

## RELATIONSHIPS AND EVOLUTION OF *MATK* IN A GROUP OF LEAFLESS ORCHIDS (*CORALLORHIZA* AND *CORALLORHIZINAE*; ORCHIDACEAE: EPIDENDROIDEAE)<sup>1</sup>

JOHN V. FREUDENSTEIN<sup>2,4</sup> AND DIANA M. SENYO<sup>3</sup>

<sup>2</sup>Ohio State University Herbarium, Department of Evolution, Ecology and Organismal Biology, 1315 Kinnear Road, Columbus, Ohio 43212 USA; and <sup>3</sup>Department of Biological Sciences, Kent State University, Kent, Ohio 44242 USA

Corallorhizinae are a small group of Old and New World temperate orchids of which a core monophyletic group comprises *Govenia*, *Cremastra*, *Aplectrum*, *Oreorchis* and the leafless *Corallorhiza*, and which according to phylogenetic analysis of nuclear ITS and plastid *matK* sequences, are related in this way: (*Govenia* (*Cremastra* (*Aplectrum* (*Oreorchis* (*Corallorhiza*))))). This hypothesis is consistent with the progressive deletion of the *trnK* intron and *matK* ORF. Frameshift-resulting indels yield a predicted loss of translation for the critical “domain X” region of *matK* and are evidence that *matK* is a probable pseudogene in *Aplectrum*, *Oreorchis*, and *Corallorhiza*. Within *Corallorhiza*, a previous hypothesis based on plastid DNA restriction site analysis is confirmed, with the thickened-labellum *C. striata* group being sister to the thin-labellum remainder of the genus, within which the circumboreal *C. trifida* is sister to the remainder, which then comprise two further sister groups: *C. maculata* + *C. bulbosa* + *C. mertensiana* and *C. odorhiza* + *C. wisteriana*. A close relationship between *C. striata* and the recently described Appalachian *C. bentleyi* is shown; in particular, *C. bentleyi* is more closely allied to a southern Mexican population of *C. striata* than it is to northern North American *C. striata* populations, suggesting that two lineages, each with Mexican and northern North American populations, exist within the *C. striata* group.

**Key words:** *Corallorhiza*; *matK*; molecular evolution; Orchidaceae; plastid DNA.

Orchidaceae are the largest family of flowering plants and have the greatest number of independent derivations of the leafless, heterotrophic habit of any angiosperm family, numbering at least 10 (Dressler, 1993; Molvray et al., 2000). The fungal associations believed to occur with all orchids make them pre-adapted for this shift. The largest temperate genus of leafless putative heterotrophs in the family is *Corallorhiza* Gagnebin, comprising 11 species that are confined to the northern hemisphere of the New World except for one, *C. trifida*, which is circumboreal. *Corallorhiza* is placed in subtribe Corallorhizinae (Calypsoeae; Chase et al., 2003) with several other genera of temperate terrestrials; the composition of the subtribe or its equivalent has varied throughout the history of orchid classification (cf. Freudenstein, 1994b). Dressler (1981) included *Aplectrum*, *Corallorhiza*, *Cremastra*, *Dactylostalix*, *Didickea*, *Ephippianthus*, *Govenia*, *Oreorchis*, and *Tipularia* in this subtribe and distinguished it from Calypsoeae, which comprised *Calypso* and *Yoania*. Subsequently (Dressler, 1993), he merged the two groups but excluded *Govenia* in part on the basis of its different root velamen type (*Cymbidium* rather than *Calanthe* type). *Govenia* occurs from Mexico to South America, which is a somewhat anomalous distribution for this subtribe, the other members of which are predominantly North American-Eurasian. Freudenstein (1994b) analyzed morphological features for the group and resolved two principal clades—one comprising *Aplectrum*, *Corallorhiza*, *Cremastra*, and *Oreorchis* (the “*Corallorhiza* clade”), and another comprising *Calypso*, *Changnienia*, *Tipularia*, and *Yoania* (the “*Calypso* clade”). There was

no resolution among the genera of the *Corallorhiza* clade, whose members share the unusual feature of a “winter leaf” (except for the leafless *Corallorhiza*). In these genera, the leaf emerges in the autumn after leaves have fallen from trees and persists through the winter, withering in spring at or before flowering. *Govenia* was resolved as part of the sister group to these two clades. More recently, a subfamily-wide analysis of molecular data confirmed the monophyly of the *Corallorhiza* clade and indicated that *Govenia* is sister to it, but left the relationship of the *Calypso* clade to this group less certain due to weak branch support (J. Freudenstein, unpublished data).

Freudenstein (1997) monographed *Corallorhiza*, and Freudenstein and Doyle (1994) and Freudenstein (1994a) analyzed morphological characters and restriction site variation in plastid DNA to produce a cladogram of species relationships for the genus. Although the pattern was well resolved, they were unable to find a molecular synapomorphy for *Corallorhiza*, and the only morphological synapomorphies were loss characters (leaves and roots), leaving open the possibility that the genus is not monophyletic.

Much morphological variation in *Corallorhiza* has been described in North America over the last 200 years, resulting in a proliferation of named taxa, most of which are not now recognized. Nonetheless, significant variation is still occasionally being uncovered. Freudenstein (1999a) named *C. bentleyi* from a single population in West Virginia; additional populations in West Virginia and Virginia have since been found. *Corallorhiza bentleyi* is similar to *C. striata* in its floral structure, but the flowers of the former are much smaller than those of eastern North American *C. striata* var. *striata* and are most similar to those from small-flowered populations of the species from southern Mexico, which are known as *C. striata* var. *involuta*. *Corallorhiza bentleyi* also flowers much later than *C. striata* (late July as opposed to early June), also resembling the southern Mexican populations.

<sup>1</sup> Manuscript received 8 September 2007; revision accepted 11 January 2008.

The authors thank K. Inoue and M. W. Chase for providing plant material and DNA, respectively, E. Greenwood and G. Salazar for field assistance, and C. Barrett and A. Chaudhuri for technical assistance. This project was supported by NSF grant DEB-0415920 to J.V.F.

<sup>4</sup> Author for correspondence (e-mail: freudenstein.1@osu.edu)

Leafless, putatively achlorophyllous angiosperms such as *Corallorhiza* often experience deletions in their plastid genomes, and in some cases this can be severe (e.g., *Epifagus*; Wolfe et al., 1992). Freudenstein and Doyle (1994) found much less deletion in the plastid genome of *Corallorhiza* than in other leafless angiosperms that have been examined. Their restriction site mapping showed that the species that were most affected were in the *C. striata* and *C. maculata* groups. DNA sequencing allows a much more precise investigation of molecular changes. The chloroplast genome locus *matK* has been used extensively in angiosperm phylogenetic reconstruction because of its relatively high substitution rate (Wolfe, 1991) in combination with a conserved overall structure. Beyond its phylogenetic utility, the locus is interesting because it resides in a group II intron in the gene for tRNA-lysine. Its substitution rates and evolutionary properties are unusual for a coding locus (Wolfe, 1991) and bear further investigation. Leafless taxa provide a unique system within which to examine the properties and changes in this gene.

