

# Molecular evolution of *rbcL* in the mycoheterotrophic coralroot orchids (*Corallorhiza* Gagnebin, Orchidaceae)

Craig F. Barrett<sup>\*</sup>, John V. Freudenstein

The Ohio State University Herbarium, Department of Evolution, Ecology, and Organismal Biology, 1315 Kinnear Road, Columbus, OH 43212, USA

Received 4 August 2007; revised 14 January 2008; accepted 14 February 2008

Available online 4 March 2008

## Abstract

The RuBisCO large subunit gene (*rbcL*) has been the focus of numerous plant phylogenetic studies and studies on molecular evolution in parasitic plants. However, there has been a lack of investigation of photosynthesis gene molecular evolution in fully mycoheterotrophic plants. These plants invade pre-existing mutualistic associations between ectomycorrhizal trees and fungi, from which they obtain fixed carbon and nutrients. The mycoheterotrophic orchid *Corallorhiza* contains both green (photosynthetic) and non-green (putatively non-photosynthetic) species. We sequenced *rbcL* from 31 accessions of eight species of *Corallorhiza* and hypothesized that some lineages would have pseudogenes resulting from relaxation of purifying selection on RuBisCO's carboxylase function. Phylogenetic analysis of *rbcL* + ITS gave high jackknife support for relationships among species. We found evidence of pseudogene formation in all lineages of the *Corallorhiza striata* complex and in some lineages of the *C. maculata* complex. Evidence includes: stop codons, frameshifts, decreased  $d_S/d_N$  ratios, replacements not observed in photosynthetic species, rate heterogeneity, and high likelihood of neutral evolution. The evolution of *rbcL* in *Corallorhiza* may serve as an exemplary system in which to study the effects of relaxed evolutionary constraints on photosynthesis genes for >400 documented fully mycoheterotrophic plant species.

© 2008 Elsevier Inc. All rights reserved.

**Keywords:** Pseudogene; Neutral evolution; Purifying selection; Corallorhizinae; Orchid

## 1. Introduction

The plastid-encoded RuBisCO (ribulose bis-phosphate carboxylase/oxygenase) large subunit gene, *rbcL*, has been used extensively in higher-level phylogenetic studies across a variety of plant taxa including the largest of all plant families, the Orchidaceae (Cameron et al., 1999). On an even larger scale, it has been used to reconstruct the phylogenetic relationships of the major groups of seed-producing plants, the spermatophytes (Chase et al., 1993). It has also been the focus of several molecular evolutionary studies in parasitic plants (summarized in Table 1). Holoparasitic plants are obligate parasites of other autotrophs and are thus nonphotosynthetic, whereas hemiparasitic plants are

facultative parasites that retain photoautotrophic capacity. The complete reliance of holoparasitic plant species on exogenous photosynthate obviates their need to perform photosynthesis, thus bringing into question the evolutionary fate of photosynthesis genes in these nonphotosynthetic plants. It has been hypothesized that rates of evolutionary change in photosynthesis genes such as *rbcL* for holoparasitic plants approach those of non-coding DNA due to a relaxation of purifying selection on these loci. Several reports on cpDNA evolution in holoparasites (Table 1) suggest that *rbcL* has become a pseudogene, either by the acquisition of premature stop codons, insertions/deletions that cause shifts in reading frame, and/or accumulation of nonsynonymous nucleotide substitutions. In other cases the species of interest retain functional copies of *rbcL* and in the most extreme cases, large portions of the plastid genome are deleted (including many photosynthesis genes) as in the orobanchaceous species *Epifagus virginiana*

<sup>\*</sup> Corresponding author. Fax: +1 614 292 1350.

E-mail address: [barrett.586@osu.edu](mailto:barrett.586@osu.edu) (C.F. Barrett).

Table 1  
Results of selected studies on *rbcL* in heterotrophic plants

Species	Family	<i>rbcL</i> status	Parasite class	Study
<b>Holoparasitic species</b>				
<i>Epifagus virginiana</i>	Orobanchaceae	Pseudogene: reduced to 434 bp	Holoparasitic	dePamphilis and Palmer (1990)
<i>Conopholis americana</i>	Orobanchaceae	Completely deleted	Holoparasitic	Colwell (1994)
<i>Cuscuta reflexa</i>	Convolvulaceae	Functional gene	Holoparasitic	Bömmer et al. (1993) and Freyer et al. (1995)
<i>Lathraea clandestina</i>	Scrophulariaceae	Functional gene: conserved length, still transcribed by a nuclear-encoded polymerase	Holoparasitic	Delavault et al. (1995)
<i>Orobanche</i> spp.	Orobanchaceae	Pseudogenes in <i>Orobanche cernua</i> and <i>O. ramosa</i> ; functional genes in <i>O. corymbosa</i> , and <i>O. fasciculata</i>	Holoparasitic	Wolfe and dePamphilis (1997)
<i>Boschniakia</i> , <i>Hyobanche</i> , and <i>Orobanche</i> spp.	Orobanchaceae	Pseudogenes	Holoparasitic, hemiparasitic, and autotrophic	Wolfe and dePamphilis (1998)
<i>Orobanche</i> spp.	Orobanchaceae	Pseudogenes in <i>O. hederiae</i> , <i>O. amethystea</i> , <i>O. minor</i> , and <i>O. loricata</i>	Holoparasitic	Benharrat et al. (2000)
<i>Orobanche</i> spp.	Orobanchaceae	Two pseudogenized copies in <i>O. cumanea</i> : 1425 bp (nucleus) and 1096 bp (plastid)	Holoparasitic	Delavault and Thalouarn (2002)
<i>Cuscuta gronovii</i> , <i>C. subinclusa</i>	Convolvulaceae	Pseudogene: conserved coding sequence, but 5' deletions in promoter	Holoparasitic	Berg et al. (2002)
<i>Orobanche</i> spp.	Orobanchaceae	36 of 37 <i>Orobanche</i> sequences were hypothesized pseudogenes	Holoparasitic	Manen et al. (2004)
<i>Harveya</i> and <i>Hyobanche</i> spp.	Orobanchaceae	Functional genes in 11 <i>Harveya</i> spp., pseudogene in one sp. ( <i>H. huttonii</i> ); pseudogenes in four <i>Hyobanche</i> spp.	Holoparasitic	Randle and Wolfe (2005)
Representative spp. within Orobanchaceae ( <i>Boschniakia</i> , <i>Epifagus</i> , and <i>Orobanche</i> spp.)	Orobanchaceae	Of 38 taxa; seven spp. had pseudogenes (two more were questionably pseudogenes), the remainder had functional copies	Holoparasitic, hemiparasitic and fully autotrophic	Young and dePamphilis (2005)
<b>Mycoheterotrophic species</b>				
<i>Burmannia</i> , <i>Gymnosiphon</i> , and <i>Geomitra</i> spp.	Representative Burmanniaceae (Dioscoreales)	Significant increases in <i>rbcL</i> substitution rates for <i>Burmannia</i> , <i>Gymnosiphon</i> , and <i>Gyromitra</i> relative to other Dioscoreales	Mycoheterotrophic	Caddick et al. (2002)
<i>Cyrtosia</i> , <i>Pseudovanilla</i> , and <i>Erythrorchis</i> spp.	Orchidaceae	Putative pseudogenes (several deletions) in <i>Cyrtosia</i> for both <i>rbcL</i> and <i>psaB</i> , but not in <i>Pseudovanilla</i> or <i>Erythrorchis</i>	Mycoheterotrophic	Cameron (2004)
<i>Cyrtosia</i> , <i>Pseudovanilla</i> , <i>Erythrorchis</i> , and <i>Lecanorchis</i> spp.	Orchidaceae	<i>Cyrtosia</i> had <i>psbB</i> and <i>psbC</i> pseudogenes; <i>Lecanorchis</i> had a <i>psbC</i> pseudogene; <i>Pseudovanilla</i> and <i>Erythrorchis</i> had apparent elevated mutation rates	Mycoheterotrophic	Cameron and Molina (2006)

(dePamphilis and Palmer, 1990) and *Conopholis americana* (Colwell, 1994).

The evolutionary path of photosynthesis genes for holoparasitic plants varies between different lineages (see references in Table 1). In some lineages, there is a gradual process of degradation from functional gene to pseudogene to partial/complete deletion. However, phylogenetic studies documenting this progression have concentrated on a few well-known groups (mostly within the Orobanchaceae), and very recent switches from autotrophy/hemiparasitism to holoparasitism may not allow sufficient time for mutations to accumulate such that pseudogenes are recognizable. Thus, it remains uncertain whether pseudogene formation or complete deletion for photosynthesis genes is the norm for holoparasites. Alternatively, some photosynthesis genes (*rbcL* for example) may function in cellular metabolism outside of the photosynthetic pathway (Wolfe and dePamphilis, 1998; Randle and Wolfe, 2005), thus

being maintained by purifying natural selection. Although ample literature exists on photosynthesis gene evolution in holo- and hemiparasitic plants, very little effort has been focused towards another, similar guild of plants: the mycoheterotrophs. Fully mycoheterotrophic plants (MHP) are invaders of pre-existing mutualistic symbioses between mycorrhizal trees or shrubs, and mycorrhizal fungi. They form tripartite symbioses in which they acquire all of their fixed carbon and nutrients directly from the fungus and therefore indirectly from the fungus' associated photobiont(s). There is a range in the degree of dependence on fungal hosts among mycoheterotrophic plant species, from partial to full mycoheterotrophy (analogous to hemi- and holoparasitism among parasitic plants). In mycoheterotrophic orchids, fungal hyphae penetrate the root or rhizome epidermis and grow into the cortical cells, where the orchid enzymatically digests and metabolizes the fungal tissue.

Over 400 species of MHP are documented, most of which are monocots (Leake, 1994; Bidartondo, 2005). One phylogenetic study on the mycoheterotroph-rich family Burmanniaceae addressed questions about molecular evolutionary rates of 18S rDNA and the mitochondrial *nad1 b-c* intron using relative rates testing, but did not observe significant evolutionary rate increases for fully mycoheterotrophic species relative to their photosynthetic relatives (Merckx et al., 2006). However, *rbcL* was not analyzed in their study, and neither of the two genes included is involved in photosynthesis. Another study by Caddick et al. (2002) documented significant substitution rate increases for *rbcL* in mycoheterotrophic *Burmannia*, *Gymnosiphon*, and *Geomitra* relative to the remaining Dioscoreales.

