GENE 08044

Extensive sequence divergence in the 3' inverted repeat of the chloroplast *rbcL* gene in non-flowering land plants and algae

(Palindromic sequences: stem-loop regions; Rubisco; *Chlamydomonas*)

Patrick J. Calie[a] and James R. Manhart[b]

[a]*Department of Microbiology and Immunology, University of Michigan, Ann Arbor, MI 48109-0620, USA; and* [b]*Department of Biology, Texas A&M University, College Station, TX 77843-3258, USA. Tel. (1-409) 845-3356*

SUMMARY

A stem-loop region is present at the 3' terminus of the chloroplast *rbcL* mRNA in all taxa surveyed to date. In spinach, this structure has been shown by others to be involved in modulating transcript stability and correct 3' terminus processing, and is a conserved feature of other flowering plant *rbcL* mRNAs. In *Chlamydomonas reinhardtii*, an analogous structure has been shown by others to serve as a transcription terminator. Our sequencing data have shown that this region is highly divergent in several non-flowering land plants, as evidenced by representatives from the ferns, conifers, 'fern-allies' and liverworts. To extend our analysis, a computer-assisted survey of the stem-loop region of the 3' flanking region of published chloroplast *rbcL* genes was undertaken. The flowering plant *rbcL* inverted repeats (IR) were remarkably conserved in sequence, allowing for precise multiple alignments of both monocot and dicot sequences within a single matrix. Surprisingly, sequences obtained from non-flowering land plants, algae, photosynthetic protists and photosynthetic prokaryotes were extremely variant, in terms of both sequence composition and thermodynamic parameters.

INTRODUCTION

A well-characterized post-transcriptional mechanism involved in modulating chloroplast gene expression is the regulation of mRNA processing and concomitant transcript stabilization through the formation of thermodynamically stable stem-loop regions at the 3'-flanking regions of many plastid-encoded mRNAs (Stern and Gruissem, 1987). These stem-loop structures are formed by intramolecular base-pairing between short IR sequences. The primary sequences of the IR from different genes within a single plant species are quite variant (Stern and Gruissem, 1987). In contrast, the IR of a single gene (e.g., that of *petD*) is often highly conserved among different flowering plant species (Stern et al., 1989).

The *rbcL* gene, encoding the large subunit of the ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco), contains an IR in its 3' terminus that is involved in the regulation of the mRNAs stability in spinach (Schuster and Gruissem, 1991). The primary sequence of the *rbcL* IR from four taxa was shown to be highly conserved (Zurawski and Clegg, 1987), as is the coding sequence of the gene in flowering plants (reviewed in Clegg, 1993). A recent analysis of the *rbcL* coding sequences of several non-flowering land plants surprisingly revealed the fact that these sequences are too divergent to test hypotheses of phylogenetic relationships among major groups of land plants (Manhart, 1994). This present study was undertaken to determine the

Correspondence to: Dr. P.J. Calie at his present address: Department of Biology, Eastern Kentucky University, Richmond, KY 40475-3100, USA. Tel. (1-606) 622-1505; Fax (1-606) 622-1020; e-mail: biocalie@acs.eku.edu

Abbreviations: bp, base pair(s); *Ch.*, *Chlamydomonas*; ΔG, free energy (kcal); GCG, Genetics Computer Group (Madison, WI, USA), IR, inverted repeat(s); *rbcL*, gene encoding the large subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco).