The objectives of this study were to analyze DNA sequence data from the nuclear and plastid genomes to (1) resolve relationships among the genera of the *Corallorhiza* clade, (2) test the relationships among the species of *Corallorhiza* that were resolved based on plastid DNA restriction site data, (3) investigate the relationships of the newly described *C. bentleyi*, and (4) further investigate changes in the plastid genome in *Corallorhiza* by focusing on *matK*.

## MATERIALS AND METHODS

The taxa examined are listed in Appendix 1. DNA was isolated from fresh or frozen leaf/inflorescence material by the CTAB method of Doyle and Doyle (1987). The *matK* and ITS sequences were generated by standard PCR methods, including amplification with *Taq* polymerase under conditions given in Goldman et al. (2001) and Freudenstein (1999b). To obtain sequences that included portions of the *trnK* gene and the entire intron, primers 2R and 3914F that are located in the *trnK* exons were used for amplification and also as sequencing primers (Goldman et al., 2001); additional sequencing primers for *matK* are given in Goldman et al. (2001). Sequencing was performed using the ABI Prism cycle sequencing (Applied Biosystems, Foster City, California, USA) or Pharmacia Autocycle sequencing kits (Pharmacia Biotech, Piscataway, New Jersey, USA) according to the manufacturer's directions, except that 1/4 reactions were used for the ABI. Reactions were run on a Pharmacia ALFexpress automated sequencer using Cy-5 labeled sequencing primers or on an ABI 3100 sequencer using the BigDye terminator reaction mix following the manufacturer's protocols. Sequences were checked and contigs assembled in Sequencher (Gene Codes, Ann Arbor, Michigan, USA) and aligned by eye for *matK* and with ClustalX (Thompson et al., 1997) followed by manual adjustment for ITS; we followed the recommendations of Simmons and Ochoterena (2000) and Freudenstein and Chase (2001) for coding simple gap characters. Cases with nested indels were coded as multistate unordered characters. Sequences were deposited in GenBank (Appendix 1). Outgroups (*Coelia* and *Calypso*) were chosen based on the results of previous higher level analyses (Freudenstein, 1994b; J. Freudenstein, unpublished data).

Full sequences comprising the 3' end of the *trnK* 5' exon, the *trnK* 5' intron, *matK*, and the *trnK* 3' intron were generated for

all accessions of *Corallorhiza*, *Oreorchis*, *Aplectrum*, *Govenia*, and for *Cremastra appendiculata* because these were important for examining the length changes that have occurred in the *trnK* locus or for providing closely related undeleted sequences for comparison. For *Cremastra unguiculata*, *Calypso*, and *Coelia*, which did not experience any length changes and were relevant only to the phylogenetic objectives of the study, only the *matK*-coding sequence was generated. Predicted protein translations were generated with the program ExpASy (<http://us.expasy.org/tools/dna.html>).

Equally weighted parsimony analysis was performed on the data sets individually and combined using the program TNT (Goloboff et al., 2003). Tree search was performed using the XMULT command, which employs sectorial search, tree fusing, and tree drifting (Goloboff, 1999). Data sets were run with and without indel characters in order to make comparisons with the model-based results. The ragged ends of sequences were excluded from the analysis. Branch lengths generated by TNT are minimum possible lengths for each branch. Jackknife support (Farris et al., 1996) was determined using 5000 replications of a thorough search strategy employing XMULT for tree search and BBREAK for generating the maximal pool of most parsimonious trees. Individual probability of deletion of 37% was assigned to each character. Uncorrected genetic distances were calculated with the program PAUP\* (Swofford, 2002) for *matK* using only the portion of the ORF that was present in all taxa.

Maximum likelihood (ML) analysis was employed using the program RAxML (Stamatakis, 2006). The ITS, *matK*, and combined analyses were run with indel characters removed. Each data set was analyzed 1000 times using randomized parsimony trees generated by the program and subsequently rearranged using the GTRMIX option, which employs a GTR +  $\Gamma$  model. For the combined data set, two partitions were specified to allow *matK* and ITS to be modeled individually. The best likelihood tree was kept and compared with the parsimony results.

## RESULTS

***matK***—Alignment of the *trnK* region, including *matK*, was straightforward; the region that was sequenced is shown in Fig. 1. Six indels were coded for the *matK* ORF. Analysis of the full *trnK/matK* matrix gave a single most parsimonious tree (CI = 0.87, RI = 0.93, length = 213; Fig. 2). The first clade within the ingroup comprises *Cremastra* and *Govenia*, supported at 93%. Among these, the two species of *Cremastra* are very strongly united (100%). The next branch along the main spine of the tree is *Aplectrum*, which is then sister to *Oreorchis* + *Corallorhiza*.

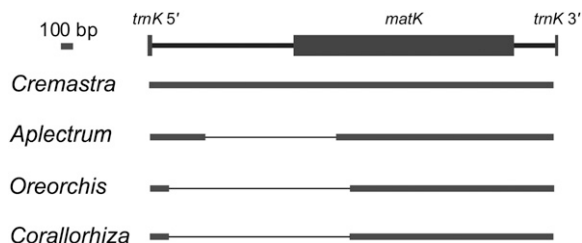


Fig. 1. A diagram of the split *trnK* gene, intron, and embedded *matK* gene. For the taxa that were sequenced, a thin black line indicates the major deletion.

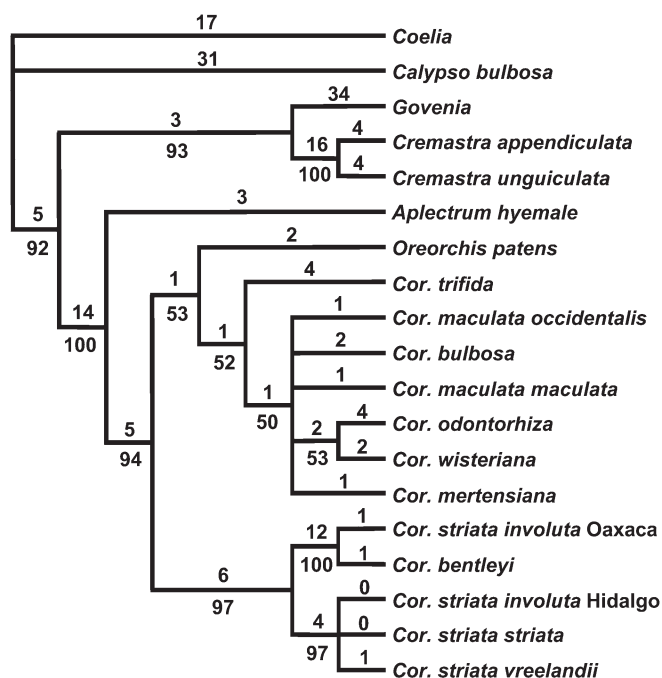


Fig. 2. The single most parsimonious tree from analysis of *matK* + *trnK* exon/intron sequence. Numbers above branches are lengths and those below are jackknife values.

