The Effect of Relaxed Functional Constraints on the Photosynthetic Gene *rbcL* in Photosynthetic and Nonphotosynthetic Parasitic Plants

*Andrea D. Wolfe**† *and Claude W. dePamphilis**

*Department of Biology, Vanderbilt University; and †Department of Plant Biology, Ohio State University

The photosynthetic gene *rbcL* has been lost or dramatically altered in some lineages of nonphotosynthetic parasitic plants, but the dynamics of these events following loss of photosynthesis and whether *rbcL* has sustained functionally significant changes in photosynthetic parasitic plants are unknown. To assess the changes to *rbcL* associated with the loss of functional constraints for photosynthesis, nucleotide sequences from nonparasitic and parasitic plants of Scrophulariales were used for phylogeny reconstruction and character analysis. Plants in this group display a broad range of parasitic abilities, from photosynthetic (''hemiparasites'') to nonphotosynthetic (''holoparasites''). With the exception of *Conopholis* (Orobanchaceae), the *rbcL* locus is present in all parasitic plants of Scrophulariales examined. Several holoparasitic genera included in this study, including *Boschniakia, Epifagus, Orobanche,* and *Hyobanche,* have *rbcL* pseudogenes. However, the holoparasites *Alectra orobanchoides, Harveya capensis, Harveya purpurea, Lathraea clandestina, Orobanche corymbosa, O. fasciculata,* and *Striga gesnerioides* have intact open reading frames (ORFs) for the *rbcL* gene. Phylogenetic hypotheses based on *rbcL* are largely in agreement with those based on sequences of the nonphotosynthetic genes *rps2* and *matK* and show a single origin of parasitism, and loss of photosynthesis and pseudogene formation have been independently derived several times in Scrophulariales. The mutations in *rbcL* in nonparasitic and hemiparasitic plants would result in largely conservative amino acid substitutions, supporting the hypothesis that functional proteins can experience only a limited range of changes, even in minimally photosynthetic plants. In contrast, ORFs in some holoparasites had many previously unobserved missense substitutions at functionally important amino acid residues, suggesting that *rbcL* genes in these plants have evolved under relaxed or altered functional constraints.

Introduction

The ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO) protein in land plants consists of eight small subunits encoded by a nuclear multigene family and eight large subunits encoded by a single chloroplast gene (reviewed in Kellogg and Juliano 1997). The function of the protein, carbon dioxide fixation, has been known for decades, but only recently have the structural features important to protein function been elucidated (Lorimer 1981; Hartman, Stringer, and Lee 1984; Igarishi, McFadden, and El-Gul 1985; Hartman et al. 1987; Lorimer, Gutteridge, and Madden 1987; Chapman et al. 1988; Knight, Andersson, and Brändén 1990; Newman and Gutteridge 1990, 1993; Ranty, Lorimer, and Gutteridge 1991; Schreuder et al. 1993*a*, 1993*b*; reviewed in Kellogg and Juliano 1997). For example, the active site is located on the large subunit and is formed by C- terminal loops of a barrel structure formed by eight α helices and eight β sheets.

The large subunit has 476 amino acid residues. Important functional sites and conserved regions of the protein have been elucidated by structural and mutational analyses (previous citations), and recently by a comparative survey of *rbcL* sequences from 499 species of seed plants (Kellogg and Juliano 1997). Kellogg and Juliano (1997) found that 35% of all residues associated with the active site were absolutely conserved (75% were highly conserved with one amino acid change observed across 499 taxa) and 24% of sites associated with intersubunit interactions were absolutely conserved (51% highly conserved). In addition, 22% of all residues show no variation among 499 seed plants, and 110 sites had changes in only one of the species assayed.

The conservative evolution of the RuBisCO large subunit can be directly attributed to the important function of carbon dioxide fixation. The mutations observed in the protein across 499 seed plants are found primarily in structurally nonconserved regions. Amino acid substitutions within structurally important regions are generally of a small subset of biochemically similar residues (often a toggling between two alternative amino acids), which would not necessarily affect the function of the protein. The conserved nature of the protein makes it useful in phylogeny reconstruction because of the slow rate of nonsynonymous substitutions (Wolfe, Li, and Sharp 1987; Chase et al. 1993). In other words, there are strong selective constraints on the protein to maintain structure and function.

Of the *rbcL* sequences for 499 species of seed plants analyzed by Kellogg and Juliano (1997), only five were in lineages with parasitic, mycoheterotrophic, or nonphotosynthetic members (Burmanniaceae, Krameriaceae, Pyrolaceae, Santalaceae, and Viscaceae; Chase et al. 1993), and *Burmannia* was the only nonphotosynthetic plant included in the analysis. Parasitic plants absorb water and/or nutrients from a host plant through a connecting tissue (haustorium); mycoheterotrophic plants parasitize fungi or other plants through a fungal intermediary. Despite the known utility of *rbcL* in phylogeny reconstruction, no studies have used this gene to examine phylogenetic relationships of parasitic plants, primarily because parasitic plants that have lost photosynthetic ability may no longer be under selective constraints to maintain the structure and function of proteins

Key words: molecular evolution, RuBisCO, Scrophulariaceae, chloroplast DNA evolution, plastid genome, cryptic pseudogene.

Address for correspondence and reprints: Andrea D. Wolfe, Department of Plant Biology, Ohio State University, 1735 Neil Avenue, Columbus, Ohio 43210-1293. E-mail: wolfe.205@osu.edu.

Mol. Biol. Evol. 15(10):1243–1258. 1998

involved in photosynthetic reactions. Therefore, *rbcL* may be altered substantially or lost, as is evident in holoparasitic plants of Scrophulariales. For example, the entire *rbcL* locus is missing in *Conopholis* (Orobanchaceae; Colwell 1994), and only a remnant of the gene remains in *Epifagus* (Orobanchaceae; dePamphilis and Palmer 1990; Wolfe, Morden, and Palmer 1992; reviewed in dePamphilis 1995). Low levels of RuBisCO activity were observed in the holoparasitic genera *Lathraea* (Scrophulariaceae; Thalouarn, Arnaud, and Renaudin 1989; Thalouarn and Renaudin 1991; Delavault, Sakanyan, and Thalouarn 1995) and *Cuscuta* (Cuscutaceae; Machado and Zetsche 1990; Haberhausen, Valentin, and Zetsche 1992), whereas no *rbcL* expression has been recorded for species of the holoparasitic genera *Epifagus, Conopholis,* and *Orobanche* (Orobanchaceae; de la Harpe, Grobbelaar, and Visser 1980; Press, Shah, and Stewart 1986; dePamphilis and Palmer 1990; Thalouarn et al. 1994). The coding region for *rbcL* is intact in *Lathraea, Cuscuta,* and at least two species of *Orobanche* (Thalouarn, Arnaud, and Renaudin 1989; Machado and Zetsche 1990; Delavault, Sakanyan, and Thalouarn 1995; Wolfe and dePamphilis 1997). In addition, Wolfe and dePamphilis (1997) found dramatically different patterns of evolution among four species of *Orobanche,* in that two New World species have intact *rbcL* reading frames, and two Old World species have pseudogenes resulting from deletions and/or insertions (indels). The two pseudogenes shared very few indels and were hypothesized to have diverged independently from a common ancestral sequence with an intact *rbcL* reading frame (Wolfe and dePamphilis 1997).

Additional concerns about using *rbcL* nucleotide sequence data for phylogeny reconstruction of parasitic plants deal with data gathering and analysis. The first concern is about the physical manipulation of the DNA of parasitic plants. Nonphotosynthetic plants may have smaller numbers of plastid DNA molecules than photosynthetic plants and are subject to high levels of sequence divergence in photosynthetic and other genes (dePamphilis and Palmer 1990; Delavault, Sakanyan, and Thalouarn 1995; dePamphilis 1995; Wolfe and dePamphilis 1997; Nickrent et al. 1997, 1998); they are thus prone to problems with contamination during polymerase chain reaction (PCR) techniques such as amplification of photosynthetic genes. Contamination has been reported for an *rbcL* sequence of *Cuscuta* (Olmstead and Palmer 1994), where host-tissue DNA, rather than parasitic plant DNA, was amplified. In addition, the *rbcL* gene of the parasitic plant may be so highly modified with deletions and insertions that it is not amplifiable or, if it can be sequenced, it is extremely difficult to align (e.g., *Epifagus* has only a 450-bp remnant of *rbcL*; Wolfe, Morden, and Palmer 1992). Another concern is with regard to analysis of DNA sequences from parasitic plants. For example, the mycoheterotrophic genus *Burmannia* has been included in several phylogenetic analyses (Gaut et al. 1992; Duvall et al. 1993; Chase et al. 1993), but its *rbcL* sequence has evolved at an accelerated rate compared with related monocots, resulting in apparent misplacement of the sequence in phylogeny reconstructions. Long-branch attraction (Felsenstein 1978) has also been noted in other parasitic plant lineages in phylogenetic trees based on 18S rDNA sequence data (Nickrent and Starr 1994; Nickrent et al. 1998). Therefore, it is uncertain that phylogeny reconstruction based on *rbcL* sequences of parasitic plants will be useful in light of the presumed relaxed functional constraints and possible increased rates of substitution.

Although it is clear that photosynthetic genes can be lost or dramatically altered in some nonphotosynthetic lineages, nothing is known of the dynamics of these events or whether important changes to the *rbcL* gene (and the large-subunit protein) begin to accrue prior to the loss of photosynthesis. At least two alternative hypotheses may explain the loss of function in *rbcL* pseudogene formation: (1) the protein is completely functional up to the loss of the reading frame via a stop codon or indel (i.e., a catastrophic loss of function), or (2) the function of the protein slowly degrades as amino acid replacements affect the structure and/or biochemical properties of the protein.

To evaluate the effect of relaxed functional constraints on the photosynthetic gene *rbcL,* we traced the pattern of mutations in a lineage of parasitic plants encompassing the entire range of parasitic ability, from nonparasitic (photosynthetic) to holoparasitic (nonphotosynthetic). Plants that are parasitic and photosynthetic are called hemiparasites, and they primarily absorb water and nutrients from their host plants. We studied plants of the Scrophulariaceae and Orobanchaceae, lineages within Scrophulariales that include nonparasitic, hemiparasitic, and holoparasitic members. It is also clear that the loss of photosynthesis has occurred independently several times in Scrophulariales based on phylogeny reconstructions of *rps2* and *matK* (dePamphilis, Young, and Wolfe 1997; unpublished data). Phylogenies based on nonphotosynthetic genes are ideal for comparison of a phylogeny reconstruction based on the photosynthetic gene *rbcL,* in that they allow us to test whether increased rates of substitution or the inclusion of *rbcL* pseudogenes will result in a tree topology that is incongruent with results from other data analyses.

Materials and Methods

Total DNAs were isolated from 27 species representing 20 genera of Scrophulariaceae and Orobanchaceae (table 1). Voucher data are given in dePamphilis, Young, and Wolfe (1997). *rbcL* was PCR-amplified using the RH1 (5′) and 1352R (3′) primers as reported in Wolfe and dePamphilis (1997). Purification of PCR product and sequencing also were as in Wolfe and dePamphilis (1997) with sequences obtained from both strands. Additional sequences of taxa from Scrophulariales were obtained from GenBank (table 1).

