Plant Molecular Biology **33:** 965–977, 1997.

Alternate paths of evolution for the photosynthetic gene *rbc*L in four nonphotosynthetic species of *Orobanche*

Andrea D. Wolfe[1,*] and Claude W. dePamphilis
Dept. of Biology, Vanderbilt University, Nashville, TN 37235, USA; [1]*Current address: Dept. of Plant Biology, Ohio State University, Columbus, OH 43210, USA (***author for correspondence)*

Received 28 February 1996; accepted in revised form 6 January 1997

Key words: Cuscuta reflexa, molecular evolution, Orobanchaceae, plastid pseudogene, probability model, ribulose-1,5-bisphosphate carboxylase/oxygenase

Abstract

We have determined the nucleotide sequence for the Rubisco large subunit from four holoparasitic species of *Orobanche*. Intact open reading frames are present in two species (*O. corymbosa* and *O. fasciculata*), whereas the remaining species (*O. cernua* and *O. ramosa*) have *rbc*L pseudogenes. Sequences for *rbc*L 5′-UTRs from species of *Orobanche* have few changes in the promoter and ribosome binding sites compared to photosynthetic higher plants. Comparison of *rbc*L 3′-UTR sequences for *Nicotiana*, *Ipomoea*, *Cuscuta*, and *Orobanche* reveal that nucleotide sequences from parasitic plants have regions capable of forming stem-loop structures, but 56–69 nt are deleted upstream of the stem-loop in the parasitic plants compared to their photosynthetic relatives. Although *rbc*L pseudogenes of *O. cernua* and *O ramosa* have many large and small deletions, few indels are shared in common, implying that their common ancestor probably had an intact *rbc*L reading frame. Intact *rbc*L reading frames in *O. corymbosa* and *O. fasciculata* retain a bias of synonymous over nonsynonymous substitutions and deduced protein sequences are consistent with potentially functional Rubisco large subunit proteins. A conservative model of random substitution processes in pseudogene sequences estimates that the probability is low ($P < 0.028$) that these sequences would retain an open reading frame by chance. Species of *Orobanche* have either had recent photosynthetic ancestors, implying multiple independent losses of photosynthesis in this genus, or the *rbc*L gene may serve an unknown function in some nonphotosynthetic plants.

Introduction

Parasitic plants lacking chlorophyll (holoparasites) do not have a functional photosynthetic apparatus and are entirely dependent on their host for reduced carbon [23, 32, 33]. Because holoparasites are not self-reliant for photosynthates, much of the plastid genome is under relaxed functional constraint [13]. The plastid genomes of several holoparasitic higher plant genera have recently been examined for genome size and/or gene content (e.g., *Cuscuta* (Cuscutaceae) [18, 19, 26]; *Epifagus* (Orobanchaceae) [14, 48]; *Lathraea* (Scrophulariaceae) [12, 42, 43]; *Conopholis* (Orobanchaceae) [5, 46, 47]; *Orobanche* (Orobanchaceae) [45]). In general, the size of the plastid genome is greatly reduced in holoparasitic plants compared to their closest autotrophic relatives and many of the bioenergetic genes are deleted or sufficiently altered to be classified as pseudogenes [5, 13, 14, 18, 19, 26, 43–48]. Similarly, the heterotrophic euglenoid, *Astasia longa*, has a plastid genome half the size of its photosynthetic relative, *Euglena gracilis* [38], and most of the genes for the photosynthetic apparatus are absent. However, the gene for the large subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco; *rbc*L) is present and expressed in the plastid genome of *A. longa* [39].

Investigations of gene expression for the large subunit of Rubisco, or carbon dioxide assimilation, have been conducted to measure relative photosynthetic

The nucleotide sequence data reported will appear in the EMBL, GenBank and DDBJ Nucleotide Sequence Databases under the accession numbers U73968, U73969, U73970 and U73971.

activity of parasitic plants [3, 12, 14, 18, 26, 32, 33, 42, 45]. The *rbc*L gene has been detected by Southern blot hybridization, polymerase chain reaction (PCR) amplification, or nucleotide sequencing in the holoparasites *Lathraea* [12, 42–44], *Cuscuta* [18, 26], and *Orobanche* [45]. *Epifagus* has retained a fragment of the *rbc*L gene [48], whereas there is no remnant of *rbc*L in *Conopholis* [5]. Surprisingly, some holoparasitic plants such as *Lathraea* [3, 12, 42] and *Cuscuta* [18, 26] have detectable Rubisco activity or *rbc*L gene transcription.

Lathraea is a genus of perennial root holoparasites, which spend up to 10 years underground before the achlorophyllous flowering shoots emerge [23]. Rubisco activity is about 20 times lower in *Lathraea* compared to other C_3 plants [3, 12]. The cause of the reduced Rubisco activity is unknown. The promoter and ribosome binding sites are intact and the *rbc*L gene encodes an open reading frame (ORF) in *Lathraea* [12].

Cuscuta comprises a large genus of annual stem holoparasites. Some species have chlorophyll in low concentration, whereas others are achlorophyllous [26]. A reduction in Rubisco activity for *C. reflexa* nearly equivalent to the enzyme activity for *Lathraea* was reported by Machado and Zetsche [26], whereas Haberhausen *et al.* [18] reported a total lack of enzyme activity. The reduced enzyme activity for *C. reflexa* was attributed to the lack of a palindromic sequence capable of forming a stem-loop structure in the 3′ untranslated region (UTR) of the *rbc*L transcript [18]. Stem-loop structures in the 3′-UTR have been shown to affect transcript stability or transcription termination of many photosynthetic genes [1, 2, 17, 35, 41].

Thalouarn *et al.* [45] detected the presence of the *rbc*L gene in the plastids of *O. hederae* and *O. minor* using PCR-amplification and Southern blot assays. However, an immunoassay directed against the Rubisco protein was negative for the two species of *Orobanche* examined. In addition, no Rubisco activity was detected for *O. crenata* and *O. ramosa* [11, 32]. The mechanisms responsible for the lack of the Rubisco protein in species of *Orobanche* are presently unknown.

In this investigation we examined the nucleotide sequence of the *rbc*L coding region and 5′- and 3′-UTRs in four species of *Orobanche*. Our goals were to determine (1) whether a *rbc*L ORF was present in *Orobanche*, and (2) whether there were mutations in the 5′- or 3′-UTRs that could affect gene expression if an ORF was detected. Here we report that ORFs for *rbc*L are present in two species of *Orobanche*, whereas insertion/deletion events (indels) led to the formation of *rbc*L pseudogenes in the other two species examined. The changes observed in the 5′- and 3′-UTRs are discussed in the context of potential affects on *rbc*L gene expression.

Materials and methods

Total DNAs were isolated from individual plants of *O. cernua*, *O. corymbosa*, *O. fasciculata* and *O. ramosa* following a large-scale modification of the CTAB extraction protocol [16]. *rbc*L was PCR-amplified using the RH1 and 1352R primers (Table 1). The 5′-UTR was PCR-amplified using *atpB* 766R × *rbc*L 1352R, and the 3′-UTR was PCR-amplified using *rbc*L 1020 × ORF106 (Table 1). A 100 μl PCR reaction for *rbc*L utilized 0.64 μm of each primer, 1× *Taq* polymerase buffer (50 mM KCl, 10 mM Tris-HCL pH 9.0, 0.1% Triton X-100; Promega, Madison, WI), 2.0 mM dNTPs, 2.0 mM $mgCl_2$, and 0.0125 units of *Taq* DNA polymerase (Promega). PCR amplifications of the 5′- and 3′-UTRs differed by utilizing 1× *Taq* Extender Buffer (100 mM KCl, 200 mM Tris-HCL pH 8.8, 100 mM $(NH_4)_2SO_4$, 20 mM $MgSO_4$, 1% Triton X-100, 1 mg/ml nuclease-free BSA; Stratagene, La Jolla, CA) and adding 0.0125 units of *Taq* Extender (Stratagene). About 1.0 μg of total DNA served as a template for each amplification. The reaction mix was overlaid with mineral oil, placed in a MJ Research thermocycler (Watertown, MA) and subjected to 35 cycles of 40 s at 94 °C, 1 min at 48 °C, 2 min + 5 s/cycle at 72 °C, after a 1.5 min denaturation at 94 °C. A 5 min extension at 72 °C followed the amplification cycle.

