbarium; (MEXU), Universidad Nacional Autónoma de México.
[e]Sequenced and assembled in Barrett and Davis (2012).

Table 2. Green Tissue Status, Lengths of Plastomes and Subregions, GC Content, and Numbers of Putatively Functional Plastid Genes for Each Accession.

Species	Green Tissue?[a]	Length (bp)	LSC (bp)	SSC (bp)	IR (bp)	%GC	No. Functional Genes[c]
Corallorhiza bulbosa	Yes	148,643	82,853	12,368	26,711	35.19	105
Co. macrantha	Yes	151,031	84,253	12,554	27,112	35.21	104
Co. maculata var. *maculata*	No	146,886	80,401	12,885	26,800	34.94	89
Co. maculata var. *mexicana*	Yes	151,506	84,349	12,671	27,243	35.18	103
Co. maculata var. *occidentalis*	No	146,595	81,363	12,368	26,432	35.14	88
Co. mertensiana	No	147,941	81,146	13,737	26,529	35.23	90
Co. odontorhiza	Yes	147,317	82,259	13,508	25,775	35.62	103
Co. striata var. *vreelandii*[d]	No	137,505	72,152	12,387	26,483	34.14	82
Co. trifida[b]	Yes	149,384	83,085	14,427	25,936	35.74	107
Co. wisteriana	Yes	146,437	82,350	11,743	26,172	35.28	103

NOTE.—%GC, G+C base percentage of each plastome.

[a]Presence of visible green tissue, based on observation of multiple populations of each species.

[b]Photosynthesis has only been demonstrated in *Co. trifida*.

[c]The number of putatively functional genes, as assessed by intact reading frames.

[d]Sequenced and assembled in Barrett and Davis (2012).

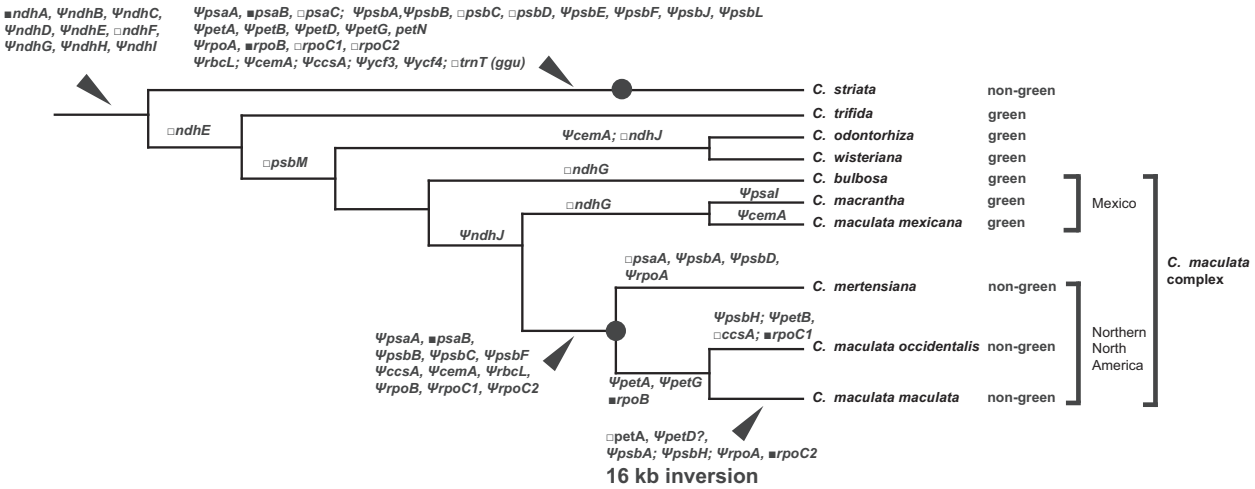

FIG. 2. Maximum-likelihood tree based on whole plastomes + partial rDNA cistron, with pseudogene (ψ) content and deletions mapped parsimoniously. ■, partial gene deletion (> 6 bp) causing frame shift or truncation (<50% of gene missing); □, complete or nearly complete gene deletion causing frame shift or truncation (> 50% of gene missing). *psa*, photosystem I; *psb*, PSII; *pet*, cytochrome; *rpo*, plastid-encoded polymerase; *ndh*, NAD(P)H dehydrogenase; *ccsA*, cytochrome C heme-binding protein A; *cemA*, chloroplast envelope membrane protein A; *rbcL*, RuBisCO large subunit; *trn*, transfer RNA; *ycf3* and *ycf4*, putative photosystem assembly factors. Dark circles on branches indicate putative transitions to strict heterotrophy.

does it correlate with chlorophyll content after accounting for phylogenetic relationships ($P_{PhyANOVA} = 0.0648$ and $P_{IC} = 0.0989$, respectively).

Structural Evolution of the Plastome: Pseudogenes and Losses

Figure 2 shows the maximum likelihood (ML) tree based on whole plastomes + rDNA (see below), with gene losses-of-function parsimoniously mapped, assuming no reversals. Nongreen taxa have significantly fewer putatively functional loci (as assessed by intact reading frames) than green *Corallorhiza* and leafy outgroups ($P_{PhyANOVA} < 0.0001$; table 2). In the subset of coralroot taxa from which chlorophyll data are available, chlrorophyll concentration is also significantly correlated with the number of putatively functional genes ($P_{IC} = 0.002$).

Plastid genes encoding subunits of the NAD(P)H dehydrogenase complex (*ndh* genes) have mostly become pseudogenes in *Corallorhiza*, with some wholly or mostly deleted. Over 50% of the gene encoding Photosystem II (PSII) protein M (*psbM*), a small gene of 105 bp in length, has been deleted in all members of *Corallorhiza* except *Co. striata vreelandii* and *Co. trifida*; it has been completely deleted in *Co. maculata* vars. *maculata* and *occidentalis*. Among members of *Corallorhiza* displaying at least some green tissue, few other pseudogenes or gene losses were detected. The *cemA* locus is

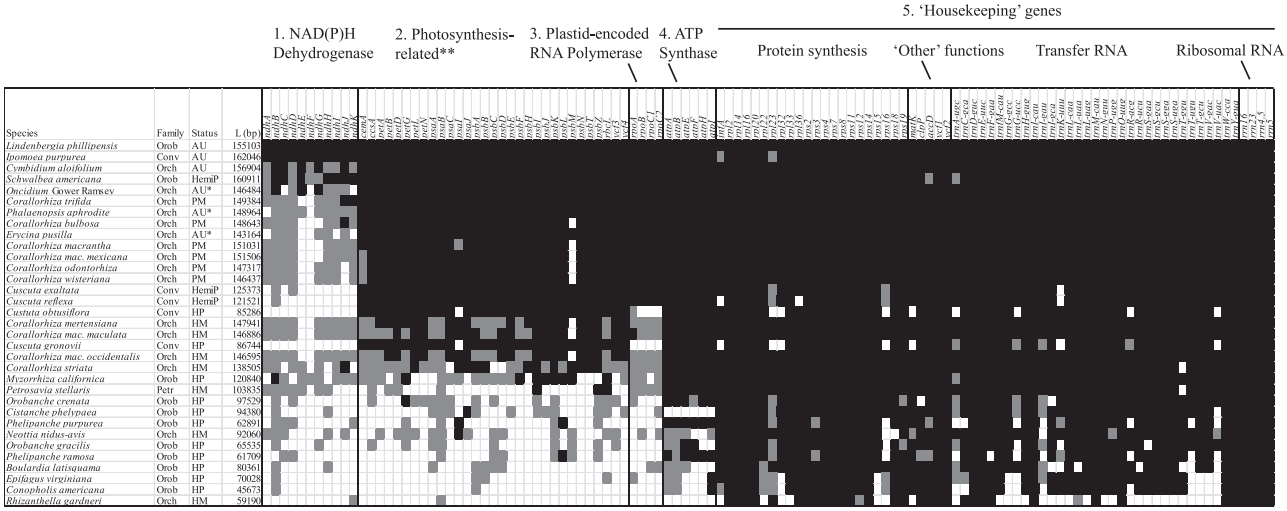

Fig. 3. A basic model of plastid genome degradation (Barrett and Davis 2012) using sequenced, exemplar plastomes from angiosperm families containing heterotrophs: Convolvulaceae (Funk et al. 2007; McNeal et al. 2007), Orobanchaceae (Wolfe et al. 1992; Wicke et al. 2013), Orchidaceae (Chang et al. 2006; Wu et al. 2010; Delannoy et al. 2011; Barrett and Davis 2012; Logacheva et al. 2011; Yang et al. 2013), and Petrosaviaceae (Logacheva et al. 2014). Plastomes are ranked by numbers of putatively functional genes (highest to lowest). Black-filled spaces represent putatively functional genes with open reading frames, whereas gray-filled spaces represent pseudogenes, and white spaces represent complete or nearly complete gene losses. "Status," trophic strategy of each species: AU, nonparasitic autotroph; HemiP, hemiparasite; HP, holoparasite; PM, partial mycoheterotroph (analogous to hemiparasite); HM, holomycotroph. *For the purposes of comparison, both leafy, green orchids (*Oncidium*, *Phalaenopsis*) are considered nonparasitic autotrophs, even though they may still rely partially on their mycorrhizal fungi. *L* (bp), total length of each plastome, in base pairs. **Photosynthesis-related excludes *atp* genes, as these may serve functions outside of photosynthesis.

a pseudogene in all but *Co. bulbosa*, *Co. macrantha*, and *Co. trifida*, whereas *psaI* has experienced a 4-bp insertion in *Co. macrantha* near the 5′-end of the gene, causing a reading frame shift (fig. 2). Grep searches of the original reads confirmed this to be the case. No pseudogenes or major deletions causing shifts in reading frame were detected in *atp* genes in any species of *Corallorhiza*.

As previously discussed in Barrett and Davis (2012), at least one member of all major photosynthesis-related gene systems in *Co. striata vreelandii* has experienced pseudogene formation or gene deletion, with the exception of genes encoding subunits of the ATP synthase complex (*atp* genes). Several loss-of-function mutations have occurred in the nongreen, northern North American members of the *Co. maculata* complex (*Co. mertensiana*, *Co. maculata maculata*, and *Co. maculata occidentalis*). These three taxa share a partial deletion in *psbB*, as well as numerous pseudogenes in the *psa* (Photosystem I), *psb* (PSII), and *rpo* (plastid-encoded RNA polymerase) complexes, evidenced by reading frame shifts and internal stop codons. They also all have pseudogenes for *ccsA* (cytochrome C biogenesis protein) and *rbcL* (RuBisCO large subunit).

Unique mutations to *Co. mertensiana* among the northern North American members of the *Co. maculata* complex include deletion of a large portion of *psaA*, and pseudogenes for *psbA*, *psbD*, and a deletion in *rpoC1*. *Corallorhiza maculata maculata* and *Co. maculata occidentalis* both contain pseudogenes for *petA* (Cytochrome f), *petG* (cytochrome b6/f complex subunit V), and *rpoB*. Mutations unique to *Co. maculata occidentalis* include pseudogenes for *psbH* and *petB*, with deletions in *ccsA* and *rpoC1*. Mutations in *Co. maculata maculata* include the evolution of pseudogenes in *psbA* and deletions in *rpoA* and *rpoC2*. Although the *petD* gene in *Co. maculata maculata* appears to bear an intact reading frame, a breakpoint of a 16-kb inversion (fig. 2) separates the two exons of this gene, which in other taxa are adjacent and thus presumably *cis*-spliced.