The *rbcL* sequences were aligned using the CLUSTAL V program (Higgins, Bleasby, and Fuchs 1992), and the results were adjusted to obtain the best alignment for each region based on conserved sequences. Most of the *rbcL* pseudogene from *Epifagus* was unalignable and was therefore not included in the phylo-

Table 1
Taxon Sampling for Scrophulariales

Family	Taxon	GenBank Accession No.	Source
Acanthaceae (ACA)[a]	*Acanthus montanus* T. Anders.[b]	L12592	Hedren et al. (1995)
	Justicia americana (L.) Vahl.[b]	L14401	Olmstead et al. (1993)
	Thunbergia usamberica Lindau[b]	L12596	Hedren et al. (1995)
Bignoniaceae (BIG)	*Catalpa* sp.[b]	L11679	Olmstead et al. (1992)
	Martinella obovata[b]	L36444	Olmstead and Reeves (1995)
	Tabebuia heterophylla[b]	L36451	Olmstead and Reeves (1995)
Callitrichaceae (CLL)	*Callitriche hermaphroditica*[b,c]	L36441	Olmstead and Reeves (1995)
	Callitriche heterophylla Pursh. emend Danley	L11681	Olmstead et al. (1993)
	Callitriche verna	L47331	Unpublished
Gesneriaceae (GES)	*Nematanthus hirsutus*[b]	L36446	Olmstead and Reeves (1995)
	Streptocarpus holstii Engl.[b]	L14409	Olmstead et al. (1993)
Lentibulariaceae (LNT)	*Pinguicula caerulea* Walter	L01942	Albert et al. (1992)
	Utricularia biflora Royb.[b]	L13190	Chase et al. (1993)
Myoporaceae (MYO)	*Myoporum mauritianum*[b]	L36445	Olmstead and Reeves (1995)
Oleaceae (OLE)	*Jasminum suavissimum* Lindl.	L01929	Albert et al. (1992)
	Ligustrum vulgare L.[b]	L11686	Olmstead et al. (1992)
	Nyctanthes arbor-tristis	U28877	Unpublished
Pedaliaceae (PED)	*Harpagophyton granidieri* Baill.	L01923	Albert et al. (1992)
	Sesamum indicum L.[b]	L14408	Olmstead et al. (1993)
Plantaginaceae (PLT)	*Plantago lanceolata*[b,c]	L36454	Olmstead and Reeves (1995)
Orobanchaceae (ORO)	*Boschniakia hookeri*[b,c]	AF026817	This study
	Boschniakia strobilacea[b,c]	AF026818	This study
	Orobanche cernua[b,c]	U73968	Wolfe and dePamphilis (1997)
	Orobanche corymbosa[b,c]	U73969	Wolfe and dePamphilis (1997)
	Orobanche fasciculata[b,c]	U73970	Wolfe and dePamphilis (1997)
	Orobanche ramosa[b,c]	U73971	Wolfe and dePamphilis (1997)
Scrophulariaceae	*Alectra orobanchoides*[b,c]	AF026819	This study
	Alectra sessiliflora[b,c]	AF026820	This study
	Alonsoa unilabiata[b,c]	AF026821	This study
	Antirrhinum majus L.[b]	L11688	Olmstead et al. (1993)
	Buchnera floridana[b,c]	AF026822	This study
	Castilleja linariaefolia[b,c]	AF026823	This study
	Celsia arcturus[b,c]	L36442	Olmstead and Reeves (1995)
	Chelone obliqua[b,c]	AF026824	This study
	Collinsia grandiflora[b,c]	AF026825	This study
	Cycnium racemosum[b,c]	AF026826	This study
	Digitalis purpurea L.[b,c]	L01902	Albert et al. (1992)
	Gratiola pilosa[b,c]	AF026827	This study
	Halleria lucida[b,c]	AF026828	This study
	Harveya capensis[b,c]	AF026829	This study
	Harveya purpurea[b,c]	AF026830	This study
	Hyobanche atropurpurea[b,c]	AF026831	This study
	Hyobanche sanguinea[b,c]	AF026832	This study
	Lathraea clandestina[b,c]	AF026833	This study
	Melampyrum lineare[b,c]	AF026834	This study
	Mimulus aurantiacus[b,c]	AF026835	This study
	Paulownia tomentosa[b,c]	L36447	Olmstead and Reeves (1995)
	Pedicularis foliosa[b,c]	AF026836	This study
	Schlegelia parviflora[b,c]	L36448	Olmstead and Reeves (1995)
	Scrophularia sp.[b,c]	L36449	Olmstead and Reeves (1995)
	Selago thunbergii[b]	L36450	Olmstead and Reeves (1995)
	Seymeria pectinata[b,c]	AF026837	This study
	Striga asiatica[b,c]	AF026838	This study
	Striga gesnerioides[b,c]	AF026839	This study
	Striga hermonthica[b,c]	AF026840	This study
	Striga passargei[b,c]	AF026841	This study
	Torenia fournieri[b,c]	AF026842	This study
	Tozzia alpina[b,c]	AF026843	This study
	Verbascum thapsus[b,c]	L36452	Olmstead and Reeves (1995)
	Veronica catenata[b,c]	L36453	Olmstead and Reeves (1995)
Solanaceae (SOL)	*Nicotiana debneyi*	D70815	Shikanai et al. (1996)
	Nicotiana tabacum L.[b,c]	S54304	Shinozaki et al. (1986)

[a] Family abbreviations used in tree diagrams in parentheses.
[b] Taxa included in K_S and K_N calculations and character mapping of amino acid substitutions.
[c] Taxa included in pam250 analysis.

genetic analysis. Amplification primer sequences were not scored. Most of the sequences included in these analyses were restricted to nucleotide positions 1–1352 on the sequence for *Nicotiana tabacum* (Solanaceae), the outgroup for this study. However, where nucleotide positions 1–1434 were available (e.g., from GenBank or where external primers were used for amplification), the entire gene sequence was included in the phylogenetic analysis. Insertions and deletions were referenced from the sequence of *N. tabacum. Nicotiana* was selected as the outgroup taxon based on its close relationship to members of Scrophulariales (Olmstead and Reeves 1995; dePamphilis, Young, and Wolfe 1997; Wolfe and dePamphilis 1997), and because it is the closest relative to Scrophulariales with an entire chloroplast genome sequence (Shinozaki et al. 1986).

PAUP 3.1.1 (Swofford 1993) was used for phylogeny reconstruction using the *rbcL* nucleotide sequences. The heuristic search used the following options: addition sequence = random with 3,000 replicates (starting number = 333); all maximum-parsimony trees were kept with each replicate; swapping algorithm = TBR with steepest descent; collapse option in effect; mulpars option in effect; and number of trees held at each step = 10. Branch lengths were calculated with the ACCTRAN transformation. A parsimony jackknife analysis (Farris et al. 1996) with 10,000 replicates and a 50% cutoff option in effect was used to assess support for the tree. A bootstrap analysis with 2,000 replicates employed the same parameters as the heuristic search except that the addition sequence option used one random replication and there was no branch swapping. Bootstrapping without the branch-swapping option in effect may underestimate the actual relative nodal support of a clade (R. G. Olmstead, University of Washington, personal communication; unpublished data), so the bootstrap values calculated in this analysis may underrepresent the phylogenetic signal present in these data.

The synonymous (K_S) and nonsynonymous (K_N) substitution rates and distances for *rbcL* from *N. tabacum* and all species of Scrophulariaceae and Orobanchaceae included in the study, and from selected taxa from other families in Scrophulariales (table 1), were calculated after a Jukes-Cantor correction using MEGA (Kumar, Tamura, and Nei 1993). Sites with potential stop codons were removed from the pseudogene sequences to facilitate computation. Gaps and missing data were removed only in the pairwise comparisons. Average K_S and K_N and K_S/K_N were calculated for *rbcL* from different categories of parasitic Scrophulariaceae/Orobanchaceae (e.g., nonparasitic, hemiparasitic, and holoparasitic) to determine whether there is an overall trend in substitution rates with increasing heterotrophy.

K_S and K_N are not defined for pseudogenes, and we did not attempt to calculate these values for pseudogene evolution. Inclusion of the pseudogene sequences aligned in the reading frame of the other genes gave us the means to assess whether inferred synonymous and nonsynonymous substitutions may have occurred prior to pseudogene formation. For example, if a nucleotide substitution was a synapomorphy for a lineage with intact reading frames and pseudogenes, it is probable that the substitution occurred prior to pseudogene formation.

The sequences for the *rbcL* coding region were translated using the computer program MacVector 4.1.5 (Eastman Kodak Co., Rochester, N.Y.). Gap positions were maintained, but nucleotide insertions from *rbcL* pseudogene sequences were not included in the translation in order to reconstruct the most conservative protein alignments.

The inferred protein sequences and hypothetical in-frame translations of pseudogenes were aligned using CLUSTAL V (Higgins, Bleasby, and Fuchs 1992). Nonconserved amino acids in the polypeptide chain were analyzed by plotting the number of character step changes (amino acid replacements) along the length of the polypeptide onto the topology generated from the strict consensus tree of a combined *rps2*/*matK* data set (unpublished data) using MacClade (Maddison and Maddison 1992). The character changes were mapped onto the tree, and histograms depicting the number of step changes per site were also calculated. Three categories were included in the histogram analysis: (1) nonparasitic plants; (2) hemiparasitic clades and hemiparasitic clades with holoparasitic members; and (3) holoparasitic clades (mostly consisting of hypothetical in-frame translations). In addition, a comparison of amino acid substitutions along the length of the RuBisCO large-subunit polypeptide was performed for taxa in this study and for 499 seed plants, following Kellogg and Juliano (1997).

To assess the effect of amino acid replacements on the structure and function of the RuBisCO large-subunit polypeptide, we conducted a pam250 match/mismatch analysis (Dayhoff, Schwartz, and Orcutt 1978) on the translated protein sequences for taxa from all categories of parasitic plants in the Scrophulariaceae and representative nonparasites of Scrophulariales (table 1). Dayhoff, Schwartz, and Orcutt (1978), using empirical data for many proteins (hemoglobins, fibrinopeptides, cytochrome *c*), constructed a table (the pam250 match/mismatch scores matrix) of amino acid substitution probabilities that shows the probabilities of the substitution of one amino acid for another. Amino acids that have similar biochemical properties are more likely to be substituted with one another than with biochemically dissimilar amino acids, and this is reflected in the values given in the pam250 matrix. These substitution probabilities are measured in terms of ''PAM'' (percent accepted mutations), with one unit representing one amino acid substitution per 100 residues.

Using the reference protein from the outgroup, *N. tabacum,* pam250 match/mismatch scores were calculated over the entire length of the polypeptide and from important structural motifs by assigning a pam250 value from the diagonal matrix scores to each amino acid in the polypeptide sequence of *N. tabacum.* For example, at amino acid position 157, *N. tabacum* has a valine residue (V; nonpolar, hydrophobic) with a pam250 match/mismatch score of 4. At position 157, species of *Alectra, Buchnera, Cycnium, Harveya, Hyobanche,* and *Striga* have a serine residue (S; uncharged, polar). The

