Proc. Natl. Acad. Sci. USA
Vol. 85, pp. 846–850, February 1988
Evolution

rbcS genes in *Solanum tuberosum*: Conservation of transit peptide and exon shuffling during evolution

(plant gene family/ribulose bisphosphate carboxylase/potato/intron)

FRANK P. WOLTER*, CHRISTIAN C. FRITZ*, LOTHAR WILLMITZER*†, JEFF SCHELL*, AND PETER H. SCHREIER*‡¶

*Max-Planck-Institut für Züchtungsforschung, D-5000 Köln 30, Federal Republic of Germany; and ‡Bayer AG, PF-A/BF, D-5090 Leverkusen, Federal Republic of Germany

Contributed by Jeff Schell, September 30, 1987

ABSTRACT **Five genes of the *rbcS* gene family of *Solanum tuberosum* (potato) were studied. One of these is a cDNA clone; the other four are located on two genomic clones representing two different chromosomal loci containing one (locus 1) and three genes (locus 2), respectively. The intron/exon structure of the three genes in locus 2 is highly conserved with respect to size and position. These genes contain two introns, whereas the gene from locus 1 contains three introns. Although in most cases the amino acid sequences in the transit peptide part of different *rbcS* genes from the same species varied considerably more than the corresponding mature amino acid sequences, one exception found in tomato and potato indicates that the transit peptide of *rbcS* could have a special function. A comparison of the *rbcS* genes of higher plants with those of prokaryotes offers suggestive evidence that introns first served as spacer material in the process of exon shuffling and then were removed stepwise during the evolution of higher plants.**

Ribulose bisphosphate carboxylase (RbuP_2 carboxylase; E.C. 4.1.1.39), the key enzyme in the dark reaction of photosynthesis (for a review, see ref. 1), is composed of eight large (LSU) and eight small (SSU) subunits in higher plants, which are both essential for the enzymatic activity (2). The corresponding genes are located in different compartments. The *rbcL* gene that codes for LSU is part of the chloroplast genome (3), leading to a copy number of several thousand genes per mesophyll cell (4). The *rbcL*-encoded mRNA is translated inside the organelle on plastid ribosomes. In contrast, the SSU is encoded in the nucleus and the *rbcS*-encoded mRNA is translated in the cytoplasm into a precursor protein containing an amino-terminal extension, named transit peptide, that mediates the transport of the polypeptide into the chloroplast and is cleaved off during or shortly after transport (for a review, see ref. 5). The mature SSU assembles in the chloroplast stroma with the LSU to form the holoenzyme. In all plants studied, several copies of *rbcS* genes have been found (6–8).

We would like to understand the reason for the existence of gene families in plants. Due to the fundamental importance of RbuP_2, the *rbcS* gene family provides a significant example with which to probe the relevance of gene families in plants. Comparative studies of gene structures and sequences give information relevant to this question. Conservation of amino acid sequence in different species can provide indications for a functional role of conserved regions in SSU proteins. We chose to study the *rbcS* genes from *Solanum tuberosum* (potato) since several members of the *rbcS* gene family from other solanaceous species are available for interspecies comparison. In addition, we have extended our comparison to *rbcS* genes of other higher plants and prokaryotes. We find different types of genes with respect to intron number, the positions of which are, however, highly conserved within higher plants. The distribution of homology between prokaryotic and eukaryotic genes leads to a scheme for the evolution of the *rbcS* genes that supports the idea of exon shuffling and stepwise removal of introns and confirms the endosymbiotic theory.

MATERIALS AND METHODS

Plant Material. The following lines from the potato collection of the Max-Planck-Institut für Züchtungsforschung were used: the diploid line HH 1201/7 (cDNA) and the haploid line AM 5793 (genomic library) (9).

DNAs and Vectors. Nuclear DNA was prepared according to Bedbrook (10). The synthesis and isolation of cDNA *rbcS* c was described as L900 by Eckes *et al.* (11). The construction of the genomic library was described elsewhere (9). Genomic clones were subcloned in pUC8 or in the M13 phage vectors mp18 and mp19 (12). Phage λ propagation, screening, and DNA isolation were performed as described by Rosahl *et al.* (9).

