Title

Maize mitochondrial plasmid S-1 sequences share homology with chloroplast gene psbA.

Journal

Genetics, 113(2)

ISSN

0016-6731

Authors

Sederoff, R R
Ronald, P
Bedinger, P
et al.

Publication Date

1986-06-01

Peer reviewed

MAIZE MITOCHONDRIAL PLASMID S-1 SEQUENCES SHARE HOMOLOGY WITH CHLOROPLAST GENE *PSBA*

RONALD R. SEDEROFF,*[1] PAMELA RONALD,[†] PATRICIA BEDINGER[†] CAROL RIVIN,[†] VIRGINIA WALBOT[†] MOLLY BLAND[‡] AND C. S. LEVINGS III[‡]

**Institute of Forest Genetics, Pacific Southwest Forest and Range Experiment Station, USDA, US Forest Service, P.O. Box 245, Berkeley, California 94701, [†]Department of Biological Sciences, Stanford University, Stanford, California 94305, and [‡]Department of Genetics, North Carolina State University, Raleigh, North Carolina 27695-7614*

Manuscript received June 20, 1985
Revised copy accepted February 24, 1986

ABSTRACT

The linear, 6397-base pair (bp), mitochondrial S-1 DNA molecule from maize contains a 420-bp segment that is homologous with the chloroplast gene (*psbA*) that codes for the quinone binding protein of photosystem II. This is the first report of a chloroplast sequence in a naturally occurring viral-like or plasmid DNA. The complete sequence of the S-1 chloroplast segment has been compared with homologous regions of six different chloroplast genes. The S-1 segment has diverged from the other genes both by length mutation and base substitution. Several of the length mutations are exact adjacent tandem duplications of 4 and 5 bp similar to "footprints" left after excision of transposable elements in maize nuclear DNA.

MITOCHONDRIA from certain cytoplasmic genotypes of maize (*Zea mays* L.) and its relatives contain unusual double-stranded linear DNA molecules with proteins covalently attached to their 5′ ends (PRING *et al.* 1977; KEMBLE and THOMPSON 1982; WEISSINGER *et al.* 1982). In the S male-sterile cytoplasm, two of these molecules, designated S-1 and S-2, are 6.4 kb and 5.4 kb, respectively. These molecules resemble linear DNA viruses, such as adenovirus and Bacillus phage ϕ29, in having proteins attached to their ends, terminal inverted repeats and long, open reading frames (LEVINGS and SEDEROFF 1983; PAILLARD, SEDEROFF and LEVINGS 1985). S-1 and S-2 are at least fivefold more abundant than a unique sequence in the mitochondrial genome, suggesting that they are self-replicating molecules.

Segments of DNA, termed "promiscuous" DNA, are transferred between membrane-bound cellular compartments (STERN and LONSDALE 1982; TIMMIS and SCOTT 1983; FARRELLY and BUTOW 1983; JACOBS *et al.* 1983). Sequences of S-1 and S-2 DNA are found integrated into the high molecular weight fraction of the mitochondrial genome (LEVINGS *et al.* 1980; SPRUILL, LEVINGS

[1] To whom correspondence should be addressed.

and SEDEROFF 1980; LONSDALE, THOMPSON and HODGE 1981; LEVINGS *et al.* 1983; SCHARDL *et al.* 1984). S-1-containing sequences are also found in the nucleus (KEMBLE *et al.* 1983). In this paper we show that S-1 hybridizes to chloroplast DNA (ctDNA). The hybridization is due to a segment of a chloroplast gene (*psbA*) in the middle of the S-1 molecule. Although chloroplast sequences are widely found in plant mitochondrial DNA (mtDNA) (STERN and PALMER 1984), this is the first report of a chloroplast sequence in a mitochondrial plasmid DNA.

MATERIALS AND METHODS

MtDNA and ctDNA were isolated from maize seedlings as previously described (PRING and LEVINGS 1978; PALMER 1983). Restriction digests were analyzed on 0.7 or 2% agarose gels in TBE buffer (0.089 M Tris-borate, 0.089 M boric acid, 0.002 M EDTA), blotted onto nitrocellulose and hybridized with nick-translated probes (MANIATIS, FRISCH and SAMBROOK 1982).