*Oreorchis* is weakly supported as embedded within *Corallorhiza* (53%) and is sister to the *C. trifida-maculata-wisteriana* clade. *Corallorhiza trifida* is sister to the remainder of the clade, which has no resolution except for *C. odontorhiza* + *C. wisteriana* (the “*C. wisteriana* group”). The other main clade of *Corallorhiza* comprises the strongly supported (97%) *C. striata* group; within this group are two strongly supported groups of accessions, one comprising the *C. striata* var. *involuta* accession (JVF 2155) from Oaxaca and *C. bentleyi* (supported by 12 character changes), the other comprising the *C. striata* var. *striata* accession (from Michigan), the *C. striata* var. *involuta* accession (JVF 2190) from Hidalgo and the accession of *C. striata* var. *vreelandii* (from New Mexico). The ML tree was identical to the parsimony tree obtained when indels were excluded (not shown). The parsimony tree run without indels differed from the tree with indels only in the placement of *Oreorchis* (it was sister to all non *C. striata* group *Corallorhiza* with indels included and formed a trichotomy with *C. trifida* and a group comprising all of the *C. maculata* and *C. wisteriana* groups of species).

The *matK* sequences had significant deletions among the members of this group. Most striking are the 5' deletions in *Aplectrum*, *Oreorchis*, and *Corallorhiza* (Fig. 1). *Cremastra appendiculata* has an ORF of 1554 bp, and *Govenia* has 1557 bp. *Aplectrum* has a deletion of 299 bp at the 5' end of the ORF (yielding a length of 1235 bp), while *Oreorchis* and all *Corallorhiza* species have a further 84-bp deletion (giving ca. 1152 bp total). The deletion extends well beyond the ORF into the 5' intron. In all of these taxa, the *trnK* 5'-coding sequence is present, as is at least a portion of the 5' end of the *trnK* intron—387 bp in *Aplectrum* and 76 bp in *Oreorchis* and all of the *Corallorhiza* accessions. This deletion pattern is in contrast to the full intron of 964 bp for *Cremastra appendiculata* and 1037 bp in *Govenia* (Fig. 1). The 3' *trnK* intron is much smaller, ap-

proximately 300 bp and was not sequenced completely here, but shows little length variation in the major portion sequenced among *Govenia*, *Cremastra*, *Aplectrum*, *Oreorchis*, and *Corallorhiza*.

There are also internal deletions and insertions in *matK* among these genera, which were coded for the phylogenetic analysis. *Govenia* has a single 3-bp insertion, the only indel detected for a species outside the *Aplectrum-Corallorhiza-Oreorchis* clade. *Aplectrum* shares a 7-bp insertion with *Oreorchis* and *Corallorhiza*, as well as a 13-bp deletion that has undergone further loss to become a 19-bp deletion in *Oreorchis* and *Corallorhiza*. *Oreorchis* and all of the accessions of *Corallorhiza* share another 7-bp deletion. The *C. trifida-maculata-wisteriana* clade is united by a 9-bp deletion. All the *C. striata* group accessions (including *C. bentleyi*) share a 1-bp deletion. Finally, the Oaxaca accession of *C. striata* var. *involuta* has a 9-bp deletion near the 3' end of the gene.

Examination of the translated *matK* sequences reveals additional insights. Translation of the outgroup sequences for *Coelia* and *Calypso* and the ingroup *Govenia* and *Cremastra* sequences gives predicted amino acid sequences that align easily with other angiosperm *matK* translations. In these taxa, *matK* represents a single ORF ending in a termination codon. Translation of *matK* in *Aplectrum*, *Oreorchis*, and *Corallorhiza* yields short reading frames because of frameshift-causing indels and major deletions. The largest of these reading frames corresponds to 241 amino acids (in *Aplectrum*), and it is not in the same frame as *Govenia* and *Cremastra*. A 56 amino acid frame is present in all taxa included except for *C. bentleyi*, *C. striata* (Oaxaca), and *C. bulbosa*, which have a stop codon 10 amino acids into the region and *C. maculata* var. *maculata*, which has a stop codon 15 amino acids into the region. No reading frame yields more than ~10 amino acids of “domain X” as defined by Mohr et al. (1993). All these stop codons are the result of G → T mutations that result in TAA codons.

The genetic distances for *matK* show some trends (Table 1). Uncorrected pairwise distances between *Govenia* and the other ingroup taxa range between 0.028 and 0.047. The smallest distances are between *Govenia* and the *Cremastra* species (0.028, 0.033). The distance increases to 0.035 with *Aplectrum* and to 0.037 and beyond with *Oreorchis* and *Corallorhiza*. The *C. striata-bentleyi* accessions had increased distances; among

TABLE 1. Uncorrected pairwise distances between *Govenia* and other ingroup accessions.

Accession	<i>matK</i>	ITS
<i>Aplectrum hyemale</i>	0.035	0.070
<i>Corallorhiza bentleyi</i>	0.047	0.106
<i>Corallorhiza bulbosa</i>	0.037	0.084
<i>Corallorhiza maculata maculata</i>	0.038	0.081
<i>Corallorhiza maculata occidentalis</i>	0.038	0.086
<i>Corallorhiza mertensiana</i>	0.038	0.084
<i>Corallorhiza odontorhiza</i>	0.039	0.091
<i>Corallorhiza striata involuta</i> 2155	0.047	0.101
<i>Corallorhiza striata involuta</i> 2190	0.041	0.091
<i>Corallorhiza striata striata</i>	0.041	0.094
<i>Corallorhiza striata vreelandii</i>	0.041	0.099
<i>Corallorhiza trifida</i>	0.039	0.083
<i>Corallorhiza wisteriana</i>	0.038	0.086
<i>Cremastra appendiculata</i>	0.033	0.070
<i>Cremastra unguiculata</i>	0.028	0.061
<i>Oreorchis patens</i>	0.037	0.081

those, *C. striata* (Oaxaca) and *C. bentleyi* had the greatest distances (both 0.047).