PCR products were isolated for sequencing by electrophoresis from 1% agarose gels onto DEAE cellulose membrane (Schleicher and Schuell SS45, Keene, NH [30]). Purified double-stranded PCR products were used directly for manual ^{35}S-dideoxy sequencing using a modification of the standard Sequenase (US Biochemicals, Cleveland, OH) protocols: 3 μl (250–1000 ng) of template DNA was denatured in a boiling water bath for 3–5 min in the presence of 1 μl of 50 μM primer, 3 μl water, and 2μl Sequenase reaction buffer; denatured DNA was then snap-chilled in a dry ice/ethanol bath. Primers used for sequencing the forward and reverse strands of *rbc*L and the 5′- and 3′-UTRs are listed in Table 1. Sequences were obtained from both strands of the coding and 5′- and 3′-UTR regions of *rbc*L. Sequences of *rbc*L and the 5′- and

Table 1. PCR amplification and DNA sequencing primers.

Primer sequence ($5' \rightarrow 3'$)	Forward primer name	Primer sequence ($5' \rightarrow 3'$)	Reverse primer name
TAACATCTCGGAAATATTCCGCCAT	766R[a]	CCGGAGCTCTTAGTAAAAGATTGGGCCGAG	3′
AYAACATAYACCACTGTCAAGARSGA	N5UTR	CTTCACAAGCAGCAGCTAGGTCAGGACTCC	Z1352R[c]
TGTTGTCAGAATTTATTGTTTTAGGG	ORA5UTR[b]	GTCCTAAAGTTCCTCCACCGAA	N1204R
ATGTCACCACAAACAGAAACTAAAGC	RH1[c]	CGCAGTAAATCAACAAAGCCCA	N1020R
CGTTACAAAGGACGATGCTACCACATCGA	Z234[c]	GATTCTTCTGTCTATCAATAACCGC	N900R
TATGTTAAAACTTTCCAAGGTCCGC	N430	TCACCTGTTTCAGCCTGTGCTTTAT	N678R
AATTGGGGTTATCTGCTAAAAACT	N530	ACGTTACCTACAATGGAAGTAAACA	N350R
TATAAAGCACAGGCTGAAACAGGTG	N674	GATTCGGCAGCTACTGCGGCCC	N158R
GCAGTTATTGATAGACAGAAGATTCATGG	N895	ACTTGCTTTAGTCTCTGTTTGTGGTGACAT	Z1R[c]
TGGGCTTTGTTGATTTACTGCG	N1020	AAATACATRCAATAGAATCTTTG	N3UTR
TTCGGTGGAGGAACTTTAGGAC	N1204	ACTACAGATCCCATACTACCCC	ACCD

[a] *atp*B primer designed by S. Hoot.
[b] *rbc*L primer used only for *O. ramosa*.
[c] *rbc*L primer designed by G. Zurawski

3′-UTRs flanking the gene were obtained from GenBank for *Nicotiana* (Solanaceae), *Ipomoea* (Convolvulaceae), and *Cuscuta* (Cuscutaceae; accession numbers Z00044, X60663 and X61698).

The *rbc*L-coding region and the 5′- and 3′-UTR sequences were aligned separately using the ClustalV program [20] running on a MacIntosh Centris 650 computer (Apple Computer; Cupertino, CA). The results were adjusted manually to obtain the best alignment for each region. 3′-UTR sequences were also converted to mRNA sequences and assayed for secondary structure using the MFOLD and PLOTFOLD programs in the Genetics Computer Group (GCG) software package version 7.0 [15].

Sequences for the coding region of *rbc*L from *Nicotiana* and the four species of *Orobanche* were analyzed using MEGA [24]. The synonymous (K_S) and nonsynonymous (K_N) substitution rates and distances were calculated after a Jukes-Cantor correction. Potential stop codons were removed from the pseudogene sequences to facilitate computation. Gaps and missing data were removed only in the pairwise comparisons.

The sequences for the *rbc*L coding region were translated using the computer program MacVector 4.1.5 (Eastman Kodak, Rochester, NY). Gap positions were maintained, but nucleotide insertions from *rbc*L pseudogene sequences were not included in the translation in order to reconstruct the most conservative protein alignments. The inferred protein sequences were aligned using ClustalV [20]. Nonconserved amino acids in the polypeptide chain were analyzed using the pam250 matrix [10] with the Rubisco large subunit of *Nicotiana* as a reference protein. The total pam250 scores for amino acid composition differences were subjected to a Wilcoxon two-sample test based on ranks in pairwise comparisons with *Nicotiana* [40].

Results and discussion

rbc*L coding region*

Here we report the first *rbc*L sequences for species of *Orobanche*. Intact *rbc*L ORFs were detected for *O. corymbosa* and *O. fasciculata*, whereas the *rbc*L sequences of *O. cernua* and *O. ramosa* have several indels resulting in frameshift mutations and premature stop codons (Fig. 1). The latter two sequences represent *rbc*L pseudogenes. The *rbc*L pseudogene sequence from *O. cernua* has a deletion of 328 bp starting at nucleotide position 42 of Fig. 1. This deletion is flanked by a 12 nt sequence (GTTGGATTCAAA; Fig. 1) on the 5′ end. The sequence adjacent to the 3′ end of the deletion is (TTTGGGTTCAAA; position 384 of Fig. 1). The similarity of those flanking regions suggests that the deletion is likely the result of a strand-slippage replication error involving TTCAAA or the entire 12 nt sequence [7, 25]. Several insertions in the *rbc*L pseudogene sequence from *O. ramosa* also appear as classic strand-slippage artifacts (e.g., nucleotide positions 132, 195, 439, 448, 512 of *Nicotiana*). However, the deletions are not flanked by repeating motifs (Fig. 1).