Phylogenetic Analyses

Relationships based on ML analyses of various data matrices were generally consistent (supplementary figs. S3 and S4, Supplementary Material online). The main difference among ML topologies resulting from these data sets is the placement of *Co. bulbosa*. For plastid coding data, *Co. bulbosa* is sister to *Co. odontorhiza* + *Co. wisteriana* (bootstrap = 97%), whereas for nuclear rDNA (partial ETS-18 S-ITS1-5.8 S-ITS2-26 S) it is placed as sister to *Co. macrantha*, *Co. maculata mexicana* (bootstrap = 92%). Whole plastomes, including introns and intergenic spacers, place *Co. bulbosa* as sister to the rest of the *Co. maculata* complex (bootstrap = 91%). Nuclear rDNA sequences place *Co. trifida* as sister to the remaining *Corallorhiza*; however, the placement of *Co. striata* as sister to the remaining *Corallorhiza* (excluding *Co. trifida*) received low support (bootstrap = 69%), suggesting little support from rDNA at the base of *Corallorhiza*.

Analyzing the entire plastome along with nuclear rDNA yields a completely resolved *Corallorhiza* tree, with 100% bootstrap support for all nodes (supplementary fig. S3, Supplementary Material online); Parsimony analysis of this data set yielded identical results to ML, including identical jackknife support values (100% for all nodes, tree not shown; see supplementary fig. S3, Supplementary Material online).

This topology was chosen for all downstream tree-based analyses. From the root of the tree (with *Cymbidium* as the outgroup), *Oncidium* and *Phalaenopsis* are successively sister to a monophyletic *Corallorhiza*, with *Co. striata vreelandii* and *Co. trifida* successively diverging from the ancestor of the remaining members of *Corallorhiza*. Nested within this clade, (*Co. odontorhiza*, *Co. wisteriana*) is sister to the *Co. maculata* complex. The Mexican *Co. bulbosa* is sister to all other members of the *Co. maculata* complex, with 100% support, whereas within the latter, the Mexican *Co. macrantha* and *Co. maculata mexicana* are sister to one another, and this clade is collectively sister to a northern North American clade of (*Co. mertensiana*, (*Co. maculata maculata*, *Co. maculata occidentalis*)).

Analyses of Selective Regime

Various branch models were constructed to explicitly test the hypothesis of different ω-ratios for different branches, specifically for nongreen *Corallorhiza*. Only loci for which all taxa had intact reading frames were included. The most basic model (M0) is a "one-ratio" model, which assumes a single ω-ratio across all branches (here, $\omega_{leafy} = \omega_{cor} = \omega_{mac\text{-}ng} = \omega_{str}$; see table 3 for branch-based abbreviations of ω-ratios). Model M1, which allows all of *Corallorhiza* to have a different ω-ratio than the leafy outgroups (ω_{leafy}, $\omega_{cor} = \omega_{mac\text{-}ng} = \omega_{str}$), has a significantly better fit for *atp*-genes (table 3). Also for *atp*-genes, a three-ratio model (M2) allowing different ω-ratios in *Corallorhiza* and additionally for nongreen *Corallorhiza* as opposed to leafy outgroups has a significantly better fit than the nested M1 model. Model M4, which is a two-ratio model specifying different ω-ratios for leafy outgroups + green *Corallorhiza* as opposed to nongreen *Corallorhiza* ($\omega_{leafy} = \omega_{cor}$, $\omega_{mac\text{-}ng} = \omega_{str}$), is highly significant compared with the one-ratio model (M0), after Bonferroni correction. Photosynthesis-related genes (excluding *atp*-genes) have a significantly better fit for models M2 (ω_{leafy}, ω_{cor}, $\omega_{mac\text{-}ng} = \omega_{str}$) and M4 ($\omega_{leafy} = \omega_{cor}$, $\omega_{mac\text{-}ng} = \omega_{str}$) than M0, but these were nonsignificant after Bonferroni correction. Regardless, the branch-specific estimate (model M4) of ω in nongreen taxa ($\omega_{mac\text{-}ng} = \omega_{str} = 0.73720$) was more than twice that of green taxa ($\omega_{leafy} = \omega_{cor} = 0.27361$), approaching a pattern expected under selective neutrality for photosynthesis-related genes ($\omega \sim 1$). "Housekeeping" genes show significantly better fit for M1 and M4, but these again were nonsignificant after Bonferroni correction for multiple comparisons. None of the branch models had a significantly better fit in nested comparisons for *ycf1* + *ycf2* or for *matK*.

Branch-site models were applied in the *atp* complex, to test for evidence of positive selection. First, for a model in which all *Corallorhiza* were specified as a foreground clade, the alternative model which allows some sites to be under positive selection ($\omega_2 > 1$) did not fit the data significantly better than the null model which allows no sites to be under positive selection ($\omega_1 = \omega_2 = 1$; $\chi^2 = 0.6$, df = 1, $P = 0.4386$). However, the alternative model had a significantly better fit than the null model when specifying nongreen *Corallorhiza* taxa as foreground branches, after Bonferroni correction for multiple comparisons ($\chi^2 = 11.4$, df = 1, $P = 0.0007$), suggesting that some sites in the *atp* complex are under positive selection.

Discussion

Chlorophyll Content of *Corallorhiza*

Some *Corallorhiza* and many heterotrophic plants are often referred to as "achlorophyllous," based on their lack of visible green coloration. However, all *Corallorhiza* included in this study contain at least some detectable levels of chlorophyll (fig. 1). Montfort and Küsters (1940) detected chlorophyll and photosynthetic activity in *Co. trifida*, which is not surprising, given the predominantly green coloration of the above-ground organs of this species; a later study by Cummings and Welschmeyer (1998) detected some levels of chlorophyll in the nongreen *Co. maculata*, as well as in another nongreen, strictly heterotrophic orchid, *Cephalanthera austiniae*. Taxa with at least some visible green tissue on average had approximately a 10-fold higher chlorophyll concentration than nongreen taxa (fig. 1). Although leafless, *Co. trifida* is the greenest of the coralroots, with green pigmentation throughout nearly the entire above-ground portion of the plant. The distribution of green tissues among other partially green coralroot species is more variable, but the common theme is that at least some green tissue is noticeable in or around the ovary.

Coupled with the fact that green or partially green coralroots also show a general lack of degradation in photosynthesis-related plastid gene systems (with the exception of the *ndh* complex and a few other "minor" examples; see below), these observations suggest that the aforementioned "green" coralroot taxa may indeed all be partially heterotrophic, relying to some small degree on photosynthetic carbon. Though it has been demonstrated that *Co. trifida* is an inefficient photosynthesizer (i.e., net photosynthetic carbon assimilation cannot compensate for net respiration), there is likely to be an adaptive reason that photosynthesis persists in this species, and possibly in others such as *Co. odontorhiza* and *Co. wisteriana*. Ironically, *Co. trifida* is the one coralroot species that might be expected to depend most heavily on photosynthesis (Zimmer et al. 2008; Cameron et al. 2009). One possibility is that these species supplement fungal carbon uptake with photosynthetic carbon in specific tissues only, mainly in the ovary, inside of which hundreds or even thousands of minute "dust seeds" are produced. This has been experimentally demonstrated in the partially mycoheterotrophic orchid *Limodorum abortivum*, with the additional findings that photosynthesis is highest in ovary tissue, and increases under experimentally induced fungal carbon limitation (Bellino et al. 2014). It is currently unknown if any temporal fungal carbon limitations occur throughout the short season when *Corallorhiza* displays aboveground growth (typically late spring for flowering to summer for seed/ovary development). Based on current knowledge, it is hypothesized that for vegetatively reduced orchids such as "green" *Corallorhiza* and *Limodorum*, compensation of carbon in and around developing ovaries seems to have some adaptive value, which in turn

Table 3. Branch-Models for Coding Regions of the Plastome in *Corallorhiza* for Which All Accessions Have Intact Reading Frames, Based on Various Configurations of Coding Data, Implemented in the CODEML Module of PAML.

Model: Omega Estimates	np[a]	ln L[b]	Λ[c]	P/P_{corr}[d]
***atp* genes**				
M0: $\omega_{leafy} = \omega_{cor} = \omega_{mac\text{-}ng} = \omega_{str} = 0.21505$	26	−9802.76868		
M1: $\omega_{leafy} = 0.15175$; $\omega_{cor} = \omega_{mac\text{-}ng} = \omega_{str} = 0.26949$	27	−9798.12794	9.281476	**/**
M2: $\omega_{leafy} = 0.15211$; $\omega_{cor} = 0.22238$; $\omega_{mac\text{-}ng} = \omega_{str} = 0.38239$	28	−9795.68827	4.879344	*/—
M3: $\omega_{leafy} = 0.15212$; $\omega_{cor} = 0.22242$; $\omega_{mac\text{-}ng} = 0.41999$; $\omega_{str} = 0.36243$	29	−9795.62404	0.128448	
M4: $\omega_{leafy} = \omega_{cor} = 0.18380$; $\omega_{mac\text{-}ng} = \omega_{str} = 0.38274$	27	−9797.32610	10.885142	***/**
M5: $\omega_{leafy} = \omega_{cor} = 0.18382$; $\omega_{mac\text{-}ng} = 0.41997$; $\omega_{str} = 0.36292$	28	−9797.26308	0.126042	
Photosynthesis genes (*pet, psa/b, rbcL*)				
M0: $\omega_{leafy} = \omega_{cor} = \omega_{mac\text{-}ng} = \omega_{str} = 0.33866$	26	−1854.62551		
M1: $\omega_{leafy} = 0.29790$; $\omega_{cor} = \omega_{mac\text{-}ng} = \omega_{str} = 0.32741$	27	−1854.48900	0.273012	
M2: $\omega_{leafy} = 0.29856$; $\omega_{cor} = 0.25277$; $\omega_{mac\text{-}ng} = \omega_{str} = 0.73721$	28	−1852.32692	4.324170	*/—
M3: $\omega_{leafy} = 0.29858$; $\omega_{cor} = 0.25278$; $\omega_{mac\text{-}ng} = 0.59352$; $\omega_{str} = 0.88556$	29	−1852.22042	0.213002	
M4: $\omega_{leafy} = \omega_{cor} = 0.27361$; $\omega_{mac\text{-}ng} = \omega_{str} = 0.73720$	27	−1852.39825	4.454520	*/—
M5: $\omega_{leafy} = \omega_{cor} = 0.27362$; $\omega_{mac\text{-}ng} = 0.59352$; $\omega_{str} = 0.88550$	28	−1852.29177	0.212952	
Houskeeping genes (*rps/l, infA, accD, clpP*, etc.)				
M0: $\omega_{leafy} = \omega_{cor} = \omega_{mac\text{-}ng} = \omega_{str} = 0.30179$	26	−25545.1250		
M1: $\omega_{leafy} = 0.29055$; $\omega_{cor} = \omega_{mac\text{-}ng} = \omega_{str} = 0.32343$	27	−25542.7794	4.691148	*/—
M2: $\omega_{leafy} = 0.29066$; $\omega_{cor} = 0.31584$; $\omega_{mac\text{-}ng} = \omega_{str} = 0.34192$	28	−25542.6470	0.264796	
M3: $\omega_{leafy} = 0.29065$; $\omega_{cor} = 0.31583$; $\omega_{mac\text{-}ng} = 0.33612$; $\omega_{str} = 0.34612$	29	−25542.6407	0.012646	
M4: $\omega_{leafy} = \omega_{cor} = 0.30267$; $\omega_{mac\text{-}ng} = \omega_{str} = 0.34230$	27	−25542.8978	4.454430	*/—
M5: $\omega_{leafy} = \omega_{cor} = 0.30269$; $\omega_{mac\text{-}ng} = 0.33612$; $\omega_{str} = 0.34668$	28	−25542.8908	0.014052	
ycf1* & *2				
M0: $\omega_{leafy} = \omega_{cor} = \omega_{mac\text{-}ng} = \omega_{str} = 0.66003$	26	−25384.8570		
M1: $\omega_{leafy} = 0.82314$; $\omega_{cor} = \omega_{mac\text{-}ng} = \omega_{str} = 0.61213$	27	−25383.1494	3.415136	
M2: $\omega_{leafy} = 0.82370$; $\omega_{cor} = 0.61073$; $\omega_{mac\text{-}ng} = \omega_{str} = 0.61543$	28	−25383.1485	0.001854	
M3: $\omega_{leafy} = 0.82338$; $\omega_{cor} = 0.61050$; $\omega_{mac\text{-}ng} = 0.59460$; $\omega_{str} = 0.63131$	29	−25383.1215	0.053994	
M4: $\omega_{leafy} = \omega_{cor} = 0.67331$; $\omega_{mac\text{-}ng} = \omega_{str} = 0.61464$	27	−25384.6635	0.386918	
M5: $\omega_{leafy} = \omega_{cor} = 0.67313$; $\omega_{mac\text{-}ng} = 0.59296$; $\omega_{str} = 0.62959$	28	−25384.6382	0.050592	
matK				
M0: $\omega_{leafy} = \omega_{cor} = \omega_{mac\text{-}ng} = \omega_{str} = 0.43481$	26	−3666.30553		
M1: $\omega_{leafy} = 0.44943$; $\omega_{cor} = \omega_{mac\text{-}ng} = \omega_{str} = 0.42324$	27	−3666.27784	0.055394	
M2: $\omega_{leafy} = 0.45060$; $\omega_{cor} = 0.38642$; $\omega_{mac\text{-}ng} = \omega_{str} = 0.53159$	28	−3665.93065	0.694374	
M3: $\omega_{leafy} = 0.45107$; $\omega_{cor} = 0.38419$; $\omega_{mac\text{-}ng} = 0.33356$; $\omega_{str} = 0.99284$	29	−3664.63804	2.585212	
M4: $\omega_{leafy} = \omega_{cor} = 0.41948$; $\omega_{mac\text{-}ng} = \omega_{str} = 0.53022$	27	−3666.08306	0.444942	
M5: $\omega_{leafy} = \omega_{cor} = 0.41857$; $\omega_{mac\text{-}ng} = 0.33352$; $\omega_{str} = 0.98601$	28	−3664.80458	2.556974	