Cameron (2004) and Cameron and Molina (2006) observed substantial deletions within four photosynthesis genes (including *rbcL*) in the mycoheterotrophic orchid genus *Cyrtosia*, but these two studies represent the only available information on mycoheterotrophic photosynthe-

sis gene evolution in orchids. Full mycoheterotrophy has evolved independently at least ten times in the Orchidaceae: the most species-rich family of plants (19,128 described species [Atwood, 1986]; ca. 25,000 estimated species Dressler, 1981), with more species of MHP than any other family (Leake, 1994; Molvray et al., 2000). It is thought that orchids, which are fully mycoheterotrophic during their early seed and protocorm stages and later become photosynthetic, are preadapted for these shifts to mycoheterotrophy (Rasmussen, 1995). The mycoheterotrophic genus *Corallorhiza* Gagnebin (Orchidaceae: Epidendroideae) and its close relatives afford a unique opportunity to study the molecular evolutionary consequences of the shift to mycoheterotrophy on photosynthesis genes. *Corallorhiza* is placed in subtribe Corallorhizinae, which according to the latest taxonomic treatment contains 10 genera (Dressler, 1981, 1993). The genus is composed of 11 rootless, reduced-leaved, mycoheterotrophic species confined to North America (Freudenstein, 1992, 1997). The exception is *C. trifida*

Table 2  
List of accessions

Species	Variety	Voucher (label in tree)	Collection locality	Habit	GenBank #	
					<i>rbcL</i>	ITS
<i>Corallorhiza striata</i>	<i>striata</i>	CFB 0002a* (MI)	Schoolcraft Co., MI, USA	ng, rl	EU391377	EU391343
<i>C. striata</i>	<i>striata</i>	JH 1a (MN)	Crow Wing Co., MN, USA	ng, rl	EU391376	EU391342
<i>C. striata</i>	<i>striata</i>	DJ 0068 (OR-2)	Wallowa Co., OR, USA	ng, rl	EU391382	EU391348
<i>C. striata</i>	<i>striata</i>	CFB 0029a (OR-1)	Lane Co., OR, USA	ng, rl	EU391387	EU391353
<i>C. striata</i>	<i>striata</i>	CFB 0039a (WA-1)	Skamania Co., WA, USA	ng, rl	EU391373	EU391339
<i>C. striata</i>	<i>striata</i>	CFB 0048a (WA-2)	Lewis Co., WA, USA	ng, rl	EU391378	EU391344
<i>C. striata</i>	<i>striata</i>	CFB 0142 (WY-2)	Sheridan Co., WY, USA	ng, rl	EU391383	EU391349
<i>C. striata</i>	<i>striata</i>	CFB 0135a (WY-1)	Natrona Co., WY, USA	ng, rl	EU391375	EU391341
<i>C. striata</i>	<i>vreelandii</i>	CFB 0008* (CA-2)	El Dorado Co., CA, USA	ng, rl	EU391371	EU391337
<i>C. striata</i>	<i>vreelandii</i>	CFB 0005a (CA-1)	Calaveras Co., CA, USA	ng, rl	EU391370	EU391336
<i>C. striata</i>	<i>vreelandii</i>	CFB 0013c (CA-3)	Tehama Co., CA, USA	ng, rl	EU391372	EU391338
<i>C. striata</i>	<i>vreelandii</i>	T191a (NM-3)	Otero Co., NM, USA	ng, rl	EU391385	EU391351
<i>C. striata</i>	<i>vreelandii</i>	T1876 (NM-2)	Otero Co., NM, USA	ng, rl	EU391384	EU391350
<i>C. striata</i>	<i>vreelandii</i>	CFB 0103e* (NM-1)	Otero Co., NM, USA	ng, rl	EU391374	EU391340
<i>C. striata</i>	<i>vreelandii</i>	CFB 0158a (CO-1)	Saguache Co., CO, USA	ng, rl	EU391379	EU391352
<i>C. striata</i>	<i>vreelandii</i>	CFB 0163a (CO-2)	Ouray Co., CO, USA	ng, rl	EU391386	EU391345
<i>C. striata</i>	<i>involuta</i>	JVF 2193 (HID-2)	Hidalgo, Mexico	ng, rl	EU391381	EU391347
<i>C. striata</i>	<i>involuta</i>	JVF 2190 (HID-1)	Hidalgo, Mexico	ng, rl	EU391380	EU391346
<i>C. striata</i>	<i>involuta</i>	JVF 2155* (OAX)	Oaxaca, Mexico	ng, rl	EU391369	EU391335
<i>C. bentleyi</i>	—	JVF 2550* (WV)	Monroe Co., WV, USA	ng, rl	EU391368	EU391334
<i>C. maculata</i>	<i>maculata</i>	CFB 0161c* (CO)	Ouray Co., CO, USA	ng, rl	EU391363	EU391329
<i>C. maculata</i>	<i>occidentalis</i>	JVF 2122* (MI)	MI, USA	ng, rl	EU391362	EU391328
<i>C. maculata</i>	<i>occidentalis</i>	CFB 0012a* (CA)	Tehama Co., CA, USA	ng, rl	EU391364	EU391330
<i>C. maculata</i>	<i>mexicana</i>	JVF 2202* (HID)	Hidalgo, Mexico	ng, rl	EU391365	EU391331
<i>C. mertensiana</i>	—	CFB 0033a* (OR)	Marion Co., OR, USA	ng, rl	EU391367	EU391333
<i>C. bulbosa</i>	—	JVF 2205* (HID)	Hidalgo, Mexico	ng, rl	EU391366	EU391332
<i>C. trifida</i>	—	JVF 2763a (MI)	Iosco Co., MI, USA	g, rl	EU391357	EU391322
<i>C. trifida</i>	—	CFB 0161a* (CO)	Clear Creek Co., CO, USA	g, rl	EU391358	EU391324
<i>C. odontorhiza</i>	—	JVF 2778a* (MI)	Benzie Co., MI, USA	sgt, rl	EU391360	EU391326
<i>C. odontorhiza</i>	—	Salazar-SLP (SLP)	San Luis Potosi, Mexico	sgt, rl	EU391359	EU391325
<i>C. wisteriana</i>	—	JVF 2758a* (OH)	Pickaway Co., OH, USA	sgt, rl	EU391361	EU391326
<i>Aplectrum hyemale</i>	—	JVF 2718*	Hocking Co., OH, USA	g, l	EU391356	EU391322
<i>Cremastra appendiculata</i>	—	Inoue 1	Japan	g, l	EU391354	EU391320
<i>Oreorchis patens</i>	—	Inoue 2	Japan	g, l	EU391355	EU391321

Note: An asterisk indicates that a particular accession was included in analyses of synonymous and nonsynonymous substitution and likelihood analyses—the remaining accessions were pruned. The specific label used for each accession in Figs. 1 and 2 is given in parentheses to the right of each collection number. For habit, g = green, ng = non-green, sgt = some green tissue, l = leafy, rl = reduced leaf.

Châtelain, which is circumboreal, green, and has retained some level of photosynthesis (Montfort and Küsters, 1940). All species of *Corallorhiza* feed upon ectomycorrhizal fungi in the families Russulaceae (*C. maculata* [Rafinesque] Rafinesque, *C. mertensiana* Bongard, *C. wisteriana* Conrad) and Thelephoraceae (*C. striata* Lindley, *C. trifida*, *C. odontorhiza* [Willdenow] Poiret, *C. wisteriana*) (Taylor and Bruns, 1997, 1999; Taylor et al., 2002, 2004). *Corallorhiza*'s closest North American relative is *Aplectrum hyemale* (Willdenow) Nuttall, a leafy, photosynthetic plant. Two other closely related species bear leaves as well: *Oreorchis* Lindley, and *Cremastra* Lindley, both from eastern Asia (Dressler, 1981; Freudenstein, 1994).

RuBisCO is one of the most thoroughly studied plant enzymes (see Kellogg and Juliano (1997) for a survey of amino acid conservation and review of mutational studies of *rbcL*), which allows in-depth comparisons to be made of amino acid changes and structural modifications in important functional domains for mycoheterotrophic orchids like *Corallorhiza*. In green plants, the RuBisCO holoenzyme is a hexadecamer composed of eight plastid-encoded large subunits (LSU) and eight nuclear-encoded small subunits (SSU; Kellogg and Juliano, 1997). The active site is composed of 20 amino acid residues (located in the LSU), while an additional 105 residues are involved in intermolecular associations; the majority of sites being either absolutely conserved or showing little variation in amino acid sequence among seed plants (Kellogg and Juliano, 1997).

In this study, we sequenced *rbcL* for representatives of four genera of subtribe Corallorhizinae (Table 2) to test hypotheses regarding the evolutionary fate of this critical photosynthesis gene. We hypothesize evolutionary patterns in *rbcL* for *Corallorhiza* will reflect those observed in some previous studies of holoparasitic plant taxa (Table 1). We predict that at least some species of *Corallorhiza* have undergone *rbcL* pseudogene formation, and will exhibit hallmark pseudogene patterns when compared to closely related members of Corallorhizinae. For mycoheterotrophs vs. photosynthetic lineages, these patterns will include: (1) presence of premature stop codons; (2) higher incidence of insertions/deletions that cause changes in *rbcL* reading frame; (3) lower ratios of synonymous vs. nonsynonymous nucleotide substitution ( $d_S/d_N$ ); (4) amino acid changes that are not observed across photosynthetic plants; and (5) variation in evolutionary rate among species.

In addition, we aim to further our understanding of relationships within Corallorhizinae, and more specifically, to resolve relationships among species and varieties of *Corallorhiza*—especially within the morphologically variable *C. striata* and *C. maculata* complexes. To aid in phylogenetic analysis, we also sequenced the nuclear ribosomal internal transcribed spacer (ITS). Lastly, we discuss the potential utility of *rbcL* sequence data for phylogeographic studies within the *C. striata* and *C. maculata* complexes.

## 2. Materials and methods

### 2.1. Taxon sampling

At least one accession from each of seven species presented in the monograph of Freudenstein (1997), plus one accession of *Corallorhiza bentleyi* (Freudenstein, 1999) were chosen for *rbcL* and ITS sequencing ( $n = 34$  accessions, Table 2). Three green, leafy taxa—*Corallorhiza*'s three closest relatives—were selected for outgroup comparison and for comparison of leafy vs. reduced-leaf *rbcL* evolution. *Corallorhiza* species not included in the analysis are *C. ekmanii* Mansfeld (Hispaniola), *C. williamsii* Correll (Central Mexico), and *C. macrantha* Schlechter (Mexico) due to their rarity and difficulty in obtaining material.