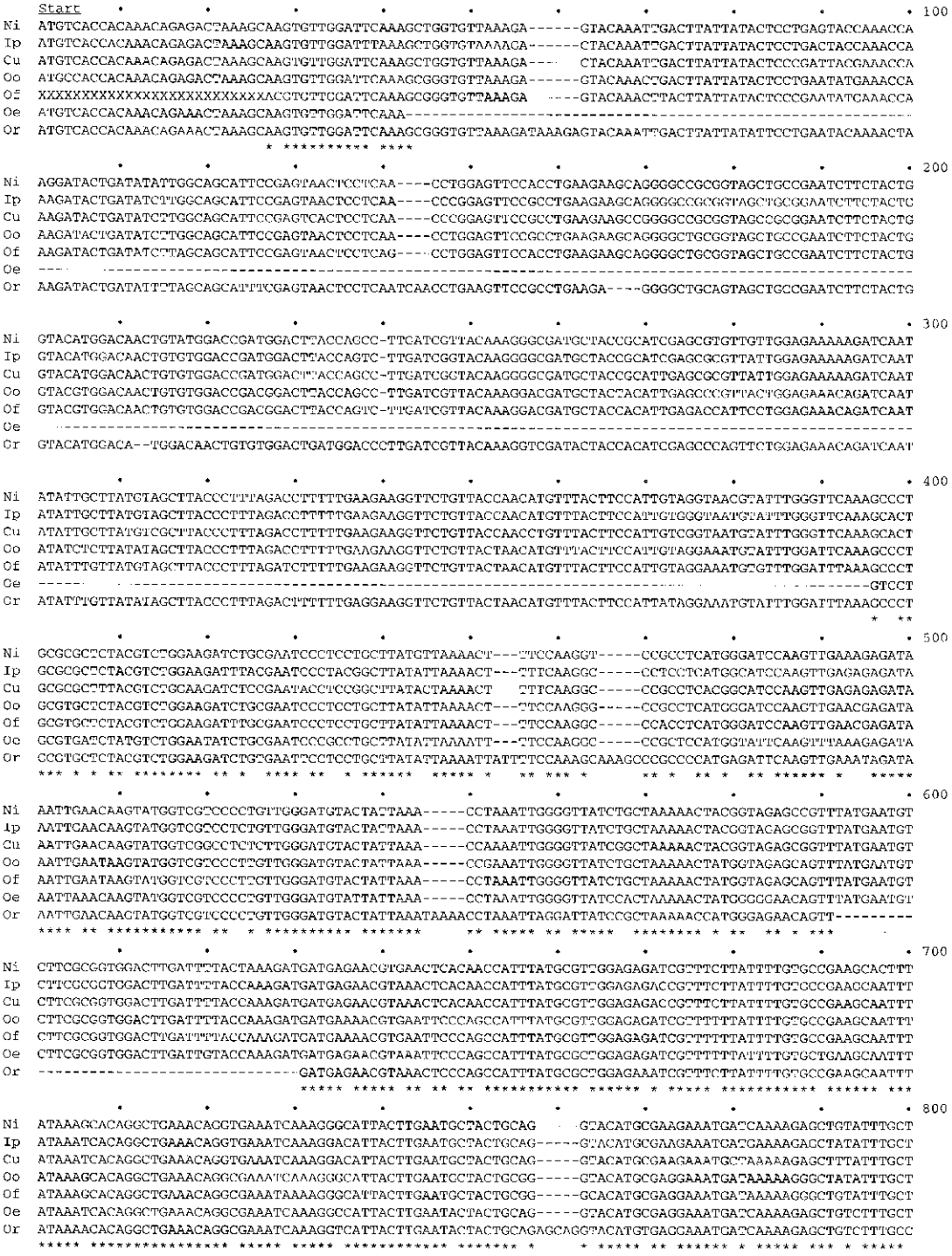

Figure 1. Sequence alignment for *rbc*L coding region from *Nicotiana* (Ni), *Ipomoea* (Ip), *C. reflexa* (Cu), *O. corymbosa* (Oo), *O. fasciculata* (Of), *O. cernua* (Oe), and *O. ramosa* (Or). An asterisk indicates sequence identity for all taxa without a gap at a particular nucleotide position.

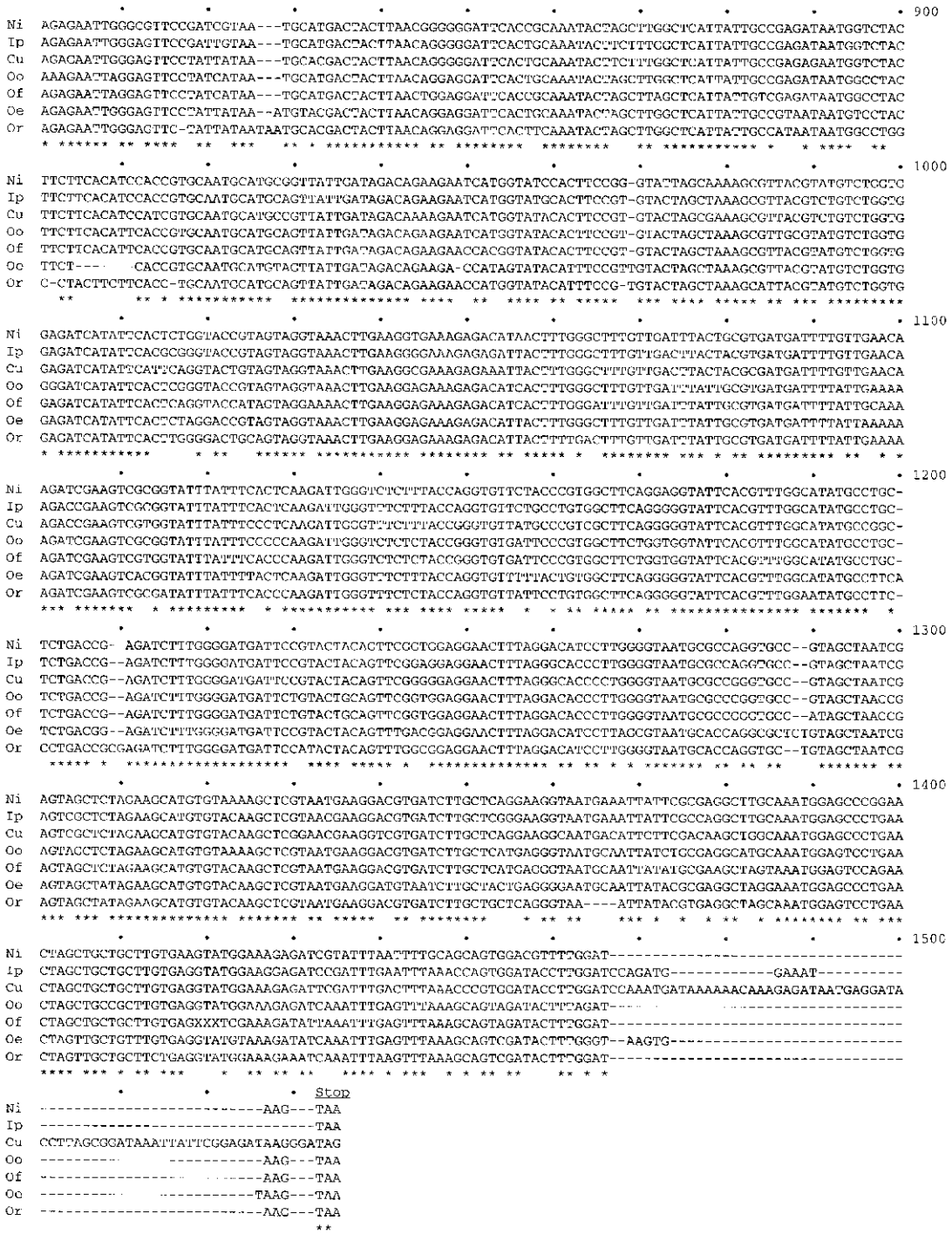

Figure 1. Continued.

There are major differences in mutations observed in the *rbc*L-coding sequences among the four species of *Orobanche* included in this study. For example, insertions found in the *rbc*L pseudogene from *O. ramosa* are primarily multinucleotide, whereas the majority of insertions observed for *rbc*L pseudogene sequence from *O. cernua* are single nucleotide (Fig. 1). No indels or premature stop codons were observed in *rbc*L sequences from *O. corymbosa* and *O. fasciculata*. *Orobanche ramosa* has twice as many indels as *O. cernua*, and there is little overlap in the location of these mutations between the two species. Taken together, these results suggest that the majority of indels observed are independently derived, and that evolution of *rbc*L within species of *Orobanche* has proceeded along divergent pathways.

Additional evidence for this conclusion comes from the comparison of K_S and K_N for each species. K_S values for species of *Orobanche* assayed ranged 0.276–0.332 and K_N values ranged 0.029–0.072 (Table 2). Whereas K_S values were evenly distributed among species with intact coding regions and pseudogene sequences, the K_N values were much higher for the pseudogene sequences compared to the sequences from species with intact reading frames. Because the mutations that would lead to the loss of Rubisco function are not shared by all species of *Orobanche* assayed, and because two species apparently have intact reading frames of *rbc*L, it is likely that the loss of Rubisco function occurred after the adaptation to heterotrophy.