Monocots

1 UUAAACUCGUCUCAAUCUgaaaaaAGAUUGAGCCGAGUUUAA   ΔG = - 18.4

2 UCGGCUCAAUCUUUUUuuucuaAAAAAGAUUGAGCCCGA   ΔG = - 21.5

3 UCGGCUCAAUCUUUUUUuuucuaAAAAAGAUUGAGCCCGA   ΔG = - 21.5

4 UCGGCUCAAUCUUUUUUuuuauUAAGAAAGAUUGAGCCCGA   ΔG = - 21.5

5 UUAAAUUCGGCCUCAAUCUUUuuagaAAAGAUUGAGCCGAAUUUAA   ΔG = - 25.1

6 UUAAAUUCGGCUCAGUCUUUUuuuuaGAAAGAUUGAGCCGAAUUUAA   ΔG = - 25.1

7 UUAAAUUCGGCUCAaUCUUUUcuaaaaAAAAGACUGAGCCGAAUUUAA   ΔG = - 23.6

Dicots

8 UUUAuuucaaUUAAacuaGGCCCAAUCUUUuaauaaAAGGAUUGaGCC   ΔG= +1.4, ΔG = -10.0

9 UUUAuuuccaUUUAAcUCGGCcCAAUCUUUuacuaaAAGGAUUGaGCCCGA   ΔG = +1.4, ΔG = -14.3

10 UUCGGCcCAAUCUUUUccuaaaaGAACGAUUGaGCCCGAA   ΔG = -15.9

11 UUAAUaaaaucuaAUUAAacUCGGCcCAAUCUUUuacuaaAAGGAUUGaGCCCGA   ΔG = +0.4, ΔG = -14.3

12 UUAAAUUGaauugCAAUUAAacUCGGCcCAAUCUUUuacuaaAAGGAUUGaGCCCGA   ΔG = -2.1, ΔG = -14.3

13 UUAAAUUGaauugCAAUUAAacUCGGCcCAAUCUUUuacuaaAAGGAUUGaGCCCGA   ΔG = -2.1, ΔG = -14.3

OR

AAAGAAGGAUAAGcaagaaCUCaaguaauUAUCCUUCGUUU   ΔG = -10.2

AAUuGAAUUgcaaUUaaacUCGGCcCANUCUUUuacuaaaAAGGAUUGaGCCGAauacAAcaaGAUUCUAUU   ΔG = -15.2

Conifer

14   AUCUGCcgauuuCACAU   ΔG = -2.4

    CUggCGAUuUCUAUAUAUUCAAUAUUCUAUCuacuuAUACAUAUUGAUAUAUAGAUAAUUCGAG   ΔG = -22.6

Fern

*15 AAAGAUAGUAGAAAAGAGAGUuuuuuuACUCUUUCUUUACUAUUCUUU   ΔG = -27.7

"Fern allies"

