In the early stages of plastome degradation in groups containing both photosynthetic and nonphotosynthetic species, plastome evolution has been studied in various heterotrophic plants, including orchids (e.g., Delannoy et al. 2011; Braukmann et al. 2013), monocots including Orchidaceae (Delannoy et al. 2011; Braukmann et al. 2013), and eudicots among angiosperms (McNeal et al. 2007; Braukmann et al. 2013). 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 Orobancheaeae (e.g., Epifagus, Wolfe et al. 1992; Conopholis, Wimpee et al. 1999; 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, tm, 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 (Freudenberg 1992, 1997; Freudenberg and Senyo 2008). The genus contains 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 var. maculata and occidentalis). Preliminary evidence of plastome degradation in Corallorhiza comes from heterologous probe experiments (Freudenberg and Doyle 1994a), the presence of Rubisco Large Subunit (rbcL) pseudogenes in some lineages but not others (Barrett and Freudenberg 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 ($P_{\text{PhyANOVA}} = 0.0146$; 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 × (Corallorhiza wisteriana) to 271.3 × (Corallorhiza odontorhiza). The length of the plastome among the Corallorhiza accessions sampled ranges from 151,506 bp in Corallorhiza macrantha to 137,505 bp in Corallorhiza 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 ($P_{\text{PhyANOVA}} = 0.06189$).

The length of the inverted repeat (IR) ranges from 25,775 bp in Corallorhiza odontorhiza to 27,243 bp in Corallorhiza 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 vreelandii has by far the shortest large single copy (LSC, 72,152 bp). This is followed by Corallorhiza maculata var. maculata (80,401 bp), Corallorhiza mertensiana (81,146 bp), and Coralloriza maculata var. occidentalis (81,363 bp); Corallorhiza macrantha (84,253 bp) and Corallorhiza maculata var. mexicana (84,349 bp) have the longest LSC regions. Thus, the range of length variation among members of the Corallorhiza complex for the LSC region (3,852 bp) is substantial. Nongreen taxa have significantly shorter LSC regions after accounting for phylogenetic relationships ($P_{\text{PhyANOVA}} = 0.0148$). Unique to Corallorhiza 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 Corallorhiza wisteriana to 14,427 bp in Corallorhiza trifida, but there is no significant difference in length between the SSC regions of the plastome for green versus nongreen taxa ($P_{\text{Cov}} = 0.7382$), nor is there any correlation between SSC length and chlorophyll content ($P_{\text{Cov}} = 0.0093$). GC content is lowest in Corallorhiza striata vreelandii (34.14%), and highest in Corallorhiza trifida (35.74%), but does not differ significantly among green versus nongreen taxa nor

![Fig. 1. Total chlorophyll concentration (chlorophyll 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.](https://academic.oup.com/mbe/article-abstract/31/12/3095/2925661)

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

<table>
<thead>
<tr>
<th>Species</th>
<th>Voucher</th>
<th>GenBank ⁴</th>
<th>#reads</th>
<th>$\times$-Cov ³</th>
</tr>
</thead>
<tbody>
<tr>
<td>Corallorhiza bulbosa</td>
<td>CFB 238</td>
<td>KM390013</td>
<td>10,757,100</td>
<td>243.8</td>
</tr>
<tr>
<td>A. Rich. &amp; Galeotti</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Co. macrantha Schtr.</td>
<td>Salazar 8191 ⁵</td>
<td>KM390017</td>
<td>9,222,076</td>
<td>136.4</td>
</tr>
<tr>
<td>Co. maculata (Raf.) Raf. var. maculata</td>
<td>JVF 2919 (OS)</td>
<td>KM390014</td>
<td>11,498,452</td>
<td>109.1</td>
</tr>
<tr>
<td>Co. maculata var. mexicana (Lindl.)</td>
<td>CFB 232 ⁴</td>
<td>KM390015</td>
<td>7,828,614</td>
<td>118.2</td>
</tr>
<tr>
<td>Co. maculata var. occidentalis (Lindl.)</td>
<td>JVF 2095 (OS)</td>
<td>KM390016</td>
<td>13,204,554</td>
<td>225.1</td>
</tr>
<tr>
<td>Co. mertensiana Bong.</td>
<td>JVF 1999 (OS)</td>
<td>KM390018</td>
<td>14,540,670</td>
<td>212.0</td>
</tr>
<tr>
<td>Co. odontorhiza (Willd.) Poir.</td>
<td>JVF 2778 (OS)</td>
<td>KM390021</td>
<td>21,861,728</td>
<td>271.3</td>
</tr>
<tr>
<td>Co. striata L. var. vreelandii (Rydbe) L.O. Williams</td>
<td>Taylor 341 (UNM)</td>
<td>JX087681</td>
<td>3,204,074</td>
<td>43.9</td>
</tr>
<tr>
<td>Co. trifida Châtel.</td>
<td>JVF 2676 (OS)</td>
<td>KM390019</td>
<td>13,617,730</td>
<td>236.9</td>
</tr>
<tr>
<td>Co. wisteriana Conrad</td>
<td>JVF 2462 (OS)</td>
<td>KM390020</td>
<td>9,992,438</td>
<td>81.0</td>
</tr>
</tbody>
</table>