The inferred amino acid sequences for taxa with intact *rbc*L-coding regions and the reconstructed 'pseudoproteins' for *O. cernua* and *O. ramosa* are represented in Fig. 2. By ignoring the inserted nucleotide sequences in the *rbc*L-coding region of *O. cernua* and *O. ramosa*, we were able to infer a sequence comparable to those of photosynthetic plants. This approach minimized reading-frame shifts that should accumulate in pseudogene sequences. It is important to note, however, that pseudogene sequences are expected to accumulate mutations that would be deleterious to functionally constrained sequences. It is relevant to note that *O. cernua* and *O. ramosa* are more closely related to each other than either is to *O. corymbosa* or *O. fasciculata* (based on a cladistic analysis of *rbc*L sequences [31]), and that several of the amino acid replacements observed are common to all species of *Orobanche* (Fig. 2). We conclude from this result that some of the inferred amino acid replacements did occur prior to pseudogene formation and that the inclusion of 'pseudoprotein' sequences in an investigation of the molecular evolution of *rbc*L is important to elucidate changes that may have led to or resulted from the loss of gene function.

When the reconstructed polypeptides from all taxa in this study were compared to the Rubisco large subunit from *Nicotiana*, 212 amino acid replacements were found and scored in the pam250 analysis. The 'pseudoproteins' from *Orobanche cernua* and *O. ramosa* have significantly different amino acid sequences compared to the other taxa assayed (Table 2, Fig. 2). However, the amino acid sequences for *C. reflexa*, *O. corymbosa*, and *O. fasciculata* are not significantly different than those of the nonparasitic genus *Nicotiana* (Table 2; Fig. 2). An examination of structural motifs [21, 22] involving the active site, and dimer-dimer, intradimer and large/small subunit interactive sites reveals that only the 'pseudoprotein' sequences for *O. cernua* and *O. ramosa* have a significant accumulation of presumably deleterious mutations in these important sites (Table 3). If the *rbc*L gene is transcribed and translated in *O. corymbosa* and *O. fasciculata*, we would expect the resulting polypeptide to have a similar structure to the Rubisco large subunit of photosynthetic higher plants.

5′- and 3′-UTRs

Secondary structures in the 5′- and 3′-UTRs of photosynthetic genes are important for regulation of gene expression [1, 2, 8, 9, 17, 34, 35, 41]. Inverted repeat sequences capable of forming stem-loop structures in the 5′-UTR of the gene coding for the 32 kDa protein of photosystem II (*psb*A) may function as nuclear protein and ribosome binding sites [8, 9]. However, Salvador *et al.* [34] found no stem-loop structures in the 5′-UTR from *rbc*L of *Chlamydomonas*, but found that the 5′-UTR interacts with the 5′ end of the coding region to stabilize mRNA transcripts. Studies of seed plant *rbc*L 5′-UTRs also reveal a lack of palindrome sequences capable of forming stem-loop structures [37, 49]. In contrast, the 3′-UTR of *rbc*L genes examined in photosynthetic higher plants have highly conserved inverted repeat sequences capable of forming stem-loop structures [4]. These structures are purported terminator regions [1, 49] or transcript stabilizers [2, 17, 35, 41].

The 5′-UTR of *Nicotiana* is 182 nucleotides long [36]. The 5′-UTR aligned sequences for the other taxa in the study range 58–202 nt as follows: *Ipomoea*, 189 nt; *C. reflexa*, 202 nt, *O. corymbosa*, 179 nt; *O. cernua*, 116 nt; *O. ramosa*, 58 nt (Fig. 3). The PCR amplification protocol for the 5′-UTR did not

Table 2. Synonymous (K_S) and nonsynonymous (K_N) substitution rates based on a Jukes-Cantor correction, and Jukes-Cantor distance values for each taxon compared to *Nicotiana*; pam250 scores for Rubisco large subunit polypeptide translated from sequence data with significance calculated from pairwise comparison of each taxon to *Nicotiana*.

	K_S	K_N	Distance	pam250 Score
Nicotiana	–	–	–	1070
Ipomoea	0.189 ± 0.026	0.025 ± 0.005	0.062	1020
Cuscuta	0.301 ± 0.034	0.031 ± 0.054	0.088	997
O. cernua	0.299 ± 0.040	0.072 ± 0.010	0.122	235***
O. corymbosa	0.276 ± 0.033	0.029 ± 0.005	0.082	994
O. fasciculata	0.332 ± 0.037	0.035 ± 0.006	0.097	961
O. ramosa	0.323 ± 0.037	0.065 ± 0.008	0.120	703***

*** Significant at $P < 0.0001$ in Wilcoxon 2-sample test based on ranks.

Table 3. Number of amino acid replacements observed within structural motifs and interactive sites. pam250 score is sum of all differences compared to *Nicotiana* including amino acid replacements and indels.

Structural motif	*Ipomoea*	*Cuscuta*	*O. corymbosa*	*O. fasciculata*	*O. cernua*	*O. ramosa*
Dimer-Dimer	1	1	1	0	6	3
Intradimer	1	1	1	3	40	11
L/S Subunit	1	1	1	2	10	11
Active Site	0	0	0	0	6	5
Δ pam250	16	8	6	30	283	133

yield a product for *O. fasciculata*. Transcription promoter sequences (-35 and -10 regions) are absent in *O. ramosa*. The -35 region of *O. cernua* has two nucleotide substitutions, whereas the -10 segment has a single nucleotide substitution. *Cuscuta* and *O. corymbosa* have the same sequence for the -35 site, and both taxa have -10 segments identical to *Nicotiana* and *Ipomoea*. All species of *Orobanche* examined have a ribosome binding site (GGAGG) [36] adjacent to the start codon of the *rbc*L reading frame except *O. cernua* has a G $\rightarrow$ A substitution for the last position.

The lack of a promoter site for *O. ramosa* and the diverged -35 site for *O. cernua* together with the truncated leader sequences for both species suggests that the *rbc*L pseudogene sequences are not expressed. *Orobanche corymbosa* has a promoter sequence comparable to *C. reflexa*, a holoparasitic plant that does express the *rbc*L gene. We predict that a *rbc*L transcript will be found for *O. corymbosa* if a plastid transcriptional apparatus is intact. However, the presence of a ribosome binding site immediately adjacent to the start codon may inhibit translation of the *rbc*L ORF in *O. corymbosa*.

Cuscuta reflexa and all species of *Orobanche* assayed had major deletions in the 3′-UTR compared to *Nicotiana* and *Ipomoea* (Fig. 4). Inverted repeat (IR) sequences capable of forming stem-loop structures were found for each taxon examined (Fig. 5) including *C. reflexa*, which purportedly is missing a 3′ IR [18]. Compared to *Nicotiana*, the stem-loop structures found for *Ipomoea* and *C. reflexa* have insertions of 25 and 33 nt, respectively. The inserted nucleotides form additional secondary structures to extend the stem-loop found in *Nicotiana* (Fig. 5). Free energy values for 3′-UTR stem-loop structures for *Ipomoea* and *C. reflexa* exceed that calculated for *Nicotiana* (-15.8, -15.7, -14.0, respectively). We conclude from this result that *Nicotiana*, *Ipomoea*, and *C. reflexa* all retain 3′-UTR sequences capable of forming appropriate stem-loop structures and that those found for *Ipomoea* and *C. reflexa* reflect the close phylogenetic relationship between the Convolvulaceae and Cuscutaceae [6]. Stem-loops in the 3′-UTR of *O. cernua* and *O. corymbosa* are very similar to the stem-loop of *Nicotiana*, whereas there are small deletions and insertions of the region for the 3′-UTRs of *O. fasciculata* and *O. ramosa* (Fig. 5). The free energy values calculated for the stem-loops of the latter two species are much lower than any of the other taxa examined in this study.