Clone pKS6.2 (obtained from CSABA KONCZ), contains the entire S-1 molecule in pBR322. The chloroplast fragment *Bam*HI fragment 8 containing the maize *psbA* gene was used for this study in a Charon 30 vector. Nucleotide sequence analysis of S-1 was carried out using the dideoxynucleotide procedures of SANGER *et al.* (1982) and M13 vectors, as previously described (LEVINGS and SEDEROFF 1983). Dot matrix analysis was performed with a computer program obtained from M. EDGELL (University of North Carolina, Chapel Hill). Alignments of the S-1 and chloroplast nucleotide sequences were done using the alignment programs of the University of Wisconsin Genetics Computer Group (UWGCG). Optimal alignments were obtained with the GAP program that followed the method of NEEDLEMAN and WUNCH (1970). We used a gap weight of 5, a length weight of 0.3, and the option of weighted ends. Homology between sequences was calculated after assigning gaps to length mutations. The extent of divergence was estimated from the number of base substitutions in pairwise combinations of the different sequences. These measurements considered the aligned regions and excluded gaps due to length mutations. The numbers of base changes in pairwise combinations were then used to generate a dissimilarity matrix (see Table 2). The dissimilarity matrix was used for average linkage cluster analysis (weighted pairs) (see Figure 2) as described by SNEATH and SOKOL (1973). For comparison of the amino acid sequences, we used the universal code and aligned the S-1 sequence with the chloroplast sequence based on the alignment of the nucleotide sequences.

RESULTS

The S-1 plasmid, the normal (N) mitochondrial genome and the chloroplast genome contain a common sequence (Figure 1). The homologies were identifed by hybridizing a nick-translated 2.3-kb *Hin*dIII restriction fragment from an internal region of S-1 with Southern blots containing *Bam*HI digests of mtDNA and ctDNA from normal and cytoplasmic male-sterile (*cms*) lines of maize (Figure 2A). The variability of hybridization to the N, *cms*-S, and *cms*-T mtDNAs reflects differences in the relative stoichiometry and extent of the S-1 homologous sequences in the three mitochondrial genotypes (LEVINGS *et al.* 1980; SPRUILL, LEVINGS and SEDEROFF 1981; LEVINGS *et al.* 1983). The probe hybridizes to a 6.9-kb fragment in normal mtDNA that represents an integrated form of R-1 (LEVINGS *et al.* 1983), a linear DNA related to S-1 (Figure 1B). The faintly hybridizing 4.7-kb band in the N mtDNA probably results from chloroplast contamination.