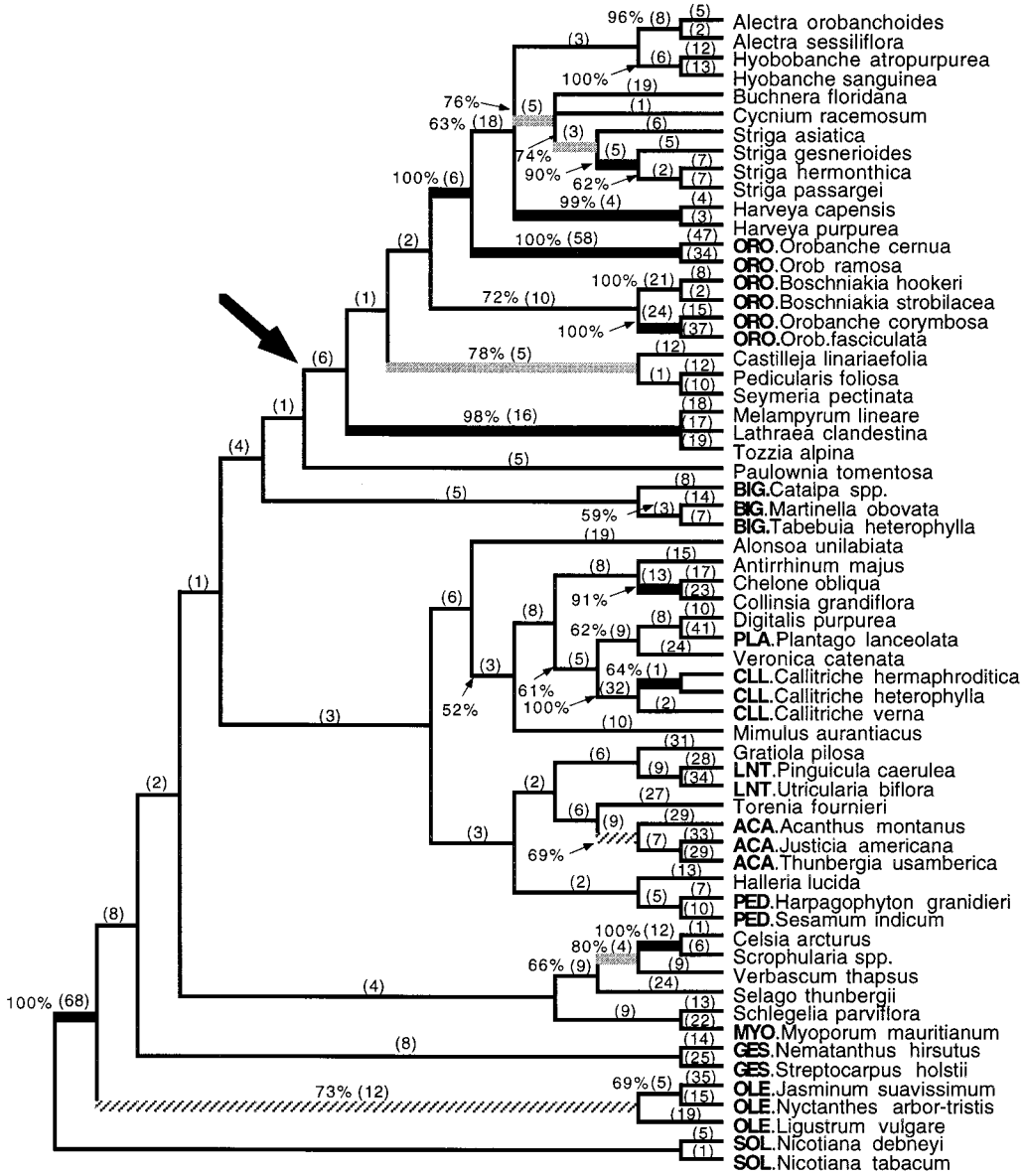

FIG. 1.—Phylogeny reconstruction based on maximum-parsimony analysis of *rbcL* nucleotide sequences; strict consensus of 18 most-parsimonious trees with 1,448 steps (CI = 0.509, RI = 0.562, RC = 0.286). An arrow indicates a monophyletic parasite clade including genera traditionally assigned to Orobanchaceae. Abbreviations for families in Scrophulariales and outgroup families: ACA = Acanthaceae, BIG = Bignoniaceae, CLL = Callitrichaceae, GES = Gesneriaceae, GLO = Globulariaceae, LEN = Lentibulariaceae, MYO = Myoporaceae, ORG = Orobanchaceae, PED = Pedaliaceae, PLA = Plantaginaceae, SOL = Solanaceae. Bootstrap values are indicated by numbers with percentage signs, number of synapomorphies for each node or terminal branch lengths are shown in parentheses and parsimony jackknife values are depicted as follows: wide black lines = 90%–100%, wide gray lines = 80%–89%, wide hashed lines = 70%– 79%.

pam250 match/mismatch score for the transition from V to S is −1. The pam250 match/mismatch score difference for these taxa at site 157 is $\Delta = 4 - (-1) = 5$. Across the length of the polypeptide, *N. tabacum* has the highest possible total pam250 match/mismatch score. If another taxon has an identical polypeptide sequence, its pam250 match/mismatch score would equal the score for *N. tabacum.* However, if a taxon has a different polypeptide sequence, the score will be lower by the differences compiled for each amino acid substitution along the length of the polypeptide. The significance of differences was assessed using a Wilcoxon two-sample test based on ranks in pairwise comparisons with *N. tabacum* (Sokal and Rohlf 1995). The assumption of this analysis is that if pam250 scores for the translated proteins from parasitic plant taxa are not significantly different from the reference protein, the inferred polypeptide should resemble a ''normal'' protein from a related photosynthetic plant. If so, this would be an indicator that the functional constraints for photosynthesis have not been relaxed in parasitic plant taxa. However, if the pam250 scores are significantly different for parasitic plant taxa, we can assume that the protein is also different (possibly nonfunctional) and that there may be relaxed functional constraints on the protein associated with the loss of photosynthesis.

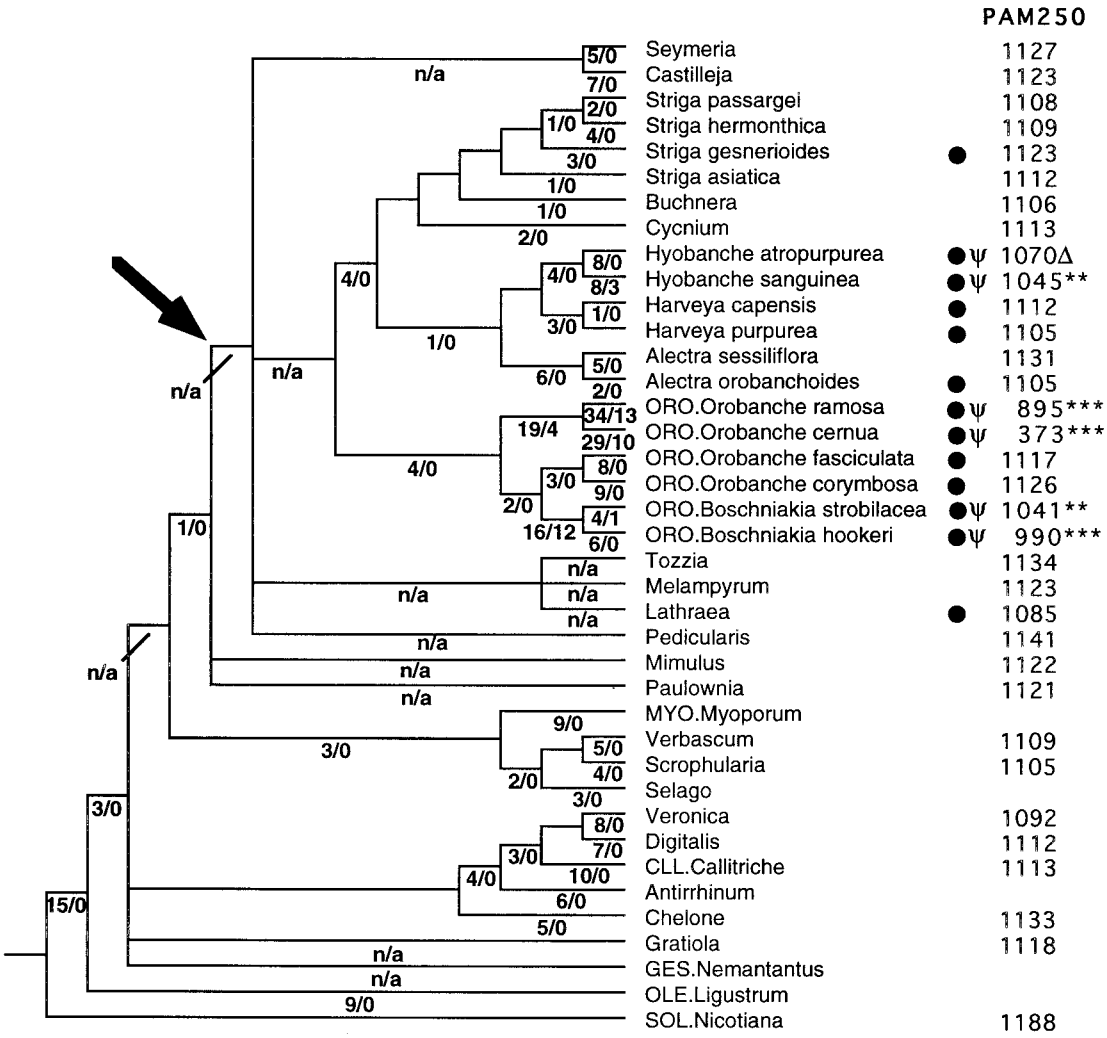

FIG. 2.—Character mapping of *rbcL* analyses onto a strict-consensus tree topology generated from a combined data set of *rps2* and *matK* sequences (dePamphilis, Young, and Wolfe 1997; unpublished data). Character mapping shows the loss of photosynthesis (●) and *rbcL* pseudogene formation (Ψ) in the Scrophulariaceae parasite clade. pam250 scores are listed to the right of taxon names (Δ = $P < 0.1$; ** = $P \leq 0.025$; *** = $P \leq 0.005$). The number of amino acid substitutions or structural mutations (indels) per branch (designated by number in box) is listed under the branch (amino acid substitutions/structural mutations). n/a = number of substitutions not calculated in polytomies. An arrow indicates a monophyletic parasite clade including genera traditionally assigned to Orobanchaceae.

Results

Phylogeny Reconstruction

Maximum-parsimony analysis of 62 *rbcL* sequences yielded 18 most-parsimonious trees with 1,448 steps (consistency index [CI] = 0.509, retention index [RI] = 0.562, rescaled consistency index [RC] = 0.286). There were two major results from this analysis. First, the strict-consensus tree (fig. 1) revealed that Scrophulariaceae, as traditionally circumscribed, are not monophyletic. The second major result was the confirmation of a single origin of parasitism in the Scrophulariales (dePamphilis, Young, and Wolfe 1997; Nickrent et al. 1998). A single clade contains all the parasites traditionally classified in Scrophulariaceae as well as Orobanchaceae.

Character Mapping—Holoparasitism and Pseudogene Formation

Lineages of holoparasitic Scrophulariaceae assayed in this study include *Alectra, Boschniakia, Harveya, Hyobanche, Lathraea, Orobanche,* and *Striga.* From mapping the distribution of holoparasitism onto the *rps2*/*matK* strict-consensus tree (fig. 2), we infer that holoparasitism has possibly arisen five times in Scrophulariaceae. It is clear that holoparasitism is derived independently in the *Hyobanche*/*Harveya* clade, the clade representing Orobanchaceae, *Alectra, Striga,* and the lineage including *Lathraea.*

Several sequences (*Hyobanche atropurpurea, Hyobanche sanguinea, Orobanche cernua, Orobanche ramosa, Boschniakia strobilacea,* and *Boschniakia hookeri*) are apparent *rbcL* pseudogenes with varying degrees of mutation (fig. 3). For example, the sequence of *H. atropurpurea* has a stop codon for amino acid position 20 but no insertions and deletions, that of *H. sanguinea* has two small deletions from nucleotide position 225 to nucleotide position 540 (including the active site), and all other pseudogene sequences include catastrophic insertions and deletions.