### 2.2. DNA extraction, amplification, and sequencing

Total DNAs were isolated from either fresh or frozen (−80 °C) floral tissue using a modified CTAB protocol (Doyle and Doyle, 1987). Accessions were PCR amplified using either primer pair *atpBF* (5'-ACA TCC AGY ACT GGG CCA ATA A-3') and *accD1290R* (5'-ACT ACG GAT CCC ATA CTA CC-3'), or *rbcL-5p* (5'-TTG TTG TGA GAA TTC TTA TTC ATG-3') and *rbcL-3p* (5'-GAG TTT CAT TGC AAT CAA TTA GG-3'). The former primer pair had annealing sites in the adjacent genes (*atpB* and *accD*) and the latter had annealing sites flanking *rbcL*'s 5' and 3' ends in the adjacent *atpB-rbcL* and *rbcL-accD* spacers, respectively. PCR was carried out on an MJ Research PTC-100 thermal cycler using primers *atpBF* and *accD1290R* with the following conditions: 3 min at 96 °C followed by 35 cycles of 96 °C (30 s), 60 °C (60 s), and 72 °C (3 min), and a final extension step at 72 °C for 7 min. Conditions for primers *rbcL-5p* and *rbcL-3p* were 96 °C (3 min) followed by 35 cycles of 96 °C (30 s), 62 °C (60 s), and 72 °C (90 s), with a final extension at 72 °C for 5 min. ITS was amplified using primers ITS1 and ITS4 of White et al. (1990). We used an initial step of 95 °C for 3 min, followed by 35 cycles (95 °C for 30 s, 55 °C for 1 min, 72 °C for 1 min), with a final extension step of 72 °C for 5 min. We amplified 1.0 µl of each genomic DNA (100 ng) in a 50 µl reaction with: 1 mM BSA, 2 mM each of the aforementioned forward and reverse primers, 2 mM Tris-HCl (pH 8.8), 0.7 mM MgCl<sub>2</sub>, 7 mM KCl, 200 µM each dNTP, and 1 U of Taq Polymerase. PCR reaction products were purified using the AmPure magnetic bead protocol (Agencourt Bioscience Corp., Beverly, MA, USA) and eluted with 30–45 µl nanopure H<sub>2</sub>O. Several internal sequencing primers were used to ensure coverage of the entire coding region, such that each nucleotide was covered by at least two primer sequences. Internal primers used for sequencing were: *rbcL1f* (5'-ATG TCA CCA CAA ACA GAA AC-3'), 285r (5'-CCA GAC GTA GAG CTC GCA G-3'), 516f (5'-TGT ACT ATT AAA CCA AAA TTG GG-3'), 880f (5'-CAG GCT GGT

ACA GTA GTG GG-3'), 1368r (5'-CTT TCC AAA TTT CAC AAG CAG CA-3'), and 1285f (5'-GCA TGT GTA CAA GCT CGT AAT GAG-3'). Dideoxy-cycle sequencing reactions were carried out on an Eppendorf Mastercycler Gradient using BigDye v3.1 terminator reaction mix and run on an ABI 3100 Genetic Analyzer (Applied Biosystems Inc., Foster City, CA, USA) in 5  $\mu$ l reactions (i.e. 1/4 the suggested volume) using instructions provided by the manufacturer. Cycle sequencing reactions were purified using the CleanSeq (Agencourt Bioscience Corp., Beverly, MA, USA) magnetic bead protocol using two 85% EtOH washes, dried, then eluted with 35  $\mu$ l Hi-Di Formamide (Fisher Biosciences). Contigs were assembled and sequences were checked exhaustively against their electropherograms using Sequencher (GeneCodes, Ann Arbor, MI, USA). Sequences were aligned using ClustalX (Thompson et al., 1997); all nucleotide substitutions and indels were double checked for ambiguities. Alignment was straightforward for both *rbcL* and ITS. Indels were coded for phylogenetic analysis using the simple gap coding method of Simmons and Ochoterena (2000). GenBank accession numbers for *rbcL* and ITS are listed in Table 2.

### 2.3. Phylogenetic analyses

Maximum parsimony (MP) and maximum likelihood (ML) searches were conducted to estimate phylogenetic relationships both within Corallorhizinae and *Corallorhiza* ( $n = 34$  total accessions). Analyses were performed using *rbcL* and nuclear ribosomal ITS individually and then in a combined analysis including insertion/deletion characters. MP analyses were conducted in PAUP\* v4b10 (Swofford, 2002) using heuristic searches with 1000 random addition sequence replicates (RAS) saving multiple trees per replicate (MULTREES), and tree bisection–reconnection (TBR) branch swapping. MP branch support was assessed using the Parsimony Jackknife (JK; Farris et al., 1996), using the same parameters described above for each of 5000 JK replicates, and the “emulate JAC” feature. This feature assigns a specific probability for each character to appear in the resampled matrix ( $1 - 1/e$ ; approx. 63%) instead of deleting a pre-specified number of characters from the original matrix. The Jackknife was chosen over the alternative, Bootstrapping, because it is less sensitive to missing and invariant characters (Farris et al., 1996; Freudenstein et al., 2004).

### 2.4. ML analyses

ML searches were run in PAUP\* v4b10 starting from a neighbor-joining (NJ) tree using 10 RAS and TBR. To test for significant rate heterogeneity among branches in the ML phylogeny, a likelihood ratio test was implemented for *rbcL*. Specifically, the null hypothesis of no difference in likelihood scores between the tree constrained to evolve in a clocklike manner and the unconstrained tree was tested. First, the log-likelihood score was calculated choos-

ing the “enforce molecular clock” option in PAUP\*, which was then repeated for the unconstrained tree. The difference in log-likelihood scores was calculated using the formula  $A = -2 (\ln L_{\text{clock constrained}} - \ln L_{\text{clock unconstrained}})$ , and significance was assessed based on a  $\chi^2$  distribution with  $(n-2)$  degrees of freedom where  $n =$  the number of taxa. Seventeen taxa were pruned from the original analysis of  $n = 34$  taxa for the ML analysis: 15 of these were various accessions of *C. striata*, and two more were additional accessions of both *C. trifida* and *C. odontorhiza* showing little or no intraspecific variation. These accessions were only included in the MP analysis for the purpose assessing phylogeographic utility of *rbcL*, and were not important for ML estimates of rate variation across *Corallorhiza* species. Furthermore, ML searches become exponentially more time consuming as more taxa are added to an analysis. The inclusion of several relatively similar *C. striata* accessions would not necessarily increase the power of our analysis, but would increase the time to completion by orders of magnitude.

“Standard” methods of model selection (i.e. hierarchical likelihood ratio testing between models with various assumptions) were not implemented, because of inconsistent model selection by these methods based on different “starting points” (Pol, 2004). The most general model with the fewest assumptions was chosen: GTR +  $\Gamma$  + I with parameters estimated from the data. Several other models were tested for both constrained and unconstrained searches—in all cases they did not affect the results of the analysis (i.e. topology). All other models are essentially special cases (i.e. more assumption-laden) of the GTR +  $\Gamma$  + I model, so we are confident that our choice of this model is warranted.

### 2.5. Molecular evolutionary analyses

MEGA v.3 (Kumar et al., 2004) was used to calculate synonymous ( $d_S$ ) and nonsynonymous ( $d_N$ ) substitution rates using the modified Nei–Gojobori method with Jukes–Cantor correction, via pairwise comparisons (Table 3). All taxa were compared with *Cremastra appendiculata*—the “leafy” outgroup taxon chosen for phylogenetic analyses—based on preliminary findings of a relatively distant relationship for this species relative to the other two leafy taxa and *Corallorhiza*.

The review of Kellogg and Juliano (1997) was used as a basis to assess deviation from observed seed plant replacement substitutions in *Corallorhiza*. In Fig. 4 of that study, the authors mapped all substitutions observed among a broad sampling of 499 seed plant species. Many of the changes observed in their study involved “togging” between two or a few chemically similar amino acids, and in some cases substitutions were observed in only one of the 499 spermatophyte taxa studied. For *Corallorhiza*, every change beyond those observed in Kellogg and Juliano (1997) was recorded, including “singleton” replacements. We hypothesize that changes seen in *Corallorhiza*

Table 3  
 $d_S/d_N$  ratios  $\pm$  standard errors (SE) for *Corallorhiza* spp. *rbcL* sequences based on pairwise comparisons with *Cremastra appendiculata*

Species	Variety	$d_S$	SE	$d_N$	SE	$d_S/d_N$
<i>Aplectrum hyemale</i>		0.021	0.008	0.004	0.002	5.250
<i>Oreorchis patens</i>		0.015	0.007	0.004	0.002	3.750
<i>C. trifida</i>		0.063	0.014	0.004	0.002	15.750
<i>C. odontorhiza</i>		0.040	0.011	0.007	0.002	5.714
<i>C. wisteriana</i>		0.044	0.011	0.006	0.002	7.333
<i>C. bulbosa</i>		0.021	0.008	0.003	0.002	7.000
<i>C. maculata</i>	<i>mexicana</i>	0.040	0.011	0.005	0.002	8.000
<i>C. maculata</i>	<i>maculata</i>	0.037	0.010	0.009	0.003	4.111
<i>C. maculata</i>	<i>occidentalis</i> (CA)	0.040	0.011	0.010	0.003	4.000
<i>C. maculata</i>	<i>occidentalis</i> (MI)	0.044	0.011	0.010	0.003	4.400
<i>C. mertensiana</i>		0.037	0.010	0.013	0.003	2.846
<i>C. striata</i>	<i>striata</i> (MI)	0.057	0.014	0.021	0.004	2.714
<i>C. striata</i>	<i>vreelandii</i> (NM-1)	0.057	0.014	0.020	0.004	2.850
<i>C. striata</i>	<i>vreelandii</i> (CA-2)	0.062	0.014	0.022	0.004	2.818
<i>C. striata</i>	<i>involuta</i> (OAX)	0.057	0.013	0.026	0.005	2.192
<i>C. bentleyi</i>		0.054	0.012	0.026	0.005	2.077

and not seen among 499 spermatophyte *rbcL* sequences represent evidence of relaxed functional constraint on *rbcL*. This is especially pertinent to sites involved in peptide associations: (1) of the large subunits, (2) between large and small subunits, (3) within the peptide (intradimer interactions), and (4) among the 20 amino acids that collectively comprise the active site—where substrate binding occurs.

## 2.6. Codon-based ML tests of selection

The PAML package (Yang, 1997) was used to test hypotheses of codon-based *rbcL* evolution among *Corallorhiza* species and relatives. The “codeml” module (Yang, 1997) was implemented, which calculates likelihood scores under various models of  $d_S$  and  $d_N$  across sites and branches, given an input tree topology and a nucleotide data matrix (Yang and Nielsen, 1998, 2000; Yang et al., 2000). It was necessary to remove all premature stop

codons and gaps from the analysis, truncating the 17-taxon aligned matrix to 1386 bases (462 codons), to avoid issues caused by missing data in some taxa that could potentially affect the outcomes of molecular evolutionary analyses. These models are based on the premise that  $d_N/d_S$  (henceforth referred to as “ $\omega$ ”, the inverse of  $d_S/d_N$  to which we have previously referred)  $< 1$  for sites under purifying selection,  $\omega = 1$  for sites evolving neutrally, and  $\omega > 1$  for sites under positive Darwinian selection. Model “M0” of Goldman and Yang (1994) assumes a single  $\omega$  across the coding sequence, with  $\omega$  estimated from the data. All subsequent models (denoted *MAi-vi*; sensu Yang et al., 2000) are further modifications thereof, due to the addition of extra parameters. We compared likelihood ratio statistics between various models of codon evolution to test for relaxed purifying selection on *rbcL*. First, model M0 (one  $\omega$  ratio) was tested vs. *M1a* (nearly neutral). The latter model assumes a proportion,  $p_0$ , of conserved sites with

Table 4  
 Models, likelihoods, and parameter estimates for *rbcL* data.