**ITS**—No polymorphisms were detected in initial sequences, so no cloning was performed for ITS sequences. The ITS data set yielded a single most parsimonious tree (CI = 0.74, RI = 0.81, length = 289; Fig. 3), which is more resolved than that from *matK*. The only points of disagreement with the *matK* pattern are the placement of *Cremastra* and *Oreorchis*. With ITS, *Cremastra* is not sister to *Govenia*, but is instead sister to *Oreorchis* + *Corallorhiza* with strong (97%) support; *Oreorchis* is strongly supported as sister to a monophyletic *Corallorhiza* rather than being embedded within it. *Corallorhiza bentleyi* is placed together with the varieties of *C. striata*. As with *matK*, *C. bentleyi* is united strongly with the Oaxaca population of *C. striata* var. *involuta*. The Hidalgo population of *C. striata* falls among the var. *striata* and var. *vreelandii* populations. The *C. striata* clade is sister to the remaining species of *Corallorhiza*, but with no jackknife support for the monophyly of the latter. *Corallorhiza trifida* is sister to two clades, the *C. maculata* group and the *C. odontorhiza-wisteriana* clade. As with *matK*, the relationships of *C. maculata*, *C. bulbosa*, and *C. mertensiana* are unresolved. The ML tree for ITS was identical to the parsimony tree just described and to that from the analysis without indels.

The genetic distances varied as with *matK*, with two exceptions. Uncorrected distances (Table 1) show relative uniformity among members of the *C. trifida-maculata-wisteriana* clade (0.081–0.086) with the exception of *C. odontorhiza*, with the increased distance of 0.091. The *C. striata* clade again shows progressively greater distance, with *C. striata* var. *striata* at 0.094, *C. striata* var. *vreelandii* at 0.099, and the sister pair of *C. bentleyi* and the Oaxaca accession of var. *involuta* at 0.106

and 0.101, respectively. The exception here was the Hidalgo accession of var. *involuta*, which had a somewhat lower value of 0.091.

The combined ITS + *matK* data set yielded a single most parsimonious tree (CI = 0.77, RI = 0.85, length = 510; Fig. 4) that provides a well-supported pattern for most resolved nodes. The ML tree was identical to the parsimony tree and to that resulting from the analysis without indels. Because the patterns from the *matK* and ITS data sets largely agreed, the combined tree simply reflects their patterns for the most part. However, the disagreement concerning the placement of *Cremastra* is here resolved in a novel way, with *Cremastra* as sister to *Aplectrum* + *Oreorchis* + *Corallorhiza*. The only clade in which relationships remain unresolved is the *C. maculata* group.

## DISCUSSION

**Phylogenetic relationships within Corallorhizinae**—The only point of disagreement between the two data sets presented here concerns the placement of the two species of *Cremastra*. A sister relationship between *Cremastra* and *Govenia*, such as indicated by the *matK* data, has not been proposed before, but was quite strongly supported (93%), as was the contrasting ITS pattern. This discrepancy could be a case of true signal difference among the data sets as a result of past hybridization or lineage sorting. Even more strongly supported by *matK* is the sister relation between *Aplectrum* and *Oreorchis* + *Corallorhiza*, while the ITS data strongly suggest rather that *Aplectrum* is sister to *Cremastra* + *Oreorchis* + *Corallorhiza*, with *Govenia* sister to this whole assemblage. The pattern of major deletions at the 5' end of the *matK* ORF and *trnK* intron may have some bearing on choice among these patterns. While the topology from *matK* is most consistent with progressive deletion in this region, from no deletion in *Cremastra* to minimal deletion in *Aplectrum* to maximal deletion in *Oreorchis* and *Corallorhiza*, the ITS pattern suggests either independent deletions in *Aplectrum* and *Oreorchis* + *Corallorhiza*, or a seemingly unlikely regain of the deleted DNA in *Cremastra*. The pattern from the combined analysis, which is taken to be the strongest hypothesis here, is consistent with progressive *matK* deletion.

The morphological analysis of Freudenstein (1994b) did not resolve the relationships among the genera of the *Corallorhiza* clade, but it did suggest a synapomorphy for *Cremastra* + *Aplectrum* + *Oreorchis* + *Corallorhiza*—the *hamulus* type of pollinium stalk (Rasmussen, 1982), which is uncommon among orchids overall. On the ITS and combined trees, the *hamulus* is a unique synapomorphy for these genera, while on the *matK* tree the *hamulus* would either have been derived twice or have transformed subsequently to the *tegula* type of stalk seen in *Govenia*. Such a transformation between these two types of pollinium stalk has not been proposed before in orchids, given the very different structure of the two types of stalks.

The combined molecular tree is consistent with the current taxonomy that is widely accepted for these genera. Maekawa (1971) considered *Cremastra* and *Aplectrum* to be congeneric and transferred both species of *Cremastra* to *Aplectrum*, but the present results indicate that this combined group would not be monophyletic, thus arguing for continued recognition of the two genera.

The pattern of geographical distribution of the genera of Corallorhizinae, when considered in the context of their phylogenetic relationships, is one of alternate Old and New World

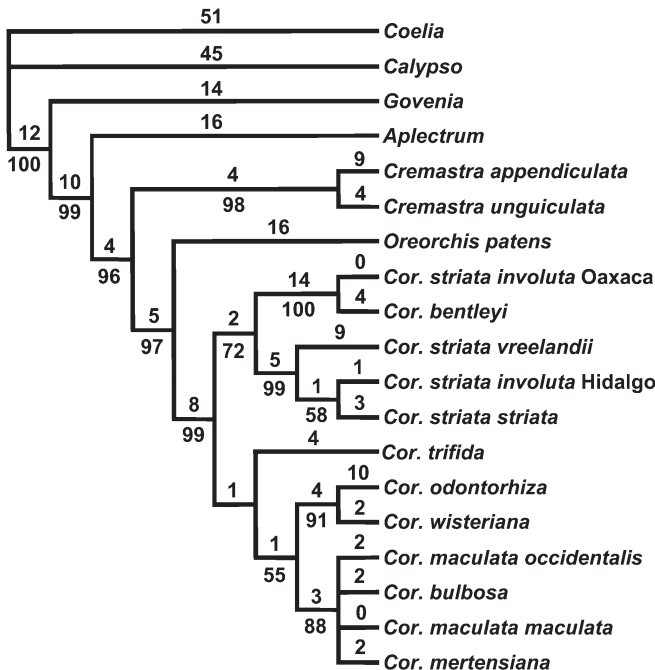


Fig. 3. The single most parsimonious tree from analysis of ITS sequence. Numbers above branches are lengths and those below are jackknife values.