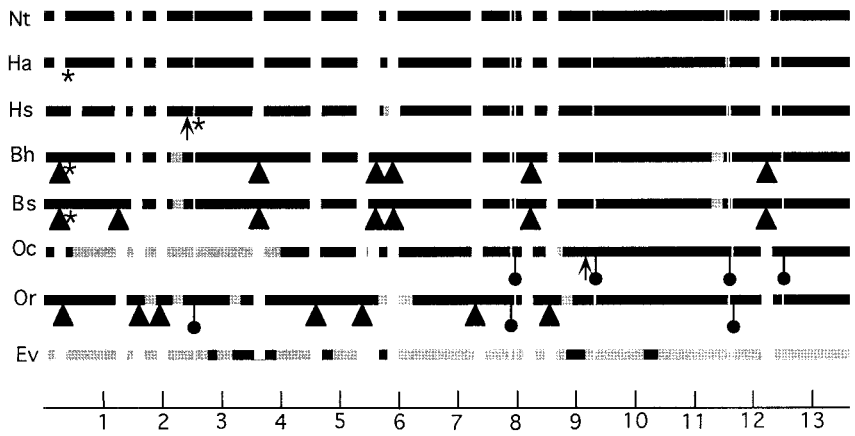

FIG. 3.—*rbcL* pseudogene map showing mutations compared with reference sequence from nonparasitic *Nicotiana tabacum* (Nt). Ha = *Hyobanche atropurpurea*; Hs = *Hyobanche sanguinea*; Bh = *Boschniakia hookeri*; Bs = *Boschniakia strobilacaea*; Oc = *Orobanche cernua*; Or = *Orobanche ramosa*; Ev = *Epifagus virginiana.* * = stop codon; black triangles = multinucleotide insertion; black circle = 1–2-bp insertion; arrow = 1-bp deletion; gray line = multinucleotide deletion, gaps = regions where insertions occur in other taxa. The scale represents the length of the *N. tabacum rbcL* nucleotide sequence divided into 100-bp segments.

Synonymous and Nonsynonymous Substitution Rates

K_S and K_N and K_S/K_N for *rbcL* from different categories of parasitic Scrophulariaceae are listed in table 2. As heterotrophy increases in Scrophulariaceae, there is an average increase in K_S. In contrast, K_N does not appear to be affected greatly until the selective constraint is removed and pseudogenes are formed. There is a bias toward synonymous substitutions in parasitic plants with *rbcL* open reading frames (ORFs), as evidenced by the high K_S/K_N ratios compared with nonparasitic Scrophulariaceae. This bias is retained even in the related holoparasitic species *Orobanche corymbosa* and *Orobanche fasciculata*, where synonymous substitutions ($K_S = 0.1149 \pm 0.0196$) significantly exceed nonsynonymous substitutions ($K_N = 0.0152 \pm 0.0038$) by a factor of 7.56 (H_O: $K_S = K_N$; $t = 4.9937$; $P \ll 0.001$; Kumar, Tamura, and Nei 1993, pp. 24–26). The inferred rate of nucleotide substitution for the *rbcL* pseudogene sequences is much higher than that for the intact reading frames, reflecting the lack of functional constraints associated with pseudogenes.

Table 2
Average Synonymous (K_S) and Nonsynonymous (K_N) Substitution Rates and K_S/K_N Ratios Based on Jukes-Cantor Corrections of *rbcL* Nucleotide Sequence Data for Categories of Parasitic Scrophulariaceae with *Nicotiana tabacum* as a Reference Protein

Parasitic Plant Category	K_S	K_N	K_S/K_N
Nonparasitic (16)	0.2022	0.0281	7.19
Photosynthetic hemiparasitic (7)	0.2227	0.0223	9.98
Nonphotosynthetic[a] hemiparasitic (11)	0.2463	0.0270	9.12
Holoparasitic[b] (6)	0.2482	0.0450	5.52

NOTE.—Numbers of taxa included in categories are shown in parentheses.

[a] Hemiparasitic lineages with holoparasitic members and holoparasitic *rbcL* ORFs (see fig. 2).

[b] *rbcL* pseudogene sequences only.

Amino Acid Substitutions

A histogram of character step changes (amino acid replacements) plotted against the tree topology depicted in figure 2 reveals a pattern of increasing amino acid substitution with increasing heterotrophic ability (fig. 4). Nonparasitic plants have many regions in the polypeptide in which no amino acid substitutions occur, whereas hemiparasites and holoparasites have increasing numbers of amino acid substitutions compared with nonparasites. Holoparasites also have more amino acid substitutions than do hemiparasites. The increased number of substitutions in the different classes of parasitic plants

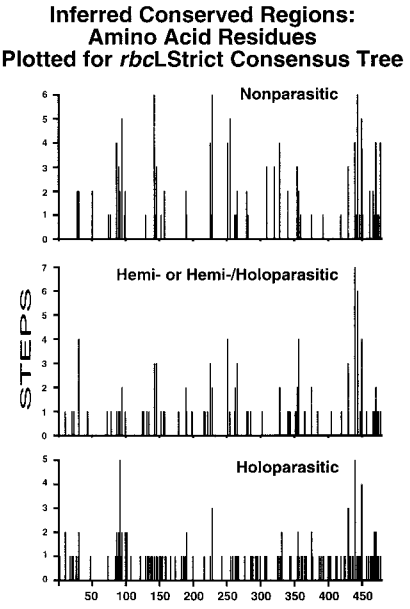

FIG. 4.—Number of steps (y-axis) for each amino acid along the length of the RuBisCO large-subunit protein (x-axis) plotted against the strict-consensus tree. Blank areas along the amino acid sequence represent conserved regions.

Table 3
Specific Amino Acid Substitutions that Differ from Those Listed for 499 Seed Plants

Amino Acid Position[a]	Amino Acids Reported for 499 Seed Plants[b]	Amino Acid Observed in this Study	Taxon	Parasitic Classification[c]
2	**S**	P	*Orobanche corymbosa*	HO
10	**S F G K N V** T Y	R	*Orobanche fasciculata*	HO
17 ˇ.....	**V** F	D	*Orobanche ramosa*	HO$^{\psi}$
20	**Y** * S	*	*Boschniakia hookeri*	HO$^{\psi}$
		*	*Boschniakia strobilacea*	HO$^{\psi}$
		*	*Hyobanche atropurpurea*	HO$^{\psi}$
22	**L I**	F	*B. hookeri*	HO$^{\psi}$
		F	*B. strobilacea*	HO$^{\psi}$
23	**T N** S	A	*Lathraea clandestina*	HO
26 βA	**T** A	I	*O. ramosa*	HO$^{\psi}$
29	**Y** L S *	N	*B. hookeri*	HO$^{\psi}$
		N	*B. strobilacea*	HO$^{\psi}$
72	**D Y**	G	*Hyobanche sanguinea*	HO$^{\psi}$
73 #	**G** R	V	*B. hookeri*	HO$^{\psi}$
		V	*B. strobilacea*	HO$^{\psi}$
78	**D N**	V	*Alectra sessiliflora*	HE
84 βC	**C** R S	Y	*O. ramosa*	HO$^{\psi}$
87 βC	**I L** F M S T	M	*B. hookeri*	HO$^{\psi}$
90	**V I L** A	S	*O. ramosa*	HO$^{\psi}$
91	**V A I L P S** T	C	*B. hookeri*	HO$^{\psi}$
91	**V A I L P S** T	T	*O. corymbosa*	HO
93	**E D G** T	K	*B. hookeri*	HO$^{\psi}$
		K	*B. strobilacea*	HO$^{\psi}$
98 βD	**I** F L M	T	*Chelone obliqua*	NP
99 βD	**A C V** P R	S	*Callitriche hermaphroditica*	NP
		S	*O. corymbosa*	HO
121 ˇ	**V E** G	I	*O. ramosa*	HO$^{\psi}$
124	**V** A	A	*H. sanguinea*	HO$^{\psi}$
129 ˇ	**A S**	T	*Plantago lanceolata*	NP
		V	*Orobanche cernua*	HO$^{\psi}$
131 ˇ βE	**R A** S H *	C	*H. sanguinea*	HO$^{\psi}$
132 βE	**A** S T V	D	*O. cernua*	HO$^{\psi}$
134 βE	**R** A	S	*H. atropurpurea*	HO$^{\psi}$
		C	*O. cernua*	HO$^{\psi}$
137 βE	**D**	Y	*O. cernua*	HO$^{\psi}$
139 βE	**R** Q	*	*O. ramosa*	HO$^{\psi}$
147 ●	**T** S	I	*O. cernua*	HO$^{\psi}$
		I	*O. ramosa*	HO$^{\psi}$
150	**G**	S	*O. ramosa*	HO$^{\psi}$
152	**P R** T	L	*H. sanguinea*	HO$^{\psi}$
		L	*O. cernua*	HO$^{\psi}$
154	**G A** C R T	E	*O. ramosa*	HO$^{\psi}$
155	**I T** M V	V	*Paulownia tomentosa*	NP
156 # ● αE	**Q**	K	*H. sanguinea*	HO$^{\psi}$
158 αE	**E D**	*	*O. cernua*	HO$^{\psi}$
159 αE	**R** G	I	*O. ramosa*	HO$^{\psi}$
166 #	**G A**	V	*B. hookeri*	HO$^{\psi}$
		V	*B. strobilacea*	HO$^{\psi}$
178 ˇ	**L** F	F	*H. atropurpurea*	HO$^{\psi}$
182 α1	**A**	T	*O. cernua*	HO$^{\psi}$
185 α1	**Y**	H	*O. ramosa*	HO$^{\psi}$
187 # α1	**R** *	G	*O. cernua*	HO$^{\psi}$
188 α1	**A** G V	T	*O. cernua*	HO$^{\psi}$
		T	*O. ramosa*	HO$^{\psi}$
190 # α1	**Y F**	D	*O. ramosa*	HO$^{\psi}$
198	**F**	Y	*H. atropurpurea*	HO$^{\psi}$
198 β2	**F**	C	*O. cernua*	HO$^{\psi}$
216 ● α2	**D** A E H	H	*L. clandestina*	HO
217 α2	**R** G	C	*H. sanguinea*	HO$^{\psi}$
221 α2	**C F S W** L	G	*L. clandestina*	HO
228	**A S** T V	T	*O. ramosa*	HO$^{\psi}$
242 ●	**A** P	T	*O. cernua*	HO$^{\psi}$
		T	*O. ramosa*	HO$^{\psi}$
258 # ● α3	**R** D K	K	*O. corymbosa*	HO
264 β4	**I**	M	*Alonsoa unilabiata*	NP
267 β4	**H** F	Y	*O. cernua*	HO$^{\psi}$