*16 UUUUAUUUAAAUUuugaaacCAUUUUAAUAAAA   ΔG = -6.2

*17 CUUgaUUUCAUUUUUCUUUCACCUuggaaaGAGAUUAGGCUGAAAGAAAAAUuGCAAaAAC   ΔG = -23.9

*18 UAUUCCUCUACUUCUGUUUUUGCUUCAAUUCCUccgaauaAGGAAUUGAAGCGAAACAaAAGUAGAGGAAUA   ΔG = -42.1

Liverworts

19 AUUUAUCAUUUUUUUCUUUUCACCCCggaacuGGGUCAAGAAAAAUGAUAAAU   ΔG = -27.4

20   UACUUUGuaaaUAAAGUA   ΔG = -3.0

    UUUUGUAAUUUCUUUUUUUACUCUGcagauuuCAUAAAAAAAGAAAAAUUACAAA   ΔG = -26.7

Green algae

21   ACAAGuuuuuuUUUUCU   ΔG = -1.0

    GAUUUUCAAGAAAAUUUACCAUGCAGGAAUCAaauuUGAUUCCUGCAUGUAAUUUUCUUGAAAGUU   ΔG = -41.5

22 UCUUCAAAAAquucuUUUUUGAGCcccauccuuuuaccAAAUUUCUaaaaaacAGAAUUUU   ΔG = -6.3, ΔG = -4.4

23 UUAAACAAAUUAAGGaaaaaCCUUAAUUUUCUUUA   ΔG = -13.6

24 AUGAUGUUUAUCUgaauaGCAUAAACAUCCUuuuuauuuuuuuuGCUCUUUuagguuAAAUACC   ΔG = -12.4, ΔG = -3.8

25 ACAAAUAUAUACUCGCCAcuuuUGGCGAGUAUAUAUUUGU   ΔG = -22.0

26 UUAUUUGAAAAUUCAUAUAUAUGUcuagACAUAUAUGUAAUUUUUAGaGUAA   ΔG = -16.1

Brown algae

27 UAAAUaccUUACAAauuAAuuAAuuuaaaUUUCUGUGUUAGGGGUUUA   ΔG = -9.2

    UUAUCaacauacGAUAA   ΔG = -1.1

    AGUUUagugcuAAACU   ΔG = -1.3

Photosynthetic bacteria

28 CUGGaCGUgCCCacaaguaagaccaGGGUGACCGCCGG   ΔG = -11.8

29 ACCaGCCGUCCcgguuGCACCGCCGGU   ΔG = -15.4

    AGAAGACGCGGAGGcucuccCCCCCGGCGUCUUUU   ΔG = -20.9

    GGGCCCgagaugGGCUC   ΔG = -7.6

Cyanobacteria

30 GCUCGGCGAGUCGAUCGUUUcugcgUAAAGCUUUCUCCCCAGC   ΔG = -17.6

    UUCAgGCAUuuucugaAUCaUGAC   ΔG = -2.7

31 GAUUGCAAGGAUGAAGGUaaauggaaaaACCuuuCAUCCUUaCAUU   ΔG = -14.7

    UGAUCUCaaGCaaauaGCggaaAGAUaCA   ΔG = -2.4

    GCCAaAaCCUCcaaACCuaUCACGGCGCUaauGACGUgUUGGC   ΔG = -17.4

Cryptomonad

32 CAAUUGGUAUuuaaauaAUACCAAUUG   ΔG = -9.0

Euglenoid

33 UAAAUuAUAAAaGUUAauuuuuUAACAUUUAUUAUUUG   ΔG = -3.3

34 AAAUUAGaauuaaaAUAAUUU   ΔG = -1.5

Fig. 1. Nucleotide sequence of *rbcL* IR. The sources of the sequences are either from the literature or were derived experimentally for this paper (indicated by an asterisk): 1, rice (Nishizawa and Hirai, 1987); 2, wheat (Terachi et al., 1987); 3, rye (Terachi et al., 1987); 4, barley (Zurawski et al., 1984); 5, mullet (Doebley et al., 1990); 6, sorghum (Doebley et al., 1990); 7, *Tripsacum dactyloides* (Morton and Clegg, 1993); 8, alfalfa (Aldrich et al., 1986b); 9, pea (Zurawski et al., 1986); 10, broadbean (Zurawski and Clegg, 1987); 11, spinach (Zurawski et al., 1981); 12, tobacco (Shinozaki et al., 1986); 13, petunia (Aldrich et al., 1986a); 14, Douglas fir (Hipkins et al., 1990); 15*, *Angiopteris evecta* (GenBank Accession No. L11052); 16*, *Equisetum arvense* (L11053); 17*, *Isoetes melanopoda* (L11054); 18*, *Lycopodium digitatum* (L11055); 19*, *Bazzania trilobata* (L11056); 20, *Marchantia polymorpha* (Ohyama et al., 1986); 21*, *Spirogyra maxima* (L11057); 22, *Chlorella* N1a (Amberg and Meints, 1991); 23, *C. ellipsoidea* (Yoshinaga et al., 1988); 24, *Chlamydomonas reinhardtii* (Dron et al., 1982); 25, *Ch. moewusii* (Yang et al., 1986); 26, *Codium fragile* (Manhart and VonderHaar, 1991); 27, *Ectocarpus siliculosus* (Valentin and Zetsche, 1990); 28, *Chromatium vinosum* (Viale et al., 1989); 29, *Rhodospirillum rubrum* (Narang et al., 1984); 30, *Anacystis* 6301 (Shinozaki et al., 1983); 31, *Anabaena* 7120 (Curtis and Haselkorn, 1983); 32, *Cryptomonas* Φ (Douglas et al., 1990); 33, *Astasia longa* (Siemeister and Hachtel, 1990); 34, *Euglena gracilis* (Gingrich and Hallick, 1985). The nt sequences reported in this paper were obtained via PCR amplification and nt sequencing as described (Manhart, 1994). Optimal base-pairing and free energy values for the IR sequences were obtained using the FOLD algorithm of the GCG Package, Version 7.0 (Devereaux et al., 1984) on a VAX computer. Those nt (including "bulges") that comprise the stem regions are underscored; loop regions are not. Stem nt involved in intramolecular pairing are represented by upper-case letters; those nt that are not (i.e., "bulges") are in lower-case letters. Free-energy values are listed to the right of each IR. The stop codons in the *Euglena* and *Astasia* coding regions are in bold-face type. The IR are listed by the order in which they follow the 3′ terminus of the *rbcL* gene.

extent of IR sequence divergence in selected taxa of non-flowering land plants, algae, protists and prokaryotes.