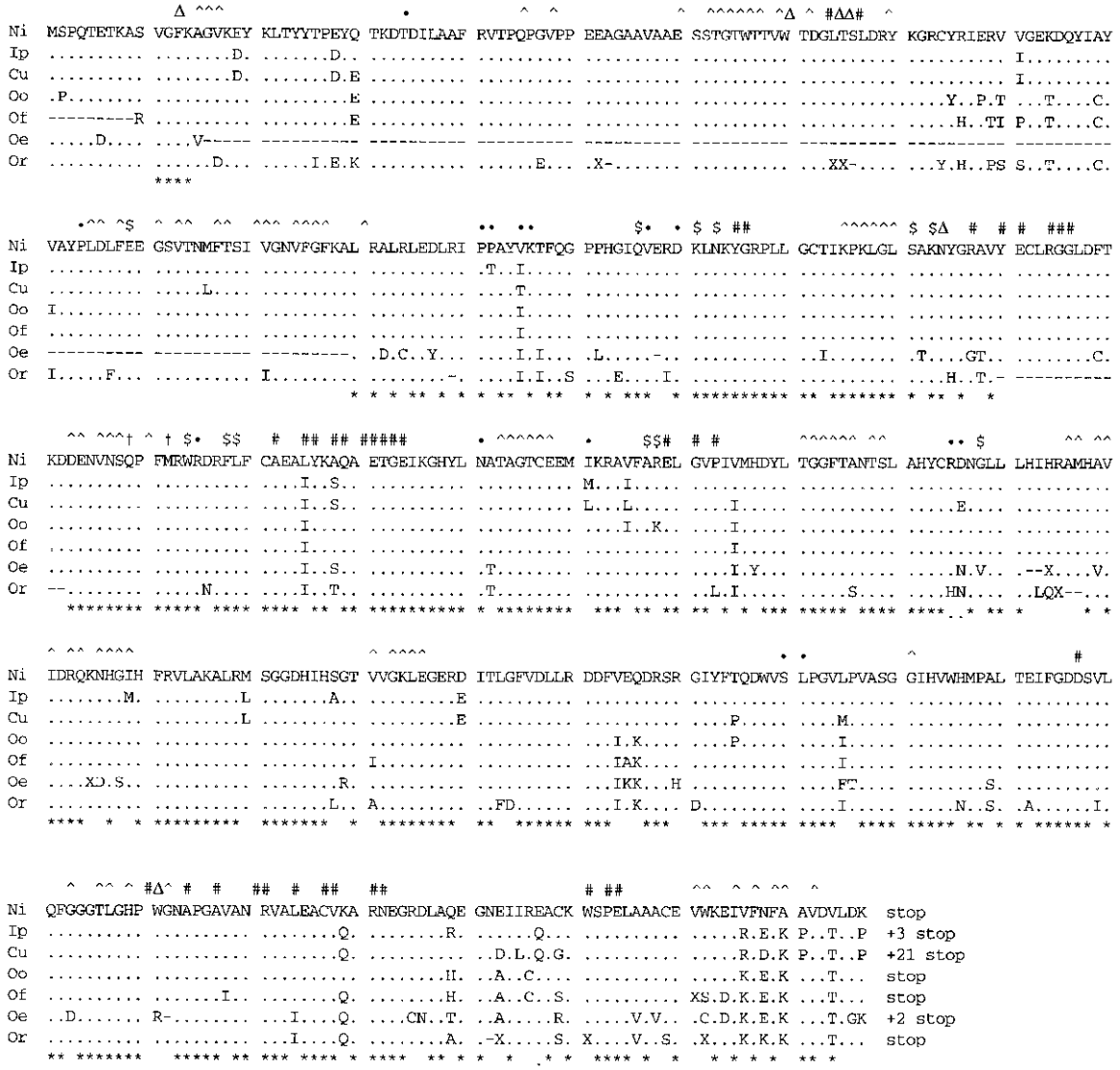

Figure 2. Rubisco large subunit amino acid alignments for *Nicotiana* (Ni), *Ipomoea* (Ip), *C. reflexa* (Cu), *O. corymbosa* (Oo), *O. fasciculata* (Of), *O. cernua* (Oe), and *O. ramosa* (Or). Sequences for *O. cernua* and *O. ramosa* are 'pseudoproteins' reconstructed by eliminating indels. Residues involved in the active site are boxed. An asterisk indicates sequence identity for all taxa without a gap at a particular amino acid position. Symbols used: ? amino acids identical to protein from *Nicotiana*; − gap; * = identical amino acids for all taxa surveyed; # large/small subunit interactive site; ^ = intradimer interactive site; • = dimer/dimer interactive site; △ = # + ^; $ = # + •; † = ^ + •.

Evidence has been presented for *rbc*L gene expression (transcription and enzyme activity) from two nonphotosynthetic plants: *C. reflexa* [26] and *Lathraea clandestina* [3, 12]. *rbc*L gene expression in *O. corymbosa* and *O. fasciculata* has not yet been examined. However, no Rubisco activity or translation products were detectable for *O. crenata*, *O. hederae*, *O. minor*, and *O. ramosa* [11, 32, 45]. Transcripts for *rbc*L were detected at highly reduced levels in *C. reflexa* compared to *Ipomoea* [26]. Haberhausen *et al.* [18] proposed that this reduction in *rbc*L gene expression is due to a lack of a 3′ stem-loop structure in *C. reflexa* and point mutations in the promoter region. Our analysis of the 3′-UTR revealed that a stem-loop structure is possible for *C. reflexa* (Fig. 5). Although the stem-loop structure is present in *C. reflexa*, there are 62 nt deleted upstream from the 3′-IR compared to the sequence from *Ipomoea*. Similarly, deletions ran-