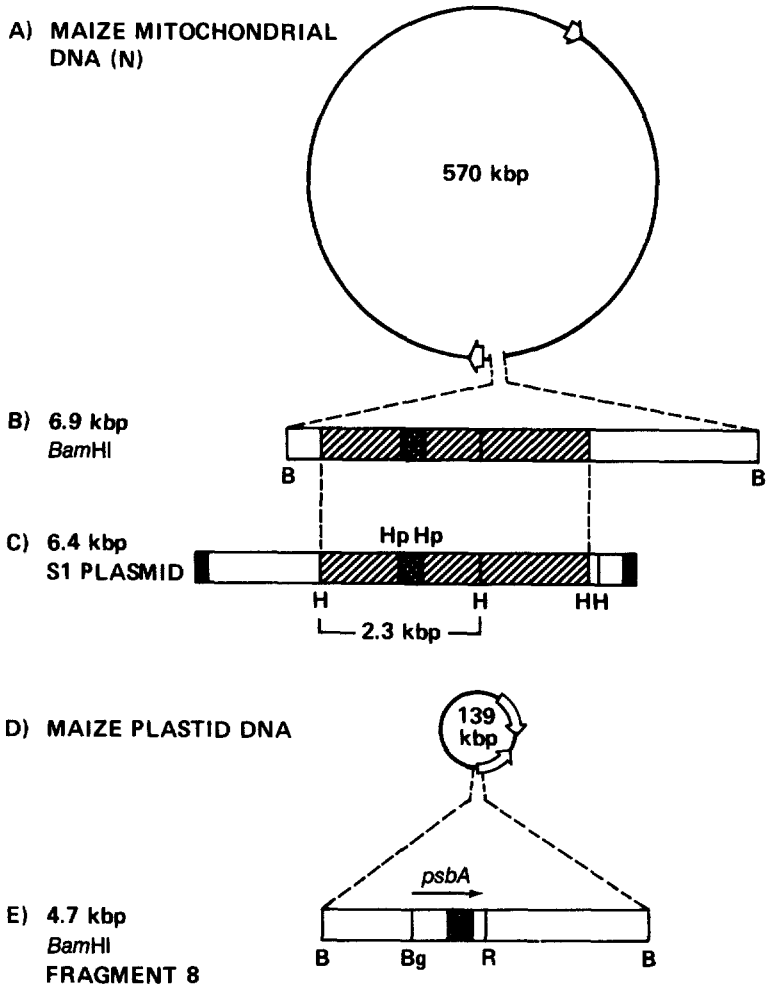

FIGURE 1.—Maps showing regions of homology between maize mitochondrial S-1 sequences and chloroplast *psbA* gene. A, Genomic map of normal (N) maize mtDNA showing the locations of the 3-kbp (kilobase pairs) direct repeats (23, 28). The open arrows indicate the direction of the repeats. B, N Mitochondrial 6.9-kbp *Bam*HI fragment containing regions homologous to the S-1 plasmid. C, Restriction endonuclease map of the plasmid S-1 (H = *Hin*dIII; B = *Bam*HI). Diagonal lines represent the region of homology shared between the S-1 plasmid and the 6.9-kbp *Bam*HI fragment. Black boxes represent the terminally inverted repeats of the S-1 plasmid. D, Genomic map of the maize chloroplast (plastid) genome showing the location of the inverted repeats (12). E, Restriction endonuclease map of the chloroplast *Bam*HI fragment 8 showing location of the *psbA* gene and direction of transcription (Bg = *Bg*lII, R = *Eco*RI, Hp = *Hpa*II; not all *EcoRI* and *Hpa*II sites are shown). Cross-hatched areas in B, C and E indicate the region of homology between *psbA* and S-1 sequences.

In *cms*-S mtDNA, a bright 6.4-kb band is observed representing hybridization to the intact S-1 DNA. Four other *Bam*HI fragments are detected at 10.4, 9.7, 4.8 and 4.5 kb. These fragments probably represent integrated S-1 sequences or sequences on the ends of the linear high molecular weight mitochondrial genome (SCHARDL *et al.* 1984). The mtDNA of *cms*-T shows a faintly

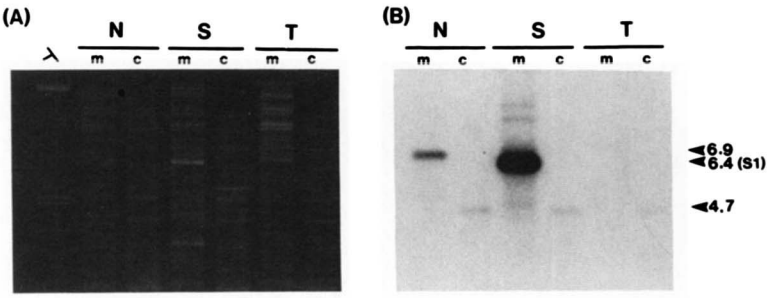

FIGURE 2.—Sequence homologies between a region of plasmid S-1 DNA and mtDNA and ctDNA from fertile and male-sterile cytoplasms of maize. A, Equimolar amounts of mitochondrial (m) and chloroplast (c) DNA were digested with *Bam*HI, electrophoresed on a 0.7% agarose gel and stained with ethidium bromide. The various cytoplasms are designated N (B37N), S (B37 *cms*-S) and T (B37 *cms*-T). B, Fragments were transferred to nitrocellulose and hybridized with a nick-translated 2.3-kb *Hin*dIII fragment internal to plasmid S-1 (Figure 1C). Fragment sizes, given in kilobases, were calculated from *Hin*dIII and *Eco*RI digests of phage lambda DNA.

hybridizing band of 7 kb. This fragment also hybridizes to the chloroplast fragment *Bam*HI fragment 8 (data not shown).

The S-1-derived 2.3-kb probe hybridizes to a single *Bam*HI fragment of 4.7 kb from each cytoplasmic genotype, to approximately the same extent (Figure 2B). This chloroplast fragment, *Bam*HI fragment 8, spans one of the junctions of the inverted repeat and the large single-copy region of the maize chloroplast genome (BEDBROOK and BOGORAD 1976; Figure 1D). No other regions of the S-1 or S-2 DNAs hybridized with ctDNA. The region of S-1 homology was mapped more precisely to a 1.2-kb *Eco*RI-*Bgl*II fragment within *Bam*HI fragment 8 (Figure 1E). This region contains the chloroplast gene *psbA*, which encodes photogene 32, a rapidly metabolized 32kd quinone binding thylakoid membrane protein that mediates electron transport between the quinones in photosystem II (RENGER 1976; MATTOO *et al.* 1981; VERMAAS *et al.* 1983).

To define the extent of S-1 homology with ctDNA, *Bam*HI fragment 8 was digested with *Hae*III and hybridized with labeled S-1. Hybridization was restricted to a 500-bp *Hae*III fragment contained in the *Eco*RI-*Bgl*II fragment (Figure 3A). To map the homologous region on S-1, labeled *Bam*HI fragment 8 was hybridized to *Hae*III and *Hpa*II digests of pKS6.2, which contains all of S-1 (Figure 3B). The chloroplast probe hybridized strongly to a 362-bp *Hpa*II fragment and a 1518-bp *Hae*III fragment. These results limit the homology to a small internal region of S-1 that does not include the terminal inverted repeats or the 1.3 kb of homology shared at one end of the S-1 and S-2 molecules (KIM *et al.* 1982).