⁴GenBank accession numbers will be included contingent upon a favorable review.

⁵Mean depth of coverage of each plastome ($\times$-cov), based on remapping the original read pool to the finished plastome.

⁶Duplicate vouchers at Universidad Nacional Autónoma de México (UNAM).

⁷Herbarium codes: (OS), Ohio State University Herbarium (UNA), University of New Mexico Herbarium (MEXU), Universidad Nacional Autónoma de México.

⁸Sequenced 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.

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

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

*Presence of visible green tissue, based on observation of multiple populations of each species.

*Photosynthesis has only been demonstrated in Co. trifida.

*The number of putatively functional genes, as assessed by intact reading frames.

*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 ycf6, 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, chlorophyll 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
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).

---

**Fig. 3.** A basic model of plastid genome degradation (Barrett and Davis 2012) using sequenced, exemplar plastomes from angiosperm families containing heterotrophs: Convulvulaceae (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.

![Plastid Genome Model](https://example.com/plastid-model.png)

**Table: Genes and Functions**

<table>
<thead>
<tr>
<th>Gene</th>
<th>Function</th>
<th>Example Taxa</th>
</tr>
</thead>
<tbody>
<tr>
<td>psbA</td>
<td>Photosystem I</td>
<td>Co. bulbosa, Co. macrantha, Co. trifida</td>
</tr>
<tr>
<td>psbB</td>
<td>Photosystem I</td>
<td>Co. striata vreelandii, Co. mertensiana, Co. maculata</td>
</tr>
<tr>
<td>rpoC1</td>
<td>ATP synthase complex</td>
<td>All taxa</td>
</tr>
<tr>
<td>ccsA</td>
<td>Cytochrome C biogenesis</td>
<td>Co. mertensiana, Co. maculata occidentalis</td>
</tr>
<tr>
<td>rbcL</td>
<td>RubisCO</td>
<td>All taxa</td>
</tr>
</tbody>
</table>

**Additional Notes**: The table includes a comprehensive overview of genes and their functions across various species, highlighting both the conservation and divergence in plastid gene composition among heterotrophic orchids. This data supports the hypothesis of plastid genome degradation in response to increased reliance on mycorrhizal fungi.
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, \( \omega_{\text{leafy}} = \omega_{\text{cor}} = \omega_{\text{mac-ng}} = \omega_{\text{str}} \); 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 (\( \omega_{\text{leafy}} = \omega_{\text{mac-ng}} = \omega_{\text{str}} \)), 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 (\( \omega_{\text{leafy}} = \omega_{\text{cor}} = \omega_{\text{mac-ng}} = \omega_{\text{str}} \)), 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 (\( \omega_{\text{leafy}} = \omega_{\text{cor}} = \omega_{\text{mac-ng}} = \omega_{\text{str}} \)) and M4 (\( \omega_{\text{leafy}} = \omega_{\text{cor}} \omega_{\text{mac-ng}} = \omega_{\text{str}} \)) than M0, but these were nonsignificant after Bonferroni correction. Regardless, the branch-specific estimate (model M4) of ω in nongreen taxa (\( \omega_{\text{mac-ng}} = \omega_{\text{str}} = 0.73720 \)) was more than twice that of green taxa (\( \omega_{\text{leafy}} = \omega_{\text{cor}} = 0.27361 \)), approaching a pattern expected under selective neutrality for photosynthesis-related genes (\( \omega \approx 1 \)). “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 corallroots, with green pigmentation throughout nearly the entire above-ground portion of the plant. The distribution of green tissues among other partially green corallroot 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 corallroots 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” corallroot 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 only corallroot 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 abortivium, 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
would cause purifying selection on the plastid- (and presumably nuclear-) encoded components of photosynthesis.