Table 3
Continued

Amino Acid Position[a]	Amino Acids Reported for 499 Seed Plants[b]	Amino Acid Observed in this Study	Taxon	Parasitic Classification[c]
279 α4	**S T** R	R	*Castilleja linariaefolia*	HE
285 ● α4	**R** Q *	Q	*L. clandestina*	HO
286 ● α4	**D** A N	N	*O. cernua*	HO$^{\psi}$
		N	*O. ramosa*	HO$^{\psi}$
288 # ●	**G R** A	V	*O. cernua*	HO$^{\psi}$
292 β5	**H**	Y	*B. hookeri*	HO$^{\psi}$
		Y	*B. strobilacea*	HO$^{\psi}$
293 β5	**I** T	L	*O. ramosa*	HO$^{\psi}$
294 active site	**H** N	Q	*O. ramosa*	HO$^{\psi}$
295 active site	**R** V	C	*B. hookeri*	HO$^{\psi}$
		C	*B. strobilacea*	HO$^{\psi}$
302 αF	**D H**	N	*H. sanguinea*	HO$^{\psi}$
306	**N I** S Y	D	*O. cernua*	HO$^{\psi}$
308	**G** D	S	*O. cernua*	HO$^{\psi}$
328 β6	**S A G** P	C	*L. clandestina*	HO
		L	*O. ramosa*	HO$^{\psi}$
329	**G**	R	*O. cernua*	HO$^{\psi}$
331	**V I**	A	*O. ramosa*	HO$^{\psi}$
341 α6	**I L S V** M	T	*L. clandestina*	HO
343 α6	**L**	F	*H. atropurpurea*	HO$^{\psi}$
		F	*H. sanguinea*	HO$^{\psi}$
		F	*O. ramosa*	HO$^{\psi}$
344 α6	**G** S V	D	*O. ramosa*	HO$^{\psi}$
351 α6	**D E L** F	V	*L. clandestina*	HO
352 α6	**D E G N**	R	*L. clandestina*	HO
358	**R A D P** G Q S	E	*Schlegelia parviflora*	NP
360	**R P** L	H	*O. cernua*	HO$^{\psi}$
365	**T A S** D L N P	P	*O. corymbosa*	HO
367 βH	**D F P S** E G	G	*B. hookeri*	HO$^{\psi}$
		G	*B. strobilacea*	HO$^{\psi}$
375	**L I** M V	F	*O. cernua*	HO$^{\psi}$
376 β7	**P A** T	T	*O. cernua*	HO$^{\psi}$
386	**H**	N	*O. ramosa*	HO$^{\psi}$
389 α7	**A** S T	S	*O. cernua*	HO$^{\psi}$
		S	*O. ramosa*	HO$^{\psi}$
392 α7	**E D** K	A	*O. ramosa*	HO$^{\psi}$
401 β8	**Q** E *	K	*B. hookeri*	HO$^{\psi}$
		K	*B. strobilacea*	HO$^{\psi}$
403 active site	**G**	D	*O. cernua*	HO$^{\psi}$
404 active site ˇ αP	**G A**	E	*H. atropurpurea*	HO$^{\psi}$
412 # ˇ	**G** E V	R	*O. cernua*	HO$^{\psi}$
418 # α8	**V A**	I	*O. fasciculata*	HO
419 α8	**A R** P	T	*H. atropurpurea*	HO$^{\psi}$
423 α8	**A S** V	D	*B. hookeri*	HO$^{\psi}$
		D	*B. strobilacea*	HO$^{\psi}$
431 α8	**R** C	H	*B. hookeri*	HO$^{\psi}$
		H	*B. strobilacea*	HO$^{\psi}$
432 # α8	**N** I Y	D	*B. hookeri*	HO$^{\psi}$
435	**R H Q** G Y	C	*O. cernua*	HO$^{\psi}$
439 αG	**Q A H I L R S T V** C K P	D	*B. hookeri*	HO$^{\psi}$
442	**N A T** I P	K	*Torenia fournieri*	NP
		D	*B. hookeri*	HO$^{\psi}$
443	**E A D N Q T V**	K	*Celsia arcturus*	NP
448	**A P R** D L T	T	*B. strobilacea*	HO$^{\psi}$
450	**K E Q T** N R S	N	*P. lanceolata*	NP
456	**A S** F	R	*Cycnium racemosum*	HE
		V	*O. ramosa*	HO$^{\psi}$
462 ˇ	**W R**	S	*O. fasciculata*	HO
464	**E A Q** K S	D	*O. fasciculata*	HO
468	**N D E I Q V** K T Y	K	*O. ramosa*	HO$^{\psi}$

[a] Boldface indicates substitutions previously known from only one other seed plant. Structural motifs follow Kellogg and Juliano (1977); ˇ indicates intradimer interactions, ● indicates dimer–dimer interactions, and # represents interactions between large and small subunits. Alpha-Beta barrel rand active site residues are also indicated.

[b] Usage follows Kellogg and Juliano (1997); boldface abbreviations represent amino acid substitutions found in two or more taxa; * = stop codon.

[c] Abbreviations for parasitic classification: NP = nonparasite, HE = hemiparasite, HO = holoparasite, HO$^{\psi}$ = holoparasite with *rbcL* pseudogene.

Table 4
Amino Acid Substitutions in Structurally Important Regions of the RuBisCO Large-Subunit Polypeptide

Taxon	Total No. of Changes[a]	α-Helix	β-Sheet	Intradimer	Between Subunits	Dimer–Dimer	Active Site	pam250 Difference[b]
Holoparasites with pseudogenes[c]								
Orobanche cernua	128	33	35	38	10	7	5	539
Orobanche ramosa	55	18	12	9	8	3	5	202
Hyobanche sanguinea	28	11	5	5	2	3	2	98
Boschniakia hookeri	32	11	11	2	6	0	2	91
Boschniakia strobilacea	30	10	10	2	6	0	2	91
Hyobanche atropurpurea	21	11	5	2	1	1	1	67
Average	49.00	15.67	13.00	9.67	5.50	2.33	2.83	181.33
Holoparasites with *rbcL* ORFs[d]								
Lathraea clandestina	19	11	5	1	1	1	0	89
Alectra orobanchoides	13	7	4	0	1	1	0	50
Harveya purpurea	14	7	5	0	1	1	0	47
Harveya capensis	14	7	5	0	1	1	0	43
Orobanche fasciculata	15	7	5	1	2	0	0	43
Orobanche corymbosa	14	7	5	0	1	1	0	37
Striga gesnerioides	13	6	5	0	1	1	0	36
Average	14.57	7.43	4.86	0.29	1.14	0.86	0.00	49.29
Hemiparasites[d]								
Striga hermonthica	13	7	4	0	1	1	0	53
Striga passargei	14	8	4	0	1	1	0	50
Striga asiatica	14	7	5	0	1	1	0	47
Cycnium racemosum	14	7	5	0	1	1	0	47
Melampyrum lineare	13	7	4	0	1	1	0	46
Buchnera floridana	13	7	4	0	1	1	0	44
Castilleja linariaefolia	13	6	5	1	1	0	0	44
Seymeria pectinata	13	6	5	0	1	1	0	40
Tozzia alpina	11	6	4	0	1	0	0	31
Alectra sessiliflora	9	4	4	0	0	1	0	25
Pedicularis foliosa	10	5	4	0	1	0	0	25
Average	12.45	6.36	4.36	0.09	0.91	0.73	0.00	41.09
Nonparasites[d]								
Gratiola pilosa	16	9	5	0	2	0	0	50
Torenia fournieri	19	11	6	1	1	0	0	45
Verbascum thapsus	8	4	4	0	0	0	0	45
Schlegelia parviflora	13	4	6	2	0	1	0	41
Alonsoa unilabiata	13	6	5	0	1	1	0	40
Paulownia tomentosa	6	2	4	0	0	0	0	36
Chelone obliqua	11	5	5	0	1	0	0	31
Average	12.29	5.86	5.00	0.43	0.71	0.29	0.00	41.14

NOTE.—Comparison includes all gap regions but excludes amino acid positions 1–9 and 451–476 because of missing data.

[a] Where total number of changes does not equal the total changes for all structural features, there are mutations in regions that have more than one structural feature (see Kellogg and Juliano 1997 for review).

[b] The pam250 difference is the net change in score between the taxon listed in the table and the reference protein from *Nicotiana tabacum* (see text).

[c,d] Plant categories with different letters denote statistically significant differences (one-way ANOVA with multiple comparisons using a Tukey-Kramer HSD) at $P < 0.05$ ($q = 2.74$, df = β, $n = 31$; SAS Institute Inc. 1995).

reflects both the quantity and the distribution of substitutions along the length of the polypeptide.

To further assess the effect of amino acid replacements on the large subunit of RuBisCO, we charted amino acid replacements in a similar fashion to Kellogg and Juliano (1997) (table 3). Only amino acid residues not observed in the Kellogg and Juliano (1997) study, or for which only a single taxon of the 499 were reported to have a particular amino acid change compared to the other 498, are included here. The amino acid residues previously recorded for 499 seed plants are listed, along with the new substitutions observed in this study. Kellogg and Juliano (1997) reported that 105 of 476 amino acid residues were absolutely conserved and an additional 110 residues have been modified in only 1 of the 499 species assayed (highly conserved sites). We found 14 new amino acid substitutions at sites in Kellogg and Juliano's (1997) absolutely conserved category, of which 13/14 are from nonphotosynthetic plants and 1 is from a nonparasite (amino acid position 264 I → M in *Alonsoa unilabiata*). For highly conserved sites (with toggling back and forth between 2 amino acid residues, or for which all but 1 of the 499 had a single residue), we found 27 sites with new amino acid substitutions and two sites at which the amino acid replacement matched the report of a single taxon change from the conserved site. Of these 29 sites, 27 of the changes were found only in nonphotosynthetic plants.

An analysis of pam250 scores (fig. 2 and table 4) for RuBisCO large-subunit polypeptide sequences re-

veals that all species of nonparasitic and hemiparasitic plants have similar scores; therefore, RuBisCO proteins from these categories of plants may be biochemically similar to the reference protein from tobacco. In addition, half of the polypeptides from holoparasitic plants are also indistinguishable from the reference protein. The only sequences with significant differences are the hypothetical in-frame translations of the *rbcL* pseudogenes (fig. 2).

Discussion

Phylogeny Reconstruction Based on *rbcL*

Studies employing *rbcL* sequencing for phylogeny reconstruction of taxa or for studies of molecular evolution in the Scrophulariaceae have been restricted to nonparasitic plants until recently (Olmstead and Reeves 1995; dePamphilis, Young, and Wolfe 1997; Wolfe and dePamphilis 1997). Olmstead and Reeves (1995) sequenced *rbcL* and *ndhF* from nine nonparasitic genera of Scrophulariaceae for phylogenetic analysis of the Scrophulariales. The ability of *rbcL* nucleotide sequence data to resolve phylogenetic relationships was less than that of *ndhF* sequences, which evolve at a slightly faster rate (Olmstead and Sweere 1994), or from combined *rbcL* and *ndhF* data sets. However, *ndhF* is not useful for phylogenetic analysis of parasitic genera of Scrophulariaceae because of the loss of the locus in most of the taxa assayed or the high level of sequence divergence among those taxa that have retained the gene (unpublished data). From the results of the *rbcL* and *ndhF* sequence analyses, two distinct clades of genera in the Scrophulariaceae were separated by taxa from other families in Scrophulariales (Olmstead and Reeves 1995). Our results confirm that Scrophulariaceae is not a monophyletic group, and suggest that the traditional Scrophulariaceae may be a large paraphyletic grade. However, our results differ from those of Olmstead and Reeves (1995) in the topological positions of the major clades.

Parasitism in the Scrophulariales

The major phylogenetic finding in our study was the agreement between results based on *rbcL* sequences and those based on other plastid genes (dePamphilis, Young, and Wolfe 1997; Nickrent et al. 1998) that all parasitic plants of Scrophulariales form a clade that includes all examined genera traditionally assigned to Orobanchaceae (fig. 1). As Cronquist (1988, p. 432) put it, ''[t]he evolutionary journey toward parasitism obviously begins with chlorophyllous root-parasites in the Scrophulariaceae; the Orobanchaceae merely occupy the house at the end of the road.'' The results presented here and those from studies using other chloroplast and nuclear gene sequences (dePamphilis, Young, and Wolfe 1997; Nickrent et al. 1998; unpublished data) reveal that Orobanchaceae is firmly nested within Scrophulariaceae. The phylogeny reconstruction based on a combined analysis of *rps2* and *matK* depicts Orobanchaceae as a clade, whereas results based on *rbcL* show Orobanchaceae as a grade in the monophyletic group of parasites (figs. 1–2). Because the holoparasitic genera of Orobanchaceae are part of the parasite clade, we suggest the interim use of the family designation Scrophulariaceae to include all of the parasitic genera of Scrophulariales until a reclassification of the order is completed.

Six nucleotide substitutions within the *rbcL* gene support the monophyly of the parasites (fig. 1). However, relative nodal support, as measured by the parsimony jackknife and bootstrap values, is very low for the parasite clade (fig. 1). In contrast, the support for this clade is high based on *rps2* data alone (bootstrap = 86%, Bremer support = 4; dePamphiis, Young, and Wolfe 1997) and a combined *rps2*/*matK* data set (bootstrap = 99%, Bremer support = 11; unpublished data). We conclude, therefore, that the phylogenetic signal in the *rbcL* data set is real and that the monophyly of the parasites is supported in phylogeny reconstructions from analyses of three plastid gene sequences.