Sequencing. The DNA sequence was determined by using the chain-termination method described by Sanger *et al.* (13) on M13-derived templates (14).

Data Handling. For sequence comparisons, the Wisconsin computer group program package (15) and the GENEXPERT package (16) were used on a VAX computer (Digital Corporation).

RESULTS

Sequence of the *rbcS* cDNA. The clone *rbcS* c isolated from a leaf-specific cDNA library (11) was subcloned in M13 vectors and the nucleotide sequence§ was determined (Fig. 1). The sequence reveals one long open reading frame of 546 nucleotides. The deduced polypeptide sequence of 181 amino acids has the typical structure of SSU with an amino-terminal extension of 58 residues similar to other transit peptides and 123 amino acids with a high homology to the mature part of SSU of other plants.

λ Clones Containing *rbcS* Genes. The potato genomic library from line AM 5793 (9) was screened with a DNA

Abbreviations: RbuP_2 carboxylase, ribulose bisphosphate carboxylase; SSU, small subunit of RbuP_2 carboxylase; LSU, large subunit of RbuP_2 carboxylase.

†Present address: IGF Berlin, Ihnestrasse 63, D-1000 Berlin 33, Federal Republic of Germany.

§The sequence reported in this paper is being deposited in the EMBL/GenBank data base (Bolt, Beranek, and Newman Laboratories, Cambridge, MA, and Eur. Mol. Biol. Lab., Heidelberg) (accession no. J03613).

¶To whom reprint requests should be addressed.

```
         .         .         .         .
ATGGCTTCCTCAATTGTCTCCTCAGCAGCCGTTGCTACCCGTAGC   45
 M  A  S  S  I  V  S  S  A  A  V  A  T  R  S
    .         .         .         .         .
AATGTTGCTCAAGCTAGCATGGTCGCACCCTTCACCGGACTCAAA   90
 N  V  A  Q  A  S  M  V  A  P  F  T  G  L  K
         .         .         .         .
TCCGCCGCGTCTTTCCCCGTTACTAAGAAGAACAACAACGTTGAC   135
 S  A  A  S  F  P  V  T  K  K  N  N  N  V  D
    .         .         .         .         .
ATTACTTCCCTTGCTAGCAACGGTGGACGTGTTAGATGCATGCAG   180
 I  T  S  L  A  S  N  G  G  R  V  R  C| M  Q
         .         .         .         .
GTGTGGCCACCAATTAACATGAAGAAATACGAGACATTGTCATAC   225
 V  W  P  P  I  N  M  K  K  Y  E  T  L  S  Y
    .         .         .         .         .
CTTCCTGACTTGTCCGATGAGCAATTGCTCAAGGAAGTTGAGTAC   270
 L  P  D  L  S  D  E  Q  L  L  K  E  V  E  Y
         .         .         .         .
CTTTTGAAAAATGGATGGGTTCCTTGCTTGGAATTCGAGACTGAG   315
 L  L  K  N  G  W  V  P  C  L  E  F  E  T  E
    .         .         .         .         .
CACGGATTTGTGTACCGTGAGCACAACAGCTCACCAGGATACTAC   360
 H  G  F  V  Y  R  E  H  N  S  S  P  G  Y  Y
         .         .         .         .
GATGGTAGATACTGGACCATGTGGAAGTTGCCTATGTTTGGGTGC   405
 D  G  R  Y  W  T  M  W  K  L  P  M  F  G  C
    .         .         .         .         .
ACTGACGGAACCCAGGTGTTGGCTGAGGTTCAAGAGGCCAAGAAT   450
 T  D  G  T  Q  V  L  A  E  V  Q  E  A  K  N
         .         .         .         .
GCGTACCCACAGGCCTGGATCCGTATTATCGGATTCGACAACGTT   495
 A  Y  P  Q  A  W  I  R  I  I  G  F  D  N  V
    .         .         .         .         .
CGTCAAGTGCAGTGCATCAGTTTCATTGCCTACAAGCCAGAAGGA   540
 R  Q  V  Q  C  I  S  F  I  A  Y  K  P  E  G

TACTAA   546
 Y  *
```