A dot matrix comparison of the nucleotide sequence of the entire *psbA* gene from spinach (ZURAWSKI *et al.* 1982) and the S-1 DNA (PAILLARD, SEDEROFF and LEVINGS 1985) delineates the region of homology more precisely (Figure 4). Two segments with 90% homology are found, a short tract of 23 bp (S-1 position 3599–3621) and a longer sequence of 252 bp (S-1 position 3711–3962). These sequences correspond to positions 712–734 and 791–1042 in the spinach *psbA* gene. At the lower criterion of homology, 75%, one large

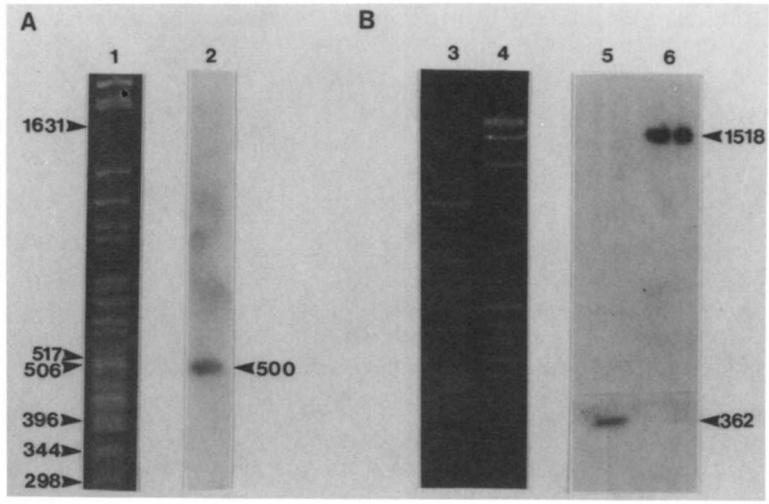

FIGURE 3.—Detection of homology between cloned chloroplast *Bam*HI fragment 8 and mitochondrial (mt) plasmid S-1 clone pKS6.2. A, Cloned *Bam*HI fragment 8 was digested with restriction enzyme *Hae*III (lane 1) and was probed with nick-translated clone pKS6.2 (lane 2). B, Clone pKS6.2 was digested with restriction enzymes *Hpa*II (lane 3) and *Hae*III (lane 4) and was probed with nick-translated cloned *Bam*HI fragment 8 (lanes 5 and 6). Samples were electrophoresed in 2% agarose gels. Fragment sizes, given in base pairs, were calculated from *Hin*fI digests of plasmid pBR322 DNA.

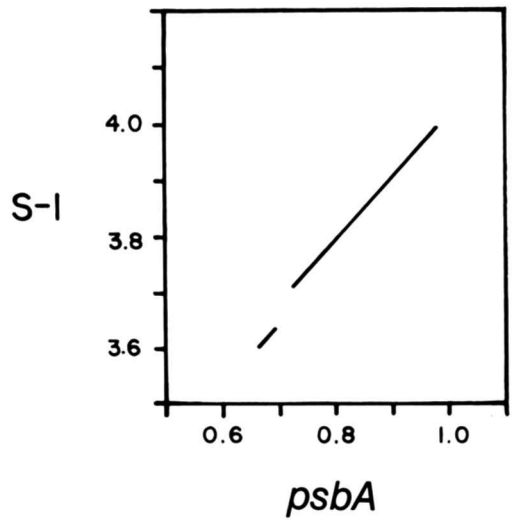

FIGURE 4.—Dot matrix plot of sequence similarity between S-1 and the *psbA* gene from spinach. Seven hundred base pairs of the S-1 sequence are shown on the ordinate plotted against the 600 bp of the spinach photogene sequence. The diagonal lines represent regions of homology ≥ 90%. Each interval represents 100 nucleotides.

region of 282 nucleotides and three small adjacent regions of 22, 27 and 35 nucleotides were observed (data not shown). These areas were located within positions 670–1067 of *psbA* and in 3563–3987 of S-1.

Homology between S-1 and the *psbA* gene is confined to the latter one-third of the coding region of *psbA*. After the stop codon TAA at position 1060–1062 of *psbA*, the homology between S-1 and *psbA* vanishes. This result is expected because the gene coding region of *psbA* is highly conserved among plant species, but flanking regions are not (HIRSCHBERG and MCINTOSH 1983; SPIELMANN and STUTZ 1983).

DISCUSSION

Sequences homologous to S-1 have been found in preparations of maize nuclear DNA by hybridization to Southern transfers (KEMBLE *et al.* 1983). The major hybridization signal to a 4.8-kb *Bam*HI fragment results from hybridization of S-1 with ctDNA contaminating the nuclear DNA preparations. The larger *Bam*HI fragments hybridizing to S-1 probably represent *bona fide* nuclear S-1-related sequences. Interestingly, another linear mitochondrial molecule, the 2.3-kb plasmid, also contains homology with the chloroplast genome (P. BEDINGER, unpublished results).