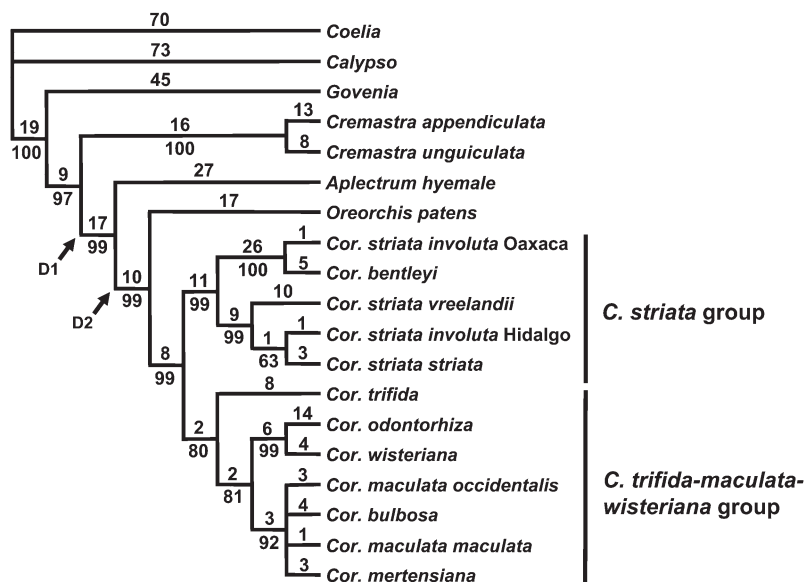


Fig. 4. The single most parsimonious tree from combined analysis of *matK*/*trnK* intron and ITS sequences. D1 and D2 designate major deletions in the *trnK* region. Numbers below branches are jackknife values and numbers above are branch lengths.

ranges, with *Govenia* occurring from Mexico to Bolivia, *Cremastra* in temperate east Asia, *Aplectrum* in eastern North America, *Oreorchis* in temperate east Asia, and *Corallorhiza* predominantly in North America (but having one circumboreal species, *C. trifida*, which does not fall at the base of the genus). The eastern North American–east Asian disjunct distribution pattern is well known and has been explained by fragmentation of a formerly more continuous temperate flora (see review in Wen, 1999). This scenario would be expected to lead to vicariant pairs, as have been described in many groups, including orchids (Boufford and Spongberg, 1983; Chen, 1983; Xiang et al., 1998b). None of the relationships depicted in the current study can be explained solely by vicariance, except perhaps for *Oreorchis*–*Corallorhiza*. This lack of correspondence to a vicariance paradigm could be due to extinction of members of vicariant pairs, such as possibly an Asian sister to *Aplectrum*, or to dispersal.

Both the ITS and *matK* patterns support the cladogram of relationships among *Corallorhiza* species based on restriction site data from the large single copy region of the plastid genome presented by Freudenstein and Doyle (1994). Two primary clades are distinguished: the *C. striata* clade, with flowers lacking a mentum (small nectar spur) and having a thickened, boat-shaped labellum with a fused pair of lamellae at its base; and the *C. trifida-maculata-wisteriana* clade, with a mentum of varying size at the summit of the ovary and a thin, flattened labellum that has two free lamellae at its base.

The least well-known species of *Corallorhiza* included here, *C. bentleyi*, has not been examined from a molecular perspective before, having only been discovered in the late 1990s. Freudenstein (1999a) described the species from West Virginia as a clear ally of *C. striata*, sharing with it the boat-shaped, thickened labellum with fused basal lamellae, and absence of a mentum. The flowers of this species are distinct from those of *C. striata* from northeastern North America because those of *C. bentleyi* are much smaller. They are most similar to the small-flowered plants of *C. striata* from southern Mexico that have been called *C. involuta* (recognized by Freudenstein [1997] as

*C. striata* var. *involuta*). The molecular evidence from both loci supports this relationship, associating the population of *C. bentleyi* examined here with a particular population of *C. striata* var. *involuta* from Oaxaca on a relatively long branch (27 total changes), while the Hidalgo accession of var. *involuta* grouped with the large-flowered *C. striata* var. *striata* and the medium-flowered *C. striata* var. *vreelandii* from the southwestern USA. This pattern suggests that two distinct lineages may exist in the *C. striata* group, each with members in Mexico and northern North America. If we were to optimize the range on the *C. striata*–*bentleyi* portion of the cladogram in Fig. 4 and considered the southwestern USA and Mexico to be one region, that southern region would be plesiomorphic and the northern North American regions would be apomorphic, representing independent gains in each clade. However, no strong conclusions about biogeography can be drawn with such sparse sampling. Freudenstein and Doyle (1994) found that the Mexican accessions of *C. maculata* were successive sisters to the northern North American accessions. If this is the case for the *C. striata* group as well, it may indicate a general pattern for the genus, and perhaps for other groups that have a montane Mexican–temperate northern North American distribution. A possible explanation for such a pattern would be northward migration following glacial retreat.

**Evolution of *matK***—*matK* is unusual among putatively coding genes in its relatively rapid overall substitution rate, the nearly equal substitution rates across positions in a codon, similar transition–transversion rates, and its low synonymous to nonsynonymous amino acid substitution rate ratio (due to higher nonsynonymous rate; Wolfe, 1991; Steele and Vilgalys, 1994; Xiang et al., 1998a). These properties are often associated with pseudogenes and some authors have suggested that *matK* might be a pseudogene in orchids and other angiosperms (Kores et al., 2000; Whitten et al., 2000; Cameron et al., 2001; Goldman et al., 2001). Li (1997, p. 182) defined pseudogenes as “DNA sequences that were derived from functional genes but have been rendered nonfunctional by mutations that prevent

their proper expression” and further added that they “are subject to no functional constraints.” While it is true that pseudogenes usually have those properties detailed for *matK*, and while those properties are unusual for a coding locus, they are not sufficient to render a locus nonfunctional. Pseudogenes typically experience frameshift mutations that cause radical amino acid change and often premature stop codons, leading to the nonfunctionality stated by Li. For example, Wolfe and dePamphilis (1997) found frameshift indels and premature stop codons in *rbcL* in two species of *Orobanchaceae* that also showed an increase in the nonsynonymous substitution rate. Although some studies have reported nontriplet indels for *matK* (e.g., Kores et al., 2000), in most cases in angiosperms *matK* indels occur in multiples of three bases (Soltis and Soltis, 1998), which is consistent with production of a protein product. Hilu and Alice (1999) reported frameshift mutations at the extreme 3' end of *matK* in grasses, but they are so close to the 3' terminus that they should not interfere with any function that the locus might have. The fact that *matK* sequences are alignable across angiosperms (Hilu et al., 2003) and within families such as Orchidaceae, suggests that most probably there is functional constraint. Moreover, Barthet and Hilu (2007) detected *matK* transcripts and probable full protein products for the maturase in rice and potato, suggesting, along with previous work (Vogel et al., 1999), probable functionality in at least some plants.