Model	$-\ln L$	parameter estimates	np	$A$	LRT	$p$
M0: one ratio	2735.927	$\omega = 0.357$	34	—	—	—
“free ratios”	2698.233	—	65	75.388	M0, df = 31	<b>&lt;0.001</b>
Site-specific models						
<i>M1a</i> : nearly neutral	2731.696	$p_0 = 0.717, p_1 = 0.283$	35	8.462	M0, df = 1	<b>&lt;0.005</b>
M2: selection	2729.705	$\omega_2 = 5.442$	37	3.982	<i>M1a</i> , df = 2	NS
Branch-site models						
(i.e. branch-specific neutrality) <i>MAi</i>	2728.914	†	37	14.026	M0, df = 2	<b>&lt;0.001</b>
<i>MAi</i>		†	37	5.564	<i>M1a</i> , df = 2	$0.05 < p < 0.1$
<i>MAii</i>	2728.781	†	37	14.292	M0, df = 2	<b>&lt;0.001</b>
<i>MAii</i>		†	37	5.830	<i>M1a</i> , df = 2	$0.05 < p < 0.1$
<i>MAiii</i>	2728.306	†	37	6.678	<i>M1a</i> , df = 2	<b>&lt;0.05</b>
<i>MAiv</i> ( $\omega_2$ fixed = 1)	2729.301	†	36	0.774	<i>Mai</i> , df = 1	NS
<i>MAv</i> ( $\omega_2$ fixed = 1)	2729.269	†	36	0.976	<i>MAii</i> , df = 1	NS
<i>MAvi</i> ( $\omega_2$ fixed = 1)	2727.287	†	36	8.818	<i>M1a</i> , df = 1	<b>&lt;0.01</b>

Note: “np” = number of parameters, “ $A$ ” =  $-2 \times$  the difference in  $\ln L$  scores, “LRT” shows other model(s) to which the model for each row was compared via likelihood ratio testing, “ $p$ ” indicates the statistical significance of the result based on a  $\chi^2$  test with “df” being the degrees of freedom (the difference in no. of parameters between the two models).  $p$ -Values in bold indicate statistical significance, and “NS” indicates nonsignificance.  $\omega_0, \omega_1$ , and  $\omega_2$  (under “parameter estimates”) represent differing  $d_N/d_S$  value estimates for different site classes or branch-site classes within the same model. † = parameter estimates listed in Appendix B.

$\omega_0$  between zero and one, and a proportion of neutral sites,  $p_1 = (1 - p_0)$ , with  $\omega_1 = 1$ . This is the most basic of the “site-specific” models. Then, model *M1a* was tested against model *M2*, the “selection” model. The latter model assumes three site classes for  $\omega$ , including an additional class  $p_2$  with  $\omega_2$  estimated from the data. Lastly, various configurations of “branch-site” models (Yang and Nielsen, 1998) were tested, denoted by “*MAi-vi*,” which allow different  $\omega$ -values along specified branches in the topology (Table 4). The rationale here was to specify neutral evolution for branches in the analysis on which nonsynonymous changes associated with pseudogene formation had occurred. Since the formulation of these models is dependent on phylogeny, each branch-site model is discussed in full detail in Section 3.7 of the results (i.e. for each model, we specify which branches in the phylogeny carry which  $\omega$ -values). A significant difference between two models would suggest a better fit of the model with the higher likelihood score. In this analysis, a statistically significant deviation from strictly purifying selection is considered to be evidence of relaxed functional constraint based on these models, not necessarily positive selection.

### 3. Results

#### 3.1. ITS

MP analysis of nuclear ITS (Fig. 1a) placed *Oreorchis* sister to *Corallorhiza* in 94% of 5000 jackknife (JK) pseudoreplicates. Although ITS failed to give a fully resolved

tree and only recovered a polytomy for interspecies relationships (JK = 100), it did recover major groupings, including *C. odorhiza* + *C. wisteriana* (JK = 84), the *C. maculata* complex (JK = 92), and the *C. striata* complex (JK = 69, not shown in Fig. 1a). Relationships within the *C. maculata* and *C. striata* groups were unresolved by ITS, but it did recover two major clades within the *C. striata* complex: one including *C. bentleyi* + *C. striata* var. *involuta* from Oaxaca, Mexico (JK = 100) and another consisting of all remaining *C. striata* (JK = 100).

#### 3.2. *rbcL*

The *rbcL* locus (Fig. 1b) resulted in a largely similar topology compared to ITS. It provided additional resolution for interspecific relationships within *Corallorhiza*, and further gave resolution within the *C. maculata* and *C. striata* species complexes. *Corallorhiza mertensiana* was placed as sister to all *C. maculata* accessions from the USA (JK = 96). A few minor differences in topology occurred (in most cases due to differences in resolution), most notably the placement of *C. bulbosa* at the base of the *rbcL* tree. Again, *C. bentleyi* and the Oaxacan accession of *C. striata* var. *involuta* formed a clade (JK = 100) sister to all remaining *C. striata*. Surprisingly, all three California accessions formed a distinct clade (JK = 100, 8 synapomorphies), sister to all remaining *C. striata* accessions (JK = 100).

The ML criterion was used under the GTR +  $\Gamma$  + I substitution model to implement a likelihood ratio test of heterogeneity for the *rbcL* gene. We rejected the null

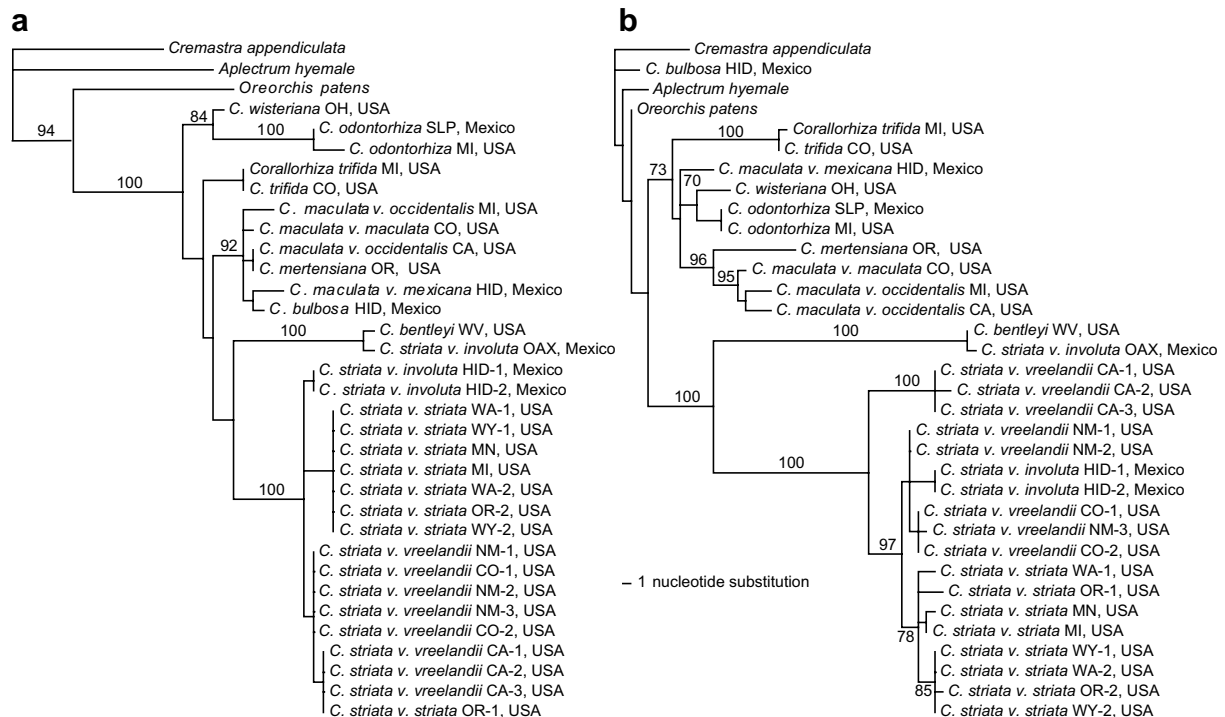


Fig. 1. (a) One of three most parsimonious gene trees based on nuclear ITS. CI = 0.91, RI = 0.95, 61 parsimony-informative sites. (b) One of two most parsimonious gene trees for plastid *rbcL*. CI = 0.89, RI = 0.98, 106 parsimony-informative sites. Numbers above branches indicate jackknife support values (JK) based on 5000 pseudoreplicates (values below 70% not shown).

hypothesis of homogeneous branch lengths for 17 taxa representative of all species and varieties of *Corallorhiza* + *Aplectrum*, *Oreorchis*, and *Cremastra* ( $\lambda = 35.599$ ,  $df = 15$ ,  $p < 0.001$ ). To be consistent between optimality criteria, we repeated an MP analysis with the same 17 taxa from the ML search using the same search parameters described for the original MP search. This was to ensure that the removal of taxa from the original analysis would not create major topological differences between the two searches: we found identical topologies for both (trees not shown).

### 3.3. Combined analysis

The combined data matrix of ITS + *rbcL* yielded a single most parsimonious tree with high support for most internal branches (Fig. 2). Seven indel characters were coded using “simple gap” presence/absence coding following the guidelines of Simmons and Ochoterena (2000). Two insertions and two deletions were observed in *rbcL* for members of the *C. striata* complex, two insertions were autapomorphies in *rbcL* for a *C. maculata* accession from Michigan, USA, and one insertion in ITS was a synapomorphy for the two *C. trifida* accessions (not shown). *Oreorchis* was sister to *Corallorhiza* (JK = 98), and *Corallorhiza* was monophyletic (JK = 99). *Corallorhiza* was composed of two well-supported clades: one represents

the *C. striata* group (JK = 100) and the other the remainder of *Corallorhiza* (JK = 89). Within *Corallorhiza*, the *C. maculata* complex was sister to *C. odontorhiza* + *C. wisteriana*, and this clade (JK = 83) was placed sister to *C. trifida* (JK = 89). *Corallorhiza maculata* var. *mexicana* was sister to *C. bulbosa* (both from Hidalgo, Mexico), but this relationship had low support (JK = 60), value not shown in Fig. 2). Two accessions of the early-flowering *C. maculata* var. *occidentalis* from Michigan and California formed a clade with low support (JK = 62, value not shown in Fig. 2), and these were sister to an accession of the late-flowering *C. maculata* var. *maculata* from Colorado (JK = 92). Sister to these was *C. mertensiana* (JK = 98) from the Pacific Northwest of North America. Thus, in this analysis, *C. bulbosa* and *C. mertensiana* both form clades with different varieties of *C. maculata* in separate regions of North America.