FIG. 1. Nucleotide sequence of the open reading frame in cDNA *rbcS* c. The deduced amino acid sequence is shown below in one-letter code. The asterisk identifies the termination codon in the cDNA. The putative processing site between transit peptide and mature SSU is marked by a horizontal line; codon 41 is underlined.

probe of *rbcS* c. Eleven positive λ clones were recovered from the library and eight of them were further characterized. Physical maps were established by restriction endonuclease and hybridization analysis. It turned out that the clones can be divided into two sets according to identical internal restriction sites and hybridization patterns. This suggests that the clones represent overlapping stretches of two different genomic loci, as shown in Fig. 2. The first set represents a chromosomal region (locus 1) containing one isolated gene, *rbcS* 1 (Fig. 2*a*). From the second set, which represents a second region (locus 2), the recombinant phage G9 contains three closely linked genes, *rbcS* 2a, 2b, and 2c (Fig. 2*b*). The distance between gene *rbcS* 2a to the middle one, *rbcS* 2b, is about 5.5 kb. The third gene, *rbcS* 2c, is located <2 kb to the other side of *rbcS* 2b. All three genes are in the same orientation and are located within a stretch of about 10 kb of DNA, whereas the single gene, *rbcS* 1, is separated from locus 2 by at least 12 kb.

Genomic *rbcS* Clones. The *rbcS* genes 1, 2a, 2b, and 2c, located on *Hin*dIII restriction fragments of 8.0, 4.3, 3.2, and 2.7 kilobase pairs (kbp), were subcloned into pUC8 and further analyzed. The DNA sequence was determined and will be published elsewhere (unpublished data). The overall structure of the *rbcS* genes is presented in Fig. 3. The first exon of the three genes of locus 2 consists of 177 base pairs (bp) and codes for 59 amino acids (transit peptide plus 2 amino acids of the mature part). The rest of the mature polypeptide is encoded by two additional exons of 135 and 228 bp leading to a potential mature SSU polypeptide of 123 amino acids. These exons are interrupted by introns located at homologous positions behind codon 2 and codon 47 of the mature part of SSU. The first intron measures 89 nucleotides in the three clustered genes, whereas the second has a variant size between 81 and 85 nucleotides. The single gene, *rbcS* 1, is interrupted by three introns. Intron 1 is located behind codon 2 of mature SSU, as in the other potato genes, but has a size of 1196 nucleotides. The first exon has an extra codon, GGC, at position 21 leading to a transit peptide of 58 amino acids. This addition is at a different position compared