Phylogenetic Utility of *rbcL* in Parasitic Plants

Concerns about using *rbcL* for phylogeny reconstruction of parasitic plants include the possibility of DNA contamination, long-branch attractions, and unavailability of sequence data for comparison with nonparasitic taxa due to absence of the gene (Olmstead and Palmer 1994). Contamination is a definite concern in dealing with photosynthetic genes of parasitic plants, and extreme measures must be used to reduce the possibility of contamination during extraction and PCR amplification. Control measures implemented in this study included acid treatment and baking of mortars and pestles, UV irradiation of solutions, temporal separation of DNA isolation for parasitic and nonparasitic plant DNA, and isolation of DNA in a fume hood away from the main laboratory. We did not see any obvious evidence of long-branch attractions in the data analysis. Instead, the topology resulting from the *rbcL* analysis was very similar to the gene trees obtained from *rps2* (dePamphiis, Young, and Wolfe 1997) and *matK* sequences (unpublished data). The absence of *rbcL* data eliminated two holoparasitic taxa from the analysis— *Epifagus*, with an unalignable remnant of the gene, and *Conopholis*, which lacks the *rbcL* locus entirely.

The agreement of the *rbcL* tree with two other plastid gene trees and the general congruence of the *rbcL* tree to preliminary 18S rDNA results (Nickrent et al. 1998) demonstrate that *rbcL* can be a useful gene for phylogeny reconstruction in parasitic plant lineages. These results suggest that other groups of parasitic and mycoheterotrophic plants may be connectable to the existing *rbcL* data set if taxon sampling is dense enough. Our success with *rbcL* may have resulted from the unique diversity of hemiparasites and holoparasites in a single lineage. Most other groups of parasitic plants and mycoheterotrophs are restricted to either hemiparasites or holoparasites, but rarely (e.g., *Pyrola*) have a continuum of parasitic abilities like that found for Scrophulariaceae (dePamphilis 1995). The placement of holoparasitic lineages with greatly altered *rbcL* pseudogenes, and with genes that lack close relatives, may remain a challenge. For example, if *O. cernua* and *O. ramosa* are used as the sole representatives of the Scrophulariaceae

parasite lineage, the topology of the entire tree changes dramatically (results not shown).

Holoparasitism and *rbcL* Pseudogene Formation

Although holoparasitism has arisen independently several times in Scrophulariaceae (fig. 2), it is clear that *rbcL* pseudogene formation is not an immediate consequence of the loss of photosynthesis in all holoparasitic lineages. Each of the five lineages of holoparasites has taxa with intact ORFs for *rbcL* (fig. 2).

In nuclear genomes, a common model of pseudogene formation is the duplication of a gene with one copy of the gene becoming nonfunctional (Hardison 1991; Gottlieb and Ford 1997). However, this mode of pseudogene formation has not been observed within chloroplast genomes. Instead, pseudogene formation is usually the result of genome rearrangements or deletions within the coding region of a gene (Morden et al. 1991; Ogihara, Terachi, and Sasakuma 1991; Shimada and Sugiura 1991; Taylor et al. 1991; Wolfe, Morden, and Palmer 1992; Downie et al. 1994; Haberhausen and Zetsche 1994). There are also hypotheses that plastid genes transferred to the nucleus or mitochondrion may represent pseudogenes (Stern and Lonsdale 1982; Makaroff and Palmer 1987; Manhart 1994). However, there are no studies that demonstrate *rbcL* pseudogene formation via the latter mechanism. *rbcL* pseudogene formation in Scrophulariales apparently does not occur via a duplication event followed by subsequent degeneration of one copy of the gene. Instead, primary pseudogene formation has occurred a minimum of three times (fig. 2), with the single *rbcL* locus becoming nonfunctional as a result of relaxed functional constraint on the protein in increasingly heterotrophic plants (see discussion below).

The most extreme pseudogene represented in figure 3 is from *Epifagus virginiana* where only a small remnant of the gene remains (Wolfe, Morden, and Palmer 1992). In the holoparasitic species *Conopholis americana*, there is not even a remnant of the gene left in the plastid genome (Colwell 1994). *Orobanche cernua* has a pseudogene that includes a deletion of 330 bp beginning at position 42, flanked by a 12-nt repeat. There is a clear trend in *rbcL* pseudogene formation in Scrophulariaceae for the accumulation of multiple and/or large deletions. There is also a trend for the accumulation of small insertions, and three of the taxa have stop codons at amino acid position 20 (fig. 3).

The *rbcL* pseudogene sequences for *B. hookeri* and *B. strobilacea* differ by only a 5-bp insertion at *Nicotiana* position 106, and both sequences share a stop codon at amino acid position 20. The most parsimonious interpretation of these mutations is that pseudogene formation occurred prior to the divergence of these two species. In contrast, the *rbcL* pseudogenes of *O. cernua* and *O. ramosa* have numerous independent mutations, suggesting that pseudogene formation occurred independently in these species. The different patterns of pseudogene formation among these taxa reflect the difference in taxonomic relationships within *Boschniakia* from those within *Orobanche*. For example, *B. hookeri* and *B. strobilacea* are morphologically very similar and parasitize closely related host plants in Ericaceae (Heckard 1993), whereas *O. cernua* and *O. ramosa* differ greatly in stem morphology (single stems vs. branched) and host preferences (sunflower vs. tomato; Musselman 1980). There are a few minor overlaps of insertions or deletions found among taxa in the Orobanchaceae lineage (fig. 3), but the majority of mutations are independent of the phylogenetic history of the taxa.

Synonymous and Nonsynonymous Substitution Rates

The increased synonymous substitution rate observed for holoparasites raises the question of whether the increase is a reflection of a genome-level phenomenon or something associated explicitly with the *rbcL* locus. Following formal relative-rates tests for the *rps2* gene, dePamphilis, Young, and Wolfe (1997) concluded that substantial differences in synonymous substitution rates among parasitic lineages may have resulted from variation in fundamental mutation or repair rates. Some parasitic lineages of Scrophulariaceae exhibiting rate heterogeneity in the *rps2* gene sequence (dePamphiis, Young, and Wolfe 1997) do not have long branch lengths in the *rbcL* tree. Lineages that show accelerated rates of evolution for *rps2* but not *rbcL* include the clade containing *Buchnera*, *Cycnium*, and *Striga* and the clade containing species of *Striga*. In both the *rps2* and *rbcL* trees, long branches are observed for the holoparasitic taxa traditionally circumscribed by Orobanchaceae. These taxa have *rbcL* pseudogenes, with the exception of two North American species of *Orobanche* (fig. 2), but have apparently functional *rps2* genes.

Overall rates of molecular evolution of plastid genes are the result of the underlying synonymous substitution rates and effects due to functional constraints on the particular gene (mostly affecting nonsynonymous substitutions; dePamphilis 1995; dePamphilis, Young, and Wolfe 1997). Relative-rates comparisons for sequences of *rps2, rbcL,* and *matK* from the parasitic continuum of Scrophulariales will be reported elsewhere. Our preliminary conclusions from the comparison of *rps2* and *rbcL* trees are that rates of molecular evolution appear to vary independently in at least some lineages of parasitic Scrophulariaceae and that underlying patterns of synonymous or nonsynonymous evolution, or both, may differ for the two genes.

It is understandable how loss of photosynthesis will have a direct impact on functional constraint for photosynthesis genes such as *rbcL*. Several lineages, including heterotrophic taxa traditionally circumscribed by Orobanchaceae, the clade including *Lathraea*, and the lineage including African holoparasites (*Striga* through *Alectra*), have numerous nucleotide substitutions (fig. 1). However, some lineages of nonparasites also have long *rbcL* branch lengths, including Callitrichaceae, Plantaginaceae, and individual taxa within the clade that includes Lentibulariaceae and Acanthaceae (fig. 1). The longer branch lengths primarily result from nonsynonymous substitutions (table 2). Callitrichaceae consist of aquatic plants, Plantaginaceae represent a shift from entomophilous to wind pollination, and Lentibulariaceae include insectivorous plants adapted to wet

habitats (Mabberley 1993). It is unclear how these life history traits might be associated with the molecular evolution of the chloroplast genome. Kellogg and Juliano (1997) noted that several regions within the *rbcL* sequence are subject to toggling back and forth between amino acid replacements, and several regions are saturated for nucleotide substitutions. These regions are, in general, not associated with highly conserved functional domains of the RuBisCO large subunit. The longer branch lengths and higher K_N values for some of the nonparasitic lineages of Scrophulariales may result from this type of phenomenon rather than from a change in the functional constraint of the protein.

Amino Acid Substitutions

Regions of the RuBisCO large-subunit polypeptide are highly conserved across seed plants, with no amino acid replacements (Kellogg and Juliano 1997) in a sample of 499 *rbcL* sequences (Chase et al. 1993). For example, important structural domains were conserved across this wide sampling of seed plants, and there are additional conserved sequences of no known function (Kellogg and Juliano 1997). For the nonparasitic Scrophulariaceae, the same pattern holds (fig. 4). However, with increasing heterotrophic ability in Scrophulariaceae, there is an apparent loss of conserved amino acid sequences in the RuBisCO large-subunit protein (fig. 4). Hemi- and holoparasitic plants of the Scrophulariaceae have more amino acid replacements in the large subunit of RuBisCO than do photosynthetic plants, and holoparasites have more substitutions than do hemiparasites (fig. 4). However, biochemically similar amino acid replacements (e.g., replacing leucine with isoleucine) often do not affect the structure and/or function of the protein and are therefore effectively neutral replacements. The exception to this is the case in which an amino acid substitution occurs in an important structural motif such as the active site. For example, table 1 of Kellogg and Juliano (1997) depicts shifts from one amino acid to a biochemically similar amino acid at active-site positions 60 (E $\rightarrow$ D), 123 (N $\rightarrow$ G, Q, S), 175 (K $\rightarrow$ R, H), 177 (K $\rightarrow$ R), 294 (H $\rightarrow$ K), 295 (R $\rightarrow$ K), 327 (H $\rightarrow$ K, R), 334 (K $\rightarrow$ R), and 379 (S $\rightarrow$ C, T); each of these substitutions results in $<1\%$ or no activity of the enzyme. Shifts involving biochemically similar residues at sites with no known structural functions show no effect on enzyme activity (e.g., sites 198 [D $\rightarrow$ E], 309 [I $\rightarrow$ M, L], and 338 [D $\rightarrow$ E]; Kellogg and Juliano 1997). Plotting the number of amino acid replacements against the strict-consensus tree gives a first indication of molecular evolution of the protein but does not elucidate the effect of molecular evolution on the structure and/or function of the protein.

Although there is an increasing amount of amino acid replacement with increasing heterotrophy (table 3 and fig. 4), the overall effect of these replacements on the protein structure as inferred from the pam250 analysis is minimal (table 4 and fig. 2). The only pam250 scores significantly different from the reference protein are pseudogene sequences. The translated protein sequences from the holoparasitic *rbcL* ORFs (*Alectra orobanchoides*, *Harveya capensis*, *Harveya purpurea*, *Lathraea clandestina*, *O. corymbosa*, *O. fasciculata*, *Striga gesnerioides*) are nearly identical to photosynthetic plants in terms of inferred structure. Furthermore, for most of the holoparasites with *rbcL* ORFs, or for hemiparasites in lineages including holoparasitic taxa, there are few additional amino acid substitutions in critical structural features of the RuBisCO large-subunit polypeptide compared with nonparasites (table 4). Holoparasites with ORFs have a slightly higher (but not significant) number of total changes in important structural motifs and pam250 score differences than do hemiparasites and nonparasites, but the overall substitution patterns are very similar to those of nonparasitic plants (table 4). Hence, despite increased synonymous and nonsynonymous substitutions in nonphotosynthetic parasites, the functionality of RuBisCO may not change dramatically with increasing heterotrophy. The exception in this study is the holoparasite *L. clandestina*, which expresses *rbcL* at low levels (Bricaud, Thalouarn, and Renaudin 1986; Thalouarn, Arnaud, and Renaudin 1989; Delavault, Sakanyan, and Thalouarn 1995). This species of *Lathraea* has sustained a number of nonconservative amino acid substitutions in important structural motifs in a pattern similar to the hypothetical in-frame translation of *Hyobanche* (table 4). The overall pam250 scores for *Lathraea* are lower than scores from other taxa with *rbcL* ORFs (fig. 2), yet the gene is minimally expressed. From the results of this analysis, we hypothesize that the low levels of RuBisCO activity in *Lathraea* may be partly a result of low expression levels (Bricaud, Thalouarn, and Renaudin 1986; Thalouarn, Arnaud, and Renaudin 1989; Delavault, Sakanyan, and Thalouarn 1995) and partly due to the function-altering mutations sustained by the protein.