NOTE.—Abbreviations for ω-ratios: ω_{leafy} = green, leafy outgroups (*Cymbidium, Oncidium, Phalaenopsis*); ω_{cor} = all *Corallorhiza*; $\omega_{mac\text{-}ng}$ = nongreen *Co. maculata* (i.e., *Co. mertensiana, Co. maculata maculata, Co. maculata occidentalis*); ω_{str} = *Co. striata*. Successively nested models were compared based on log-likelihood ratio tests (M1 vs. M0, M2 vs. M1, M3 vs. M2, M4 vs. M0, M5 vs. M4); all comparisons had 1 degree of freedom.

*$P < 0.05$, **$P < 0.01$, ***$P < 0.001$, "—" = not significant (where indicated), otherwise, blank space indicates no significance.

[a]The number of free parameters for a particular model.

[b]Log-likelihood of the data for a particular model.

[c]The chi-square distributed, log-likelihood ratio test statistic used to evaluate significant differences in model fit.

[d]P values for uncorrected/Bonferroni corrected χ^2 tests, where $P_{corr} = \alpha/\#$ branch parameters in the model being tested.

would cause purifying selection on the plastid- (and presumably nuclear-) encoded components of photosynthesis.

The more puzzling finding applies to nongreen taxa; nongreen coralroots had surprisingly consistent (albeit relatively minute) chlorophyll concentrations, with very little variation within or among species (fig. 1), despite having been sampled from different parts of their respective geographic ranges. Chlorophyll has been detected in other strictly heterotrophic plants (Cummings and Welschmeyer 1998), including members of Orobanchaceae and the monotropoid Ericaceae. The question naturally arises: Why would putatively nonphotosynthetic taxa continue to produce chlorophyll? There are several hypotheses as to why this might be the case, including photoprotection/avoidance of light damage to DNA and other cellular components, involvement of the chlorophyll biosynthesis machinery in other biochemical pathways such as plastid-nuclear signaling, or prevention of the buildup of chlorophyll precursors that may lead to increased oxidative damage (discussed in Wickett et al. 2011, and references therein). In nearly all cases studied, strictly parasitic species

(including the nongreen coralroots) that still synthesize detectable levels of chlorophyll do so at reduced levels. The fact that all nongreen coralroots contain relatively low chlorophyll concentrations means that these pigments are masked, and thus the overall coloration of these species tends to be red, purple, or brown, most likely due to anthocyanins occurring at higher concentrations (Freudenstein and Doyle 1994a; personal observation). Occasionally, yellow variants occur among populations of the various coralroot species (Freudenstein 1997; Barrett CF, personal observation). These are presumably recurrent anthocyanin-deficient mutants (yellow coloration may be due to persistence of chlorophylls and other yellow/orange carotenoid pigments), though no specific studies have been carried out to demonstrate the causes of this phenomenon in *Corallorhiza*. Regardless, reduced chlorophyll concentrations and the resulting apparent lack of visible green coloration in *Corallorhiza* are excellent indicators of relaxed selective constraints on photosynthetic machinery, and this is likely to be the case for many other parasite-containing taxa.

Characteristics of *Corallorhiza* Plastomes

With perhaps the exception of *Co. striata vreelandii*, plastomes among coralroot taxa do not altogether display major size reduction that would be expected for a group leafless parasites, and thus can be said to occupy the early stages along the path to a highly reduced or "minimal" plastid genome (Barbrook et al. 2006; Krause 2008). In a previous study, it was hypothesized that members of the *Co. striata* complex would have the most reduced plastomes in the genus in terms of overall size (Barrett and Davis 2012), but even the plastome of *Co. striata vreelandii* (the representative sequenced in that study, and included here) only represents an approximately 6% total genome size reduction relative to the leafy *Phalaenopsis* (Chang et al. 2006). In fact, *Co. macrantha* and *Co. maculata mexicana* (table 2) both have larger plastomes than two of the three leafy outgroup orchid taxa: *Phalaenopsis* (148,964 bp) and *Oncidium* (146,484 bp). Thus, the shift to leaflessness per se in *Corallorhiza* does not appear to be associated with a reduction in plastome size for the genus overall (at least for partially green taxa), although an examination of plastomes for the closest relatives of *Corallorhiza* (e.g., *Oreorchis*, *Aplectrum*, *Cremastra*, *Govenia*, *Calypso*) in tribe Calypsoeae will help clarify this.

A reduction in plastome size due to deletions and mutations resulting in pseudogenes is expected to be associated with a decrease in GC content, as once-functional (and relatively GC-rich) genes become riddled with loss-of-function mutations over time as a result of relaxed photosynthesis-associated selective pressures. These coding regions begin to resemble noncoding DNA, which typically has a lower GC content (e.g., Bernardi 1989; Glémin et al. 2014). This is apparently the case for *Co. striata vreelandii* (34.02% GC; table 2), which is 1.16% lower than the mean GC content for the remaining coralroot plastomes (35.28%). Mean GC content for the nongreen members of the *Co. maculata* complex is 35.10%, and no different than that found in the green coralroots. This suggests that large-scale differences in plastid GC content may take longer to manifest, and thus likely come about later in the process of plastid modification relative to other processes such as photosynthesis gene deletions and other initial gene mutations causing loss of function. This lends additional support to the notion that *Co. striata vreelandii* is farther "down the path" of plastid genome degradation relative to the nongreen members of the *Co. maculata* complex.

More broadly, the leafy orchids *Phalaenopsis aphrodite* and the hybrid *Oncidium* Gower Ramsey have GC contents of 35.6% and 37.32%, respectively (Chang et al. 2006; Wu et al. 2010); these values are very similar to what is found in *Co. trifida* and other green coralroots, thus green coralroots do not seem to deviate much from other leafy orchids. Members of holo- and hemiparasitic Orobanchaceae (Wicke et al. 2013) represent an interesting clade for comparison with *Corallorhiza* in terms of GC content. Orobanchaceae range from 38.08% GC in the hemiparasitic *Schwalbea* to 31.09% in the holoparasitic *Phelipanche purpurea*. Variation in GC content among plastomes of *Corallorhiza* lies completely within the bounds of that observed in Orobanchaceae, which is hypothesized to be in the relatively more advanced stages of degradation.

Structural Evolution of the Plastome: Pseudogenes and Losses

ndh Genes

Degradation of the plastid-encoded *ndh* complex is advanced in *Corallorhiza*, as has been observed in other groups (e.g., Chang et al. 2006; Funk et al. 2007; McNeal et al. 2007; Braukmann et al. 2009; Blazier et al. 2011; Braukmann and Stefanovic 2012; Braukmann et al. 2013; Iles et al. 2013; Peredo et al. 2013; Wicke et al. 2013), including some of the leafy, photosynthetic orchid plastomes included in figure 3. Interestingly, there is variation among the orchid plastomes sequenced to date with respect to the status of *ndh* genes (Chang et al. 2006; Wu et al. 2010; Delannoy et al. 2011; Barrett and Davis 2012; Logacheva et al. 2011; Yang et al. 2013), suggesting multiple independent degradation pathways in this complex in the orchid family.

The ancestor of *Corallorhiza* likely already experienced some degree of degradation in the *ndh* complex, as evidenced by several shared pseudogenes. Sequencing of additional orchid plastomes could reveal an evolutionary transition from a functional, plastid-encoded *ndh* complex in leafy relatives (e.g., *Aplectrum*, *Oreorchis*, *Cremastra*, *Govenia*), to a largely degraded state, as observed in *Corallorhiza*. However, there is even some variation in the degree of degradation among certain members of *ndh* within *Corallorhiza*, which in part, reflects the plastid phylogeny (fig. 2). This pattern includes a major deletion in *ndhE* in all taxa but *Co. striata vreelandii*, a shared pseudogene for *ndhJ* for all members of the *Co. maculata* complex besides *Co. bulbosa* (which here is sister to the remaining members of the *Co. maculata* complex; fig. 2), a loss of the same gene in (*Co. odontorhiza*, *Co. wisteriana*), and hypothesized parallel deletions of the *ndhG*

pseudogene in *Co. bulbosa* and the clade of (*Co. macrantha*, *Co. maculata mexicana*).

The plastid-encoded *ndh* complex is believed to be dispensable, for a number of reasons (reviewed in Martín and Sabater 2010; Wicke et al. 2011, 2013), which may differ for various groups of plants displaying evidence of degradation in this gene complex. Generally, the *ndh* complex is believed to be essential for photosynthetic function in environments with highly variable light intensities by functioning in regulation of cyclic electron transport, and also in mitigating the effects of photo-oxidative damage under high light intensities (Martín et al. 2009; Martín and Sabater 2010). In the ancestor of *Corallorhiza*, selective constraints on the *ndh* complex may have been lifted by a combination of decreased dependence upon photosynthetic carbon as a result of a shift to heavy dependence upon fungal carbon (a similar situation is also observed in carnivorous plants using animals as a carbon source in the family Lentibulariaceae; Wicke et al. 2014), and a tendency to grow in dark, shaded, mature forests, where light intensity is always low. This may also occur in other orchid lineages (including leafy, photosynthetic orchids), due to high dependence on fungal carbon as seedlings, and also the tendency to live in habitats with relatively low or at least static amounts of light intensity (e.g., a shaded rainforest understory or floor).