Two major clades emerge within the *C. striata* complex. In the first of these, *Corallorhiza bentleyi* from West Virginia, USA was sister to a small-flowered accession of *C. striata* var. *involuta* from Oaxaca, Mexico (JK = 100). They shared 46 total synapomorphies, indicating a very close relationship. Interestingly, two other small-flowered, sister accessions of *C. striata* var. *involuta* from Hidalgo, Mexico (HID-1 and HID-2) were part of a larger clade with members of *C. striata* var. *vreelandii* from the southwestern USA, although this relationship received low sup-

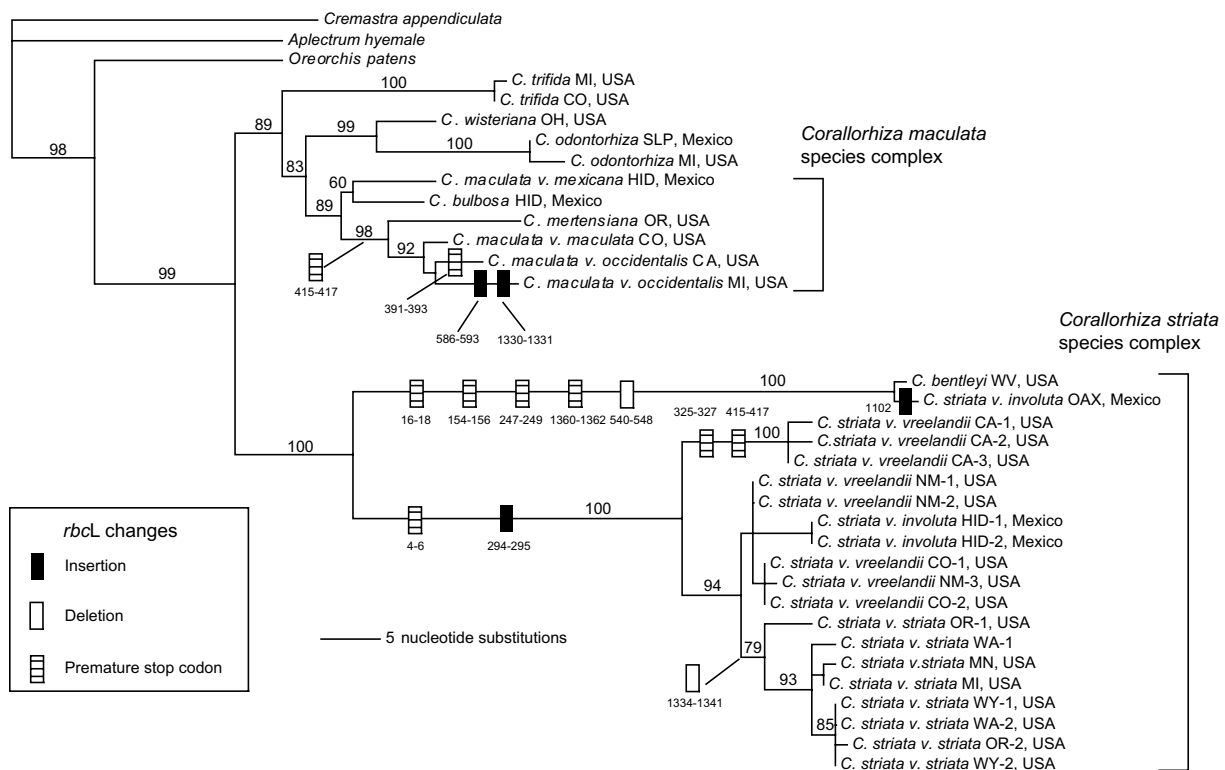


Fig. 2. Single most parsimonious gene tree for the combined ITS + *rbcL* data matrix (including seven indel characters). CI = 0.90, RI = 0.93, 172 parsimony-informative sites. Numbers above branches indicate jackknife support values (JK) based on 5000 pseudoreplicates (values below 70% not shown). Major gene structural changes for *rbcL* are mapped on the phylogeny, with their respective positions in the aligned *rbcL* matrix indicated.

port. Lastly, the large-flowered and strictly northern North American-distributed *C. striata* var. *striata* formed a clade (JK = 79) consisting of six different *rbcL* sequence types.

### 3.4. Premature stop codons and indels

Major structural changes have occurred in the *rbcL* gene for both *C. maculata* and *C. striata*. Fig. 2 shows the most parsimonious construction for a stop codon (positions 415–417) within the *C. maculata* group, shared by *C. mertensiana* and *C. maculata* vars. *maculata* and *occidentalis*. An additional stop codon lies on the branch leading to the California accession of var. *occidentalis*, while two insertions (8 bp, positions 586–593; and 2 bp, positions 1330–1331) are unique to the Michigan accession of var. *occidentalis*. *Corallorhiza striata* shows a number of premature stop codons, insertions and deletions (Fig. 1). The second codon (bases 4–6) for all *C. striata* accessions excluding *C. bentleyi* + *C. striata* var. *involuta* (Oaxaca) is a premature stop codon. This same set of accessions share a 2-base insertion of ‘AT’ at alignment positions 294–295. Within this clade, all California accessions share two additional stop codons, at DNA positions 325–327 and 415–417. Putative members of *C. striata* var. *striata* (i.e. members of the ‘northern’ clade) share an eight base deletion at positions 1334–1341 in the aligned matrix. The branch for *C. bentleyi* and *C. striata* var. *involuta* (Oaxaca) clade shares four stop codons, three at the 5′ and one at the 3′ ends of the gene. In addition, these two taxa share a 9-base deletion (positions 540–548). Although it is possible for indels in multiples of three to exist while still maintaining an intact reading frame, this indel occurs such that it cannot be aligned to preserve an intact reading frame in the 3′ region beyond.

### 3.5. Comparisons of synonymous vs. nonsynonymous substitution

Table 3 lists  $d_N$ ,  $d_S$ , and  $d_S/d_N$  ratios for *Corallorhiza* species in comparison to the leafy species *Cremastra appendiculata*. The photosynthetic *Corallorhiza trifida* showed an exceptionally high value ( $d_S/d_N = 15.750$ ), due to a high number of synonymous changes. *Corallorhiza wisteriana*, *C. odontorhiza*, *C. bulbosa*, and *C. maculata* var. *mexicana* had lower values (5.714–8.000). The values decreased for vars. *maculata*, *occidentalis*, and *C. mertensiana* (2.846–4.111), and even more so for the *C. striata* complex (2.077–2.850).

### 3.6. Amino acid changes in *Corallorhiza* RuBisCO vs. 499 seed plants

Using Kellogg and Juliano (1997) as a basis to assess inferred RuBisCO amino acid changes in *Corallorhiza*, we observed patterns suggestive of pseudogene formation in members of both the *C. striata* and *C. maculata* complexes (Appendix A). Except for a single site in *C. wisteriana* and

*C. odontorhiza*, all other sequences (excluding the *C. striata* and *C. maculata* accessions mentioned above) showed no evidence of amino acid change outside of the changes present in 499 sequences representative of the entire phylogenetic breadth of the spermatophytes. The *C. striata* complex showed the highest degree of change. *Corallorhiza bentleyi* and *C. striata* var. *involuta* (Oaxaca) shared 17 changes not observed in any of the seed plants in Kellogg and Juliano (1997), including two changes at sites involved in intradimer interactions, one change at a large subunit–large subunit (LSU–LSU) interaction site, and one change at a large–small subunit (LSU–SSU) interaction site. *Corallorhiza striata* vars. *vreelandii* (CA and NM) and *striata* had 13, 12 and 12 changes, respectively. Two of these were at sites involved with intradimer interactions for all three accessions, and *C. striata* var. *vreelandii* (NM accession) had a change at one site involved in LSU–SSU interactions. In the *C. maculata* group, *C. mertensiana* showed eight changes not observed in RuBisCO of any other seed plant in Kellogg and Juliano (1997), with two of these at intradimer interaction sites. *Corallorhiza maculata* vars. *occidentalis* (MI and CA accessions) and *maculata* had seven, eight, and five changes, respectively, one of which was a shared change by all three at an intradimer interaction site. No changes occurred at any of the 20 amino acid residues that form the active site.

No unexpected changes (i.e. changes seen in *Corallorhiza* but in none of the 499 plants surveyed by Kellogg and Juliano (1997)) occurred in the leafy taxa *Cremastra appendiculata*, *Aplectrum hyemale*, or *Oreorchis patens*. Likewise, none occurred in the green, photosynthetic *Corallorhiza trifida*, and only one shared change occurred in *C. odontorhiza* + *wisteriana*—two species with at least some green and putatively photosynthetic tissue.

### 3.7. Codon-based ML analysis

Branch-site models  $MAi-vi$  refer to variations of “model A” of Yang and Nielsen (1998).  $MAi$  and  $MAiv$  specify the branches leading to (*C. bentleyi* + *C. striata* var. *involuta* Oaxaca), (*C. striata* vars. *striata* and *vreelandii*), and (*C. maculata* vars. *occidentalis* and *maculata* + *C. mertensiana*). The main difference between these two models is that in the latter,  $\omega_2 = 1$  instead of being estimated from the data. Models  $MAii$  and  $MAv$  exclude the latter branch (*C. maculata* vars. *occidentalis* and *maculata* + *C. mertensiana*).  $MAiii$  and  $MAvi$  specify the same branches as  $MAi$  and  $MAiv$ , plus all internal branches below these (i.e. toward the tips of the tree). In models  $MAiv-vi$   $\omega_2$  is fixed at 1. Appendix B shows parameter estimates of the branch-site models.