Five different genes coding for 32kd quinone binding proteins have been sequenced. Four are from the dicotyledons, spinach (*Spinacea oleracea*) (ZURAWSKI *et al.* 1982), pigweed (*Amaranthus hybridus*) (HIRSCHBERG and MCINTOSH 1983), *Nicotiana debneyi* (ZURAWSKI *et al.* 1982), and soybean (*Glycine max*) (SPIELMAN and STUTZ 1983). The fifth gene is from a green alga (*Chlamydomonas reinhardtii*) (ERICKSON, RAHIRE and ROCHAIX 1984). In addition to these functional genes, a "32kd"-like sequence has been found in spinach chloroplast DNA; this is thought to be a related gene associated with the photosystem II assembly (ALT *et al.* 1984). We have compared these sequences with the S-1 segment to estimate the extent of divergence from these chloroplast sequences and to learn about the mechanisms responsible for sequence changes (Figure 5).

The sequence of the S-1 segment can be aligned with all the chloroplast sequences for most of its length; however, significant discontinuities occur at the 5′ end of the segment. These discontinuities are a consequence of length mutations not found in dicot or algal sequenes. We have used the GAP program for alignment of sequences to investigate the number and extent of length mutations that may have occurred in the S-1 segment. The alignments show four gaps that can account for the length variation (Figure 5). Alignment of the sequences without regard to gap penalties gives similar results, except for one region where a more complex alignment could be considered (see legend to Figure 5). Most of the length changes appear to result from insertions of nucleotides (Table 1). Only one deletion of 4 bp is indicated by the computer-based alignment.

Closer examination of the sequences at the site of the gaps indicates that insertions are predominantly duplications of adjacent sequences. For example, the insertion of GAATC at positions 122–126 in Figure 5 is adjacent to the identical sequence at 117–121. The gap at position 145–153 is created by two duplications of adjacent sequences. The first five bases GCTAG at positions 145–149 are an exact duplication of the previous five aligned bases at positions

140–144. The remaining four bases in the gap are a duplication of the four adjacent nucleotides at the right end of the gap; that is, TTTC occurs at positions 150–153 in the gap and at positions 154–157 immediately adjacent to the gap. The insertion of four bases at positions 127–130 has the sequence TTCC and is adjacent to the sequence ATCC. A large insertion of 14 bases at positions 84–97 also shows evidence for adjacent duplication. Most of the insertion can be accounted for by an exact triplication of the sequence CACGG. Superimposed on this triplication is a further repetition of CCCC at positions 92–95.

Short direct repeats leading to length polymorphisms are common in organellar genomes. In a study of length mutation in human mtDNA, CANN and WILSON (1983) found 14 direct repeats in regions of length polymorphism. Some of the repeats in human mtDNA resemble the adjacent direct repeats observed in the S-1 segment. Short, adjacent direct repeats associated with length mutations have also been found in coding and flanking (noncoding) regions of the mitochondrial 18S and 5S ribosomal genes when maize, wheat and *Zea diploperennis* are compared (B. GWYNN *et al.*, unpublished results). Similar short repeats have been observed in comparisons of the upstream noncoding regions of the *atpB* and *rbcL* genes in barley and maize chloroplast DNA (ZURAWSKI, CLEGG and BROWN 1984). Moreover, comparison of the spacer region of the chloroplast rRNA genes in tobacco and maize also shows short adjacent direct repeats (TAKAIWA and SUGIURA 1982). Further examples of this phenomenon occur in nuclear genes. Noncoding regions of the β-hemoglobin gene cluster show short direct repeats associated with length mutation (EFSTRATIADIS *et al.* 1980).

STREISINGER *et al.* (1966) first showed that frameshift mutations could result from adjacent repeats of specific nucleotides. Short, adjacent direct repeats may arise by mutation during repair or replication. Misalignments of multiple direct repeats can product duplications or deletions. Misalignment mutagenesis is well charaterized in bacteria in which short direct repeats can cause mutational hot spots (FARABAUGH *et al.* 1978; DRAKE, GLICKMAN and RIPLEY 1983).

The short tandem repeats found in the S-1 *psbA* segment also resemble repeats generated in maize by transposable elements (SAEDLER and NEVERS 1985; DORING and STARLINGER 1984). Integration of *Mu*, *Spm* and *Ac* are known to generate a short repeat of the "target" site at the ends of the inserted element. After excision of the element, a short tandem duplication usually remains at the site of the insertion-excision event. The resulting duplications are often inexact. SAEDLER and NEVERS (1985) have proposed that the variations result from errors in repair or from template switching that may accompany the excision process.