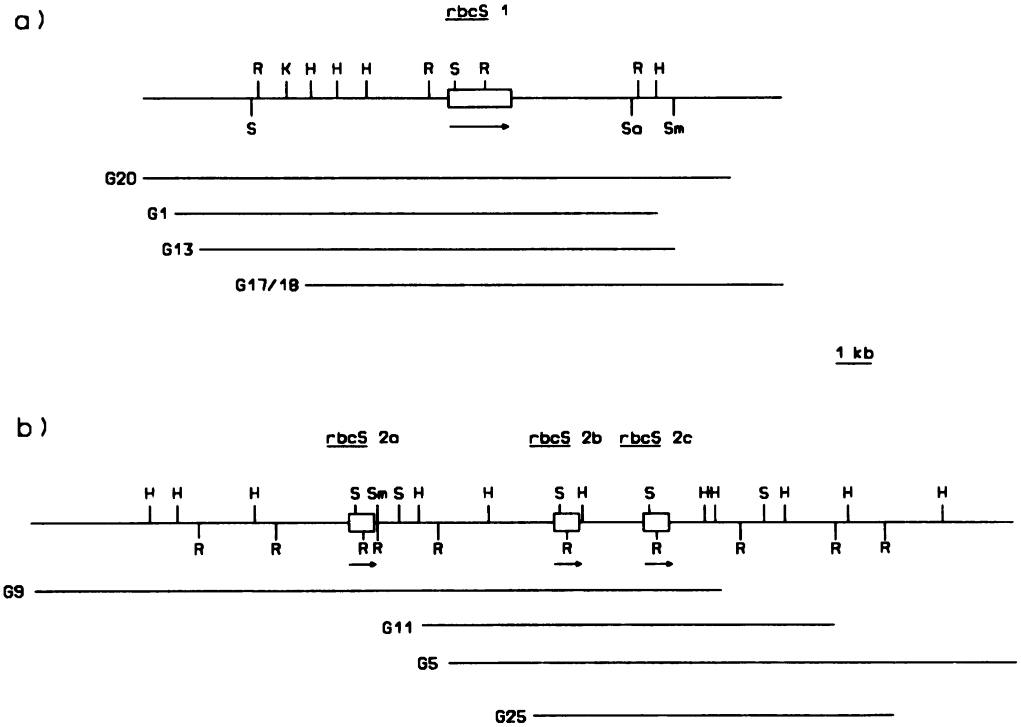

FIG. 2. Restriction map of two loci containing potato *rbcS* genes. (*a*) Maps of clones G20, G1, G13, and G17/18 containing the *rbcS* 1 gene. (*b*) Clones G9, G11, G5, and G25 are presented with the genes *rbcS* 2a, *rbcS* 2b, and *rbcS* 2c. Restriction sites are marked as indicated: R, *Eco*RI; S, *Sph* I; H, *Hin*dIII; K, *Kpn* I; Sa, *Sal* I; Sm, *Sma* I. The genes are shown as boxes and the direction of transcription is marked with arrows. kb, Kilobase.

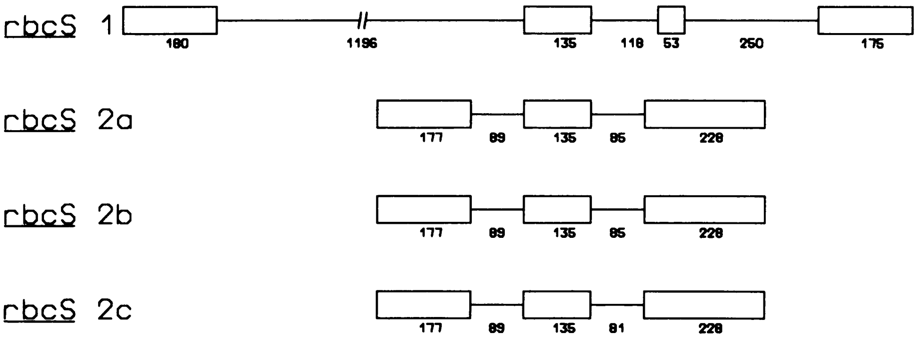

FIG. 3. Intron/exon structure of the four *rbcS* genes as predicted by consensus sequences (17, 18). The orientation of the genes is 5′ → 3′ from left to right. The sizes are given in bp. Exons are marked as boxes; introns are marked as lines.

to that found in the cDNA clone (*rbcS* c). Intron 2 is at the same position as the second intron in the other potato *rbcS* genes but is 118 bp long. The additional intron in *rbcS* 1 (intron 3) measures 249 nucleotides. This intron splits codon 65, whereas all other introns are located between codons. It appears that *rbcS* c is not represented by any of the four genomic clones and therefore represents a fifth potato gene.

DISCUSSION

We have isolated and characterized five *rbcS* genes of potato, one as a cDNA and four as genomic clones. Three of these genes are organized as direct repeats in one locus. They contain two introns of similar length. The other gene, *rbcS* 1, is separated by at least 12 kbp from this cluster and shows a different structure. It has three introns and the first of them is 1196 bp long. This gene has an additional codon in the first exon due to an insertion at position 21 of the deduced transit peptide. An additional codon is also found in the potato cDNA behind codon 41. There is only one other gene within the Solanaceae with such a codon insertion. This is a tomato gene (*rbcS-1*) that contains the same codon insertion at the identical position to that found in the potato cDNA. In both genes the region around the additional codon has the sequence 5′ GAAGAACAACAACAA 3′, showing twice a repetition of GAA and three times a repetition of CAA. This sequence resembles a structure observed in revertant genes after visitation by a mobile DNA element (19). During insertion of the element a short direct repeat is created at the border of the insertion that is partially retained after the element has transposed away. It seems conceivable to imagine mutations in *rbcS* genes created by the action of a yet unidentified transposable element in Solanaceae. This insertion provides circumstantial evidence indicating that these two genes, although present in different species, are derived from a common ancestor.