Retention of *rbcL* in Holoparasitic Plants

Open reading frames for the *rbcL* gene have been retained in 7 of the 13 holoparasitic taxa examined in this study. Furthermore, the gene is expressed in *L. clandestina* (Bricaud, Thalouarn, and Renaudin 1986; Thalouarn, Arnaud, and Renaudin 1989; Delavault, Sakanyan, and Thalouarn et al. 1994), although at low levels. Expression of the *rbcL* gene has also been noted in the holoparasitic genus *Cuscuta* (Cuscutaceae; Haberhausen, Valentin, and Zetsche et al. 1992; Machado and Zetsche 1990) and in the heterotrophic alga *Astasia longa* (Siemeister and Hachtel 1990). Taken together, the retention of *rbcL* ORFs, analyses of their evolutionary patterns (Wolfe and dePamphilis 1997), and the expression of the gene in several heterotrophic lineages suggest that the gene remains functional after the loss of photosynthesis.

Wolfe and dePamphilis (1997) suggested three alternative hypotheses for the retention of *rbcL* in holoparasitic plant lineages: (1) the plant requires RuBisCO activity for low levels of autotrophic carbon fixation; (2) the oxygenase activity may be functional in the glycolytic pathway; and (3) ORFs are maintained not by functional constraint but by chance, implying that loss of

photosynthesis has been too recent for sufficient accumulation of deleterious mutations.

Some parasitic plants of Scrophulariaceae and other parasitic plant families perform photosynthesis only during a discrete phase of the life cycle, such as seedling development or seed maturation (Kuijt 1969, p. 95). These plants would technically be hemiparasites because of the minimal retention of photosynthetic pigments and some capacity for autotrophic carbon fixation, but they do not resemble other hemiparasites with well-developed green leaves and normal photosynthetic abilities (e.g., *Castilleja* or *Pedicularis*). For example, the hemiparasitic genus *Tozzia* spends several years of its life cycle as an achlorophyllous underground parasite, but then a green flowering shoot emerges (and after seed set, the entire plant dies; Kuijt 1969, p. 95). Such cryptic hemiparasitism would easily be confused with holoparasitism, and it may be that the retention of *rbcL* in so many holoparasitic Scrophulariales, as well as in *Cuscuta* and many mycoheterotrophic monocots (M. Chase, personal communication), may be due to this life history trait.

Physiological data are helpful in elucidating whether cryptic hemiparasitism is the only possible explanation for the retention of *rbcL* ORFs in holoparasites in Scrophulariaceae. *Striga* belongs to the African parasite clade (fig. 1), which includes several genera with one or more holoparasitic species. *Striga hermonthica* and *S. gesnerioides* have chlorophyll and intact *rbcL* ORFs, but physiological data have demonstrated that these hemi- and holoparasites, respectively, obtain most or all of their reduced carbon from their host plants (Press, Smith, and Stewart 1991) and could presumably survive without fixing any carbon themselves. In Scrophulariaceae, the presence of chlorophyll is apparently not enough of an indicator to determine whether a plant is functionally a hemiparasite or a holoparasite. Similarly, the presence of an *rbcL* ORF is not a sufficient indicator that the plant is functionally photosynthetic.

The question remains, then, whether the oxygenase activity or some completely unexpected function of RuBisCO is important in parasitic plants. RuBisCO oxygenase activity results from a competitive reaction from photorespiration and is involved in glycolate metabolism (e.g., glycine and serine biosynthesis; Miziorko and Lorimer 1983). To invoke oxygenase activity as a necessary function of glycolate metabolism in parasitic plants implies that amino acid biosynthesis via this pathway is crucial for parasitic plants. However, physiological studies (Press, Shah, and Stewart 1986) have demonstrated that nitrogen reduction and the distribution of glutamine synthetase isoforms, which function in nitrate reduction, are drastically reduced with increasing heterotrophy. Press, Shah, and Stewart (1986) argued, ''[t]hese reductions in the capacity of parasitic angiosperms to assimilate inorganic nitrogen ions reflect the plant's access to a reduced organic nitrogen supply. It is likely therefore that they may also lack the ability, or demonstrate a reduced ability to interconvert amino acids, since they are readily available from the host xylem and phloem tissue via the haustorium.'' With the ability to absorb reduced nitrogen compounds from a host plant, it seems unlikely that glycolate metabolism resulting from photorespiration is critical in parasitic plants unless amino acid availability from the host is insufficient.

In our previous examination of four species of *Orobanche* (Wolfe and dePamphilis 1997), we demonstrated that stochastic factors alone are unlikely to explain the retention of *rbcL* ORFs in the holoparasites *O. corymbosa* and *O. fasciculata.* This conclusion was based on a mathematical model that estimated the probability of retention of an intact ORF after the loss of photosynthesis. It is also consistent with the observation reported here that there are many more synonymous substitutions than nonsynonymous substitutions, implying a functional constraint on the *rbcL* gene product in these species. We also examined the 5′ and 3′ untranslated regions of *rbcL* for the four species of *Orobanche* and found that in *O. corymbosa,* but not necessarily in *O. fasciculata,* the promoter, ribosome-binding region, and termination sequences were similar to those of photosynthetic plants or to those of parasitic plants known to have *rbcL* gene expression. It is certainly possible that the sequence from *O. fasciculata* represents a cryptic pseudogene—one in which the reading frame is still intact but the sequence is not expressed due to changes in the upstream or downstream control regions. If so, then the implication is that initial degradation of the gene may be slow, but once a stop codon or frame-altering indel occurs, the gene structure is no longer constrained. This certainly appears to be a plausible hypothesis when reviewing the information gained from this study: (1) there is an increase in the rate of synonymous substitutions with increasing heterotrophy, possibly due to increased rates of error in plastid DNA replication or repair in parasitic plants (dePamphilis, Young, and Wolfe 1997); (2) there is an increase in amino acid replacement with increasing heterotrophy, but the replacement pattern is largely neutral until functional constraints have been released; (3) heterotrophic plants may retain a viable *rbcL* gene or ORF after the loss of photosynthesis as a result of either chance or functional constraints for an unknown, nonphotosynthetic function of RuBisCO in some parasites; and (4) with a small disruptive mutation to the *rbcL* locus (e.g., *Hyobanche*), there is a dramatic increase in accumulated mutations. With the relaxed functional constraints associated with the adaptation to heterotrophy, it is apparent that there is no cost to the loss of the gene for many holoparasitic plants, but neither is there a substantial cost to maintaining an ORF.

Acknowledgments

We thank K. Steiner, G. Sallé, J. Alison, A. Batten, M. Wetherwax, L. Musselman, W. Wetschnig, C. Morden, and J. Palmer for assistance in collecting plant material or providing DNAs; N. Young, J. Leebens-Mack, L. M. Bowe, E. Kellogg, M. Chase, R. Olmstead, and D. Crawford for helpful discussions or comments on the manuscript; D. McCauley and L. Leege for help on the statistical analyses; J. Wenzel for assistance with phy-

logenetic analyses; and Mr. Tim Atkinson for technical assistance. Financial support for this research was provided by an NSF postdoctoral fellowship (BIR-9303630) to A.D.W., and by NSF grant DEB 9120258 to C.W.D.

LITERATURE CITED

ALBERT, V. A., S. E. WILLIAMS, and M. W. CHASE. 1992. Carnivorous plants: phylogeny and structural evolution. Science **257**:1491–1495.

BRICAUD, C. H., P. THALOUARN, and S. RENAUDIN. 1986. Ribulose 1,5-bisphosphate carboxylase activity in the holoparasite *Lathraea clandestina* L. J. Plant Physiol. **125**:367–370.

CHAPMAN, M. S., S. W. SUH, P. M. G. CURMI, D. CASCIO, W. W. SMITH, and D. S. EISENBERG. 1988. Tertiary structure of plant RuBisCO: domains and their contacts. Science **241**:71–74.

CHASE, M. W., D. E. SOLTIS, R. G. OLMSTEAD et al. (42 coauthors). 1993. Phylogenetics of seed plants: an analysis of nucleotide sequences from the plastid gene *rbcL*. Ann. Mo. Bot. Gard. **80**:528–580.

COLWELL, A. 1994. Genome evolution in a non-photosynthetic plant, *Conopholis americana*. Ph.D. dissertation, Washington University, St. Louis.

CRONQUIST, A. 1988. The evolution and classification of flowering plants. 2nd edition. The New York Botanical Garden, New York.

DAYHOFF, M. O., R. M. SCHWARTZ, and B. C. ORCUTT. 1978. A model of evolutionary change in proteins. Pp. 345–352 *in* M. O. DAYHOFF, ed. Atlas of protein sequence and structure. Maryland National Biomedical Research Foundation, Silver Springs.

DE LA HARPE, A. C., N. GROBBELAAR, and J. H. VISSER. 1980. The ultrastructure of the chloroplast and the chlorophyll content of various South African parasitic flowering plants. Z. Pflanzenphysiol. **100**:85–90.

DELAVAULT, P., V. SAKANYAN, and P. THALOUARN. 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. 1995. Genes and genomes. Pp. 177–205 *in* M. C. PRESS and J. D. GRAVES, eds. Parasitic plants. Plenum Press, Chapman and Hall, London.

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.

DEPAMPHILIS, C. W., N. D. YOUNG, and A. D. WOLFE. 1997. Evolution of plastid gene *rps2* in a lineage of hemiparasitic and holoparasitic plants: many losses of photosynthesis and complex patterns of rate variation. Proc. Natl. Acad. Sci. USA **94**:7367–7372.

DOWNIE, S. R., D. S. KATZ-DOWNIE, K. H. WOLFE, P. J. CALIE, and J. D. PALMER. 1994. Structure and evolution of the largest chloroplast gene (ORF2280): internal plasticity and multiple gene loss during angiosperm evolution. Curr. Genet. **25**:367–378.

DUVALL, M. R., M. T. CLEGG, M. W. CHASE et al. (11 coauthors). 1993. Phylogenetic hypotheses for the monocotyledons constructed from *rbcL* sequence data. Ann. Mo. Bot. Gard. **80**:607–619.

FARRIS, J. S., V. A. ALBERT, M. KALLERSJO, D. LIPSCOMB, and A. G. KLUGE. 1996. Parsimony jackknifing outperforms neighbor-joining. Cladistics **12**:99–124.

FELSENSTEIN, J. 1978. Cases in which parsimony and compatibility methods will be positively misleading. Syst. Zool. **27**:401–410.

GAUT, B. S., S. V. MUSE, W. D. CLARK, and M. T. CLEGG. 1992. Relative rates of nucleotide substitution at the *rbcL* locus of monocotyledonous plants. J. Mol. Evol. **35**:292–303.

GOTTLIEB, L. D., and V. W. FORD 1997. A recently silenced, duplicate *PgiC* locus in *Clarkia.* Mol. Biol. Evol. **14**:125–132.

HABERHAUSEN, G., K. VALENTIN, and K. ZETSCHE. 1992. Organization and sequence of photosynthetic genes from the plastid genome of the holoparasitic flowering plant *Cuscuta reflexa*. Mol. Gen. Genet. **232**:154–161.

HABERHAUSEN, G., and K. ZETSCHE. 1994. Functional loss of all *ndh* genes in an otherwise relatively unaltered plastid genome of the holoparasitic plant *Cuscuta reflexa*. Plant Mol. Biol. **24**:217–222.