Photosynthesis-Related Genes (excluding ndh, rpo, *and* atp)

Nongreen members of *Corallorhiza*—that is, those with relatively reduced levels of chlrorophyll (fig. 1)—all display evidence of degradation of genes directly involved in photosynthetic processes (fig. 2; e.g., *psa/b, pet, rbcL*). Of these taxa, *Co. striata vreelandi* displays the highest number of degraded genes (table 2; figs. 2 and 3), as reported in Barrett and Davis (2012), followed by northern North American members of the *Co. maculata* complex (*Co. mertensiana, Co. maculata maculata, Co. maculata occidentalis*). Both of these complexes (*Co. striata, Co. maculata*) have independently undergone degradation among genes in the following genes/gene complexes (omitting *ndh*, see above): *pet, psa/b, rpo, ccsA, cemA*, and *rbcL*. It is highly unlikely that transcript editing of any kind can compensate for these changes, and it can be assumed that these changes are irreversible. The possibility exists, however slightly, that all deleted genes or pseudogenes among the plastomes of these taxa are redundant and that functional copies are encoded in the nuclear genome, or that genes have been "regained" through intergenomic or horizontal transfer (e.g., Iorizzo et al. 2012; Straub et al. 2013; Wicke et al. 2013).

Based on the degraded states of photosynthesis-related genes among these plastomes, it can be concluded that the aforementioned nongreen taxa are strict parasites of fungi, incapable of photosynthesis, and thus the terms "holomycotroph" or "full mycoheterotroph" are indeed appropriate. No studies of photosynthetic activity have been carried out in *Corallorhiza* aside from those in the "greenest" coralroot, *Co. trifida* (Zimmer et al. 2008; Cameron et al. 2009). It will be beneficial to measure photosynthetic rates (or determine the lack thereof) along with plastid-gene transcript levels among multiple individuals of both green and nongreen coralroot taxa, preferably growing in close proximity to one another and in similar habitats, to definitively characterize photosynthetic capacity in the genus.

Though green members of *Corallorhiza* do not display significant degradation of photosynthesis-related genes (*psa, psb, pet, rbcL*, etc.), there are a few important exceptions, which may or may not affect the efficacy of photosynthesis in these taxa. First, members of *Corallorhiza* excluding *Co. trifida* and *Co. striata* have experienced a large deletion in *psbM* (PSII protein M). This low molecular weight peptide is believed to play a role in maintaining the stability of PSII dimers; *psbM*-deficient tobacco mutants displayed impaired stability of PSII, but were still able to function (Umate et al. 2007; Kawakami et al. 2011). Photosynthesis has only been explicitly measured and demonstrated in *Co. trifida* (Zimmer et al. 2008; Cameron et al. 2009), which has a putatively functional copy of *psbM*. Thus, it is possible that either 1) other "green" coralroots (*Co. odontorhiza, Co. wisteriana, Co. bulbosa, Co. macrantha, Co. maculata mexicana*) carry out inhibited photosynthesis, 2) this gene has been transferred to the nucleus or mitochondrion and its product is imported into the plastid, allowing stabilization of the PSII dimer complex and some level of photosynthetic function, or 3) a nonfunctional *psbM* gene has no effect on photosynthetic function.

Another partially green coralroot, *Co. macrantha* (restricted to high elevation forests in southern Mexico), has a pseudogenized copy of *psaI* (Photosystem I Reaction Center Subunit VIII) due to a 4-bp insertion near the 5′-end of the gene relative to other coralroots, causing a reading frame shift. The psaI protein is believed to function in the stabilization of a nuclear-encoded subunit of the same complex (psaL) in *Arabidopsis thaliana*, and has also been shown to bind to the Light-harvesting Complex II (Jensen et al. 2007); it was suggested that the latter putative function is redundant. In *Corallorhiza*, this might explain why *psaI* in *Co. macrantha* has apparently become a pseudogene. Thus, among the green coralroots, *Co. macrantha* may represent yet another, very early step in the transition to the loss of photosynthesis, or at least a reduction in photosynthetic efficacy, but this remains a hypothesis to be tested.

rpo Genes

There is evidence for *rpo* pseudogenes in both the nongreen *Co. striata* and *Co. maculata* complexes, based on large deletions, reading frame shifts, and internal stop codons. Plastid genes are transcribed by two polymerases in monocots: A nuclear-encoded polymerase and a plastid-encoded polymerase (Liere et al. 2011). The plastid-encoded RNA polymerase is believed to preferentially transcribe photosynthesis-related genes, whereas the *rpoB-C1-C2* operon itself is transcribed by a nuclear-encoded RNA polymerase (Hajdukiewicz et al. 1997; reviewed in Liere et al. 2011; Krause 2012). In addition to transcribing photosynthesis-related genes for the most part, PEP also tends to be more active in later leaf developmental stages (Hajdukiewicz et al. 1997; reviewed in Liere et al. 2011).

Based on this latter finding, it might be expected that all members of *Corallorhiza* should show some degradation in *rpo* genes, as leaves do not fully develop in any species within this genus (i.e., only the basal leaf sheath persists, but the blade, or lamina, never develops), presumably obviating the role of PEP in transcription of photosynthesis-related genes (e.g., *psa, psb*) in leaves. However, green coralroots have intact reading frames for all *rpo* genes, whereas nongreen coralroots show varying degrees of degradation in the *rpo* complex. As with photosynthesis-related genes in *Corallorhiza* (figs 2 and 3), it does not appear to be the transition to leaflessness per se that has resulted in degradation of the *rpo* complex, but more likely the shift from partial heterotrophy (and the associated need to carry out photosynthesis, whatever small amount) to strict heterotrophy. An interesting parallel occurs in the carnivorous family Lentibulariaceae, in which *rpo* genes display significant departures from purifying selection (Wicke et al. 2014). An interesting follow-up would be to determine whether degradation first begins among photosystem genes (*psa, psb*), diminishing the importance of having a functional PEP complex, or whether the opposite is the case, in which a dysfunctional (or nonexistent) PEP has contributed to relaxed selective pressures associated with photosynthesis-related genes (i.e., if *psa* and *psb* genes can no longer be transcribed, or are done so by an inefficient PEP, then is there any benefit to maintaining them?). Regardless, the degradation of *rpo* genes in nongreen *Corallorhiza* closely mirrors that in genes directly involved in photosynthesis (*psa/b*, *rbcL*, etc.).

Relationships in *Corallorhiza*

Complete plastome sequences of *Corallorhiza* yielded a highly resolved, highly supported tree, consistent with previous studies (Freudenstein and Doyle 1994a, 1994b; Barrett and Freudenstein 2008; Freudenstein and Senyo 2008). Here for the first time, relationships among the six taxa that comprise the *Co. maculata* complex (as recognized in Freudenstein 1997) are fully resolved, with high support for *Co. bulbosa* as sister to the rest of the complex (bootstrap = 91%), whereas previous studies could not resolve relationships in this complex (Barrett and Freudenstein 2008; Freudenstein and Senyo 2008). Complete plastomes also confirm the sister relationship of *Co. striata* to the rest of *Corallorhiza*, with *Co. trifida* (moving away from the base of the tree), then *Co. odontorhiza* and *Co. wisteriana* diverging, successively. Depending on the data set analyzed, *Co. bulbosa* appears in a few different places, most likely due to homoplasy and few variable characters available to resolve the position of this species in smaller data sets, as evidenced by short internal branches (supplementary fig. S3, Supplementary Material online). Thus, inclusion of complete plastomes (pseudogenes, spacers, and introns) resolves the otherwise uncertain position of this species. Nuclear rDNA, on the other hand, places *Co. bulbosa* as sister to (*Co. macrantha, Co. maculata mexicana*) with moderate support, a similar finding to the combined analysis of Barrett and Freudenstein (2008) based on plastid *rbcL* + nuclear ITS, though that study did not include *Co. macrantha*. Combining all nuclear and plastid data places *Co. bulbosa* as sister to the remaining members of the *Co. maculata* complex, with 100% bootstrap support (supplementary fig. S3, Supplementary Material online). The complete plastome + rDNA data set fully resolves relationships among the major species complexes of *Corallorhiza* (excluding members not sampled from some complexes).

In particular, relationships within the *Co. maculata* complex shed light on a phylogeographically intriguing aspect of plastome degradation. All Mexican members of the *Co. maculata* complex have at least some visible green tissue and are presumed to be partially mycoheterotrophic, based on conservation of plastid-encoded photosynthetic apparatus (for the most part, see fig. 2). Those members of the *Co. maculata* complex that occur in northern North America have no visible green tissue (two of these species were shown to have highly reduced chlorophyll levels; fig. 1) and display evidence of degradation in these genes. The topology of the *Co. maculata* complex based on combined plastome + rDNA data (supplementary fig. S3, Supplementary Material online, fig. 2) lends support to a previous hypothesis that this complex (and possibly the genus *Corallorhiza* itself) originated in southern Mexico (Freudenstein and Doyle 1994a, 1994b; Barrett and Freudenstein 2008; Freudenstein and Senyo 2008), and expanded into northern North America, with the ancestor of *Co. mertensiana*, *Co. maculata maculata*, and *Co. maculata occidentalis* undergoing a shift from partial to strict heterotrophy (and continuing this trend among extant lineages), and thus presumably losing the ability to carry out photosynthesis. This suggests that deeper sampling among individuals in members of both the *Co. maculata* complex and the *Co. striata* complex (which is here represented only by *Co. striata vreelandii*) will yield a much clearer picture of plastome degradation in *Corallorhiza*, on a phylogeographic scale.

Selective Regime among Genes with Open Reading Frames

Model M4, allowing green and nongreen taxa to have different ω-ratios ($\omega_{leafy} = \omega_{cor}$; $\omega_{mac\text{-}ng} = \omega_{str}$), fit the data significantly better than the one-ratio model for both housekeeping genes and photosynthesis-related genes, suggesting different selective scenarios for nongreen taxa relative to green taxa, but these comparisons are nonsignificant when corrected for multiple comparisons (table 3). For housekeeping genes, ω-values for green versus nongreen taxa did not differ by much ($\omega_{leafy} = \omega_{cor} = 0.30267$; $\omega_{mac\text{-}ng} = \omega_{str} = 0.34230$; table 3). However, this difference is much more pronounced for photosynthesis-related genes ($\omega_{leafy} = \omega_{cor} = 0.27361$; $\omega_{mac\text{-}ng} = \omega_{str} = 0.73720$), with the ω estimate for nongreen taxa being approximately 2.6× that of green taxa, and approaching a ω value expected under neutral evolution ($\omega \sim 1$). Thus, it is evident that even though some photosynthesis genes in nongreen taxa still have intact reading frames, there may be a relaxation of selective pressure associated with these genes.

For *atp* genes, Models M1 and M4 fit the data significantly better than does M0, even after correcting for multiple comparisons, suggesting different selective regimes in *Corallorhiza*

versus leafy, green outgroups, or at least in nongreen *Corallorhiza* versus green *Corallorhiza* and leafy outgroups. For example, in model M1, the ω value for *Corallorhiza* is nearly twice that for leafy orchids ($\omega_{leafy} = 0.15175$; $\omega_{cor} = \omega_{mac\text{-}ng} = \omega_{str} = 0.26949$), and in model M4, nongreen *Corallorhiza* have ω values over twice those for green *Corallorhiza* and leafy orchids. Further investigations using branch-site models, allowing some sites in the "foreground" branches to be under positive selection, suggest evidence for positive selection in nongreen *Corallorhiza* for the *atp* complex.