If this were the case, this pair of genes would provide an excellent test to verify the idea of concerted evolution suggested recently by Pichersky *et al.* (8). According to this hypothesis the mature parts of SSU are selected to a uniform amino acid sequence within a species. Two genes derived from a common progenitor and now located in different species should therefore change in their mature parts and should as a result be more similar to the corresponding genes

A

```
                  1        10        20        30        40        50           dif.
                  .        .         .         .         .         .            a.a.

rbcS-1(tom)                 A                                                   1
rbcS c(pot)       MASSIVSSAAVATRSNVAQA.SMVAPFTGLKSAASFPVTKKNNNVDITSLASNGGRVRC
rbcS 1(pot)           VI        T  T  G  I          T   SR .Q L    I            13
rbcS 2a(pot)          VM        G G                     SR .Q L    I            10
rbcS 2b(pot)                    A G       G       T     SR .Q L    I            10
rbcS 2c(pot)          VM        G G               T     SR .Q L    I            11
```

B

```
                  1        10        20        30        40        50        60        70        80
                  .        .         .         .         .         .         .         .         .

rbcS-1(tom)                                  S I                 R      N
rbcS c(pot)       MQVWPPINMKKYETLSYLPDLSDEQLLKEVEYLLKNGWVPCLEFETEHGFVYREHNSSPGYYDGRYWTMWKLPMFGCTDA
rbcS 1(pot)                            T                                NHK
rbcS 2a(pot)                                                            N K
rbcS 2b(pot)                                                            N K
rbcS 2c(pot)                                                            N K
```

```
                  81       90        100       110       120    dif.
                  .        .         .         .         .      a.a.

rbcS-1(tom)                  K      V                       F   7
rbcS c(pot)       TQVLAEVQEAKNAYPQAWIRIIGFDNVRQVQCISFIAYKPEGY*
rbcS 1(pot)                C KS                                 7
rbcS 2a(pot)             E   K                                  4
rbcS 2b(pot)             E   K                                  4
rbcS 2c(pot)             E   K                                  4
```

FIG. 4. Comparison of the amino acid sequences of the *rbcS* genes from *rbcS* c to that of the other potato (pot; *S. tuberosum*) *rbcS* genes and *rbcS-1* of tomato (tom; *Lycopersicon esculentum*). The number of amino acids different to *rbcS* c (dif. a.a.) is given at the end of each sequence. (*A*) Comparison for the transit peptide part. (*B*) Comparison for the mature part of SSU.

within their species than to each other. Their transit peptide parts, however, not subjected to such a stringent selection may serve to measure the evolutionary distance of genes from a common origin. Indeed, we can show in a comparison between the potato and the tomato SSUs (Fig. 4) that the mature part encoded by the potato cDNA is at least as close and in three cases much closer to the other potato genes than to its counterpart in tomato (*rbcS-1*). For the transit peptide part, in contrast, the divergence is higher between the potato genes. This divergence is 10–13 amino acids for 58 residues or one exchange every 5 residues, contrasted by only 4–7 for the 123 amino acids of the mature part, which is one exchange every 20 residues. Our data are in good agreement with previous observations; they underline and give further evidence to support the hypothesis of concerted evolution for the mature polypeptides in a given species and they show the expected high divergence in the transit peptide.

The comparison between the transit peptides encoded by the potato cDNA and its homologue in tomato, however, shows only 1 amino acid difference. One exchange in 58 amino acids gives a value much lower than for the mature parts of these genes (7/123). We interpret the high conservation of the transit peptides of two genes in two species as indicating an important and specific function for these particular polypeptides. Since it has been shown by *in vitro* uptake experiments that precursors with very different transit peptide sequences are still competent in transport through a given chloroplast envelope (for a review, see ref. 5), we feel that the function causing the conservation in the transit peptide is not critical for transport in general but might be important for transport during certain phases of development—e.g., when functional chloroplasts are not established or in specific cells or organelles of the adult plant. Another possibility might be that this sequence plays a role in the communication process between chloroplast and nucleus. A similar finding confirming our observation has been described by Poulsen *et al.* (20) comparing two genes from *Nicotiana tabacum* and *Nicotiana plumbaginifolia*. These genes are different in seven positions in the mature part but contain transit peptides that are entirely identical.