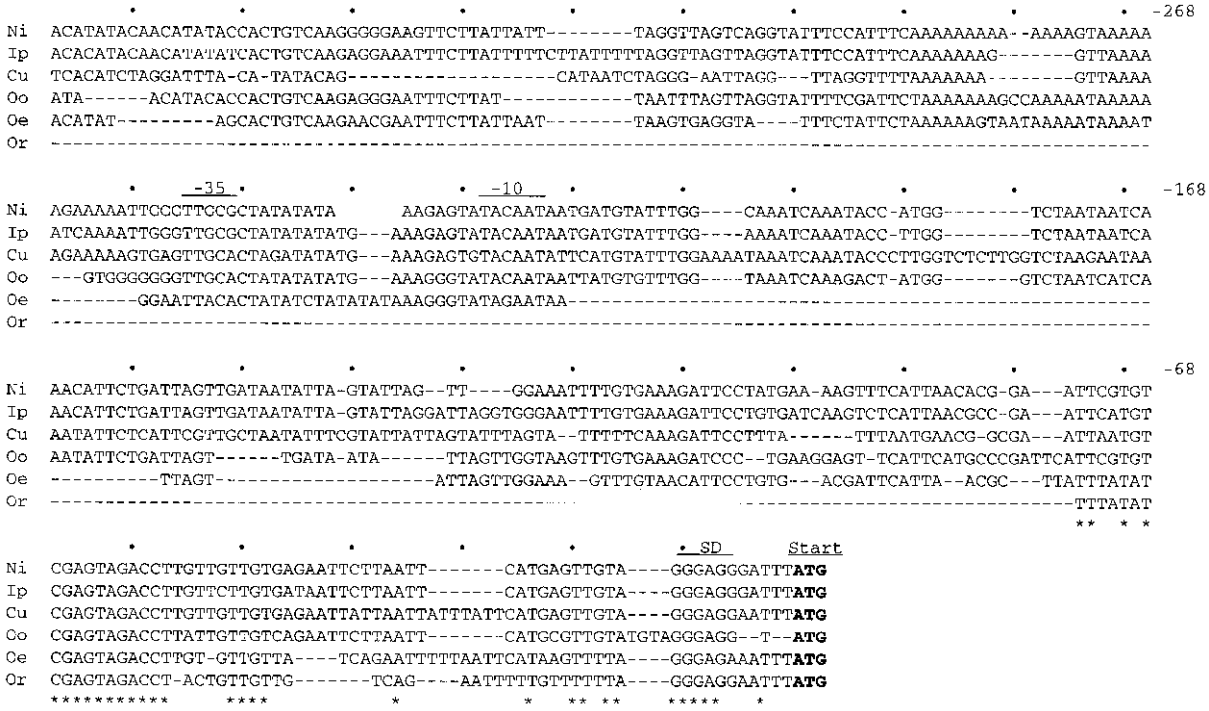

Figure 3. Sequence alignment for 5′-UTR. Taxon abbreviations same as in Fig. 1; transcriptional start site (|), −35, −10, and ribosome binding (SD) motifs indicated. Asterisk indicates sequence identity for all taxa without a gap at a particular nucleotide position. Start codon of ORF indicated.

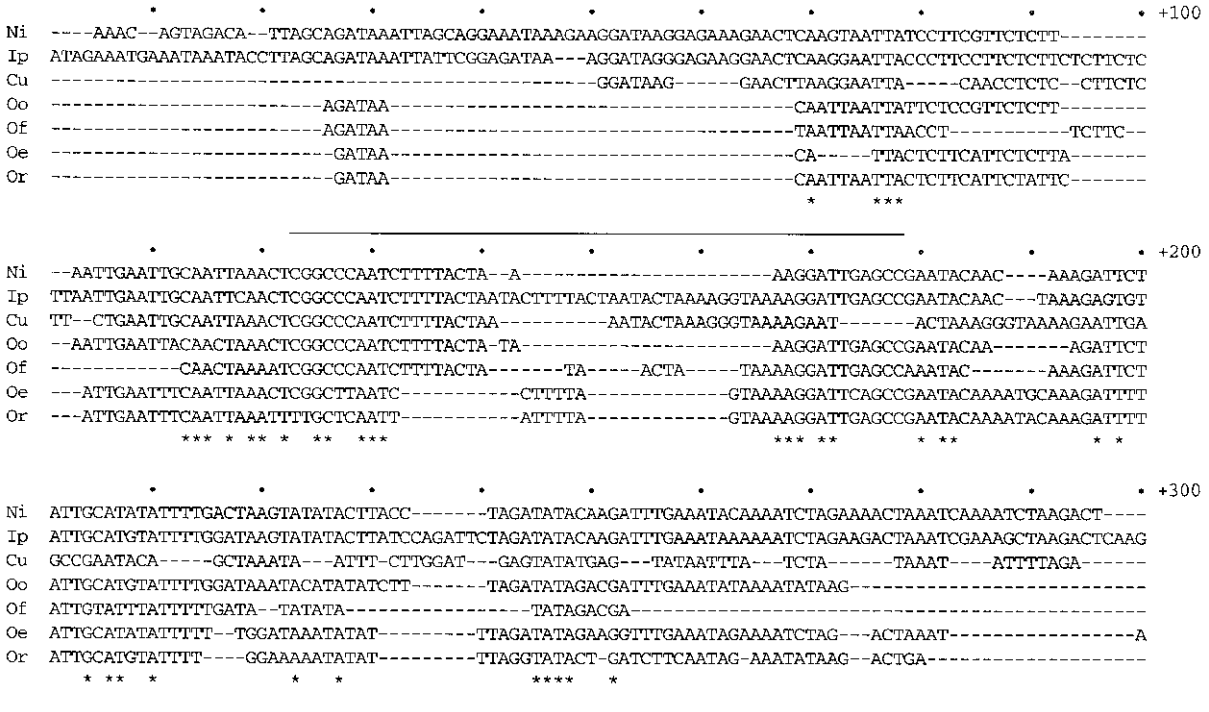

Figure 4. Sequence alignment for 3′-UTR. Taxon abbreviations same as in Fig. 1; sequence for stem-loop structure overlined. Asterisk indicates sequence identity for all taxa without a gap at a particular nucleotide position.

Nicotiana	UCGGCCcAAUCUUUUACUAAAAGGAUUGaGCCGA	ΔG = -14.0
Ipomoea	UCGGCcCAAUCUUUUACUAAUACUUUUACUAAUACUAAAAGGUA*****AAAGGAUUGaGCCGA	ΔG = -15.8
Cuscuta	UCGGCcCAAU*CUUUUACUAAAAUACUAAAGGGUAAAAGAAUACUAAAGGGUAAAAGaAUUGaGCCGA	ΔG = -15.7
O. cernua	UCGGCUuAAUCCUUUUAGUAAAAGGAUUcAGCCGA	ΔG = -16.2
O. corymbosa	UCGGCcCAAUCUUUUACUAUAAAGGAUUGaGCCGA	ΔG = -15.1
O. fasciculata	GGCcCAAUCUUUUACUAUAACUAUAAAAGGAUUGaGCC	ΔG = -9.1
O. ramosa	GCUCAAUUaUUUUAGUAAAAgGAUUGAGC	ΔG = -7.9

Figure 5. mRNA sequences for stem-loop structures and free energy values calculated using MFOLD program of GCG software package. Taxon abbreviations: same as in Fig. 1.

ging 56–69 nt are found upstream of the 3′-UTR stem-loop structures for all species of *Orobanche* examined. It is possible that the distance from the stop codon to the stem-loop is an integral part of the regulatory mechanism(s) involved in *rbc*L gene expression (e.g., transcription termination, processing or stability; recognition-site for RNA-binding protein). If so, the low level of *rbc*L gene expression observed for *C. reflexa* may result from the truncated distance from the stop codon to the stem-loop structure.

Why is *rbc*L retained in holoparasitic plant plastid genomes? As noted above, the *rbc*L gene has been retained in a viable form in some holoparasitic plants (e.g., *Cuscuta*, *Lathraea*) and as an open reading frame in at least two species of *Orobanche*. The gene is also present with intact coding regions in several other non-photosynthetic genera of the Scrophulariaceae (Wolfe and dePamphilis, unpublished). Although major deletions in the plastid genomes of heterotrophic genera of the Orobanchaceae, Cuscutaceae and *A. longa* have been observed, *rbc*L is present in the majority of taxa surveyed, whereas the other photosynthetic and chlororespiratory gene classes are not [5, 13, 14, 19, 28, 48].

There are at least three alternative hypotheses for the retention of *rbc*L in a holoparasitic plant: (1) the plant still requires the Rubisco protein for low levels of autotrophic carbon fixation; (2) the oxygenase activity may function in glycolate metabolism; (3) stochastic events have resulted in its maintenance and/or the loss of photosynthesis was recent and sufficient time has not yet passed for the accumulation of deleterious mutations.

Rubisco has a carboxylase and oxygenase activity. The carboxylase activity is responsible for the fixation of carbon dioxide into a C_3 intermediate followed by the C_6 sugar glucose, whereas the oxygenase activity of photorespiration is a competitive reaction involved in glycolate metabolism (e.g., glycine and serine biosynthesis) [27]. Although most photosynthetic hemiparasites are immediately recognizable by the presence of green leaves and measurable photosynthetic activity [23], some species may maintain very low levels of photosynthetic pigment production and photosynthesis, or perform photosynthesis only during a discrete phase of the life cycle (e.g., seed maturation or seedling development). Such plants might be termed ‘cryptic hemiparasites’, and could appear to be holoparasitic at first analysis. This may explain the retention of a minimally expressed *rbc*L in *Cuscuta reflexa* [18].