In one of the S-1 segment repeats, the expected repeated sequence at position 131–134 is TTCC, because all the dicot genes contain the adjacent TTCC sequence; however, the duplication is TTCCATCC. The A at position 131 in the sequence may be a consequence of excision-dependent variation. Alterna-

```
          1                                                          60
S-1       CAGGGAAACC ggtGAAAAat gAa....gAA TctgGGTTcg ccATTCGGTc AAGAGGAAGA
SPINACH   CAGGGAAACC ACAGAAAATG AATCTGCTAA TGAAGGTTAC AGATTCGGTG AAGAGGAAGA
SOYBEAN   CAGGGAAACC ACAGAAAATG AATCTGCTAA TGAAGGTTAC AGATTtGGTG AAGAGGAAGA
PIGWEED   CAGGGAAACC ACAGAAAATG AATCTGCTAA cGAAGGTTAC AGATTCGGTc AAGAGGAAGA
DEBNEYI   CAGGGAAACC ACAGAAAATG AATCTGCTAA TGAAGGTTAC AGATTCGGTG AAGAGGAAGA
CHLAMYD   CcGtGAAACa ACtGAAAAcG AATCaGCTAA cGAAGGTTAC cGtTTCGGTc AAGAaGAAGA
32KDLIKE  gctattcAtg gtgctAccgt tgaaaatact TtAtttgaAg AtggTgatgG ggcAaatAcA
Consensus cagggaaacc acagaaaatg aatctgctaa tgaaggttac agattcggtg aagaggaaga

          61                                                         120
S-1       gtCaagTAgT AggTGTgGCT GCTcacggca ccccccggCAc GGTTcTTTaG GTCGATgaAT
SPINACH   AACtTATAAT A.TcGTAGCT GCT.......  .......CAT GGTTATTTTG GTCGATTGAT
SOYBEAN   AACcTATAAT A.TTGTAGCT GCT.......  .......CAT GGTTATTTTG GcCGATTGAT
PIGWEED   AACtTATAAc A.TcGTAGCT GCT.......  .......CAT GGTTATTTTG GTCGATTGAT
DEBNEYI   AACcTATAAc A.TcGTAGCc GCT.......  .......CAT GGTTATTTTG GcCGATTGAT
CHLAMYD   AACtTAcAAc A.TTGTAGCT GCT.......  .......CAT GGTTAcTTTG GTCGtcTaAT
32KDLIKE  ttCcgggctT ttaacccaac cCaagctgaa gaaacttatT caaTggTcac cgCtAaccgc
Consensus aac-tataat a.ttgtagct gct.......  .......cat ggttattttg gtcgattgat

          121                                                        180
S-1       CgaatcttcC aTCCAATAcG CTAGgctagt ttcTTTCcAC AACTCTCGTT CTTTACACga
SPINACH   C.........  TTCCAATATG CTAG......  ...TTTCAAC AACTCTCGTT CTTTACACTT
SOYBEAN   C.........  TTCCAATATG CaAG......  ...TTTCAAC AAtTCTCGTT CTTTACAtTT
PIGWEED   C.........  TTCCAATATG CTAG......  ...TTTCAAC AACTCTCGTT CTTTACACTT
DEBNEYI   C.........  TTCCAATATG CTAG......  ...TTTCAAC AACTCTCGTT CgTTACACTT
CHLAMYD   C.........  TTCCAATAcG C.........  ttcTTTCAAC AACTCTCGTT CaTTACACTT
32KDLIKE  ttttggtccC aaatctTtgG g...gttgct tttTccaAta AACgtTgGTT acaTttctTT
Consensus c.........  ttccaatatg ctag......  ...tttcaac aactctcgtt ctttacactt

          181                                                        240
S-1       CTgaTTgGCT GCTTGGCCTG TAGTAGGgAT cTGGaTcACT GCTTTAGGTA TTAGtACTAT
SPINACH   CTTCTTAGCT GCTTGGCCTG TAGTAGGTAT TTGGTTTACT GCTTaAGGTA TTAGtACTAT
SOYBEAN   CTTCTTAGCT GCTTGGCCTG TAGTAGGTAT TTGGTTTACc GCTTTAGGTA TcAGcACTAT