The observation of different intron numbers in the potato genes and the availability of prokaryotic and eukaryotic sequences led us to extend our comparison to all known *rbcS* genes in order to contribute to the discussion of the evolutionary history and possible function of introns. The prokaryotic SSU sequences from cyanobacteria have apparent homology to those of higher plants in two discrete blocks with about 40% homology for the first and about 35% for the second part (see refs. 21 and 22). In the case of *rbcS* genes with three introns (e.g., *rbcS* 1 of potato) these blocks are encoded by the second and fourth exon. The regions encoded by the first and third exon have no homologue in the prokaryotic proteins (see Fig. 5). The *rbcS* genes with three introns can be described as follows. Exon 1 has no homology to prokaryotic *rbcS* genes and codes for the transit peptide sequence as well as for two N-terminal amino acids of mature SSU. Exon 2 shows homology to prokaryotic genes, thus pointing to the basic functional importance of this part of the SSU protein. The third exon is not homologous to prokaryotic SSU sequences and no specific function could thus far be correlated with this domain. However, recent experiments with mutated SSU polypeptides lacking the sequence of exon 3 suggest a role of this part in assembly of SSU with LSU (H. Bohnert, personal communication). The last exon again shows homology to prokaryotic SSU sequences. There is therefore an alternating row of regions with and without homology to prokaryotic genes, and a change in homology is reflected structurally by introns interrupting the coding sequence.

This conformation of gene structure for the genes with three introns can best be explained by embedding it into the endosymbiotic theory implying that chloroplasts are the offspring of prokaryotic autotrophic symbionts (23). It is assumed that present nuclear-encoded chloroplast proteins originated from the prokaryotic organelle genome. New evidence in favor of this thesis was recently presented (24, 43). An arrangement whereby an *rbcS* gene resides at its original location, 3′ to the *rbcL* gene as part of a small operon, can still be found in cyanobacteria (21, 22), in endosymbiotic cyanelles (25), and perhaps in plastids of non-green algae (26, 27). During evolution the *rbcS* genes were transferred from the endosymbiont into the nucleus of the host (28). After this transfer the gene product had to fulfill new functions. First, the need to cross the chloroplast envelope required the addition of a domain for transport and subsequent processing. Second, new signals might have been required to coordinate the expression and assembly of the two subunit genes now located in different compartments. For genes with three introns we describe two exons without homology to their prokaryotic counterparts. We propose that these exons contain new domains incorporated into the gene to facilitate its new functions. In addition, the observation that these new domains are separated from the old ones by introns strongly suggests that introns were used as spacer material in this process of "exon shuffling" (29).

A comparison of different eukaryotic *rbcS* genes shows a very high homology between their coding sequences. The genes carry three, two, or one intron and the position of introns is conserved. We suggest that after the successive construction of an original eukaryotic gene introns have no particular function and may have been lost stepwise during evolution. Removal of introns has been shown for other genes, such as insulin (30) and triosephosphate isomerase (31, 32), and might be a general tendency during evolution (33). Under this aspect the different types of *rbcS* genes can be classified as shown in Fig. 5. Type 0 consists of uninterrupted genes as found in prokaryotic organisms. The next type with three introns, type III, is represented by the structure of *rbcS* 1 in potato. All four genes of this type known up to now are found in Solanaceae: *S. tuberosum*, *Petunia hybrida* (7, 34), *N. tabacum* (35) and *N. plumbaginifolia* (20). Type II genes, which include all other *rbcS* genes of dicotyledonous plants, show two introns at the positions of intron one and two in type III genes. Type Ia and Ib are genes with one intron, being either at the position of the first or the second intron of the type III genes. These