Table 4 shows the results of several likelihood ratio tests based on different models of codon evolution.  $\omega$  was heterogeneous among lineages, as shown by comparison of model  $M0$  (one ratio for  $\omega$ ) with the free ratios model ( $\lambda = 75.388$ ,  $p < 0.001$ ). Further, the nearly neutral model ( $M1a$ ) fit the data significantly better than model  $M0$

( $A = 8.462$ ,  $p < 0.005$ ). Comparison of the selection model  $M2$  did not show a significantly better fit than the nearly neutral model  $M1a$  (Table 4). Branch-site model A (Yang and Nielsen, 1998) was compared with models  $M0$  and  $M1a$ , with various branch specifications (Table 4; see Appendix B for parameter estimates). These configurations varied by the inclusion or exclusion of the branch leading to *C. mertensiana* + *C. maculata* vars. *occidentalis* and *maculata* (models  $MAi$  and  $MAiv$ ;  $MAii$  and  $MAv$ , respectively), and in addition by specification of branches within the *C. striata* and *C. maculata* groups (models  $MAiii$  and  $MAvi$ , respectively). When compared with model  $M0$ , all branch specifications of model A tested showed a significantly better fit (Table 4). Models  $MAiii$  ( $A = 6.678$ ,  $p < 0.05$ ) and  $MAvi$  ( $A = 8.818$ ,  $p < 0.01$ ) fit the data significantly better when compared to model  $M1a$  (as suggested in the PAML manual, [<http://abacus.gene.ucl.ac.uk/software/paml.html>]). The latter model ( $MAvi$ ) assumes neutral evolution ( $\omega = 1$ ) along branches leading to and within the *C. striata* and *C. maculata* (US accessions) clades, and purifying selection along all other branches of the phylogeny in Fig. 1b.

## 4. Discussion

### 4.1. Relationships within *Corallorhiza* and phylogeographic utility of *rbcL*

The combined matrix of ITS and *rbcL* yielded a phylogeny concordant with the morphological analysis Freudenstein (1994). The latter analysis failed to resolve relationships among *C. mertensiana*, *C. bulbosa*, and varieties of *C. maculata*, which here are resolved further with slightly increased taxon sampling. The existence of geographically structured sequence variation within the *C. striata* and *C. maculata* groups for *rbcL* illustrates its potential for phylogeographic study in combination with other high-variation loci.

These data provide evidence that a geographic barrier may have played a role in the patterns of diversification in the *C. maculata* group. *Corallorhiza bulbosa* and *C. maculata* var. *mexicana* may have diverged in Mexico, with *C. mertensiana* + *C. maculata* vars. *occidentalis* and *maculata* diverging in northern North America. These findings are in agreement with the restriction site analysis of Freudenstein and Doyle (1994) in that Mexican and US accessions diverge at the base of the group, and that *C. mertensiana* is placed sister to the US and Canadian accessions of *C. maculata* vars. *occidentalis* and *maculata*. However, this remains a hypothesis to be more explicitly tested. Several geographically associated *C. striata* clades emerge from our analysis (Figs. 1 and 2): a northern clade composed of large-flowered var. *striata* (JK = 79), a California clade of var. *vreelandii* (JK = 100), a variable but weakly supported US/Mexican “Cordilleran” clade of var. *vreelandii* and two accessions of var. *involuta* (JK = 58, value not shown in Fig. 2), and *C. bentleyi* + var. *involuta* from Oax-

aca, Mexico (JK = 100). The amount of *rbcL* variation observed in *C. maculata* and especially in *C. striata* is surprising. This locus has typically been useful for higher-level systematic studies, usually above the genus level (Cameron et al., 2001; Chase et al., 1993), including across the entire Orchidaceae (Cameron et al., 1999). The fact that changes in *C. striata rbcL* are geographically correlated suggests that they have occurred relatively recently, and thus are appropriate for phylogeographic study. Further sampling of populations, additional DNA loci, and analysis of morphological variation are all currently underway to address more specific phylogeographic hypotheses for these groups.

### 4.2. Evolution of *rbcL* in *Corallorhiza*

Evidence from this study indicated that *rbcL* has become or is in the process of becoming a pseudogene within some *C. maculata* lineages and within all *C. striata* lineages. The evidence for this includes: premature stop codons (Fig. 2), reading frame shifts due to non-triplet insertions and deletions (Fig. 2), decreased  $d_S/d_N$  ratios (Table 3), replacement substitutions not observed in other photosynthetic species (Appendix B), significant rate heterogeneity (results, Section 3.2), and high likelihood of neutral evolution among specific branches of the *rbcL* phylogeny (Table 4). Taken together, these observations support an evolutionary trajectory of pseudogenization. It is highly improbable that changes in *rbcL* are sequencing errors, because the great majority of them are synapomorphies for closely related lineages (39 of 50 AA changes listed in Appendix A occur in >1 accession). Also, it seems unlikely that the *rbcL* sequences generated in this study represent the result of some past intergenomic gene transfer event (e.g. plastid-to-nucleus, see Wolfe and Randle, 2004), since the primers initially used to amplify *rbcL* were anchored in genes ca. 1 kb from each end of *rbcL* and no evidence of multiple copies was observed. However, these observations alone are not sufficient to rule out gene transfer, and an important future research objective will be to localize *rbcL* in *Corallorhiza* and relatives. Lastly, RNA editing of plastid transcripts must be considered. This is also unlikely to be the case in *Corallorhiza*, due to the fact that the main type of change observed in plants is cytosine-to-uracil conversion (Bock, 2000; Freyer et al., 1997), which would not be sufficient to replace other types of nonsynonymous substitutions or multi-base deletions.

Our findings are in accordance with Wolfe and dePamphilis (1997, 1998) and Randle and Wolfe (2005) although some of the *Boschniakia*, *Hyobanche*, and *Orobanchae* pseudogenes from those studies showed “catastrophic” deletions on the scale of hundreds of bases. We did not find extensive portions of coding sequence deleted in *Corallorhiza rbcL*, but nonetheless are convinced that the 5' stop codons plus the insertions/deletions observed in *C. striata* and *C. maculata* represent pseudogenes. The prevalence of nonsynonymous changes in these two lineages not

observed across the diversity of photosynthetic spermatophytes (Kellogg and Juliano, 1997) further suggests the presence of *rbcL* pseudogenes. More specifically, many of these changes occurred at sites that show absolutely no variation among photosynthetic spermatophytes, including several sites involved in intramolecular association, LSU–LSU association, and LSU–SSU association. However, none of the 20 amino acids associated with the RuBisCO LSU active site (Knight et al., 1990) suffered replacement substitutions for any taxa in our analysis. This does not necessarily imply functional constraint on *rbcL* in all accessions: these sites may have, by chance alone, not experienced mutations. However, it may not be entirely meaningful in a biological sense to speak of substitution in sequences without open reading frames. Nonetheless, we investigated these properties following the rationale that if the sequences for some reason maintained high  $d_S/d_N$  ratios, the possibility of functional constraint could not be abandoned—this was certainly not the case for the aforementioned *Corallorhiza* lineages. We chose not to treat  $d_S/d_N$  ratios statistically, due to the issue of common ancestry. The changes observed in the *C. striata* lineages (Fig. 2) occur mainly along the two deepest branches, followed by divergence of several *rbcL* sequence types (Figs. 1b and 2). Thus, treating several individuals that share changes along a single ancestral branch as replicates for the purpose of reaching a suitable sample size for statistical analysis represents a violation of data independence: shared mutations are not random samples. This being stated, there was a clear trend of decreased  $d_S/d_N$  ratios for the *C. striata* and *C. maculata* groups (Table 3).

According to a molecular clock analysis using a likelihood ratio test (Section 3.2), and branch lengths observed in Fig. 1b, evolutionary rates differed across branches in the *Corallorhiza rbcL* gene tree. However, we interpret these findings with caution, given the pervasive nature of rate heterogeneity in previous phylogenetic studies of *rbcL* and other plastid genes like *matK* in both photosynthetic (Bosquet et al., 1992; Zhong et al., 2002 [Rhizophoraceae]; Huang and Price, 2003 [Gnetaceae]) and nonphotosynthetic plant lineages (references in Table 1). Rates were conspicuously elevated in the *C. striata* group, most notably along the branches leading to *C. bentleyi* + var. *involuta* (Oaxaca) and the remainder of *C. striata* (Fig. 1b). Interestingly, the lineage comprising *C. bentleyi* + *C. striata* var. *involuta* (Oaxaca) and the lineage leading to the remainder of *C. striata* have followed independent paths of *rbcL* degradation (Fig. 2). The former clade shares 46 synapomorphic changes, and the latter 28 (with 10 changes shared between these two sister clades; Fig. 1b). The former two branches are longer than any seen within *Corallorhiza*, and each harbor unique structural changes (stop codons and frameshifts; Fig. 2). Furthermore, no “drastic” nonsynonymous substitutions (Appendix A) were shared between these lineages. Thus, it is evident that pseudogene formation has occurred after the divergence of the two lineages.

Likelihood ratio tests of different codon substitution models support a deviation from purifying selection (Table 4, see Appendix B for model parameters). The “nearly neutral” model showed a significantly better fit than the “one ratio” model, as did the “free ratios” model. These findings are not surprising, due to the unrealistic assumption of a single  $\omega$  across all sites and lineages. The variation in  $\omega$  among lineages and/or sites observed in this study is not sufficient to identify adaptive (positive) evolution, because it cannot be distinguished from the effect of relaxed purifying selection (Yang and Bielawski, 2000). Therefore, it was necessary to more specifically test for neutral evolution along selected branches and at a proportion of sites using branch-site models. Model *M<sub>Avi</sub>* showed the best fit of any branch-site model ( $p < 0.01$ ). In this model, we specified branches where evidence of pseudogene formation was observed (leading to and within all accessions of both the *C. striata* and *C. maculata* groups, excluding var. *mexicana* and *C. bulbosa*; refer to Fig. 2), and set  $\omega = 1$  [neutral evolution], while the remaining branches evolved at  $0 < \omega < 1$  [purifying selection]. These findings support our hypothesis of neutral evolution along specific branches and purifying selection on *rbcL* carboxylase function throughout the lineages that comprise the rest of the tree. The “selection” model, *M<sub>2</sub>*, did not fit the data significantly better than the nearly neutral model. Given the structural changes (Fig. 2) to *rbcL* among the *C. striata* and *C. maculata* lineages, it is unlikely that positive selection has caused the observed evolutionary patterns. If this were the case, we would expect to observe a significantly higher likelihood for model *M<sub>2</sub>* (selection)—we did not observe this pattern in our analysis.