PIGWEED   CTTCTTAGCT GCTTGGCCgG TAaTcGGTAT TTGGTTTACT GCTTTgGGTA TTAGtACTAT
DEBNEYI   CTTCcTAGCT GCTTGGCCTG TAGTAGGTAT cTGGTTTACc GCTTTAGGTA TcAGcACTAT
CHLAMYD   CTTCTTAGCT GCTTGGCCgG TAaTcGGTAT TTGGTTcACT GCTTTAGGTt TatcaACTAT
32KDLIKE  aTgtTatttg taccaGtaac cgGTttaTgg aTGagTgctc ttggagtagt cggtctggcT
Consensus cttcttagct gcttggcctg tagtaggtat ttggtttact gctttaggta ttag-actat

          241                                                        300
S-1       GGCaTTCAAC cTAAATGGTT TCAATTTCAA CCAATCTGTA GTTGATAGcC AAGGTCGcGT
SPINACH   GGCTTTCAAC tTAAATGGTT TCAATTTCAA CCAATCTGTA GTTGATAGTC AAGGTCGTGT
SOYBEAN   GGCTTTCAAC tTAAATGGTT TCAATTTCAA CCAATCcGTA GTTGATAGTC AAGGTCGTGT
PIGWEED   GGCTTTCAAC cTAAAcGGTT TCAAcTTCAA CCAATCTGTA GTTGATAGTC AAGGTCGTGT
DEBNEYI   GGCTTTCAAC cTAAATGGTT TCAATTTCAA CCAATCTGTA GTTGAcAGTC AAGGcCGTGT
CHLAMYD   GGCaTTCAAC tTAAAcGGTT TCAAcTTCAA CCAATCaGTA GTaGActcaC AAGGTCGTGT
32KDLIKE  ttgaacCtAC gTgccTatga cttcgTTtcc Caggaaatcc GTgcAgctga AgatcCtgaa
Consensus ggctttcaac -taaatggtt tcaatttcaa ccaatctgta gttgatagtc aaggtcgtgt

          301                                                        360
S-1       tATTAATACT TGGGCTGATA TCATcAACCG TGCTAAtCTT GGTATGGAgG TAATGCAcGA
SPINACH   AATTAATACT TGGGCTGATA TCATTAACCG TGCTAACCTT GGTATGGAAG TAATGCATGA
SOYBEAN   AATTAATACc TGGGCTGATA TtATTAACCG aGCTAACCTT GGTATGGAAG TAATGCATGA
PIGWEED   AATTAAcACc TGGGCTGATA TCATTAACCG TGCTAACCTT GGTATGGAAG TtATGCATGA
DEBNEYI   AATTAATACT TGGGCTGATA TtATTAACCG TGCTAACCTT GGTATGGAAG TtATGCATGA
CHLAMYD   AcTaAAcACT TGGGCaGAcA TCATcAACCG TGCTAACtTa GGTATGGAAG TAATGCAcGA
32KDLIKE  ttTgAAactT TttaCaccaA aaATattCtc TtaaAcgagg GtatccGtgc Ttggatggcg
Consensus aattaatact tgggctgata tcattaaccg tgctaacctt ggtatggaag taatgcatga
```

```
            361                                                        420
S-1         ACGTAATGCT CACAACTTCC CTCTAGACCT AGCaGCTcTt GAAGtTCCAT CTcttAATGG
SPINACH     ACGTAATGCT CAtAACTTCC CTCTAGACCT AGCTGCTATt GtAGCTCCAT CTACAAATGG
SOYBEAN     ACGTAATGCT CAtAAtTTCC CTCTAGAtCT AGCTGCgATc GAcGCTCCAT CTAttAATGG
PIGWEED     ACGTAATGCT CAtAACTTCC CTCTAGACtT AGCTGCTATc GAAGCTCCAT CTACAAATGG
DEBNEYI     ACGTAATGCT CACAACTTCC CTCTAGACCT AGCTGCTATc GAAGCTCCAT CTACAAATGG
CHLAMYD     gCGTAAcGCT CACAACTTCC CTCTAGACtT AGCTtCaAct aActCTagcT C...AAAca
32KDLIKE    gCtcAAga.. .tCAgCcTCa tgaaAacCtT AtaTtCccTg aggaggttcT accCcgtgGa
Consensus   acgtaatgct cacaacttcc ctctagacct agctgctat- gaagctccat ctacaaatgg

            421         432
S-1         ATAA..... ..
SPINACH     ATAA..... ..
SOYBEAN     ATAA..... ..
PIGWEED     ATAA..... ..
DEBNEYI     ATAA..... ..
CHLAMYD     cTAA..... ..
32KDLIKE    Aacgctcttt aa
Consensus   ataa..... ..
```