The changes described here in *Corallorhiza matK* sequences are the most extreme of any documented for angiosperms and would support the assertion that if *matK* is a pseudogene in any angiosperm, it is one in *Aplectrum*, *Oreorchis*, and *Corallorhiza*. Typical *matK* ORFs are approximately 1500–1600 bp (Soltis and Soltis, 1998; Hilu et al., 2003); the gene in *Epifagus* is reduced to 1320 bp due to deletions in the 5' end of the gene (Wolfe et al., 1992). Other leafless species of *Orobanchaceae* do not have unusually small *matK* ORFs, nor do they have nontriplet indels (Young et al., 1999). The *Corallorhiza* and *Oreorchis matK* sequences examined here are 1152 bp, also due to 5' deletion, and are thus the smallest *matK* sequences reported to date. Although *Epifagus* has significant 5' deletion, its changes are not so extensive that the reading frame for the “domain X” region at the 3' end is affected. This region was indicated to be most important for the DNA-binding function of *matK* by Mohr et al. (1993), although Young and dePamphilis (2000) did not observe a higher level of constraint on this region than on the rest of the gene. In *Calypso*, *Govenia*, and the two *Cremastras*, a single large ORF is present for *matK*, as in most other angiosperms. In contrast, internal length changes have altered the reading frames in such a way that there is no large reading frame in *Aplectrum*, *Oreorchis*, and *Corallorhiza*, because of the presence of stop codons. In all of these, only smaller potential ORFs are present, and these are not in the same frame as those in *Cremastra* and *Govenia*, meaning that the amino acid sequence produced would be radically different in *Aplectrum*, *Oreorchis*, and *Corallorhiza* and that, unless a phenomenon such as RNA editing (cf. Gott and Emeson, 2000) is in operation here, even the majority of the “domain X” portion of the gene cannot be correctly translated. RNA editing in plants is limited, as far as currently known, to C → U and U → C changes, with no addition or removal of bases having been detected (Gray, 2003; Tillich et al., 2006), so it is unlikely that this phenomenon could restore the reading frame in these taxa. Hence, these may be the first carefully documented *matK* pseudogenes.

Because of the broader history of *matK* and its antecedents, the evolution of the *matK* ORF cannot be dissociated from changes in the *trnK* intron in which it resides (Chuang and Hu, 2004; Hausner et al., 2006). Chloroplast genome-encoded group II introns, such as that in *trnK*, have self-splicing ability (Michel et al., 1989), but may require chaperoning from a maturase such as the product of *matK* (Hess et al., 1994). The fact that *matK* does exist rarely as a freestanding ORF, without associated *trnK* gene, as in *Epifagus* and *Adiantum capillis-veneris* (Wolf et al., 2003), suggests that this function may be broader than just assisting splicing of the *trnK* intron (Ems et al., 1995). In this study we did not sequence the entire *trnK* exons, but the 3' portion of the 5' exon that we did sequence was present and highly conserved in all accessions. The *trnK* intron shows quite a different pattern, however. Although large in *Govenia* and *Cremastra* (>900 bp), the 5' portion of the *trnK* intron is drastically reduced in *Aplectrum*, *Oreorchis*, and *Corallorhiza*, paralleling the changes seen in *matK*.

Although it is tempting to try to associate the sequence changes observed in *matK* in *Corallorhizinae* with the loss of leaves, the situation is not straightforward. First, because the large 5' deletions and frameshift indels occur in the leaf-bearing *Aplectrum* and *Oreorchis* as well as in *Corallorhiza*, they precede the loss of leaves (phylogenetically). This pattern could indicate that changes in the plastid genome foreshadow physical leaf loss in the shift to increased heterotrophy. However, the role of *matK* and its known variation in other plants need to be considered as well because it is not at all clear that the gene should be expected to be nonfunctional in nonphotosynthetic plants, given that its presumed maturase role is not necessarily related to photosynthesis. Accordingly, Wolfe et al. (1992) found that *matK* is one of the few genes remaining in the plastome of the leafless direct (haustorial) parasite *Epifagus*, even though the *trnK* gene in which it normally resides has been lost. *Epifagus* retains this ORF in spite of its plastid genome being much more deleted overall than the plastid genome of *Corallorhiza* (dePamphilis and Palmer, 1990; Freudenstein and Doyle, 1994). It is clear that, if indeed related to increased heterotrophy, the changes seen in *Corallorhiza* and allies in *matK* indicate a very different situation with respect to its role than in *Epifagus*.

Although a normal, functional *matK* product could not be produced in *Aplectrum*, *Oreorchis*, or *Corallorhiza*, the rate of base substitution, as assessed by genetic distance, is not very different in these taxa with respect to *Govenia* than it is in *Cremastra*. This raises the question of why, in the most convincing case where *matK* could be postulated to be a pseudogene, the rate has not increased substantially and sequence conservation seems to remain the rule. Perhaps it is just that the group is still in the early stages of the shift to heterotrophy and that the various taxa provide windows into positions on a continuum of change. While the rate of substitution for *matK* is elevated overall somewhat in the *C. striata* clade, it is only in *C. striata* var. *involuta* (JVF 2155) and *C. bentleyi* that we begin to see a notable increase in substitution rate. This pattern of substitution rate increase in the *C. striata* clade also holds for ITS, which is consistent with the observation of Freudenstein and Doyle (1994) that the greatest degree of plastid DNA deletion in *Corallorhiza* was among the *C. striata* accessions, particularly in *C. striata* var. *involuta*. This range of substitution rates within *Corallorhiza*, in combination with the molecular changes seen even in related leafy taxa, argues for the importance of this group as a natural system in the early stages of plastid genome change, from which we can continue to learn about the transformation from autotrophy to heterotrophy in plants.