We interpret our findings of pseudogenes in some lineages of *Corallorhiza* but not in others as being indicative of the early stages of pseudogene formation as a result of the switch from autotrophy to mycoheterotrophy. In this sense, *Corallorhiza* provides a useful window into the process of photosynthetic gene degradation as a result of the relaxed functional constraint of photosynthesis. We have observed that pseudogene formation in the *Corallorhiza* lineages presented in this study correlates roughly with the absence of visible green coloration. The three leafy taxa (*Cremastra*, *Oreorchis* and *Aplectrum*) plus *Corallorhiza trifida*, *C. wisteriana*, and *C. odontorhiza* all have some level of green pigmentations, although in the latter two species this is spatially localized within the plant. *Aplectrum* has a winter leaf and performs photosynthesis over the winter months. *Corallorhiza trifida* has the most visible green pigmentation in the genus, ranging from very bright green when found in closed-canopy forest, to greenish yellow or brown in more open situations like peat bogs/muskeg. *Corallorhiza odontorhiza* and *C. wisteriana* both often have dull to bright green coloration in their developing ovaries. Ratios of  $d_S/d_N$  in *rbcL* are approximately four times

lower in the non-green *C. maculata* when compared to the green *C. trifida* and seven times lower for *C. striata*, driven in the latter by a 5- to 6.5-fold increase in  $d_N$  (Table 3). Although *C. trifida* is the only species in which photosynthesis has been demonstrated (Montfort and Küsters, 1940), this pattern suggests that *C. trifida*, *C. odontorhiza*, and *C. wisteriana* species may still perform photosynthesis. The presence of green pigmentation in specific tissues of *C. wisteriana* and *C. odontorhiza* indicate localized expression of chlorophyll. Further, Cummings and Welschmeyer (1998) found low levels of chlorophyll A and B in *C. maculata*, suggesting that lack of visible green pigment in *C. maculata* is due to the masking effect of various flavonoid-derived anthocyanins over chlorophyll, rather than a complete lack of chlorophyll. Although it cannot be assumed that either the preservation of RuBisCO function (i.e. photosynthesis) or the apparent presence of green pigmentation are directly causative, it may be the case that they are correlated within some plant-parasitic and mycoheterotrophic plant lineages. An obvious next step will be to determine chlorophyll A and B concentrations and photosynthetic activity in *Corallorhiza* species.

We provide the first specific investigation of synonymous and nonsynonymous evolutionary changes in plastid loci for mycoheterotrophic plants. Cameron (2004) and Cameron and Molina (2006) observed putative pseudogenes for four photosynthesis loci (*rbcL*, *psaB*, *psbB*, and *psbC*) in fully mycoheterotrophic species of the vanilloid orchid genus *Cyrtosia* (see Table 1). For example, *Cyrtosia septentrionalis* had two *rbcL* deletions (287 and 8 bp), while *psaB* had 17 deletions (ca. 8 bp each). Thus, the degree of

*rbcL* deletion observed by Cameron (2004) for *Cyrtosia* is more extensive than that observed in the *Corallorhiza maculata* and *C. striata* complexes. Overall, a similar trend of photosynthesis gene degradation emerges from both of these mycoheterotrophic taxa. Leake (1994) estimated that 209 species of putative mycoheterotrophic orchids exist, distributed across 45 genera, 15 of which contain both mycoheterotrophs and non-mycoheterotrophs. Thus it remains to be explored whether or not *Corallorhiza* is a “typical” mycoheterotrophic genus that encapsulates the molecular evolutionary processes that have occurred or are occurring in other taxa that contain mycoheterotrophs.

### Acknowledgments

The authors wish to thank D. Jolles, J. Horky, T. Todsen, G. Salazar, and K. Inoue for providing specimens. We thank the following US Department of Agriculture Forest Service regional botanists for their assistance in obtaining permission to collect orchids: D. Golnick, F. Duran, T. Prendusi, A. Kratz, and D. Harris. We thank K. Tignor (Virginia Department of Agriculture) for permission to collect non-destructive samples of the endangered *Corallorhiza bentleyi*. We also thank A. Wolfe, D. Jolles, J. Morawetz, E. Rothacker, S. Krosnick, S. Bentley, and D. Taylor for helpful feedback. This research was funded by the National Science Foundation (Grant DEB-0415920) and the American Orchid Society (Graduate Student Research Award).

### Appendix A

Replacement substitutions in *rbcL* unique to *Corallorhiza* spp. and not observed across 499 spermatophyte taxa (Kellogg and Juliano, 1997)

Residue	AA	Changed to	Residue function	Structural motif	Accessions
2	S	Stop (TAA)			ben, inv
6	E	Stop (TGA)			str, vreCA, vreNM
9	A,K	V			ben, inv
38	A	T		$\beta$ -Sheet B	ben, inv
46	P,L	S			ben, inv
52	E	Stop (TGA)		$\alpha$ -Helix B	ben, inv
56	A,V,P,S	T		$\alpha$ -Helix B	ben, inv
57	V	L		$\alpha$ -Helix B	str, vreNM
58	A	T		$\alpha$ -Helix B	mer
82	G,A,W	V			vreCA
83	R	Stop (TGA)		$\beta$ -Sheet C	ben, inv
89	R,A,K,P,S,T,V	I		$\beta$ -Sheet C	mer, mac, occCA, occMI
109	E	Stop (TAA)			vreCA
126	G	C	Intradimer interaction		mer
131	R	Stop (TGA)		$\beta$ -Sheet E	occCA
139	R	Stop (TGA)		$\beta$ -Sheet E	vreCA, mer, mac, occCA, occMI

**Appendix A.** (continued)

Residue	AA	Changed to	Residue function	Structural motif	Accessions
217	R,G	L		$\alpha$ -Helix 2	mer
239	Y,C	H		$\beta$ -Sheet 3	occMI
243	T,A,S	I		Loop 3	ben, inv
244	A,R,G,S,T	V		Loop 3	str, vreNM
260	L,F	S		$\alpha$ -Helix 3	occCA, occMI
276	A,S	T		$\alpha$ -Helix 4	str, vreCA, vreNM
281	A,S	F		$\alpha$ -Helix 4	mac, occCA, occMI
289	L,P,S	I			str, vreCA, vreNM
296	A	T	Intradimer interaction		mer
304	Q,H	P	Intradimer interaction		mac, occCA, occMI
312	R	N	Intradimer interaction	$\alpha$ -Helix 5	ben, inv
312	R	C	Intradimer interaction	$\alpha$ -Helix 5	str, vreCA, vreNM
329	G	D			occCA
336	E	G		Loop 5	mer
337	G	R			ben, inv
339	R	H		$\alpha$ -Helix 6	str, vreCA, vreNM
340	D,E,G,N,Q,L	K		$\alpha$ -Helix 6	vreCA
343	L	F		$\alpha$ -Helix 6	str, vreCA
347	V,D,G,H,I	Y		$\alpha$ -Helix 6	ben, inv
350	R	H		$\alpha$ -Helix 6	str, vreCA, vreNM
352	D,E,G,N	Y		$\alpha$ -Helix 6	str, vreCA, vreNM
360	R,P,L	C			str, vreCA, vreNM
369	V,A,L	I	Dimer–dimer interaction		wi, od
370	S	F	Dimer–dimer interaction		ben, inv
375	L	P			ben, inv
383	H	N			str, vreCA, vreNM
395	G,R	W			mer
405	G	E		$\alpha$ -Helix P	ben, inv
407	L,I	S	Intradimer interaction	$\alpha$ -Helix P	mer
418	V,A	L	LSU–SSU interaction	$\alpha$ -Helix 8	ben, inv
421	R,S	W	LSU–SSU interaction	$\alpha$ -Helix 8	vreNM
441	G,D,S	V		$\alpha$ -Helix G	ben, inv
448	A,P,R,D,L,T	V			ben, inv
457	A,S	D			mac, occCA, occMI

Note: Accessions used in this analysis were abbreviated as follows: ben = *C. bentleyi*, inv = *C. striata* var. *involuta* (JVF 2155, Oaxaca), str = *C. striata* var. *striata* (CFB 0002a, Michigan), vreCA = *C. striata* var. *vreelandii* (CFB 0008, California), vreNM = *C. striata* var. *vreelandii* (CFB 0103e, New Mexico), wi = *C. wisteriana* (JVF 2758a, Ohio), od = *C. odorhiza* (JVF 2778a, Michigan), mer = *C. mertensiana* (CFB 0033a, Oregon), mac = *C. maculata* var. *maculata* (CFB 0161a, Colorado), occMI = *C. maculata* var. *occidentalis* (JVF 2122, Michigan), and occCA = *C. maculata* var. *occidentalis* (CFB 0012a, California).

**Appendix B**

Parameter estimates of branch-site codon models

Model/site class	0	1	2a	2b
MAi				
Proportion	0.648	0.207	0.110	0.035
Background $\omega$	0.048	1.000	0.048	1.000
Foreground $\omega$	0.048	1.000	2.797	2.797
MAii				
Proportion	0.063	0.210	0.103	0.033
Background $\omega$	0.493	1.000	0.049	1.000
Foreground $\omega$	0.493	1.000	3.097	3.096

(continued on next page)

## Appendix B. (continued)

Model/site class	0	1	2a	2b
MAiii				
Proportion	0.675	0.146	0.147	0.032
Background $\omega$	0.059	1.000	0.059	1.000
Foreground $\omega$	0.059	1.000	3.245	3.245
Maiv				
Proportion	0.447	0.166	0.282	0.105
Background $\omega$	0.017	1.000	0.017	1.000
Foreground $\omega$	0.017	1.000	1.000	1.000
MAv				
Proportion	0.440	0.165	0.287	0.108
Background $\omega$	0.017	1.000	0.017	1.000
Foreground $\omega$	0.017	1.000	1.000	1.000
MAvi				
Proportion	0.422	0.119	0.358	0.101
Background $\omega$	0.022	1.000	0.022	1.000
Foreground $\omega$	0.022	1.000	1.000	1.000