EXPERIMENTAL AND DISCUSSION

(a) Conservation of the flowering plant *rbcL* IR

The results of our computer analysis on the *rbcL* stem-loop regions found in all taxa surveyed are summarized in Fig. 1. The flowering plant IR are very highly conserved among themselves, relative to the other taxa surveyed. The monocot structures have relatively high free energy values (ranging from $\Delta G = -18.4$ for rice to -25.1 for mullet and sorghum) in contrast to the dicot structures ($\Delta G = -10.0$ for alfalfa to -14.3 for pea, spinach, tobacco and petunia). The only differences among the flowering plant IR sequences were three nt substitutions within the core sequences (Fig. 2). The overall length of the IR did vary, due to expansion and contraction of the different IR in the flanking regions distal to the loop. However, with the exception of these three substitutions, the monocot and dicot sequences were remarkably similar, allowing for a successful alignment within a single matrix (Fig. 2).

In the flowering plant core IR nucleotide sequence GGCmCAAUCUUU(N_3-N_8)AARGAMUGaGCC (M-(m) = A,C; N = A,C,G,T; R = A,G) only the underlined base is divergent in the dicots surveyed (Fig. 2). For petunia there are two possible structures (Fig. 1). Based on comparisons with tobacco, which resides in the same family as petunia (Solanaceae), we would propose that the first representation best exemplifies the true in vivo structure. All the dicots examined (with the exception of broadbean) contain an additional IR sequence in an A+U rich region (Fig. 1). Given the predicted thermodynamic instability of the stem-loop structures ($\Delta G = +1.4$ in alfalfa and pea, $+0.4$ in spinach and -2.1 in tobacco and petunia) these structures are unlikely to exist in vivo, and are more likely fortuitous occurrences.

The A+T composition of land plant *rbcL* IR analyzed in this paper ranges from 54 (dicots) to 100% (*Equisetum arvense*). Land plant chloroplast DNAs, in general, tend to be rather A+T rich, as exemplified by tobacco, 62% (Shinozaki et al., 1986); rice, 61% (Hiratsuka et al., 1989); *Marchantia polymorpha*, 71% (Ohyama et al., 1986). This suggests that new IR can arise over time due to the accumulation of random nt changes in noncoding spacer regions. This could be a mechanism by which new IR arose in plastid genomes. Furthermore, the intergenic spacer regions downstream from *rbcL* in several grasses contain two regions of secondary structure that are mutational hotspots, one of these being the transcript-stabilizing structure (Morton and Clegg, 1993). Thus, mechanisms

254

```
1    UUAAACUCGGCUCAAUCUuuuuc.........AGAUUGAGCCGAGUUUAA
2         UCGGCUCAAUCUUUUUuuucua.AAAAAGAUUGAGCCGA
3         UCGGCUCAAUCUUUUUuuucua.AAAAAGAUUGAGCCCGA
4         UCGGCUCAAUCUUUUUUuuuauAAGAAAGAUUGAGCCGA
5    UUAAAUUCGGCUCAAUCUUUUuaga......AAAGAUUGAGCCGAAUUUAA
6    UUAAAUUCGGCUCAGUCUUUUUuuuua....GAAAGAUUGAGCCGAAUUUAA
7    UUAAAUUCGGCUCAaUCUUUUcuaaaa...AAAAGAcUCAGCCGAAUUUAA
8         GGCcCAAUCUUUuaaU.....aaAAGGAUUGAGCC
9         UCGGCcCAAUCUUUuacu.....aaAAGGAUUGaGCCGA
10        UUCGGCaCAAUCUUUUccu..aaaaGAACGAUUGaGCCGAA
11        UCGGCcCAAUCUUUUuacu.....aaAAGGAUUGaGCCGA
12        UCGGCcCAAUCUUUUuacu.....aaAAGGAUUGaGCCGA
13        UCGGCcCAAUCUUUUuacu.....aaAAGGAUUGaGCCGA
                          or
13 AAUuGAAUUgcaaUUaaacUCGGCcCAAUCUUUUacu.....aaAAGGAUUGaGCCCGAauacAAca
     aaGAUUCuAUUAAaGAAGGAUAAgGAGaaagaaCUCaaguaaUUAUCCUUCgUU
```

Fig. 2. Alignment of flowering plant *rbcL* major IR. Numerals in the left margin indicate the taxonomic origin of the IR sequences, and follow the convention in the legend to Fig. 1. For *Petunia hybrida* (No. 13) two different structures are predicted; both are shown. Upper case letters indicate those nt involved in intramolecular base pairing; lower case letters within the stem regions compose bulges, and those between the stem regions compose loops. Nucleotides conserved between IR are underscored; gaps were introduced to maximize alignments.

for generating new IR, as well as modifying pre-existing IR, are operating within the chloroplast.