The more puzzling finding applies to nongreen taxa; non-green coralloids have 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
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., Oreocharis, Aplectrum, Cremas tra, 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; Glém in 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, Oreocharis, Cremas tra, 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 chlorophyll (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 psal (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 psal 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 psal 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 (rps, rpsb), 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 (ωleafy = ωcore, ωmac-ng = ωatr), 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 (ωleafy = ωcore = 0.30267; ωmac-ng = ωatr = 0.34230; table 3). However, this difference is much more pronounced for photosynthesis-related genes (ωleafy = ωcore = 0.27361; ωmac-ng = ωatr = 0.73270), with the ω estimate for nongreen taxa being approximately 2.6× that of green taxa, and approaching a ω value expected under neutral evolution (ω ~ 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 $\omega$ value for Corallorhiza is nearly twice that for leafy orchids ($\omega_{\text{leafy}} = 0.15175$; $\omega_{\text{Corallorhiza}} = 0.26949$), and in model M4, nongreen Corallorhiza have $\omega$ 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–GCU 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, tmD–guc, tmE–uuc, trnL–cau, trnM–cau, trnQ–uug, trnY–gua, rna16, rm23, rm4.5, and rm5. 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 idiosyncratic 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 CO₂ 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., Orobancheae, 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 corallorhizas 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 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/μl 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_PRL.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 Ca. 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], 18S rDNA, ITS1 [Internal Transcribed Spacer], 5.8S rDNA, ITS2, 26S 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, 18S rDNA, ITS1, 5.8S rDNA, ITS2, 26S rDNA; excluding 5S 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: KM390033-KM390012. Available outgroup sequences for Cymbidium, Oncidium, and Phalaenopsis were mined from GenBank for the purposes of downstream phylogenetic analyses (18S rDNA—Cy. farberi, GenBank accession number JN418934; ITS—Cy. aloifolium JF729014, Phalaenopsis aphrodite KM390012, 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 bi-direction–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+I 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+I 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 taxon 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 ultrametric 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 ($\omega_{Leafy}$) than the remaining leafy outgroups sequences ($\omega_{Leaf}$). A three-ratio model was then run (M2), allowing separate $\omega$ values for $\omega_{Leaf}$, $\omega_{Leaf}$, and nongreen Corallorhiza ($\omega_{Leaf} = \omega_{Leaf} = \omega_{Leaf}$). 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 $\omega_{Leaf} = \omega_{Leaf}$ and $\omega_{Leaf} = \omega_{Leaf}$ Whereas M5 (three ratios) allows $\omega_{Leaf}$ and $\omega_{Leaf}$ 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 ($\Lambda$) 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 $\chi^2$ 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 $\alpha/m$, 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 $\omega$ 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 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 $\omega_{Leaf} < 1$ for a proportion of sites and $\omega_{Leaf} = 1$ for the remaining sites (no positive selection), whereas the alternative model allows $\omega_{Leaf} > 1$ (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 $\chi^2$ 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, rrm, tm, 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 accesses 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 Ellington, 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

tobacco alters the Q(B) site properties and the electron flow within photosystem II. J Biol Chem. 282:9758–9767.