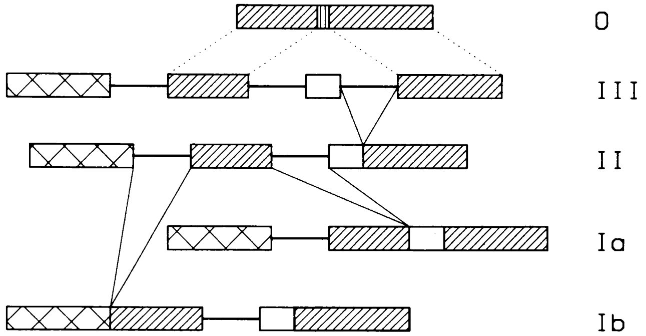

FIG. 5. Structure of different *rbcS* gene types. Type 0: prokaryotic, intronless (cyanobacteria); type III: three-intron genes (four of the Solanaceae *rbcS* genes); type II: two-intron genes, the last intron is missing (all other dicotyledonous plants); type I: one-intron genes (monocotyledonous plants); type Ia: the second intron is missing (wheat, corn); type Ib: the first intron is missing (Lemna). Exons are shown as boxes; introns are shown as lines. The common exons between prokaryotes and higher plants are dotted lines. The new exons are open (middle) or crossed (first exon, transit peptide). The few amino acids found between the two common regions in the prokaryotic genes are marked by vertical stripes.

genes are exclusively found in monocotyledonous plants. Wheat and corn have *rbcS* genes with an intron in position one (36, 37), whereas the genes of Lemna have an intron at the second position (38).

The genomic structure of *rbcS* genes of *Chlamydomonas*, a primitive eukaryotic alga, has been published (39). There are two genes in this organism located in the nucleus, both of which contain three introns. The existence of type III *rbcS* genes in this primitive organism supports our view that this gene type, indeed, is the most archaic one. The positions of the introns, however, have shifted slightly in comparison to those of higher plants. This phenomenon is known from other gene families and termed "intron sliding" (40). When aligning these *rbcS* genes with the higher plant type III genes, good homology is found only for exons two and four, whereas it is poor for exons one and three. Intron sliding and low homology for the two exons, which we have suggested to be new for eukaryotic genes, can be explained in two ways. First, the *rbcS* genes of higher plants and *Chlamydomonas* may have evolved from a common ancestor, which already had the three intron structure. The differences then purely reflect divergence that accumulated since separation, which according to 5S RNA comparisons seems to have occurred very early in evolution (41). The second possibility is that the addition of introns and the following exon shuffling occurred independently. Functional requirements and domain integrity, however, enforced the addition of exons to happen at similar but not identical positions, as suggested by Rogers (42), and the new exons are similar in function but not in primary sequence. Comparative investigations of the corresponding structures in other lower plant taxa may help to decide between these alternatives.

In summary, we postulate that the structure of the small subunit genes of RbuP_2 carboxylase is the result of two counteracting processes working sequentially during evolution. First, introns were introduced before or during exon shuffling, adding new domains for new functions. Later, these introns were lost stepwise, leading to a more streamlined gene structure.

We thank S. Rosahl for the genomic potato library and P. Eckes for the cDNA clone.