Zurawski and Clegg (1987) first observed that both the sequence and structure of IR present at the 3' end of chloroplast mRNAs were highly conserved between different angiosperm species. This observation led them to the rather prescient conclusion that, among other possible functions, the IR could serve as either mRNA processing signals or mRNA degradation endpoints. Subsequent biochemical studies of the role of different IR in spinach (e.g., those of *petD*, *rbcL*) have elucidated their functions as specific substrates for a complex series of biochemical processing events (Stern and Gruissem, 1987; Chen and Stern, 1991). It is reasonable to assume, given the highly conserved nature of these elements in other flowering plants, that similar biochemical processing events are present in these systems as well, and that the *rbcL* transcript in these taxa is subject to similar processing events as in the spinach transcript. Presumably alterations in the primary sequence or secondary structure (i.e., topology) would drastically alter the efficiency and accuracy of the processing events.

The flowering plant *rbcL* IR surveyed to date (rice, wheat, alfalfa, pea, broadbean, spinach, tobacco and petunia) conform to the current orthodoxy, i.e., IR of the same gene from different taxa will have highly conserved sequences, reflective of the conservation of the mRNA processing apparatus across phylogenetic boundries. Thus, these IR sequences have been subjected to considerable functional constraint, in terms of their evolutionary history. This observation is quite consistent with the biochemical studies mentioned above, in which precise nt sequences and/or stem-loop topologies are essential for the precise interaction of post-transcriptional regula-

tory factors. Thus, those principles regarding mRNA processing and stabilization derived from spinach can most likely be applied to the flowering plant chloroplast in general.

(b) Sequence divergence of non-flowering land plant *rbcL* IR

The *rbcL* IR of the non-flowering land plant representatives (i.e., a conifer, Douglas fir; a fern, *Angiopteris evecta*; three fern allies, *Equisetum arvense*, *Isoetes melanopoda*, and *Lycopodium digitatum*; and two liverworts, *Bazzania trilobata* and *Marchantia polymorpha*) cannot be aligned with either themselves or with the angiosperm sequences (Fig. 1). Free energy values ranged from $\Delta G = -6.2$ (*Equisetum arvense*) to $\Delta G = -42.1$ (*Lycopodium digitatum*), the lowest and highest thermodynamic values seen in the land plant IR to date. Thus, in marked contrast to the situation in the flowering plants, the regulatory role of these IR is not reflected in their sequence conservation. If they do serve the same function as their flowering plant counterparts, then we must infer a rapid rate of evolution in the transcript stabilizing/processing mechanism for the *rbcL* mRNA. Or, these non-flowering land plant IR could be involved in a different process entirely, such as transcriptional termination.

(c) Sequence divergence of algal, photosynsthetic protist and prokaryote *rbcL* IR

The representatives sampled for the algae, photosynthetic protists and prokaryotes likewise are all highly divergent in nt sequence and thermodynamic parameters (Fig. 1). For example, the green algae, represented here by the genera *Chlorella*, *Chlamydomonas*, *Spirogyra* and *Codium*, range from $\Delta G = -6.3$ (*Chlorella* N1a) to $\Delta G = -41.5$ (*Spirogyra maxima*). Most remarkably, two species of the same genus, *Ch. reinhardtii* and *Ch. moewusii*, differ greatly in their primary sequence and free energy values (Fig. 1). Taxa from two other major algal lineages, *Ectocarpus siliculosus* (brown algae) and *Cryptomonas* Φ (cryptomonad algae) contained likewise similar divergent IR (Fig. 1).