## LITERATURE CITED

- BARTHET, M. M., AND K. W. HILU. 2007. Expression of *matK*: Functional and evolutionary implications. *American Journal of Botany* 94: 1402–1412.
- BOUFFORD, D. E., AND S. A. SPONGBERG. 1983. Eastern Asian–eastern North American phytogeographical relationships—a history from the time of Linnaeus to the twentieth century. *Annals of the Missouri Botanical Garden* 70: 423–439.
- CAMERON, K. C., M. W. CHASE, W. R. ANDERSON, AND H. G. HILLS. 2001. Molecular systematics of Malpighiaceae: Evidence from plastid *rbcL* and *matK* sequences. *American Journal of Botany* 88: 1847–1862.
- CHASE, M. W., K. M. CAMERON, R. L. BARRETT, AND J. V. FREUDENSTEIN. 2003. DNA data and Orchidaceae systematics: A new phylogenetic classification. In K. M. Dixon, S. P. Kell, R. L. Barrett, and P. J. Cribb [eds.], *Orchid conservation*, 69–89. Natural History Publications, Kota Kinabalu, Sabah, Malaysia.
- CHEN, S.-C. 1983. A comparison of orchid floras of temperate North America and eastern Asia. *Annals of the Missouri Botanical Garden* 70: 713–723.
- CHUANG, S.-L., AND J.-M. HU. 2004. The evolution of chloroplast *matK* genes, including identification of new homologues from *Ophioglossum petiolatum* and two lycophytes. *Taiwania* 49: 273–287.
- DEPAMPHILIS, C. W., AND J. D. PALMER. 1990. Loss of photosynthetic and chlororespiratory genes from the plastid genome of a parasitic flowering plant. *Nature* 348: 337–339.
- DOYLE, J. J., AND J. L. DOYLE. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin* 19: 11–15.
- DRESSLER, R. L. 1993. *Phylogeny and classification of the orchid family*. Cambridge University Press, Cambridge, UK.
- EMS, S. C., C. W. MORDEN, C. K. DIXON, K. H. WOLFE, C. W. DEPAMPHILIS, AND J. D. PALMER. 1995. Transcription, splicing and editing of plastid RNAs in the nonphotosynthetic plant *Epifagus virginiana*. *Plant Molecular Biology* 29: 721–733.
- FARRIS, J. S., V. A. ALBERT, M. KÄLLERSJÖ, D. L. LIPSCOMB, AND A. G. KLUGE. 1996. Parsimony jackknifing outperforms neighbor-joining. *Cladistics* 12: 99–124.
- FREUDENSTEIN, J. V. 1994a. Character transformation and relationships in *Corallorhiza* (Orchidaceae: Epidendroideae). II. Structural features and phylogenetic analysis. *American Journal of Botany* 81: 1458–1467.
- FREUDENSTEIN, J. V. 1994b. Gynostemium structure and relationships of the Corallorhizinae (Orchidaceae: Epidendroideae). *Plant Systematics and Evolution* 193: 1–19.
- FREUDENSTEIN, J. V. 1997. A monograph of *Corallorhiza*. *Harvard Papers in Botany* 10: 5–52.
- FREUDENSTEIN, J. V. 1999a. A new species of *Corallorhiza* from West Virginia, USA. *Novon* 9: 511–513.
- FREUDENSTEIN, J. V. 1999b. Relationships and character transformation in Pyroloideae (Ericaceae) based on ITS sequences, morphology, and development. *Systematic Botany* 24: 398–408.
- FREUDENSTEIN, J. V., AND M. W. CHASE. 2001. Analysis of mitochondrial *nad1b-c* intron sequences in Orchidaceae: Utility and coding of length-change characters. *Systematic Botany* 26: 643–657.
- FREUDENSTEIN, J. V., AND J. J. DOYLE. 1994. Character transformation and relationships in *Corallorhiza* (Orchidaceae: Epidendroideae). I. Plastid DNA. *American Journal of Botany* 81: 1449–1457.
- GOLDMAN, D. H., J. V. FREUDENSTEIN, P. J. KORES, M. MOLVRAY, D. C. JARRELL, W. M. WHITTEN, K. M. CAMERON, AND M. W. CHASE. 2001. Phylogenetics of Arethuseae (Orchidaceae) based on plastid *matK* and *rbcL* sequences. *Systematic Botany* 26: 670–695.
- GOLOBOFF, P. 1999. Analyzing large data sets in reasonable times: Solutions for composite optima. *Cladistics* 15: 415–428.
- GOLOBOFF, P., J. FARRIS, AND K. NIXON. 2003. TNT: Tree analysis using new technology. Program and documentation available from the authors and at website <http://www.zmuc.dk/public/phylogeny/TNT/>.
- GOTT, J. M., AND R. B. EMESON. 2000. Functions and mechanisms of RNA editing. *Annual Review of Genetics* 34: 499–531.
- GRAY, M. W. 2003. Diversity and evolution of mitochondrial RNA editing systems. *International Union of Biochemistry and Molecular Biology Life* 55: 227–233.
- HAUSNER, G., R. OLSON, D. SIMON, I. JOHNSON, E. R. SANDERS, K. G. KAROL, R. M. MCCOURT, AND S. ZIMMERLY. 2006. Origin and evolution of the chloroplast *trnK* (*matK*) intron: A model for evolution of group II intron RNA structures. *Molecular Biology and Evolution* 23: 380–391.
- HESS, W. R., B. HOCH, P. ZELTZ, T. HÜBSCHMANN, H. KÖSSEL, AND T. BÖRNER. 1994. Inefficient *rpl2* splicing in barley mutants with ribosome-deficient plastids. *Plant Cell* 6: 1455–1465.
- HILU, K. W., AND L. A. ALICE. 1999. Evolutionary implications of *matK* indels in Poaceae. *American Journal of Botany* 86: 1735–1741.
- HILU, K. W., T. BORSCH, K. MÜLLER, D. E. SOLTIS, P. S. SOLTIS, V. SAVOLAINEN, M. W. CHASE, M. P. POWELL, L. A. ALICE, R. EVANS, H. SAUQUEST, C. NEINHUIS, T. A. B. SLOTTA, J. G. ROHWER, C. S. CAMPBELL, AND L. W. CHATROU. 2003. Angiosperm phylogeny based on *matK* sequence information. *American Journal of Botany* 90: 1758–1776.
- KORES, P. J., P. WESTON, M. MOLVRAY, AND M. W. CHASE. 2000. Phylogenetic relationships within Diurideae: Inferences from plastid *matK* DNA sequences. In K. L. Wilson and D. A. Morrison [eds.], *Monocots: Systematics and evolution*, 449–456. CSIRO Publishing, Collingwood, Victoria, Australia.
- LI, W.-H. 1997. *Molecular evolution*. Sinauer, Sunderland, MA.
- MAEKAWA, F. 1971. *The wild orchids of Japan in colour*. Seibundo-shinkousha, Tokyo, Japan.
- MICHEL, F., K. UMESONO, AND H. OZEKI. 1989. Comparative and functional anatomy of group II catalytic introns—a review. *Gene* 82: 5–30.
- MOHR, G., P. S. PERLMAN, AND A. M. LAMBOWITZ. 1993. Evolutionary relationships among group II intron-encoded proteins and identification of a conserved domain that may be related to maturase function. *Nucleic Acids Research* 21: 4991–4997.
- MOLVRAY, M., P. J. KORES, AND M. W. CHASE. 2000. Polyphyly of mycoheterotrophic orchids and functional influences on floral and molecular characters. In K. L. Wilson and D. A. Morrison [eds.], *Monocots: Systematics and evolution*, 441–448. CSIRO Publishing, Collingwood, Victoria, Australia.
- RASMUSSEN, F. N. 1982. The gynostemium of the neottiid orchids. *Opera Botanica* 65: 1–96.
- SIMMONS, M. P., AND H. OCHOTERENA. 2000. Gaps as characters in sequence-based phylogenetic analyses. *Systematic Biology* 49: 369–381.
- SOLTIS, D. E., AND P. S. SOLTIS. 1998. Choosing an approach and an appropriate gene for phylogenetic analysis. In D. E. Soltis, P. S. Soltis, and J. J. Doyle [eds.], *Molecular systematics of plants II: DNA sequencing*, 1–42. Kluwer, Boston, Massachusetts, USA.
- STAMATAKIS, A. 2006. RaxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics (Oxford, England)* 22: 2688–2690.
- STEELE, K. P., AND R. VILGALYS. 1994. Phylogenetic analyses of Polemoniaceae using nucleotide sequences of the plastid gene *matK*. *Systematic Botany* 19: 126–142.
- SWOFFORD, D. L. 2002. PAUP\*: Phylogenetic analysis using parsimony (\*and other methods), version 4.0b. Sinauer, Sunderland, MA, USA.
- THOMPSON, J. D., T. J. GIBSON, F. PLEWNIAC, F. JEANMOUGIN, AND D. G. HIGGINS. 1997. The Clustal\_X windows interface: Flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research* 25: 4876–4882.
- TILLICH, M., P. LEHWARK, B. R. MORTON, AND U. G. MAIER. 2006. The evolution of chloroplast DNA editing. *Molecular Biology and Evolution* 23: 1912–1921.
- VOGEL, J., T. BÖRNER, AND W. R. HESS. 1999. Comparative analysis of splicing of the complete set of chloroplast group II introns in three higher plant mutants. *Nucleic Acids Research* 27: 3866–3874.
- WEN, J. 1999. Evolution of eastern Asian and eastern North American disjunction distributions in flowering plants. *Annual Review of Ecology and Systematics* 30: 421–455.