1. Miziorko, H. M. & Lorimer, G. H. (1983) *Annu. Rev. Biochem.* **52**, 507–535.
2. Andrews, T. J., Lorimer, G. H. & Pierce, J. (1986) *J. Biol. Chem.* **261**, 12184–12188.
3. Coen, D. M., Bedbrook, J. R., Bogorad, L. & Rich, A. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5487–5491.
4. Bohnert, H. J., Crouse, E. J. & Schmitt, J. M. (1982) in *Nucleic Acids and Proteins in Plants II*, eds. Parthier, B. & Boulter, D. (Springer, Berlin), pp. 475–530.
5. Schmidt, G. W. & Mishkind, M. L. (1986) *Annu. Rev. Biochem.* **55**, 879–912.
6. Wimpee, C. F., Stiekema, W. J. & Tobin, E. M. (1983) in *Plant Molecular Biology*, ed. Goldberg, R. B. (Liss, New York), pp. 391–401.
7. Dean, C., Elzen van den, P., Tamaki, S., Dunsmuir, P. & Bedbrook, J. (1985) *EMBO J.* **4**, 3055–3061.
8. Pichersky, E., Bernatzky, R., Tanksley, S. D. & Cashmore, A. R. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 3880–3884.
9. Rosahl, S., Eckes, P., Schell, J. & Willmitzer, L. (1986) *Mol. Gen. Genet.* **202**, 368–373.
10. Bedbrook, J. (1981) *Plant Mol. Biol. Newsl.* **2**, 24.
11. Eckes, P., Schell, J. & Willmitzer, L. (1985) *Mol. Gen. Genet.* **199**, 216–224.
12. Yanisch-Perron, C., Vieira, J. & Messing, J. (1985) *Gene* **33**, 109–119.
13. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
14. Schreier, P. H. & Cortese, R. (1979) *J. Mol. Biol.* **129**, 169–172.
15. Devereux, J., Haeberli, P. & Smithies, O. (1984) *Nucleic Acids Res.* **12**, 387–395.
16. Stüber, K. (1986) *Nucleic Acids Res.* **14**, 317–326.
17. Mount, S. M. (1982) *Nucleic Acids Res.* **10**, 459–472.
18. Brown, J. W. S. (1986) *Nucleic Acids Res.* **14**, 9549–9559.
19. Saedler, H. & Nevers, P. (1985) *EMBO J.* **4**, 585–590.
20. Poulsen, C., Fluhr, R., Kauffman, J. M., Boutry, M. & Chua, N. H. (1986) *Mol. Gen. Genet.* **205**, 193–200.
21. Nierzwicki-Bauer, S. A., Curtis, S. E. & Haselkorn, R. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 5961–5965.
22. Shinozaki, K. & Sugiura, M. (1983) *Nucleic Acids Res.* **11**, 6957–6964.
23. Margulis, L. (1970) *Origin of Eukaryotic Cells* (Yale Univ. Press, New Haven, CT).
24. Shih, M. C., Lazor, G. & Goodman, H. M. (1986) *Cell* **47**, 73–80.
25. Heinhorst, S. & Shively, J. M. (1983) *Nature (London)* **304**, 373–374.
26. Steinmüller, K., Kaling, M. & Zetsche, K. (1983) *Planta* **159**, 308–313.
27. Reith, M. & Cattolico, R. A. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 8599–8603.
28. Akazawa, T., Takabe, T. & Kobayashi, H. (1984) *Trends Biochem. Sci.* **9**, 380–383.
29. Gilbert, W. (1978) *Nature (London)* **271**, 501.
30. Perler, F., Efstratiadis, A., Lomedico, P., Gilbert, W., Kolodner, R. & Dodgson, J. (1980) *Cell* **20**, 555–566.
31. Marchioni, M. & Gilbert, W. (1986) *Cell* **46**, 133–141.
32. McKnight, G. L., O'Hara, P. J. & Parker, M. L. (1986) *Cell* **46**, 143–147.
33. Doolittle, W. F. (1978) *Nature (London)* **272**, 581–582.
34. Tumer, N. E., Clark, W. G., Tabor, G. J., Hironaka, C. M., Fraley, R. T. & Shah, D. M. (1986) *Nucleic Acids Res.* **14**, 3325–3342.
35. Mazur, B. J. & Chui, C.-F. (1985) *Nucleic Acids Res.* **7**, 2373–2386.
36. Broglie, R., Coruzzi, G., Lamppa, G., Keith, B. & Chua, N.-H. (1983) *Biotechnology* **1**, 55–61.
37. Lebrun, M., Waksman, G. & Freyssinet, G. (1987) *Nucleic Acids Res.* **15**, 4360.
38. Tobin, E. M., Wimpee, C. F., Karlin-Neumann, G. A., Silverthorne, J. & Kohorn, B. (1985) in *Molecular Biology of the Photosynthetic Apparatus*, eds. Steinback, K. E., Bonitz, S., Arntzen, C. J. & Bogorad, L. (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY), pp. 373–380.
39. Goldschmidt-Clermont, M. & Rahire, M. (1986) *J. Mol. Biol.* **191**, 421–432.
40. Rogers, J. (1985) *Nature (London)* **315**, 458–459.
41. Hori, H., Lim, B.-L. & Osawa, S. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 820–823.
42. Rogers, J. (1986) *Trends Genet.* **2**, 223.
43. Martin, W. & Cerff, R. (1986) *Eur. J. Biochem.* **159**, 323-331.