Holoparasitic plants receive reduced carbon from their host [33] eliminating the necessity for carbon fixation. Siemeister and Hachtel [39] suggested photorespiration from Rubisco activity may be the reason *rbc*L is retained and expressed in the heterotrophic alga *Astasia longa*. Press *et al.* [32] reported that photorespiration decreases with increasing parasitic ability in higher plants. Clearly, Rubisco activity is maintained in some holoparasitic plant genera, and several others have intact *rbc*L coding regions. Although unlikely, it is possible that the oxygenase activity of Rubisco is maintained in parasitic plants until the adaptation to heterotrophy is complete, and acquisition of host amino acids reduces the need for a glycolate pathway.

To determine whether the retention of *rbc*L in *O. corymbosa* and *O. fasciculata* could be due to stochastic factors, we introduce a probability model to assess the likelihood of a gene with no functional constraint (e.g., loss of photosynthetic ability) retaining an ORF. We assume: (1) that *O. corymbosa* and *O. fasciculata* had a nonphotosynthetic ancestor, and that all *rbc*L sequence divergence has occurred without

Table 4. Jukes-Cantor distances calculated for all taxa included in study from 1431 nt of *rbc*L sequence. Gap sites and missing data removed only in pairwise comparisons. Distances in upper right, standard errors in lower left.

OTU's	*Nicotiana*	*Ipomoea*	*Cuscuta*	*O. cernua*	*O. corymbosa*	*O. fasciculata*	*O. ramosa*
Nicotiana	–	0.0620	0.0883	0.1216	0.0821	0.0968	0.1199
Ipomoea	0.0068	–	0.0485	0.1260	0.0955	0.1124	0.1294
Cuscuta	0.0082	0.0060	–	0.1505	0.1115	0.1249	0.1497
O. cernua	0.0113	0.0115	0.0128	–	0.1075	0.1152	0.0836
O. corymbosa	0.0079	0.0086	0.0094	0.0106	–	0.0381	0.0995
O. fasciculata	0.0087	0.0095	0.0101	0.0111	0.0053	–	0.1104
O. ramosa	0.0100	0.0104	0.0113	0.0094	0.0090	0.0096	–

photosynthetic constraint; and (2) that the probability of any random mutations generating a stop codon is 4/63 and the probability of any random mutation *not* introducing a stop codon is 59/63. If these assumptions are correct and all mutations occurring to these *rbc*L sequences are independent, then, if the sequences are pseudogenes, the probability of retaining an ORF without a stop codon is:

$P = (59/63)^n$, where n is the number of mutational (nucleotide substitution) differences between them.

*rbc*L in Scrophulariaceae/Orobanchaceae taxa has 1434 nucleotides; 1431 nucleotides excluding the terminal stop codon. From Table 4, we find that there are ca. 54.52 (1431×0.0381) nucleotide differences after correcting for multiple substitutions between the *rbc*L sequences of *O. corymbosa* and *O. fasciculata*.

$$P = (59/63)^{54.52} = 0.0279$$

Therefore, the probability of retaining a *rbc*L ORF in these two species of *Orobanche* under a loss of functional constraint is 2.79%. This number is actually an overestimate of the probability of retention because the model does not factor in the possibility of indels. Although the probability of retention under these circumstances is low, it is not outside the realm of possibility that preservation of *rbc*L in these two species is the result of stochastic factors.

The Orobanchaceae is estimated to have diverged from the common ancestor of its sister group (Scrophulariaceae) some $5–50 \times 10^6$ years ago [14], whereas Cuscutaceae and Convolvulaceae (the closest photosynthetic relatives of Cuscutaceae) diverged sometime in the last $15–55 \times 10^6$ years [14, 29]. The major difference between the two groups of holoparasites is that genera of Orobanchaceae are all root parasites spending a large part of their life cycle underground, whereas the species of Cuscutaceae are all stem parasites, germinating in the soil and twining up their host stems before initiating haustoria. All genera of the Orobanchaceae are nonphotosynthetic, but several species of Cuscutaceae have a minimal capacity for photosynthesis [23, 26]. Although the estimated time since divergence for both lineages of parasitic plants overlap, the time since adaptation to holoparasitism has been sufficient for *Epifagus* and *Conopholis* (Orobanchaceae) to have deleted all photosynthetic genes. However, photosynthetic genes have been retained in functional form in some species of Cuscutaceae. The hypothesis of too little time since the loss of photosynthesis does not entirely account for the retention of the *rbc*L gene in *O. corymbosa* and *O. fasciculata* if they shared the same nonphotosynthetic ancestor as did other taxa of Orobanchaceae (e.g., *O. cernua* and *O. ramosa*). Phylogenetic reconstructions based on *rps*2 and *rbc*L gene sequences have not yet been able to resolve the monophyly of genera within Orobanchaceae and among species of *Orobanche* (dePamphilis *et al.*, unpublished; Wolfe and dePamphilis, unpublished). If loss of photosynthesis occurred independently among genera of Orobanchaceae, and if the genus *Orobanche* is not monophyletic, the hypothesis of a recent loss of photosynthesis could explain the retention of *rbc*L in *O. corymbosa* and *O. fasciculata*.

If all of the above hypotheses are rejected, then it is also possible that *rbc*L DNA sequence, transcript, or peptide may be required for some additional function unrelated to either of the known carboxylase or oxygenase activities. We are presently unable to select among these hypotheses, but suggest that at least one must correctly identify why *rbc*L has been retained in some holoparasitic taxa.

Acknowledgements

We thank Drs N. Young, J. Leebens-Mack, B. Swalla, B. Eickmeyer, C. Birky, E. Kellogg, M. Chase, and R. Doolittle for helpful discussions or comments on the manuscript, D. McCauley for help on the statistical analyses, and Mr Tim Atkinson for technical assistance. Financial support for this research was provided by a NSF postdoctoral fellowship (BIR-9303630) to A.D.W., and by NSF grant (DEB-9120258) and a Vanderbilt University Research Council grant to C.W.D.