The IR of the closely related organisms *Euglena gracilis* (a photosynthetic protist) and *Astasia longa* (a non-photosynthetic protist), are remarkable for several reasons. First, the IR are encompassed within the coding region of the *rbcL* mRNA (Fig. 1). No other secondary structures were detected downstream from the *rbcL* gene in either taxon. Second, these structures have the lowest free energy values (*Euglena gracilis*, $\Delta G = -1.5$; *Astasia longa*, $\Delta G = -3.3$) of any systems to date. As with the structures in petunia, we question whether or not these could actually exist in vivo. If they do not, then these are the only *rbcL* mRNAs lacking a 3' in vivo IR structure.

Not unexpectedly, the photosynthetic prokaryotes *Chromatium vinosum*, *Rhodospirillum rubrum*, *Anacystis* 6301 and *Anabaena* 7120, (representing different evolutionary lineages) bore little resemblance to each other, and none to the sampled eukaryotes.

(d) Significance of IR sequence divergence

The *rbcL/psaB* transcript IR in *Chlamydomonas reinhardtii* serve not as transcript stabilizers or post-transcriptional processing signals, but instead as transcription terminators (Blowers et al., 1993). Experimentally, it has been shown that the spinach chloroplast RNA polymerase will recognize *E. coli* IR transcription terminators in vitro (Chen and Orozco, 1988). Thus, if the situation in *Ch. reinhardtii* is representative of the green algae, then these IR serve a biochemical function (transcription termination) that could be tolerant of greater sequence divergence than are their angiosperm counterparts. If these structures in systems outside the green algae are involved in the same general processes (e.g., transcriptional termination), it then appears that extreme flexibility in primary sequence of the IR is tolerated. Further experimental investigations are required (*i*) to define the in vivo role of the non-angiosperm *rbcL* IR, and (*ii*) to determine the effects of sequence alterations on the functionality of these regulatory elements.

The flowering plant IR are, most likely, homologous, given their high level of sequence conservation. It is difficult to determine, however, the true evolutionary origins of the non-angiosperm land plant IR. If, for example, the non-flowering land plant IR are homologous with the flowering plant sequences, then the differences observed between the two sets are due to accumulated nt substitutions in the divergent IR. However, as exemplified by petunia, IR can arise de novo in the A+T-rich spacer regions of chloroplast genomes. Thus, it is possible that the non-flowering land plant IR arose independently within different lineages over time, perhaps replacing pre-existing IR that were subsequently lost. The present data do not allow us to distinguish between these two possibilities.

(e) Conclusions

(*1*) The *rbcL* 3' IR is highly conserved in the flowering plants surveyed to date, reflecting the documented in vitro biochemical function of this structure in the spinach mRNA, namely transcript stabilization and correct 3' terminus processing. Given this high level of sequence conservation across lineages representing at least 150 million years of evolutionary distance, we conclude that these are homologous structures in the flowering plants.

(*2*) In the surveyed non-flowering land plants, algae, photosynthetic protists and prokaryotes, these IR are extremely divergent in both nt sequence and predicted free-energy value. This is in marked contrast to the situation in the angiosperms.

(*3*) The maintenance of these structures over the evolutionary distance encompassed by the lineages surveyed herein (with the possible exception of *E. gracilis* and *A. longa*) strongly suggests that they serve an essential purpose in the *rbcL* mRNA. In spinach (Stern et al., 1987) and the other surveyed angiosperms this function is as a transcript stabilizer and a 3' terminus processing recognition sequence. In the green alga *Ch. reinhardtii*, this structure in vivo serves as a transcription terminator (Blowers et al, 1993). Further experimental evidence is required to define the in vivo role of these regulatory elements in the non-angiosperm taxa that contain the *rbcL* gene.

REFERENCES

Aldrich, J. Cherney, B. Merlin, E. and Palmer, J.: Sequence of the *rbcL* gene for the large subunit of ribulose bisphosphate carboxylase-oxygenase from petunia. Nucleic Acids Res. 14 (1986a) 9534.

Aldrich, J. Cherney, B. Merlin, E. and Palmer, J.: Sequence of the *rbcL* gene for the large subunit of ribulose biphosphate carboxylase-oxygenase from alfalfa. Nucleic Acids Res. 14 (1986b) 9535.

Amberg, S.M. and Meints, R.H.: Nucleotide sequence of the genes for ribulose-1,5- bisphosphate carboxylase/oxygenase large subunit and ribosomal protein S14 from a *Chlorella*-like green alga. J. Phycol. 27 (1991) 753–758.