## References

- Atwood, J.T., 1986. The size of the Orchidaceae and the systematic distribution of the epiphytic orchids. *Selbyana* 9, 171–186.
- Benharrat, H., Delavault, P., Theodet, C., Figureau, C., Thaluarn, P., 2000. *rbcL* plastid pseudogene as a tool for *Orobanchae* (subsection *Minores*) identification. *Plant Biol.* 2, 34–39.
- Berg, S., Krause, K., Kupinska, K., 2002. The *rbcL* genes of two *Cuscuta* species, *C. gronovii* and *C. subinclusa*, are transcribed by the nuclear-encoded plastid RNA polymerase (NEP). *Planta* 219, 541–546.
- Bidartondo, M.I., 2005. Tansley review: the evolutionary ecology of myco-heterotrophy. *New Phytol.* 167, 335–352.
- Bock, R., 2000. Sense from nonsense: how the genetic information of chloroplasts is altered by RNA editing. *Biochemie* 82, 549–557.
- Bömmer, D., Haberhausen, G., Zetsche, K., 1993. A large deletion in the plastid DNA of the holoparasitic flowering plant *Cuscuta reflexa* concerning two ribosomal proteins (*rpl2*, *rpl23*) one transfer RNA (*trnI*) and an ORF 2280 homologue. *Curr. Genet.* 24, 171–176.
- Bosquet, J., Strauss, S., Doerksen, A., Price, R., 1992. Extensive variation in evolutionary rate of *rbcL* gene sequences among seed plants. *Proc. Natl. Acad. Sci. USA* 89, 7844–7848.
- Caddick, L., Rudall, P., Wilkin, P., Hedderson, T., Chase, M., 2002. Phylogenetics of Dioscoreales based on combined analyses of morphological and molecular data. *Bot. J. Linn. Soc.* 138, 123–144.
- Cameron, K., 2004. Utility of plastid *psaB* sequences for investigating intrafamilial relationships within the Orchidaceae. *Mol. Phyl. Evol.* 31, 1157–1180.
- Cameron, K., Molina, M., 2006. Photosystem II gene sequences of *psbB* and *psbC* clarify the phylogenetic position of *Vanilla* (Vanilloideae, Orchidaceae). *Cladistics* 22, 239–248.
- Cameron, K., Chase, M., Whitten, M., Kores, P., Jarrell, D., Albert, V., Yukawa, T., Hills, H., Goldman, D., 1999. A phylogenetic analysis of the Orchidaceae: evidence from *rbcL* nucleotide sequences. *Am. J. Bot.* 86, 208–224.
- Cameron, K., Chase, M., Anderson, W., Hills, H., 2001. Molecular systematics of Malpighiaceae: evidence from plastid *rbcL* and *matK* sequences. *Am. J. Bot.* 88, 1847–1862.
- Chase, M.W. et al., 1993. Phylogenetics of seed plants: an analysis of nucleotide sequences from the plastid gene *rbcL*. *Ann. Missouri Bot. Gard.* 80, 528–580.
- Colwell, A., 1994. Genome evolution in a non-photosynthetic plant, *Conopholis americana*. Ph.D. dissertation, Washington University.
- Cummings, M., Welschmeyer, N., 1998. Pigment composition of putatively achlorophyllous angiosperms. *Plant Syst. Evol.* 210, 105–111.
- Delavault, P., Thaluarn, P., 2002. The obligate root parasite *Orobanchae cumana* exhibits several *rbcL* sequences. *Gene* 297, 85–92.
- Delavault, P., Sakanyan, V., Thaluarn, P., 1995. Divergent evolution of two plastid genes, *rbcL* and *atpB*, in a non-photosynthetic parasitic plant. *Plant Mol. Biol.* 29, 1071–1079.
- dePamphilis, C.W., Palmer, J.D., 1990. Loss of photosynthetic and chlororespiratory genes from the plastid genome of a parasitic flowering plant. *Nature (London)* 348, 337–339.
- Doyle, J., Doyle, J., 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem. Bull.* 19, 11–15.
- Dressler, R., 1981. *The Orchids: Natural History and Classification*. Harvard University Press, Cambridge, MA, USA.
- Dressler, R., 1993. *Phylogeny and Classification of the Orchid Family*. Cambridge University Press, Cambridge, UK.
- Farris, J., Albert, V., Källersjö, M., Lipscomb, D., Kluge, A., 1996. Parsimony jackknifing outperforms neighbor-joining. *Cladistics* 12, 99–124.
- Freudenstein, J.V., 1992. Systematics of *Corallorhiza* and the Corallorhizinae (Orchidaceae). Ph.D. dissertation, Cornell University, Ithaca, NY.
- Freudenstein, J.V., 1994. Character transformation and relationships in *Corallorhiza* (Orchidaceae: Epidendroideae). II. Morphological variation and phylogenetic analysis. *Am. J. Bot.* 81, 1458–1467.
- Freudenstein, J.V., 1997. A monograph of *Corallorhiza* (Orchidaceae). *Harv. Pap. Bot.* 10, 5–51.
- Freudenstein, J.V., 1999. A new species of *Corallorhiza* (Orchidaceae) from West Virginia, USA. *Novon* 9, 511–513.
- Freudenstein, J., Doyle, J., 1994. Plastid DNA, morphological variation and the phylogenetic species concept: The *Corallorhiza maculata* (Orchidaceae) complex. *Syst. Bot.* 19, 273–290.
- Freudenstein, J., van den Berg, C., Goldman, D., Kores, P., Molvray, M., Chase, M., 2004. An expanded plastid DNA phylogeny of Orchidaceae and analysis of jackknife branch support strategy. *Am. J. Bot.* 91, 149–157.
- Freyer, R., Neckermann, R., Maier, R., Kossel, H., 1995. Structural and functional analysis of plastid genomes from parasitic plants: loss of an intron within the genus *Cuscuta*. *Curr. Genet.* 27, 580–586.
- Freyer, R., Kiefer-Meyer, M.-C., Kossel, H., 1997. Occurrence of RNA editing in all major lineages of land plants. *Proc. Natl. Acad. Sci. USA* 94, 6285–6290.

- Goldman, N., Yang, Z., 1994. A codon-based model of nucleotide substitution for protein-coding DNA sequences. *Mol. Biol. Evol.* 11, 725–736.
- Huang, J., Price, R., 2003. Estimation of the age of extant *Ephedra* using chloroplast *rbcL* sequence data. *Mol. Biol. Evol.* 20, 435–440.
- Kellogg, E., Juliano, N., 1997. The structure and function of RuBisCO and their implications for systematic studies. *Am. J. Bot.* 84, 413–428.
- Knight, S., Andersson, I., Brändén, C., 1990. Crystallographic analysis of ribulose 1,5 biphosphate carboxylase from spinach at 2.4 Å resolution. *J. Mol. Biol.* 213, 112–160.
- Kumar, S., Tamura, K., Nei, M., 2004. MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. *Brief. Bioinform.* 5, 150–163.
- Leake, J., 1994. Tansley Review No. 69. The biology of myco-heterotrophic ('saprophytic') plants. *New Phytol.* 127, 171–216.
- Manen, J., Habashi, C., Jeanmonod, D., Park, J., Schneeweiss, G., 2004. Phylogeny and intraspecific variability of holoparasitic *Orobanche* (Orobanchaceae) inferred from plastid *rbcL* sequences. *Mol. Phyl. Evol.* 33, 482–500.
- Merckx, V., Schols, P., Maas-van de Kamer, H., Maas, P., Huysmans, S., Smets, E., 2006. Phylogeny and evolution of Burmanniaceae (Dioscoreales) based on nuclear and mitochondrial data. *Am. J. Bot.* 93, 1684–1698.
- Molvray, M., Kores, P., Chase, M., 2000. Polyphyly of mycoheterotrophic orchids and functional influences on floral and molecular characters. In: Wilson, K., Morrison, D. (Eds.), *Monocots: Systematics and Evolution*. CSIRO Publishing, Collingwood, Victoria, Australia, pp. 441–448.
- Montfort, C., Küsters, E., 1940. Saprophytismus und photosynthese. I. Biochemische physiologische Studien an Humus-Orchideen.. *Botanisches Archiv.* 40, 571–633.
- Pol, D., 2004. Empirical problems of the hierarchical likelihood ratio test for model selection. *Syst. Biol.* 53, 949–962.
- Randle, C., Wolfe, A., 2005. The evolution and expression of *rbcL* in holoparasitic sister-genera *Harveya* and *Hyobanche* (Orobanchaceae). *Am. J. Bot.* 92, 1575–1585.
- Rasmussen, H., 1995. *Terrestrial Orchids: From Seed to Mycoheterotrophic Plant*. Cambridge University Press, Cambridge, UK.
- Simmons, M., Ochoterena, H., 2000. Gaps as characters in sequence-based phylogenetic analyses. *Syst. Biol.* 49, 369–381.
- Swofford, D., 2002. *PAUP\*: Phylogenetic Analysis Using Parsimony\* (and Other Methods)*. Sinauer, Sunderland, MA, USA.
- Taylor, D., Bruns, T., 1997. Independent, specialized invasions of ectomycorrhizal mutualism by two nonphotosynthetic orchids. *Proc. Natl. Acad. Sci. USA* 94, 4510–4515.
- Taylor, D., Bruns, T., 1999. Population, habitat and genetic correlates of mycorrhizal specialization in the 'cheating' orchids *Corallorhiza maculata* and *Corallorhiza mertensiana*. *Mol. Ecol.* 8, 1719–1732.
- Taylor, D., Bruns, T., Leake, J., Read, D., 2002. Mycorrhizal specificity and function in myco-heterotrophic plants. In: van der Heijden, M., Sanders, I. (Eds.), *The Ecology of Mycorrhizas*, vol. 157. Springer, Berlin, pp. 375–414.
- Taylor, D., Bruns, T., Hodges, S., 2004. Evidence for mycorrhizal races in a cheating orchid. *Proc. R. Soc. B.* 271, 35–42.
- Thompson, J., Gibson, T., Plewniak, F., Jeanmougin, F., Higgins, D., 1997. The CLUSTALX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* 25, 4876–4882.
- White, T., Bruns, T., Lee, S., Taylor, J., 1990. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In: Innis, M., Gelfand, D., Sninsky, J., White, T. (Eds.), *PCR Protocols: A Guide to Methods and Applications*. Academic Press, Inc., New York, pp. 315–322.
- Wolfe, A., dePamphilis, C., 1997. Alternate paths of evolution for the photosynthetic gene *rbcL* in four nonphotosynthetic species of *Orobanche*. *Plant Mol. Biol.* 33, 965–977.
- Wolfe, A., dePamphilis, C., 1998. The effect of relaxed functional constraints on the photosynthesis gene in photosynthetic and non-photosynthetic parasitic plants. *Mol. Biol. Evol.* 15, 1243–1258.
- Wolfe, A., Randle, C., 2004. Recombination, heteroplasmy, haplotype polymorphism, and paralogy in plastid genes: implications for plant molecular systematics. *Syst. Bot.* 29, 1101–1120.
- Yang, Z., 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. *Comp. Appl. Biosci.* 13, 555–556.
- Yang, Z., Bielawski, J., 2000. Statistical methods for detecting molecular adaptation. *Trends Ecol. Evol.* 15, 496–503.
- Yang, Z., Nielsen, R., 1998. Synonymous and nonsynonymous rate variation in nuclear genes of mammals. *J. Mol. Evol.* 46, 409–418.
- Yang, Z., Nielsen, R., 2000. Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. *Mol. Biol. Evol.* 17, 32–43.
- Yang, Z., Nielsen, R., Goldman, N., Pedersen, A.-M., 2000. Codon-substitution models for heterogeneous selection pressure at amino acid sites. *Genetics* 155, 431–449.
- Young, N., dePamphilis, C., 2005. Rate variation in parasitic plants: correlated and uncorrelated patterns among plastid genes of different function. *BMC Evol. Biol.* 16 (1–10).
- Zhong, Y., Zhao, Q., Huang, Y., Hasegawa, M., 2002. Detecting evolutionary rate heterogeneity among mangroves and their close terrestrial relatives. *Ecol. Lett.* 5, 427–432.