References

1. Blowers AD, Klein U, Bogorad L: Functional in vivo analyses of the 3′ flanking sequences of the *Chlamydomonas reinhardtii* chloroplast *rbc*L and *psa*B genes. Mol Gen Genet 238: 339–349 (1993).
2. Brawerman G: Determinants of messenger RNA stability. Cell 48: 5–6 (1987).
3. Bricaud CH, Thalouarn P, Renaudin S: Ribulose 1,5-bisphosphate carboxylase activity in the holoparasite *Lathraea clandestina* L. J Plant Physiol 125: 367–370 (1986).
4. Calie PJ, Manhart JR: Extensive sequence divergence in the 3′ inverted repeat of the chloroplast *rbc*L gene in non-flowering land plants and algae. Gene 146: 251–256 (1994).
5. Colwell A: Genome evolution in a non-photosynthetic plant, *Conopholis americana*. Ph.D. dissertation, Washington University (1994).
6. Cronquist A: An Integrated System of Classification of Flowering Plants. Columbia University Press, New York (1993).
7. Cummings MP, King LM, Kellogg EA: Slipped-strand mispairing in a plastid gene: *rpo*C2 in grasses (Poaceae). Mol Biol Evol 11: 1–8 (1994).
8. Danon A, Mayfield SP: ADP-dependent phosphorylation regulates RNA-binding in vitro: implications in light-modulated translation. EMBO J 13: 2227–2235 (1994).
9. Danon A, Mayfield SPY: Light regulated translational activators: identification of chloroplast gene specific mRNA binding proteins. EMBO J 10: 3993–4001 (1991).
10. Dayhoff MO, Schwartz RM, Orcutt BC: A model of evolutionary change in proteins. In: MO Dayoff (ed), Atlas of Protein Sequence and Structure, vol 4, pp. 345–352. Silver Springs, MD National Biomedical Research Foundation (1978).
11. de la Harpe AC, Grobbelaar N, Visser JH: The ultrastructure of the chloroplast and the chlorophyll content of various South African parasitic flowering plants. Z Pflanzenphysiol 100: 85–90 (1980).
12. Delavault P, Sakanyan V, Thalouarn P: Divergent evolution of two plastid genes, *rbc*L and *atp*B, in a non-photosynthetic parasitic plant. Plant Mol Biol 29: 1071–1079 (1995).
13. dePamphilis CW: Genes and genomes. In: Press MC, Graves JD (eds) Parasitic Plants, pp. 177–205. Chapman and Hall, London (1995).
14. dePamphilis CW, Palmer JD: Loss of photosynthetic and chlororespiratory genes from the plastid genome of a parasitic flowering plant. Nature 348: 337–339 (1990).
15. Devereux J, Haberli P, Smithies O: A comprehensive set of sequence analysis programs for the VAX. Nucl Acids Res 12: 387–395 (1984).
16. Doyle JJ, Doyle JL: A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochem Bull 19: 11–15 (1987).
17. Gruissem W: Chloroplast gene expression: how plants turn their plastids on. Cell 56: 161–170 (1989).
18. Haberhausen G, Valentin K, Zetsche K: Organization and sequence of photosynthetic genes from the plastid genome of the holoparasitic flowering plant *Cuscuta reflexa*. Mol Gen Genet 232, 154–161 (1992).
19. Haberhausen G, Zetsche K: 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 (1994).
20. Higgins DG, Bleasby AJ, Fuchs R: Clustal V. Improved software for multiple sequence alignment. Comp Appl Biosci 8: 189–191 (1992).
21. Kellogg EA, Juliano ND. The structure and function of rubisco and their implications for systematic studies. Am J Bot 84, in press.
22. Knight S, Andersson I, Brändén CI: Crystallographic analysis of ribulose 1,5-bisphosphate carboxylase from spinach at 2.4 Å resolution. J Mol Biol 215: 113–160 (1990).
23. Kuijt J: The Biology of Parasitic Flowering Plants. University of California Press, Berkeley/Los Angeles, CA (1969).
24. Kumar S, Tamura K, Nei M: MEGA: Molecular Evolutionary Genetics Analysis, version 1.01. The Pennsylvania State University, University Park, PA 16802 (1993).
25. Levinson G, Gutman GA: Slipped-strand mispairing: a major mechanism for DNA sequence evolution. Mol Biol Evol 4: 203–221 (1987).
26. Machado MA., Zetsche K: A structural, functional and molecular analysis of plastids of the holoparasites *Cuscuta reflexa* and *Cuscuta europaea*. Planta 181: 91–96 (1990).
27. Miziorko HM, Lorimer GH: Ribulose-1,5-bisphosphate carboxylase-oxygenase. Annu Rev Biochem 52: 507–535 (1983).
28. Morden CW, Wolfe KH, dePamphilis CW, Palmer JD: Plastid translation and transcription genes in a non-photosynthetic plant: intact, missing and pseudogenes. EMBO J 10: 3281–3288 (1991).
29. Muller, J: Fossil pollen records of extant angiosperms. Bot Rev 47: 1–142 (1981).
30. Nickrent DL: From field to film: rapid sequencing methods for field-collected plant species. Bio/technology 16: 470–475 (1994).
31. Nickrent DL, Duff RJ, Colwell AE, Wolfe AD, Young ND, Steiner KE, dePamphilis CW. Molecular phylogenetic and evolutionary studies of parasitic plants. In: Soltis DE, Soltis PS, Doyle JJ (eds) Molecular Systematics of Plants II, Ch. 14. Chapman and Hall, New York (1997).
32. Press MC, Shah N, Stewart GR: The parasitic habit: trends in metabolic reductionism. In: ter Borg SJ (ed) Biology and Control of Orobanche, pp. 96–106. LH/VPO, Wageningen, Netherlands (1986).
33. Press MC, Smith S, Stewart GR: Carbon acquisition and assimilation in parasitic plants. Functional Ecol 5: 278–283 (1991).
34. Salvador ML, Klein U, Bogorad L: 5′ sequences are important positive and negative determinants of the longevity of *Chlamydomonas* chloroplast gene transcripts. Proc Natl Acad Sci USA 90: 1556–1560 (1993).

35. Schuster G, Gruissem W: Chloroplast mRNA 3′ end processing requires a nuclear-encoded RNA-binding protein. EMBO J 10: 1493–1502 (1991).
36. Shinozaki K, Sugiura M: The nucleotide sequence of the tobacco chloroplast gene for the large subunit of ribulose-1,5-bisphosphate carobylase/oxygenase. Gene 20: 91–102 (1982).
37. Shinozaki K, Sugiura M: Sequence of the intercistronic region between the ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit and the coupling factor b subunit gene. Nucl Acids Res 10: 4923–4934 (1982).
38. Siemeister G, Hachtel W: A circular 73 kb DNA from the colourless flagellate *Astasia longa* that resembles the chloroplast DNA of *Euglena*: restriction and gene map. Curr Genet 15: 435–441 (1989).
39. Siemeister G, Hachtel W: Structure and expression of a gene encoding the large subunit of ribulose-1,5-bisphosphate carboxylase (*rbc*L) in the colourless euglenoid flagellate *Astasia longa*. Plant Mol Biol 14: 825–833 (1990).
40. Sokal RR, Rohlf FJ: Biometry, 3rd ed. W. H. Freeman and Co., New York (1995).
41. Stern DB, Gruissem W: Control of plastid gene expression: 3′ inverted repeats act as mRNA processing and stabilizing elements, but do not terminate transcription. Cell 51: 1145–1157 (1987).
42. Thalouarn P, Arnaud M-C, Renaudin S: Evidence of ribulose-bisphosphat carboxylase in the Scrophulariaceae holoparasite *Lathraea clandestina* L. CR Acad Sci Paris 309 (Série III) 275–280 (1989).
43. Thalouarn P, Renaudin S: Polymerase chain reaction evidence of the *rbc*L gene in the Scrophulariaceae holoparasite *Lathraea clandestina* L. Comparison with the autotroph *Digitalis purpurea* L. and hemiparasite *Melampyrum pratense* L. CR Acad Sci Paris 309: 381–387 (1991).
44. Thalouarn P, Theodet C, Renaudin S: Evidence of plastid and nuclear genes for the large and small subunits of Rubisco in the Scrophulariaceae holoparasite *Lathraea clandestina* L. Comparison with the autotroph *Digitalis purpurea* L. and hemiparasite *Melampyrum pratense* L. CR Acad Sci Paris 312 (Sèrie III) 1–6 (1991).
45. Thalouarn P, Theodet C, Russo N, Delavault P: The reduced plastid genome of a non-photosynthetic angiosperm *Orobanche hederae* has retained the *rbc*L gene. Plant Physiol Biochem 32: 233–242 (1994).
46. Wimpee CF, Morgan R, Wrobel R: An aberrant plastid ribosomal RNA gene cluster in the root parasite *Conopholis americana*. Plant Mol Biol 18: 275–285 (1992).
47. Wimpee CF, Morgan R, Wrobel RL: Loss of transfer RNA genes from the plastid 16S–23S ribosomal RNA gene spacer in a parasitic plant. Curr Genet 21: 417–422 (1992).
48. Wolfe KH, Morden CW, Palmer JD: Function and evolution of a minimal plastid genome from a nonphotosynthetic parasitic plant. Proc Natl Acad Sci USA 89: 10648–10652 (1992).
49. Zurawski G, Clegg MT: Evolution of higher-plant chloroplast DNA-encoded genes: implications for structure-function and phylogenetic studies. Annu Rev Plant Physiol 38: 391–418 (1987).