The plastid-encoded *atp* genes play a critical role in photosynthesis by both generating ATP and translocating protons across the Thylakoid Membrane (Evron et al. 2000; McCarty et al. 2000; Allen et al. 2011). However, in taxa that display evidence of loss-of-function for photosystem and other photosynthesis-related genes, the fact that *atp* genes are preserved as intact reading frames suggests that they may play additional roles in plastid function (discussed in Wicke et al. 2013). This is bolstered by the fact that this complex shows evidence of positive selection in strictly heterotrophic, nongreen taxa, but not for *Corallorhiza* as a whole. This would support the hypothesis that the *atp* complex has either taken on additional, as-of-yet unknown roles in holomycotrophic plastids, or at a minimum has fallen under increased selective pressure to adapt to the overall "alternative lifestyle" associated with strict fungal heterotrophy in nonphotosynthetic plastids (including in amyloplasts, chromoplasts, elaioplasts, proteinoplasts, etc.). A future research objective is to examine the potential biological roles and molecular evolutionary patterns of the *atp* complex in nonphotosynthetic plants and their green, photosynthetic relatives, as members of this gene complex are often preserved and putatively functional among various heterotrophic angiosperms (see fig. 3 and references therein, including discussion in Wicke et al. 2013).

Comparison of *Corallorhiza* with Other Parasitic Angiosperms, Using the Model of Barrett and Davis (2012)

A comparison of the plastomes of several heterotrophic taxa that have been sequenced to date shows that *Corallorhiza* is not quite in the "advanced" stages of plastid genome degradation, relative to several taxa in Orobanchaceae (e.g., Wicke et al. 2013), or other orchids such as *Neottia* (Logacheva et al. 2012) and *Rhizanthella* (Delannoy et al. 2011) (fig. 3). Interpreted in the context of the basic model presented in Barrett and Davis (2012), members of *Corallorhiza* fall into categories 1 and 3: Taxa with some visible green tissue display pseudogenes/deletions mainly in the *ndh* complex (category 1; gene-class categories are listed at the top of fig. 3), with a few additional changes, whereas those with no green tissue have experienced pseudogenization and deletions in *ndh*, photosynthesis-related, and *rpo* genes, but not in *atp* genes or other "housekeeping" genes (category 3, excepting the loss of a redundant $trnT^{GGU}$ in *Co. striata vreelandii*; Barrett and Davis 2012; figs. 2 and 3). Thus, nonparasitic autotrophs and partial heterotrophs roughly fall either into category 0 (no degradation of plastid genes) or category 1 (degradation of the *ndh* complex only). More specifically, all partial heterotrophs fall into category 1, with evidence of degradation of the *ndh* complex.

Strict heterotrophs fall into more "advanced" categories. Several, but not all, members of Orobanchaceae and two orchids (*Neottia* and *Rhizanthella*) can be classified in category 5, in which there is evidence of degradation of *atp* genes and at least some evidence of degradation in "housekeeping" genes (e.g., *rpl*, *rps*, *trn*, etc.). There are some genes for which no taxa have experienced pseudogenization or loss, including: *rpl2*, *rpl16*, *rpl20*, *rps2*, *rps4*, *rps8*, *rps11*, *ycf1*, *ycf2*, *trnD-guc*, *trnE-uuc*, *trnI-cau*, *trnM-cau*, *trnQ-uug*, *trnY-gua*, *rrn16*, *rrn23*, *rrn4.5*, and *rrn5*. This suggests that they may be indispensable for plastid function at least in the angiosperms included in this comparison, and suggests that a "minimal set" of genes might be necessary for functional plastid "housekeeping" and perhaps other functions outside of photosynthesis and ATP production.

Overall, no taxa fell in categories 2 or 4, suggesting that degradation of the *rpo* genes occurs more or less simultaneously with degradation in photosynthesis-related genes, and that degradation of *atp* genes occurs along with that in "housekeeping" genes (also discussed in Wicke et al. 2013). Thus, the model of Barrett and Davis (2012) could be modified, in light of more recently sequenced plastomes including those of *Corallorhiza*, having five general categories: 1) no degradation, 2) degradation in the *ndh* complex only, 3) degradation of the *ndh* complex and photosynthesis-related genes including *rpo* genes, and 4) degradation in all plastid gene systems, including "housekeeping" and *atp* genes, and 5) complete or nearly complete loss of the plastid genome (as is hypothesized to have occurred in some strict heterotrophs such as *Corynaea* and *Rafflesia* [Nickrent et al. 1997; Molina et al. 2014]).

There is some degree of deviation from the Barrett and Davis (2012) model, as might be expected given the potential for somewhat idiosynchratic patterns of plastome evolution in unrelated groups. The obvious exception to this model would be the sequenced members of the genus *Cuscuta*, in the family Convolvulaceae. One of the best-studied parasitic systems in terms of plastome evolution, *Cuscuta* contains members with "extreme" plastome reduction relative to closely related leafy species (approximately 50% total size reduction in *Co. gronovii* and *Co. obtusiflora* relative to the leafy *Ipomoea*) (fig. 3 and references therein). Yet, both highly reduced species of *Cuscuta* retain most if not all of the genes for photosynthesis, but have degraded RNA polymerase complexes and also show evidence of pseudogenes/deletions of some housekeeping genes. A further survey based on probe-hybridization in *Cuscuta* with greatly increased taxon sampling by Braukmann et al. (2013) suggests that this genus encapsulates multiple transitions to highly degraded plastid genomes, with members of subgenus *Grammica* displaying even more advanced stages of degradation. These taxa were not included in figure 3, as they are based on hybridization, and have not yet been sequenced to generate complete

plastomes. Recycling of respiratory CO_2 has been offered as an explanation for the retention of putatively functional genes in the *psa*, *psb*, and *pet* complexes (McNeal et al. 2007), underscoring the importance of studying the physiology of strict heterotrophs on a taxon-by-taxon basis, and also highlighting how little is known about the basic physiology and functional genomics of these plants.

Conclusion

All members of *Corallorhiza* produce detectable levels of chlorophyll based on the current sample, but nongreen, strictly heterotrophic members have on average tenfold lower concentrations relative to green, partially heterotrophic members. Genomic data allowed assembly of complete plastomes as well as partial rDNA operons, which when combined, fully resolve relationships among the major species complexes in *Corallorhiza*. Based on complete plastomes, *Corallorhiza* can be said to span the "early" stages of plastid genome degradation as a result of relaxed selective pressures on photosynthetic function. *Corallorhiza* is useful in providing a model system to understand the earliest stages in the transition to a "minimal plastome," as are observed in other heterotrophic plants (e.g., Orobanchaceae, other orchids like *Rhizanthella*). There have been at least two independent transitions to strict heterotrophy from partial heterotrophy in the genus (*Co. maculata* and *Co. striata* complexes). Nongreen coralroots also display evidence of relaxed selective pressures on photosynthesis-related genes which have retained open reading frames. Members of the *atp* complex display relaxed purifying selection among *Corallorhiza* as a whole, but more specifically display evidence for positive selection for nongreen, strictly heterotrophic taxa, although for reasons currently unknown.

Generally, there is a clear path of plastid genome degradation in the heterotrophic angiosperm lineages studied thus far (with some deviations from the overall model, e.g., some *Cuscuta*). Moreover, it can be hypothesized that there is a stepwise progression to a "minimal" plastome, and *Corallorhiza* occupies categories 2 (partially heterotrophic taxa, which display degradation of the *ndh* complex only, and a few other pseudogenes) and 3 (strictly heterotrophic taxa; including degradation of the *ndh* complex, photosynthesis-related genes, and *rpo* genes [excluding *atp* genes]) in a modified model of the degree of plastome degradation among angiosperms. Future studies will include 1) deeper sampling of complete plastomes within species complexes of *Corallorhiza* to gain a phylogeographic perspective on plastome degradation, 2) direct assessment of the relative photosynthetic capabilities of members of each coralroot taxon in the field, and 3) quantification of transcript levels for plastid-encoded genes within each species, to gain a clearer understanding of gene expression, and to test the hypothesis of the existence of duplicate, functional copies of pseudogenized or deleted plastid genes.

Materials and Methods

Plant Material, DNA Isolation, Library Preparation, and PE Sequencing

Plant material was collected and preserved either in silica gel or frozen at −80 °C. Total genomic DNAs were extracted using the cetyltrimethylammonium bromide (CTAB) method (Doyle JJ and Doyle JL 1987) from 1.0 g frozen or 0.2–0.5 g silica-dried material, and quantified using a NanoDrop2000 spectrophotometer (Thermo Scientific, Waltham, MA). Samples with total DNA concentrations $> 100\,ng/\mu l$ were selected and run on a 1.5% agarose gel to check for high molecular weight, nondegraded DNA. Samples were shipped on ice to Cold Spring Harbor Laboratory (CSHL, Woodbury, NY), where they were sheared to approximately 450 bp, followed by library preparation and barcoding for PE Illumina sequencing. The exception was *Co. odontorhiza*, for which library preparation and sequencing were completed at the University of Missouri–Columbia using the Illumina TruSeq Library Preparation Kit (Illumina, Inc., San Diego, CA); *Co. odontorhiza* DNAs were sheared to approximately 320 bp. All samples were run on Illumina HiSeq 2000 machines to generate 100-bp PE reads. Indexed samples were pooled and run with accessions from other projects, with a total of 40 or 20 accessions in a single lane on the CSHL and University of Missouri machines, respectively. Voucher information and characteristics of each of the resulting data sets and completed plastomes are listed in table 1.

Quantification of Chlorophyll Content

Approximately 0.1 g of −80 °C frozen tissue was harvested from the outer wall of the developing ovary and weighed for chlorophyll quantification. Ovary tissue was chosen: 1) because this is the only organ that is visibly green in some taxa (e.g., *Co. odontorhiza*, *Co. wisteriana*) and 2) to be consistent between taxa. Chlorophylls were extracted in 80% ice-cold acetone, stored at −80 °C, and chlorophylls *a* and *b* were quantified on a Cary 4000 UV-VIS spectrophotometer (Agilent Technologies, Inc., Santa Clara, CA; Department of Plant Cellular & Molecular Biology, Ohio State University) and their concentrations were calculated using formulas of Porra (2002) (Lima D, personal communication). Data from chlorophylls a and b were pooled to give total chlorophylls. Although individual accessions used to quantify chlorophyll concentration among species were not the same as those used for sequencing, multiple individuals from across each species' geographic range were included to account for variation in chlorophyll content among individuals within species. *Corallorhiza bentleyi* is included here (an endangered member of the *Co. striata* complex), but was not sequenced. Conversely, Mexican-endemic members of the *Co. maculata* complex (*Co. macrantha*, *Co. bulbosa*, *Co. maculata* var. *mexicana*) were not included in the chlorophyll analysis due to rarity and thus lack of fresh/frozen material. Voucher specimens for plants used for chlorophyll analysis are listed in supplementary figure S5, Supplementary Material online.

Data Processing

The resulting paired, FASTQ data were processed using a custom UNIX script ("process_fastq.sh") that calls several freely available Perl scripts for the purposes of 1) quality filtering/trimming and adaptor removal ("IlluQC_PRLL.pl," "TrimmingReads.pl," Patel and Jain 2012), 2) "shuffling" PE files into a single file ("ShuffleSequences_fastq.pl," part of the VELVET suite of tools [E. Cabot, University of Wisconsin-Madison; Zerbino and Birney 2008]), and 3) conversion to FASTA format ("fastq2fasta.pl," B. Knaus 2009 [http://brianknaus.com/software/srtoolbox/, last accessed May 1, 2014]. Reads were trimmed from their 3′-ends discarding bases under a quality threshold of PHRED score of 20 (corresponding to an error rate of 1/100 or greater), and also discarding trimmed reads below 61 bp in length.