Blowers, A.D., Klein, U. and Bogorad, L.: Functional in vivo analyses of the 3' flanking sequences of the *Chlamydomonas reinhardtii* chloroplast *rbcL* and *psaB* genes. Mol. Gen. Genet. 238 (1993) 339–349.

Chen, H.-C. and Stern, D.B.: Specific ribonuclease activities in spinach chloroplasts promote mRNA maturation and degradation. J. Biol. Chem. 266 (1991) 24205- 24211.

Chen, L.-J. and Orozco Jr., E.M.: Recognition of prokaryotic transcription terminators by spinach chloroplast RNA polymerase. Nucleic Acids Res. 16 (1988) 8411–8431.

Clegg, M.T.: Chloroplast gene sequences and the study of plant evolution. Proc. Natl. Acad. Sci. USA 90 (1993) 363–367.

Curtis, S.E. and Haselkorn, R.: Isolation and sequence of the gene for the large subunit of ribulose-1,5-bisphosphate carboxylase from the

cyanobacterium *Anabaena* 7120. Proc. Natl. Acad. Sci. USA 80 (1983) 1835–1839.

Devereux, J., Haberli, P. and Smithies, O.: A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res. 12 (1984) 387–395.

Doebley, J., Durbin, M., Golenberg, E.M., Clegg, M.T. and Ma, D.P.: Evolutionary analysis of the large subunit of carboxylase (*rbcL*) nucleotide sequence among the grasses (Gramineae). Evolution 44 (1990) 1097–1108.

Douglas, S.E., Durnford, D.G. and Morden, C.W.: Nucleotide sequence of the gene for the large subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase from *Cryptomonas* Φ: Evidence supporting the polyphyletic origin of plastids. J. Phycol. 26 (1990) 500–508.

Dron, M., Rahire, M. and Rochaix, J.-D.: Sequence of the chloroplast DNA region of *Chlamydomonas reinhardii* containing the gene of the large subunit of ribulose bisphosphate carboxylase and parts of its flanking genes. J. Mol. Biol. 162 (1982) 775–793.

Gingrich, J.C. and Hallick, R.B.: The *Euglena gracilis* chloroplast ribulose-1,5-bisphosphate carboxylase gene: I. Complete DNA sequence and analysis of the nine intervening sequences. J. Biol. Chem. 260 (1985) 16156–16161.

Hipkins, V.D., Tsai, C.-H. and Strauss, S.H.: Sequence of the gene for the large subunit of ribulose-1,5-bisphosphate carboxylase from a gymnosperm, Douglas fir. Plant Mol. Biol. 15 (1990) 505–507.

Hiratsuka, J., Shimada, H., Whittier, R., Ishibashi, T., Sakamoto, M., Mori, M., Kondo, C., Honji, Y., Sun, C.-R., Meng, B.-Y., Li, Y.-Q., Kanno, A., Nishizawa, Y., Hirai, A., Shinozaki, K. and Sugiura, M.: The complete sequence of the rice (*Oryza sativa*) chloroplast genome: Intermolecular recombination between distinct tRNA genes accounts for a major plastid DNA inversion during the evolution of the cereals. Mol. Gen. Genet. 217 (1989) 185–194.

Manhart, J.R.: Phylogenetic analysis of green plant *rbcL* sequences. Mol. Phylogenet. Evol. (1994) in press.

Manhart, J.R. and VonderHaar, R.A.: Intron revealed by nucleotide sequence of large subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase from *Codium fragile* (Chlorophyta): phylogenetic analysis. J. Phycol. 27 (1991) 613–617.

Morton, B.R. and Clegg, M.T.: A chloroplast DNA mutational hotspot and gene conversion in a noncoding region near *rbcL* in the grass family (Poaceae). Curr. Genet. 24 (1993) 357–365.

Narang, F., McIntosh, L. and Somerville, C.: Nucleotide sequence of the ribulosebisphosphate carboxylase gene from *Rhodospirillum rubrum*. Mol. Gen. Genet. 193 (1984) 220–224.

Nishizawa, Y. and Hirai, A.: Nucleotide sequence and expression of the gene for the large subunit of rice ribulose-1,5-bisphosphate carboxylase. Jpn. J. Genet. 62 (1987) 389–395.