HARDISON, R. C. 1991. Evolution of globin gene families. Pp. 272–289 *in* R. K. SELANDER, A. G. CLARK, and T. S. WHITTAM, eds. Evolution at the molecular level. Sinauer, Sunderland, Mass.

HARTMAN, F. C., R. S. FOOTE, F. W. LARIMER et al. (12 coauthors). 1987. Function of active site residues of ribulose bisphosphate carboxylase/oxygenase. Pp. 9–20 *in* D. VON WETTSTEIN and N.-H. CHUA, eds. Plant molecular biology. Plenum Press, New York, N.Y.

HARTMAN, F. C., C. D. STRINGER, and E. H. LEE. 1984. Complete primary structure of ribulosebisphosphate carboxylase/oxygenase from *Rhodospirillum rubrum*. Arch. Biochem. Biophys. **232**:280–295.

HECKARD, L. R. 1993. Orobanchaceae. *In* J. C. HICKMAN, ed. The Jepson manual: higher plants of California.

HEDREN, M., M. W. CHASE, and R. G. OLMSTEAD. 1995. Relationships in the Acanthaceae and related families as suggested by cladistic analysis of *rbcL* nucleotide sequences. Plant Syst. Evol. **194**:93–109.

HIGGINS, D. G., A. J. BLEASBY, and R. FUCHS. 1992. CLUSTAL V. Improved software for multiple sequence alignment. Comput. Appl. Biosci. **8**:189–191.

IGARISHI, Y., B. A. MCFADDEN, and T. EL-GUL. 1985. Active site histidine in spinach ribulosebisphosphate carboxylase/oxygenase modified by diethyl pyrocarbonate. Biochemistry **24**:3957–3962.

KELLOGG, E. A., and N. D. JULIANO. 1997. The structure and function of RuBisCO and their implications for systematic studies. Am. J. Bot. **84**:413–428.

KNIGHT, S., I. ANDERSSON, and C.-I. BRÄNDÉN. 1990. Crystallographic analysis of ribulose 1,5-bisphosphate carboxylase from spinach at 2.4 Å resolution. J. Mol. Biol. **215**: 113–160.

KUIJT, J. 1969. The biology of parasitic flowering plants. University of California Press, Berkeley and Los Angeles.

KUMAR, S., K. TAMURA, and M. NEI. 1993. MEGA: molecular evolutionary genetics analysis. Version 1.01. The Pennsylvania State University, University Park.

LORIMER, G. 1981. Ribulose bisphosphate carboxylase: amino acid sequence of a peptide bearing the activator carbon dioxide. Biochemistry **20**:1236–1240.

LORIMER, G., S. GUTTERIDGE, and M. W. MADDEN. 1987. Partial reactions of ribulose bisphosphate carboxylase: their utility in the study of mutant enzymes. Pp. 21–31 *in* D. VON WETTSTEIN and N.-H. CHUA, eds. Plant molecular biology. Plenum Press, New York.

MABBERLEY, D. J. 1993. The plant-book. A portable dictionary of the higher plants. Cambridge University Press, Cambridge, England.

MACHADO, M. A., and K. ZETSCHE. 1990. A structural, functional and molecular analysis of plastids of the holoparasites *Cuscuta reflexa* and *Cuscuta europaea*. Planta **181**:91–96.

MADDISON, W. P., and D. R. MADDISON. 1992. MacClade. Sinauer, Sunderland, Mass.

MAKAROFF, C. A., and J. D. PALMER. 1987. Extenisve mitochondrial specific transcription of the *Brassica campestris* mitochondrial genome. Nucleic Acids Res. **15**:5141–5156.

MANHART, J. R. 1994. Phylogenetic analysis of green plant *rbcL* sequences. Mol. Phylogenet. Evol. **3**:114–127.

MIZIORKO, H. M., and G. H. LORIMER. 1983. Ribulose-1,5-bisphosphate carboxylase-oxygenase. Annu. Rev. Biochem. **52**:507–535.

MORDEN, C. W., K. H. WOLFE, C. W. DEPAMPHILIS, and J. D. PALMER. 1991. Plastid translation and transcription genes in a non-photosynthetic plant: intact, missing and pseudogenes. EMBO J. **10**:3281–3288.

MUSSELMAN, L. J. 1980. The biology of *Striga, Orobanche*, and other root-parasitic weeds. Annu. Rev. Phytopathol. **18**: 463–489.

NEWMAN, J., and S. GUTTERIDGE. 1990. The purification and preliminary X-ray diffraction studies of recombinant *Synechococcus* ribulose-1,5-bisphosphate carboxylase/oxygenase from *Escherichia*. J. Biol. Chem. **265**:15154–15159.

———. 1993. The X-ray structure of *Synechococcus* ribulose-bisphosphate carboxylase/oxygenase—activated quaternary complex at 2.2 Å resolution. J. Biol. Chem. **268**:25876–25886.

NICKRENT, D. L., R. J. DUFF, A. E. COLWELL, A. D. WOLFE, N. O. YOUNG, K. E. STEINER, and C. W. DEPAMPHILIS. 1998. Molecular phylogenetic and evolutionary studies of parasitic plants. Pp. 211–241 *in* D. E. SOLTIS, P. S. SOLTIS and J. J. DOYLE, eds. Plant molecular systematics II. Chapman and Hall, New York.

NICKRENT, D. L., Y. OUYANG, R. J. DUFF, and C. W. DEPAMPHILIS. 1997. Do nonasterid holoparasitic flowering plants have plastid genomes? Plant Mol. Biol. **34**:717–729.

NICKRENT, D. L., and E. M. STARR. 1994. High rates of nucleotide substitution in nuclear small-subunit (18S) rDNA from holoparasitic flowering plants. J. Mol. Evol. **39**:62–70.

OGIHARA, Y., T. TERACHI, and T. SASAKUMA. 1991. Molecular analysis of the hot spot region related to length mutations in wheat chloroplast DNAs. I. Nucleotide divergence of genes and intergenic spacer regions located in the hot spot region. Genetics **129**:873–884.

OLMSTEAD, R. G., B. BREMER, K. M. SCOTT, and J. D. PALMER. 1993. A parsimony analysis of the Asteridae sensu lato based on rbcL sequences. Ann. Mo. Bot. Gard. **80**:700–722.

OLMSTEAD, R. G., H. J. MICHAELES, K. M. SCOTT, and J. D. PALMER. 1992. Monophyly of the Asteridae and identification of their major lineages inferred from DNA sequences of rbcL. Ann. Mo. Bot. Gard. **79**:249–265.

OLMSTEAD, R. G., and J. D. PALMER. 1994. Chloroplast DNA systematics: a review of methods and data analysis. Am. J. Bot. **81**:1205–1224.

OLMSTEAD, R. G., and P. A. REEVES. 1995. Evidence for the polyphyly of the Scrophulariaceae based on chloroplast *rbcL* and *ndhF* sequences. Ann. Mo. Bot. Gard. **82**:176–193.

OLMSTEAD, R. G., and J. A. SWEERE. 1994. Combining data in phylogenetic systematics: an empirical approach using three molecular data sets in the Solanaceae. Syst. Biol. **43**: 467–481.

PRESS, M. C., N. SHAH, and G. R. STEWART. 1986. The parasitic habit: trends in metabolic reductionism. Pp. 96–106 *in* S. J. TER BORG, ed. Biology and control of Orobanche. LH/VPO, Wageningen, the Netherlands.

PRESS, M. C., S. SMITH, and G. R. STEWART. 1991. Carbon acquisition and assimilation in parasitic plants. Funct. Ecol. **5**:278–283.

RANTY, B., G. LORIMER, and S. GUTTERIDGE. 1991. An intradimeric crosslink of large subunits of spinach ribulose-1,5-bisphosphate carboxylase/oxygenase is formed by oxidation of cysteine 247. Eur. J. Biochem. **200**:353–358.

SAS INSTITUTE INC. 1995. JMP. Version 3.1. Cary, N.C.

SCHREUDER, H. A., S. KNIGHT, P. M. G. CURMI, I. ANDERSSON, D. CASCIO, R. M. SWEET, C.-I. BRÄNDÉN, and D. EISENBERG. 1993*a*. Crystal structure of activated tobacco RuBisCO complexed with the reaction-intermediate analogue 2-carboxy-arabinitol 1,5-bisphosphate. Protein Sci. **2**:1136–1146.

SCHREUDER, H. A., S. KNIGHT, P. M. G. CURMI, I. ANDERSSON, D. CASCIO, C.-I. BRÄNDÉN, and D. EISENBERG. 1993*b*. Formation of the active site of ribulose-1,5- bisphosphate carboxylase/oxygenase by a disorder-order transition from the unactivated to the activated form. Proc. Natl. Acad. Sci. USA **90**:9968–9972.

SHIKANAI, T., C. H. FOYER, and H. DULIEAU. 1996. A point mutation in the gene encoding the Rubisco large subunit interferes with holoenzyme assembly. Plant Mol. Biol. **31**: 399–403.

SHIMADA, H., and M. SUGIURA. 1991. Fine structural features of the chloroplast genome: comparison of the sequenced chloroplast genomes. Nucleic Acids Res. **19**:983–995.

SHINOZAKI, K., M. OHME, M. TANAKA et al. (23 co-authors). 1986. The complete nucleotide sequence of the tobacco chloroplast genome: its gene organization and expression. EMBO J. **5**:2043–2049.

SIEMEISTER, G., and W. HACHTEL. 1990. Structure and expression of a gene encoding the large subunit of ribulose-1,5-bisphosphate carboxylase (*rbcL*) in the colourless euglenoid flagellate *Astasia longa.* Plant Mol. Biol. **14**:825–833.

SOKAL, R. R., and F. J. ROHLF. 1995. Biometry. 3rd ed. W. H. Freeman, New York.

STERN, D. B., and D. M. LONSDALE. 1982. Mitochondrial and chloroplast genomes of maize have a 12-kilobase DNA sequence in common. Nature **299**:698–702.

SWOFFORD, D. L. 1993. PAUP: phylogenetic analysis using parsimony. Version 3.1.1. Illinois Natural History Survey, Champaign.

TAYLOR, G. W., K. H. WOLFE, C. W. MORDEN, C. W. DEPAMPHILIS, and J. D. PALMER. 1991. Lack of a functional plastid tRNA-Cys gene is associated with loss of photosynthesis in a lineage of parasitic plants. Curr. Genet. **20**:515–518.

THALOUARN, P., M.-C. ARNAUD, and S. RENAUDIN. 1989. Evidence of ribulose-bisphosphate carboxylase in the Scrophulariaceae holoparasite *Lathraea clandestina* L. C. R. Acad. Sci. Paris **309**:275–280.

THALOUARN, P., and S. RENAUDIN. 1991. Polymerase chain reaction evidence of the *rbcL* gene in the Scrophulariaceae holoparasite *Lathraea clandestina* L. Comparison with the autotroph *Digitalis purpurea* L. and hemiparasite *Melampyrum pratense* L. C. R. Acad. Sci. Paris **309**:381–387.

THALOUARN, P., C. THEODET, N. RUSSO, and P. DELAVAULT. 1994. The reduced plastid genome of a non-photosynthetic angiosperm *Orobanche hederae* has retained the *rbcL* gene. Plant Physiol. Biochem. **32**:233–242.

WOLFE, A. D., and C. W. DEPAMPHILIS. 1997. Alternate paths of evolution for the photosynthetic gene *rbcL* in four non-photosynthetic species of *Orobanche*. Plant Mol. Biol. **33**: 965–977.

WOLFE, K. H., W.-H. LI, and P. M. SHARP. 1987. Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. Proc. Natl. Acad. Sci. USA **84**:9054–9058.

WOLFE, K. H., C. W. MORDEN, and J. D. PALMER. 1992. Function and evolution of a minimal plastid genome from a non-photosynthetic parasitic plant Proc. Natl. Acad. Sci. USA **89**:10648–10652.

PAMELA SOLTIS, reviewing editor

Accepted June 26, 1998