Read Mapping

PE reads were then mapped five times iteratively in Geneious v.6 (created by Biomatters, Inc., Auckland, New Zealand; http://www.geneious.com, last accessed December 1, 2013) using automatically determined settings to a combined file of 1) the *Phalaenopsis aphrodite* plastome (GenBank accession number NC_007499; Chang et al. 2006), 2) the *Phoenix dactylifera* mitochondrial genome (NC_016740; Fang et al. 2012), and 3) the *Asclepias syriaca* nuclear ribosomal DNA consensus operon (JF312046; Straub et al. 2011). The purpose of read mapping was to filter or "screen" reads that matched the reference genomes, and to get initial, crude estimates of coverage to be used in downstream de novo assemblies.

Plastome Assembly

Mean coverage estimates from Geneious were used as input for de novo plastome assembly ("-exp_cov" parameter). De novo assembly was completed using both the VELVET v.1.210 assembler (Zerbino and Birney 2008) and the Geneious v.6 de novo assembler. VELVET assemblies were generated using a variety of hash lengths (i.e., "word match," or "kmer length") ranging from 51 to 81 for each complete data set, with specified expected coverage estimates taken from Geneious mapping (specified to be in approximate kmer coverage; see Zerbino and Birney 2008). Coverage cutoff values were input as roughly one-half of the specified expected coverage estimates, minimum contig length was set to 300 bp, and insert size set to 450 bp (320 bp for *Co. odontorhiza*), and the resulting contigs were assembled in Sequencher v.6 (GeneCodes, Ann Arbor, MI). The resulting "draft de novo assembly" from VELVET was used as a reference to map reads in Geneious, as described above (five iterations, automatically determined settings). Read mappings were visually scanned in Geneious to check for areas of low coverage or potential misassemblies. Read pairs that mapped successfully to the draft reference were used in a second round of de novo assembly, this time in Geneious with automatically determined settings, saving 100 consensus contigs. This method allowed construction of complete draft plastomes with relative ease.

Contigs from both assemblies (Geneious, Velvet) were merged in Sequencher to form complete plastomes (parameters = "dirty data" or "large gap," minimum overlap = 20 bp, minimum similarity 70–90%). Any apparent "gaps" among contigs resulting from one software platform were usually covered by a contig from the other, allowing assembly of completed plastomes (Large Single Copy–Inverted Repeat B–Small Single Copy). In a few cases, VELVET scaffolds contained strings of Ns between 30 and 100 bp, mostly associated with homopolymer repeats ($>$ 10 bp). These were resolved by 1) the Geneious de novo assembly (typically) and 2) UNIX "grep" searches with flanking sequences of 30–50 bp used as search terms taken from each end of the string of Ns. Grep results were converted to FASTA format using a custom Perl script, to build a "bridge" of reads. This was done iteratively until the region in question was resolved, although most cases were resolved with a single iteration. Grep searches were applied to all cases of ambiguity between contigs from the two de novo assembly methods, as a quality control. Polymerase chain reaction amplification and Sanger sequencing were not necessary due to high read coverage (table 1).

IR boundaries for each draft plastome were determined by using the first and last approximately 20 bp of the draft plastome as search terms (reverse complement) in Sequencher to find matches, and marking their match points as features. Then, the first approximately 700 bp and last approximately 1–2,000 bp of the draft plastome were copied, reverse complemented, and manually aligned to the IR using the marked features from above as reference points. IR boundaries were determined informatically using PE information (i.e., read pairs that spanned the IR boundaries). Because the orientation of the ends of the LSC and SSC can vary, the plastomes were often in different configurations; single copy regions were reoriented to correspond to the "standard" model of gene order to simplify downstream comparison. Special attention was paid to the possibility of significant modification or even loss of the IR (either partial or complete), and also to the possibility of other genomic rearrangements or structural modifications. In other words, when building draft plastomes, it was not assumed that all accessions displayed the "standard" plastome structure found in angiosperms (i.e., LSC, IRb, SSC, IRa).

Plastome Annotation

Plastid genomes were annotated by first aligning the set of coding loci (coding sequences [CDS]) containing introns and all tRNAs from *Phalaenopsis* to each draft plastome (minus IRa) in Sequencher, to aid in intron/exon and tRNA boundary determination. Annotations were conducted in DOGMA (Wyman et al. 2004), which uses BLASTX searches against a database of annotated plastomes and amino acid translations, and allows manual determination of start/stop codons. DOGMA BLASTX parameters were as follows: genetic code = plant plastid, *e* value = 5, percent cutoff for protein coding genes = 55, and percent cutoff for tRNAs = 80. The resulting GenBank feature table was exported from the DOGMA server and imported along with the finished plastid FASTA file into SEQUIN (National Center for Biotechnology Information), where the annotation was validated and completed for

submission to GenBank. A large inversion in *Co. maculata* var. *maculata* was manually reversed for the purposes of phylogenetic analyses and other whole-plastome comparisons, but the plastome for that accession was annotated with the inversion as assembled de novo. Plastome maps were drawn with GenomeVx (Conant and Wolfe 2008) and manually edited in Adobe Illustrator CS6 (Adobe, Inc., San Jose, CA).

Ribosomal DNA Assembly and Annotation

Nuclear ribosomal DNA reads were filtered by mapping reads from *Co. odontorhiza* to a nearly complete rDNA operon reference (partial ETS [External Transcribed Spacer], 18 S rDNA, ITS1 [Internal Transcribed Spacer], 5.8 S rDNA, ITS2, 26 S rDNA) from *A. syriaca*, described above, as well as mapping to an ITS-spanning sequence from *Co. odontorhiza* from GenBank (accession number EU391326). Other regions were not assembled (e.g., 5 S rDNA). Resulting reads that mapped to each reference were assembled de novo in Geneious, and resulting contigs were merged through overlap in Sequencher. This draft assembly was used for a second round of read mapping (original *Co. odontorhiza* reads) and de novo assembly in Geneious, resulting in a single consensus contig. The consensus contig was based on 50% majority per site, bearing in mind that the rDNA operon exists in numerous-repeated copies that often differ from one another, appearing as polymorphism in read pileups. The *Co. odontorhiza* rDNA consensus operon was then used as a reference in read mapping for all other *Corallorhiza* species, followed by individual de novo assemblies, as above.

Nearly complete ribosomal DNA operons (partial ETS, 18 S rDNA, ITS1, 5.8 S, ITS2, 26 S rDNA; excluding 5 S rDNA) were annotated by first aligning finished contigs in Geneious (using the MUSCLE v.3.8.31 plugin in Geneious; Edgar 2004), and transferring the annotation features from *Co. odontorhiza* to the aligned *Corallorhiza* sequences. Then, the entire alignment was annotated in SEQUIN for submission to GenBank, by adding each feature to *Co. odontorhiza*, and subsequently propagating the features to the remaining sequences. Reads were again mapped in Geneious to each finished, annotated rDNA cistron to check for misassemblies. GenBank accession numbers for rDNA partial operons are: KM390003-KM390012. Available outgroup sequences for *Cymbidium*, *Oncidium*, and *Phalaenopsis* were mined from GenBank for the purposes of downstream phylogenetic analyses (18 S rDNA—*Cy. farberi*, GenBank accession number JN418934; ITS—*Cy. aloifolium* JF729014, *Phalaenopsis aphrodite* AY391543, *O. andradeanum* FJ565598).

Data Matrix Assembly and Sequence Alignment

Annotated plastomes were imported into Geneious with subsequent removal of the second copy of the IR (IRa), and aligned with MAFFT v.7.157 (Katoh et al. 2002) using the "fast-progressive" method (gap opening penalty = 6; offset = 0.1), followed by manual adjustment. The alignment also included complete plastomes of *Cy. aloifolium* (GenBank accession number NC_021429), *Oncidium* Gower Ramsey (GenBank accession number NC_014056), and *Phalaenopsis aphrodite* (GenBank accession number NC_017049). A short inversion of the *petN-psbM* region in *Cymbidium* was removed for that accession in downstream analyses. Annotated CDS were exported from Geneious and data from each locus were saved as separate FASTA files. Each file was submitted for codon-based alignment to the MACSE web server (Ranwez et al. 2011; http://mbb.univ-montp2.fr/MBB/subsection/softExec.php?soft=macse, last accessed May 30, 2014), which allows alignment of both intact open reading frames and incomplete sequences (including pseudogenes with modified reading frames). Alignments were adjusted manually when necessary and imported back into Geneious or SequenceMatrix (Vaidya et al. 2011), where they were concatenated into larger matrices. Protein-coding genes that contained pseudogenes (shifted reading frames, internal stop codons) in at least one taxon were excluded. Data matrices are deposited in supplementary fig. S6, Supplementary Material online.

Phylogenetic Analyses

Phylogenetic analyses were conducted using both Parsimony and ML. Parsimony searches were conducted in TNT (Goloboff et al. 2003, 2008) with 100 random addition starting sequences, saving 100 trees per replicate, and using tree bisection–reconnection branch swapping on the pool of saved trees. Support was assessed through 2,000 Jackknife pseudoreplicates, with the above search parameters. Model fit for ML analyses was assessed for the complete plastome alignment in MEGA v.6 (Tamura et al. 2013), which yielded GTR+Γ as the best-fit model, based on the corrected Akaike Information Criterion (AICc) metric (Burnham and Anderson 2002). All phylogenetic analyses were conducted in RAxML v.8.0.20 (Stamatakis 2006). Three separate searches were initiated for each matrix analyzed from different random starting seeds using an unpartitioned GTR+ Γ model to ensure convergence on the same topology. A codon-partitioned model was not chosen due to the possibility of relaxed functional constraints among some coding regions for some taxa, which could possibly represent model violations. Support was assessed with 1,000 standard bootstrap pseudoreplicates. *Cymbidium* was chosen as the outgroup taxon.

Plastid Pseudogenes, Genomic Deletions, and Genomic Attributes

The presence of pseudogenes and whole gene/regional deletions were initially assessed visually in DOGMA by identifying interrupted BLASTX reading frames. All coding loci aligned in MACSE (see above) were imported into Geneious, and sequences were translated to assess pseudogene status through the presence of frame shifts and/or premature stop codons. Loci with interrupted reading frames were considered to be pseudogenes, unless there was an alternative start or stop codon within three codon positions of the other taxa in the alignment. In addition, the alignment of whole annotated plastomes (described above) was used to visualize any larger genomic deletions spanning CDS boundaries, or rearrangements.

Hypotheses of differences in plastid genomic attributes among nongreen versus green taxa, and correlations between chlorophyll content and genomic attributes were assessed among taxa in using two phylogeny-based analytical approaches. First, a Phylogenetic Analysis of Variance (Garland et al. 1993) was used to test for differences in continuous plastome variables (e.g., length, GC content) among green versus nongreen taxa (i.e., a discrete grouping variable), taking phylogenetic relationships into account. Analyses were run in the R package "geiger" (Harmon et al. 2008) with 10,000 simulations each using the "aov.phylo" function based on an ultrametic tree generated through Nonparametric Rate Smoothing (Sanderson 1997). These analyses were run for a complete data set of 13 taxa (three leafy, green outgroup taxa, and ten species of *Corallorhiza*). A second set of analyses was conducted, on a subset of six taxa of *Corallorhiza* for which chlorophyll data were available, to mitigate problems with missing chlorophyll data. In these analyses, continuous genomic variables (length, number of putatively functional genes, GC content) were correlated with chlorophyll content (continuous), using Phylogenetically Independent Contrasts to account for phylogenetic relatedness among taxa (Felsenstein 1985). The latter analyses were conducted under a Brownian motion model with the PDAP package (Midford et al. 2005) in Mesquite v.2.75 (Maddison WP and Maddison DR 2011).