Ohyama, K., Fukuzawa, H., Kohchi, T., Shirai, H., Sano, T., Sano, S., Umesono, K., Shiki, Y., Takeuchi, M., Chang, Z., Aota, S., Inokuchi, H. and Ozeki, H.: Chloroplast gene organization deduced from complete nucleotide sequence of liverwort *Marchantia polymorpha* chloroplast DNA. Nature 322 (1986) 572–574.

Schuster, G., and Gruissem, W.: Chloroplast mRNA 3′ end processing requires a nuclear-encoded RNA binding protein. EMBO J. 10 (1991) 1493–1502.

Shinozaki, K., Ohme, M., Tanaka, M., Wakasugi, T., Hayshida, N.,

Matsubayasha, T., Zaita, N., Chungwongse, J., Obokata, J., Yamaguchi-Shinozaki, K., Ohto, C., Torazawa, K., Meng, B.Y., Sugita, M., Deno, H., Kamogashira, T., Yamada, K., Kusuda, J., Takaiwa, F., Kata, A., Tohdoh, N., Shimada, H. and Sugiura, M.: The complete nucleotide sequence of the tobacco chloroplast genome. EMBO J. 5 (1986) 2043–2049.

Shinozaki, K., Yamada, C., Takahata, N. and Sugiura, M.: Molecular cloning and sequence analysis of the cyanobacterial gene for the large subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase. Proc. Natl. Acad. Sci. USA 80 (1983) 4050–4054.

Siemeister, G and Hachtel, W.: 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 (1990) 825–833.

Stern, D.B. and Gruissem, W.: Control of plastid gene expression: 3′ inverted repeats act as mRNA processing and stabilizing elements, but do not terminate transcription. Cell 51 (1987) 1145–1157.

Stern, D.B., Jones, H. and Gruissem, W.: Function of plastid 3′ mRNA inverted repeats. J. Biol. Chem. 264 (1989) 18742–18750.

Terachi, T., Ogihara, Y. and Tsunewaki, K.: The molecular basis of genetic diversity among cytoplasms of *Triticum* and *Aegilops*. VI. Complete nucleotide sequences of the *rbcL* genes encoding H- and L-type rubisco large subunits in common wheat and *Ae. crassa* 4x. Jpn. J. Genet. 62 (1987) 375–387.

Valentin, K. and Zetsche, K.: Rubisco genes indicate a close phylogenetic relation between the plastids of Chromophyta and Rhodophyta. Plant Mol. Biol. 15 (1990) 575–584.

Viale, A.M., Kobayashi, H. and Akazawa, T.: Expressed genes for planttype ribulose-1,5-bisphosphate carboxylase/oxygenase in the photosynthetic bacterium *Chromatium vinosum*, which possesses two complete sets of the genes. J. Bacteriol. 171 (1989) 2391–2400.

Yang, R.C.A., Dove, M., Seligy, V.L., Lemieux, C., Turmel, M. and Narang, S.A.: Complete nucleotide sequence and mRNA-mapping of the large subunit gene of ribulose-1,5-bisphosphate carboxylase oxygenase (Rubisco) from *Chlamydomonas moewusii*. Gene 50 (1986) 259–270.

Yoshinaga, K., Ohta, T., Suzuki, Y. and Sugiura, M.: *Chlorella* chloroplast DNA sequence containing a gene for the large subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase and a part of the possible gene for the β′ subunit of RNA polymerase. Plant Mol. Biol. 10 (1988) 245–250.

Zurawski, G. and Clegg, M.T.: Evolution of higher-plant chloroplast DNA-encoded genes: implications for structure function and phylogenetic studies. Annu. Rev. Plant Physiol. 38 (1987) 391–418.

Zurawski, G., Clegg, M.T. and Brown, A.H.D.: The nature of nucleotide sequence divergence between barley and maize chloroplast DNA. Genetics 106 (1984) 735–749.

Zurawski, G., Whitfeld, P.R. and Bottomley, W.: Sequence of the gene for the large subunit of ribulose-1,5-bisphosphate carboxylase from pea chloroplasts. Nucleic Acids Res. 14 (1986) 3975–3976.

Zurawski, G., Perrot, B., Bottomley, W. and Whitfeld, P.R.: The structure of the gene for the large subunit of ribulose-1,5-bisphosphate carboxylase from spinach chloroplast DNA. Nucleic Acids Res. 9 (1981) 3251–3270.