- WHITTEN, W. M., N. H. WILLIAMS, AND M. W. CHASE. 2000. Subtribal and generic relationships of Maxillarieae (Orchidaceae) with emphasis on Stanhopeinae: Combined molecular evidence. *American Journal of Botany* 87: 1842–1856.
- WOLF, P. G., C. A. ROWE, R. B. SINCLAIR, AND M. HASEBE. 2003. Complete nucleotide sequence of the chloroplast genome from a leptosporangiate fern, *Adiantum capillus-veneris* L. *DNA Research* 10: 59–65.
- WOLFE, A. D., AND C. W. DEPAMPHILIS. 1997. Alternate paths of evolution for the photosynthetic gene *rbcL* in four nonphotosynthetic species of *Orobanchae*. *Plant Molecular Biology* 33: 965–977.
- WOLFE, K. H. 1991. Protein-coding genes in chloroplast DNA: Compilation of nucleotide sequences, data base entries, and rates of molecular evolution. In L. Bogorad and I. K. Vasil [eds.], *The photosynthetic apparatus: Molecular biology and operation*, vol. 7B, Cell culture and somatic cell genetics of plants, 467–482. Academic Press, San Diego, California, USA.
- WOLFE, K. H., C. W. MORDEN, AND J. D. PALMER. 1992. Function and evolution of a minimal plastid genome from a nonphotosynthetic parasitic plant. *Proceedings of the National Academy of Sciences, USA* 89: 10648–10652.
- XIANG, Q.-Y., D. E. SOLTIS, AND P. S. SOLTIS. 1998a. Phylogenetic relationships of Cornaceae and close relatives inferred from *matK* and *rbcL* sequences. *American Journal of Botany* 85: 285–297.
- XIANG, Q.-Y., D. E. SOLTIS, AND P. S. SOLTIS. 1998b. The eastern Asian and eastern and western North American floristic disjunction: Congruent phylogenetic patterns in seven diverse genera. *Molecular Phylogenetics and Evolution* 10: 178–190.
- YOUNG, N. D., AND C. W. DEPAMPHILIS. 2000. Purifying selection detected in the plastid gene *matK* and flanking ribozyme regions within a group II intron of nonphotosynthetic plants. *Molecular Biology and Evolution* 17: 1933–1941.
- YOUNG, N. D., K. E. STEINER, AND C. W. DEPAMPHILIS. 1999. The evolution of parasitism in Scrophulariaceae/Orobanchaceae: Plastid gene sequences refute an evolutionary transition series. *Annals of the Missouri Botanical Garden* 86: 876–893.

APPENDIX 1. List of taxa examined, voucher specimens, origin, and GenBank numbers. All vouchers are at BH unless otherwise specified.

Taxon	Collection	Locality	<i>matK</i>	ITS
<i>Aplectrum hyemale</i> (Willd.) Torr.	JVF 2391 (OS)	NC, USA	—	EF525675
<i>Aplectrum hyemale</i> (Willd.) Torr.	MW Chase 104 (K)	—	EF525693	—
<i>Calyso bulbosa</i> (L.) Oakes	MW Chase 490 (K)	—	EF525690	AF521076
<i>Coelia triptera</i> (Sm.) Steud.	MW Chase 324 (K)	—	AF263643	AF260151
<i>Corallorhiza bentleyi</i> Freudenstein	JVF 2550 (OS)	WV, USA	EF525706	EF525688
<i>Corallorhiza bulbosa</i> A. Rich. & Gal.	JVF 2205	Mexico, Mexico	EF525699	EF525681
<i>Corallorhiza maculata</i> (Raf.) Raf. var. <i>maculata</i>	JVF 2209	MI, USA	EF525700	EF525682
<i>Corallorhiza maculata</i> var. <i>occidentalis</i> (Lindl.) Ames	JVF 1885	CO, USA	EF525697	EF525679
<i>Corallorhiza mertensiana</i> Bongard	JVF 2288	OR, USA	EF525704	EF525686
<i>Corallorhiza odontorhiza</i> (Poir.) Nutt.	JVF 2214	MI, USA	EF525701	—
<i>Corallorhiza odontorhiza</i> (Poir.) Nutt.	JVF 2439	MI, USA	—	EF525683
<i>Corallorhiza striata</i> Lindl. var. <i>striata</i>	JVF 2226	MI, USA	EF525702	EF525684
<i>Corallorhiza striata</i> var. <i>involuta</i> (Greenman) Freudenstein	JVF 2155	Oaxaca, Mexico	EF525696	EF525678
<i>Corallorhiza striata</i> var. <i>involuta</i> (Greenman) Freudenstein	JVF 2190	Hidalgo, Mexico	EF525698	EF525680
<i>Corallorhiza striata</i> var. <i>vreelandii</i> (Rydb.) L. O. Williams	Todsén 191a	NM, USA	EF525705	EF525687
<i>Corallorhiza trifida</i> Chatelain	JVF 2087	MI, USA	—	EF525677
<i>Corallorhiza trifida</i> Chatelain	JVF 2090	MI, USA	EF525695	—
<i>Corallorhiza wisteriana</i> Conrad	Dressler, s.n.	FL, USA	EF525703	EF525685
<i>Cremastra appendiculata</i> (D. Don) Makino	Inoue, s.n.	Japan	EF525691	EF525673
<i>Cremastra unguiculata</i> Finet	Inoue, s.n.	Japan	EF525692	EF525674
<i>Govenia</i> sp.	MW Chase 146 (K)	Mexico	EF525690	EF525672
<i>Oreorchis patens</i> (Lindl.) Lindl.	Inoue, s.n.	Japan	EF525694	EF525676