Model-Based Analyses of Selective Regime

Changes in selective regime among species of *Corallorhiza* for codon-based matrices were assessed with the CODEML module in PAML v.4.8 (Yang 1997, 2007), using "branch" models. Only loci for which all taxa had open reading frames were included in the analyses. Various matrices are listed in table 3. The objective was to test hypotheses of different ω ratios for 1) *Corallorhiza* versus leafy outgroups and 2) nongreen *Corallorhiza* versus green *Corallorhiza* + leafy outgroups. To accomplish this, different branch models were fit to the data, using the topology resulting from the ML analysis of whole plastomes + rDNA. Six different branch models were tested. First, a model in which all branches evolve under one ω ratio (Model M0) was tested, where $\omega = dN/dS$, the ratio of nonsynonymous substitutions per nonsynonymous site to synonymous substitutions per synonymous site. A two-ratio model (M1) was then run in which all branches of *Corallorhiza* were allowed to have a different ω ratio (ω_{cor}) than the remaining leafy outgroups sequences (ω_{leafy}). A three-ratio model was then run (M2), allowing separate ω values for ω_{leafy}, ω_{cor}, and nongreen *Corallorhiza* ($\omega_{str} = \omega_{mac\text{-}ng}$; *Co. striata*; *Co. mertensiana*, *Co. maculata* var. *maculata* and *occidentalis*, respectively), followed by a four-ratio model (M3) with additionally separate ω ratios for *Co. striata* (ω_{str}) and the nongreen members of the *Co. maculata* complex ($\omega_{mac\text{-}ng}$). The latter three models (M1–M3) specifically test hypotheses of whether *Corallorhiza* and particularly nongreen *Corallorhiza* display departures from purifying selection. Two additional models were run to test whether the nongreen *Corallorhiza* differ significantly from green *Corallorhiza* and leafy outgroup taxa. Specifically, Model M4 (two ratios) sets $\omega_{leafy} = \omega_{cor}$, and $\omega_{str} = \omega_{mac\text{-}ng}$, whereas M5 (three ratios) allows ω_{str} and $\omega_{mac\text{-}ng}$ to differ. Successively nested models were compared based on log-likelihood ratio tests (M1 vs. M0, M2 vs. M1, M3 vs. M2, M4 vs. M0, M5 vs. M4), where the likelihood ratio test statistic (Λ) was calculated as two times the difference in log-likelihood between the two models, with degrees of freedom calculated as the difference in the number of free parameters for the models, and tested against the χ^2 distribution. *P* values for multiple comparisons to the same data set were Bonferroni-corrected (Dunn 1959) as suggested by Yang (2007), with the corrected significance level corresponding to α/m, where m is the number of branches being tested. Genes ycf1 and ycf2 were analyzed separately because they have extraordinarily high substitution rates (e.g., Barrett et al. 2013), as was matK based on the finding of a pseudogenized copy in a previous study (Freudenstein and Senyo 2008).

To further investigate the ω values among branches for the *atp* complex, two sets of branch-site models were tested against the data. In the first, all *Corallorhiza* were set as the foreground branch, and in the second, nongreen *Corallorhiza* (*Co. striata vreelandii*, *Co. maculata* var. *maculata* and *occidentalis*, and *Co. mertensiana*) were set as foreground branches. For both sets of branch models, two alternative models were tested. The "null" model specifies $\omega_0 < 1$ for a proportion of sites and $\omega_1 = \omega_2 = 1$ for the remaining sites (no positive selection), whereas the alternative model allows $\omega_2 > 1$ (positive selection on some sites). The alternative model (some sites under positive selection) was then tested against the null model (no sites under positive selection), using a χ^2 distribution as above, with 1 degree of freedom (i.e., the alternative and null models differ by one parameter).

Comparison with Other Parasitic Angiosperm Lineages

Plastid gene content (losses, pseudogenes, putatively functional genes) was compared among diverse lineages of angiosperms containing heterotrophs, including newly sequenced *Corallorhiza*. Taxa and genes were sorted by the numbers of putatively functional genes and gene functional class, respectively, to illustrate the different stages of plastid genome degradation. Specifically, genes were arranged according to the model of Barrett and Davis (2012), in which degradation of plastid functional gene classes is hypothesized to have occurred roughly in the following order: 1) NAD(P)H dehydrogenase genes (*ndh*) → 2) "photosynthesis"-related genes, including *pet*, *psa*, *psb*, *cemA*, *ccsA*, *rbcL*, *ycf3*, and *ycf4* → 3) Plastid-encoded RNA Polymerase genes (*rpo*) → 4) ATP synthase genes (*atp*) → and finally, 5) "housekeeping" genes (*rpl*, *rps*, *rrn*, *trn*, *matK*, *clpP*, *infA*, *accD*, *ycf1*, and *ycf2*). Though not a formal test in the strict sense, the use of this model does allow an assessment of how advanced the sequenced accessions of *Corallorhiza* are, when compared with other heterotrophic angiosperms; it is hypothesized that *Corallorhiza* occupies the early stages of this model. Technically, *atp* genes function in the generation of ATP and movement of

protons across the Thylakoid Membrane, and thus play a crucial role in photosynthesis, but this functional class was kept separate from photosynthesis genes due to the possibility that ATP synthase serves roles outside of photosynthesis.

Supplementary Material

Supplementary figures S1–S5 are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

This work was supported by the National Science Foundation (awards DEB-0830020 to Jerrold I. Davis and DBI 1110443 to J.C.P.), the California State University Program for Education and Research in Biotechnology (CSUPERB) to C.F.B., the California State University, Los Angeles Center for Effective Teaching and Learning Faculty-Student Mentorship award to C.F.B. and J.L., and the California State University Council on Ocean Affairs, Science, and Technology to C.S. The authors are grateful to: Jerrold Davis, Patrick Edger, Ryan Ellingson, Gwynne Lim, Jeffrey Morawetz, and Chris Randle for feedback and laboratory/bioinformatics support; the staff at Cold Spring Harbor Laboratory (Woodbury, NY) for sample preparation and sequencing; Daniel Lima and Richard Sayre for assistance with chlorophyll spectroscopy and data interpretation; Gerardo Salazar (UNAM) for assistance in collecting Mexican specimens; Maria Logacheva and three anonymous reviewers for critical insight that improved the paper.

References

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

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

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, Davis JI, Leebens-Mack J, Conran JG, Stevenson DW. 2013. Plastid genomes and deep relationships among the commelinid monocot angiosperms. *Cladistics* 29:65–87.

Barrett CF, Freudenstein JV. 2008. Molecular evolution of *rbcL* in the mycoheterotrophic coralroot orchids (*Corallorhiza* Gagnebin, Orchidaceae). *Mol Phylogenet Evol.* 47:665–679.

Bellino A, Alfani A, Selosse M-A, Guerrieri R, Borghetti M, Baldantoni D. 2014. Nutritional regulation in mixotrophic plants: new insights from *Limodorum abortivum*. *Oecologia* 175:875–885.

Bernardi G. 1989. The isochore organization of the human genome. *Annu Rev Genet.* 23:637–661.

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

Blazier J, Guisinger MM, Jansen RK. 2011. Recent loss of plastid-encoded *ndh* genes within *Erodium* (Geraniaceae). *Plant Mol Biol.* 76:263–272.

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

Braukmann T, Stefanovic S. 2012. Plastid genome evolution in mycoheterotrophic Ericaceae. *Plant Mol Biol.* 79:5–20.

Braukmann TWA, Kuzmina M, Stefanovic S. 2009. Loss of all plastid *ndh* genes in Gnetales and conifers: extent and evolutionary significance for the seed plant phylogeny. *Curr Genet.* 55:323–337.

Burnham KP, Anderson DR. 2002. Avoiding pitfalls when using information-theoretic methods. *J Wildl Manage.* 66:912–918.

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

Chang CC, Lin HC, Lin IP, Chow TY, Chen HH, Chen WH, Cheng CH, Lin CY, Liu SM, Chang CC, et al. 2006. The chloroplast genome of *Phalaenopsis aphrodite* (Orchidaceae): comparative analysis of evolutionary rate with that of grasses and its phylogenetic implications. *Mol Biol Evol.* 23:279–291.

Conant GC, Wolfe KH. 2008. GenomeVx: simple web-based creation of editable circular chromosome maps. *Bioinformatics* 24:861–862.

Cummings MP, Welschmeyer NA. 1998. Pigment composition of putatively achlorophyllous angiosperms. *Plant Syst Evol.* 210: 105–111.

Delannoy E, Fujii S, Colas des Francs-Small C, Brundrett M, Small I. 2011. Rampant gene loss in the underground orchid *Rhizanthella gardneri* highlights evolutionary constraints on plastid genomes. *Mol Biol Evol.* 28:2077–2086.

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

Dunn OJ. 1959. Confidence intervals for the means of dependent, normally distributed variables. *J Am Stat Assoc.* 54:613–621.

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32:1792–1797.

Evron Y, Johnson EA, McCarty RE. 2000. Regulation of proton flow and ATP synthesis in chloroplasts. *J Bioenerg Biomembr.* 32:501–506.

Fang YJ, Wu H, Zhang TW, Yang M, Yin YX, Pan LL, Yu XG, Zhang XW, Hu SNA, Al-Mssallem IS, et al. 2012. A complete sequence and transcriptomic analyses of date palm (*Phoenix dactylifera* L.) mitochondrial genome. *PLoS One* 7:e37164.

Felsenstein J. 1985. Phylogenies and the comparative method. *Am Nat.* 125:1–15.

Freudenstein JV. 1992. Systematics of *Corallorhiza* and the Corallorhizinae (Orchidaceae) [Ph.D. dissertation]. [Ithaca (NY)]: Cornell University.

Freudenstein JV. 1997. A monograph of *Corallorhiza* (Orchidaceae). *Harv Pap Bot.* 10:5–51.

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. The Proceedings of the Fourth International Conference on Monocot Systematics. Aarhus (Denmark): Aarhus University Press. p. 25–37.

Freudenstein JV, Doyle JJ. 1994a. Character transformation and relationships in *Corallorhiza* (Orchidaceae, Epidendroideae). 1. Plastid DNA. *Am J Bot.* 81:1449–1457.

Freudenstein JV, Doyle JJ. 1994b. Plastid DNA, morphological variation, and the phylogenetic species concept—the *Corallorhiza maculata* (Orchidaceae) complex. *Syst Bot.* 19:273–290.

Freudenstein JV, Senyo DM. 2008. Relationships and evolution of *matK* in a group of leafless orchids (*Corallorhiza* and Corallorhizinae; Orchidaceae : Epidendroideae). *Am J Bot.* 95:498–505.

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.* 7:45.

Galagan JE, Calvo SE, Borkovich KA, Selker EU, Read ND, Jaffe D, FitzHugh W, Ma LJ, Smirnov S, Purcell S, et al. 2003. The genome sequence of the filamentous fungus *Neurospora crassa*. *Nature* 422: 859–868.

Garland T, Dickerman AW, Janis CM, Jones JA. 1993. Phylogenetic analysis of covariance by computer simulation. *Syst Biol.* 42:265–292.

Glémin S, Clément Y, David J, Ressayre A. 2014. GC content evolution in coding regions of angiosperm genomes: a unifying hypothesis. *Trends Genet.* 30:263–270.

Goloboff PA, Farris JS, Kallersjo M, Oxelman B, Ramirez MJ, Szumik CA. 2003. Improvements to resampling measures of group support. *Cladistics* 19:324–332.

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

Hajdukiewicz PTJ, Allison LA, Maliga P. 1997. The two RNA polymerases encoded by the nuclear and the plastid compartments transcribe distinct groups of genes in tobacco plastids. *EMBO J.* 16:4041–4048.
Harmon LJ, Weir JT, Brock CD, Glor RE, Challenger W. 2008. GEIGER: investigating evolutionary radiations. *Bioinformatics* 24:129–131.
Iles WJD, Smith SY, Graham SW. 2013. A well-supported phylogenetic framework for the monocot order Alismatales reveals multiple losses of the plastid NADH dehydrogenase complex and a strong long-branch effect. In: Wilkin P, Mayo S, editors. Early events in monocot evolution. Cambridge: Cambridge University Press. p. 1–28.
Iorizzo M, Grzebelus D, Senalik D, Szklarczyk M, Spooner D, Simon P. 2012. Against the traffic: the first evidence for mitochondrial DNA transfer into the plastid genome. *Mob Genet Elements.* 2:261.
Jensen PE, Bassi R, Boekema EJ, Dekker JP, Jansson S, Leister D, Robinson C, Scheller HV. 2007. Structure, function and regulation of plant photosystem I. *Biochim Biophys Acta.* 1767:335–352.
Katinka MD, Duprat S, Cornillot E, Metenier G, Thomarat F, Prensier G, Barbe V, Peyretaillade E, Brottier P, Wincker P, et al. 2001. Genome sequence and gene compaction of the eukaryote parasite *Encephalitozoon cuniculi*. *Nature* 414:450–453.
Katoh K, Misawa K, Kuma Ki, Miyata T. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* 30:3059–3066.
Kawakami K, Umena Y, Iwai M, Kawabata Y, Ikeuchi M, Kamiya N, Shen J-R. 2011. Roles of PsbI and PsbM in photosystem II dimer formation and stability studied by deletion mutagenesis and X-ray crystallography. *Biochim Biophys Acta.* 1807:319–325.
Knaus BJ. 2010. Short read toolbox. Available from: http://brianknaus.com.
Krause K. 2008. From chloroplasts to "cryptic" plastids: evolution of plastid genomes in parasitic plants. *Curr Genet.* 54:111–121.
Krause K. 2012. Plastid genomes of parasitic plants: a trail of reductions and losses. In: Bullerwell C, editor. Organelle genetics. Berlin (Germany): Springer. p. 79–103.
Kuijt J. 1969. The biology of flowering parasitic plants. Berkeley (CA): University of California Press.
Lawrence JG. 2005. Common themes in the genome strategies of pathogens. *Curr Opin Genet Dev.* 15:584–588.
Leake JR. 1994. The biology of myco-heterotrophic (Saprophytic) plants. *New Phytol.* 127:171–216.
Leake JR, Cameron DD. 2010. Physiological ecology of mycoheterotrophy. *New Phytol.* 185:601–605.
Liere K, Weihe A, Borner T. 2011. The transcription machineries of plant mitochondria and chloroplasts: composition, function, and regulation. *J Plant Physiol.* 168:1345–1360.
Logacheva MD, Schelkunov MI, Nuraliev MS, Samigullin TH, Penin AA. 2014. The plastid genome of mycoheterotrophic monocot *Petrosavia stellaris* exhibits both gene losses and multiple rearrangements. *Genome Biol Evol.* 6:238–246.
Logacheva MD, Schelkunov MI, Penin AA. 2011. Sequencing and analysis of plastid genome in mycoheterotrophic orchid *Neottia nidus-avis*. *Genome Biol Evol.* 3:1296–1303.
Maddison WP, Maddison DR. 2011. Mesquite: a modular system for evolutionary analysis. version 2.75. Available from: http://mesquiteproject.org.
Martín M, Funk HT, Serrot PH, Poltnigg P, Sabater B. 2009. Functional characterization of the thylakoid *ndh* complex phosphorylation by site-directed mutations in the *ndhF* gene. *Biochim Biophys Acta.* 1787:920–928.
Martín M, Sabater B. 2010. Plastid *ndh* genes in plant evolution. *Plant Physiol Biochem.* 48:636–645.
McCarty RE, Evron Y, Johnson EA. 2000. The chloroplast ATP synthase: a rotary enzyme? *Annu Rev Plant Physiol Plant Mol Biol.* 51:83–109.
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, Bennett JR, Wolfe AD, Mathews S. 2013. Phylogeny and origins of holoparasitism in Orobanchaceae. *Am J Bot.* 100:971–983.
Merckx V, Freudenstein JV. 2010. Evolution of mycoheterotrophy in plants: a phylogenetic perspective. *New Phytol.* 185:605–609.
Midford PE, Garland T Jr, Maddison WP. 2005. PDAP package of Mesquite. version 1.07. Available from: http://mesquiteproject.org/.
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.* 31:793–803.
Montfort C, Küsters E. 1940. Saprophytismus und photosythese. I. Biochemische physiologische studien an humus-orchideen. *Botanisces Archiv* 40:571–633.
Moran NA. 2002. Microbial minimalism: genome reduction in bacterial pathogens. *Cell* 108:583–586.
Morrison HG, McArthur AG, Gillin FD, Aley SB, Adam RD, Olsen GJ, Best AA, Cande WZ, Chen F, Cipriano MJ, et al. 2007. Genomic minimalism in the early diverging intestinal parasite *Giardia lamblia*. *Science* 317:1921–1926.
Nickrent DL, Yan OY, Duff RJ, dePamphilis CW. 1997. Do nonasterid holoparasitic flowering plants have plastid genomes? *Plant Mol Biol.* 34:717–729.
Patel RK, Jain M. 2012. NGS QC Toolkit: a toolkit for quality control of next generation sequencing data. *PLoS One* 7:e30619.
Peredo EL, King UM, Les DH. 2013. The plastid genome of *Najas flexilis*: adaptation to submersed environments is accompanied by the complete loss of the NDH complex in an aquatic angiosperm. *PLoS One* 8:e68591.
Porra RJ. 2002. The chequered history of the development and use of simultaneous equations for the accurate determination of chlorophylls a and b. *Photosyn Res.* 73:149–156.
Protasio AV, Tsai IJ, Babbage A, Nichol S, Hunt M, Aslett MA, De Silva N, Velarde GS, Anderson TJC, Clark RC, et al. 2012. A systematically improved high quality genome and transcriptome of the human blood fluke *Schistosoma mansoni*. *PLoS Negl Trop Dis.* 6:e1455.
Randle CP, Wolfe AD. 2005. The evolution and expression of *rbcL* in holoparasitic sister-genera *Harveya* and *Hyobanche* (Orobanchaceae). *Am J Bot.* 92:1575–1585.
Ranwez V, Harispe S, Delsuc F, Douzery EJP. 2011. MACSE: Multiple Alignment of Coding SEquences accounting for frameshifts and stop codons. *PLoS One* 6:e22594.
Sanderson MJ. 1997. A nonparametric approach to estimating divergence times in the absence of rate constancy. *Mol Biol Evol.* 14:1218–1231.
Smith DR, Lee RW. 2014. A plastid without a genome: evidence from the nonphotosynthetic green algal genus *Polytomella*. *Plant Physiol.* 164:1812–1819.
Stamatakis A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22:2688–2690.
Straub SCK, Cronn RC, Edwards C, Fishbein M, Liston A. 2013. Horizontal transfer of DNA from the mitochondrial to the plastid genome and its subsequent evolution in milkweeds (Apocynaceae). *Genome Biol Evol.* 5:1872–1885.
Straub SCK, Fishbein M, Livshultz T, Foster Z, Parks M, Weitemier K, Cronn RC, Liston A. 2011. Building a model: developing genomic resources for common milkweed (*Asclepias syriaca*) with low coverage genome sequencing. *BMC Genomics* 12:211.
Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. 2013. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol Biol Evol.* 30:2725–2729.
Tsai IJ, Zarowiecki M, Holroyd N, Garciarrubio A, Sanchez-Flores A, Brooks KL, Tracey A, Bobes RJ, Fragoso G, Sciutto E, et al. 2013. The genomes of four tapeworm species reveal adaptations to parasitism. *Nature* 496:57–63.
Umate P, Schwenkert S, Karbat I, Dal Bosco C, Mlcochova L, Volz S, Zer H, Herrmann RG, Ohad I, Meurer J. 2007. Deletion of *psbM* in

tobacco alters the Q(B) site properties and the electron flow within photosystem II. *J Biol Chem.* 282:9758–9767.
Vaidya G, Lohman DJ, Meier R. 2011. SequenceMatrix: concatenation software for the fast assembly of multi-gene datasets with character set and codon information. *Cladistics* 27:171–180.
Westwood JH, Yoder JI, Timko MP, dePamphilis CW. 2010. The evolution of parasitism in plants. *Trends Plant Sci.* 15:227–235.
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: 3711–3725.
Wicke S, Schäferhoff B, dePamphilis CW, Müller KF. 2014. Disproportional plastome-wide increase of substitution rates and relaxed purifying selection in genes of carnivorous Lentibulariaceae. *Mol Biol Evol.* 31:529–545.
Wicke S, Schneeweiss GM, dePamphilis CW, Muller KF, Quandt D. 2011. The evolution of the plastid chromosome in land plants: gene content, gene order, gene function. *Plant Mol Biol.* 76: 273–297.
Wickett NJ, Honaas LA, Wafula EK, Das M, Huang K, Wu BA, Landherr L, Timko MP, Yoder J, Westwood JH, et al. 2011. Transcriptomes of the parasitic plant family Orobanchaceae reveal surprising conservation of chlorophyll synthesis. *Curr Biol.* 21:2098–2104.
Wimpee CF, Wrobel RL, Garvin DK. 1991. A divergent plastid genome in *Conopholis americana*, an achlorophyllous parasitic plant. *Plant Mol Biol.* 17:161–166.
Wolfe AD, dePamphilis CW. 1998. The effect of relaxed functional constraints on the photosynthetic gene *rbcL* in photosynthetic and nonphotosynthetic parasitic plants. *Mol Biol Evol.* 15:1243–1258.
Wolfe KH, Morden CW, Palmer JD. 1992. Function and evolution of a minimal plastid genome from a nonphotosynthetic parasitic plant. *Proc Natl Acad Sci U S A.* 89:10648–10652.
Wu FH, Chan MT, Liao DC, Hsu CT, Lee YW, Daniell H, Duvall MR, Lin CS. 2010. Complete chloroplast genome of *Oncidium* Gower Ramsey and evaluation of molecular markers for identification and breeding in Oncidiinae. *BMC Plant Biol.* 10:68.
Wyman SK, Jansen RK, Boore JL. 2004. Automatic annotation of organellar genomes with DOGMA. *Bioinformatics* 20:3252–3255.
Yang JB, Tang M, Li HT, Zhang ZR, Li DZ. 2013. Complete chloroplast genome of the genus *Cymbidium*: lights into the species identification, phylogenetic implications and population genetic analyses. *BMC Evol Biol.* 13:84.
Yang ZH. 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput Appl Biosci.* 13:555–556.
Yang ZH. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol.* 24:1586–1591.
Yoder OC, Turgeon BG. 2001. Fungal genomics and pathogenicity. *Curr Opin Plant Biol.* 4:315–321.
Zerbino DR, Birney E. 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res.* 18:821–829.
Zimmer K, Meyer C, Gebauer G. 2008. The ectomycorrhizal specialist orchid *Corallorhiza trifida* is a partial myco-heterotroph. *New Phytol.* 178:395–400.