FIGURE 5.—Nucleotide sequence comparisons of the S-1 segment and six chloroplast sequences. Sequences have been aligned for the S-1 segment and the homologous regions of the *psbA* photogene from spinach, soybean, pigweed, *N. Debneyi, C. reinhardtii* and the 32kd-like gene from spinach. The alignment was established by the GAP program using the default values given in MATERIALS AND METHODS. An alternative alignment is of interest for nucleotides at positions 19–38.

```
S-1             AAAATGAAGAATCTG--------GGTT
ct consensus    AAA-TGAA---TCTGCTAATGAAGGTT
```

tively, the mutation may have occurred independently after the duplication. If it occurred before the duplication event, the expected sequence would be ATCCATCC. More complex events may take place. For example, two regions in the S-1 segment appear to result from adjacent double duplications, *i.e.*, region 117–134 and region 140–157. Perhaps multiple insertion-duplication events have taken place in a single transposition process.

With a reasonable alignment of the S-1 sequence and the chloroplast sequences, excluding length mutation, it is possible to estimate that the extent of nucleotide substitution that has taken place in the evolution of these sequences. We have made pairwise comparisons of the five *psbA* sequences, the 32kd-like sequence and the S-1 segment. The substitutional mutations in the comparisons of S-1 with the chloroplast sequences show a low level of transitional bias. For example, equal numbers of transitions and transversions occur between S-1 and pigweed or between S-1 and *N. debneyi*; however, between pigweed and *N. debneyi* there is a fivefold bias in favor of transitions. Similarly, a high level of transitional bias is found in the comparisons between the other dicots. Transitional bias is known to decrease over evolutionary time (BROWN *et al.* 1982). In addition, the increase in transversions may reflect a change or loss of function of the S-1 segment in the mitochondrion. The total number of nucleotide differences between all the pairs of sequences, excluding gaps resulting from length mutations, are shown as a dissimilarity matrix (Table 2) and as an average linkage cluster analysis with weighted pair-groups (SNEATH and SOKOL 1973) (Figure 6).

TABLE 1

Length mutations in the S-1 chloroplast segment

Position	Sequence	Type of change
24–27	CTGC	Deletion of four bases (flanked by GAA on both sides)
72	G	Insertion of one base (duplication of G)
84–97	CACGG	Adjacent direct triplication of five bases includes insertion 92–95 (see text)
92–95	CCCC	Insertion of four bases (within 84–97)
122–126	GAATC	Adjacent direct duplication of five bases
127–130	TTCC	Insertion of four bases
145–149	GCTAG	Adjacent direct duplication of five bases
150–153	TTTC	Adjacent direct duplication of four bases
Alternative alignment: for positions 19–38		
19	A	Insertion of one base
24–26	GAA	Adjacent duplication of three bases
31–38	GTAATGAA	Deletion of eight bases

Length mutations are listed for S-1 sequence using the alignments presented in Figure 5. Numerical positions correspond to the numbers in Figure 5.

The S-1 nucleotide sequence has greater similarity to the sequences of the dicots than the *psbA* sequence from Chlamydomonas has with dicots. This implies that the sequence in S-1 was derived from a higher plant. The S-1 sequence may well be derived from the chloroplast of maize or a close relative, but this possibility cannot be evaluated until the sequence of the 32kd gene from maize chloroplast becomes available.

We have examined the nucleotide sequence of the S-1-*psbA* segment and the distribution of mutations for clues concerning its possible expression in mitochondria (Figure 5). In the left end of the segment, length mutations cause shifts in the normal reading frame of the *psbA* gene at positions 24–27 (or 19–38 for the alternative alignment), 72 and 84–102. The remaining insertions do not affect the reading frame. Two potential in-frame TGA stop codons (FOX and LEAVER 1981; HIESEL and BRENNICKE 1983) appear in the sequence at positions 116 and 182. In the rightmost two-thirds of the segment, 94% homology exists between the deduced amino acid sequence of S-1 and *psbA*. A possible downstream translation initiation codon within the same reading frame, at position 239 (Figure 5) could initiate the production of a polypeptide of 62 amino acids. This polypeptide would correspond to a distinct domain of the photogene 32 protein that is hypothesized to extend out of the thylakoid membrane (RAO, HARGRAVE and ARGOS 1983). The right half of the segment has no length mutations and a low frequency of amino acid substitutions. One explanation of this distribution would be that the right half of the molecule serves some function. An alternative explanation would require polarity of mutation during or after integration of the segment.

Segments of promiscuous DNA with high levels of similarity to known genes may identify segments of DNA that have recently entered the mitochondria or the nucleus and may, thereby, provide a means to study the changes in

TABLE 2

Dissimilarity matrix of nucleotide substitutions between S-1 and ctDNA sequences

	Soy.	Spin.	Deb.	Pig.	Chlamy	32kd-like
S-1	71	59	66	66	93	240
Soybean	—	22	25	32	76	194
Spinach	—	—	15	17	64	185
Debneyi	—	—	—	24	69	186
Pigweed	—	—	—	—	55	181
Chlamy	—	—	—	—	—	185

Values presented represent the number of nucleotide substitutions between pairs of sequences, corrected for gaps (length mutations), so that the length of sequences compared is always the same (392 bp).

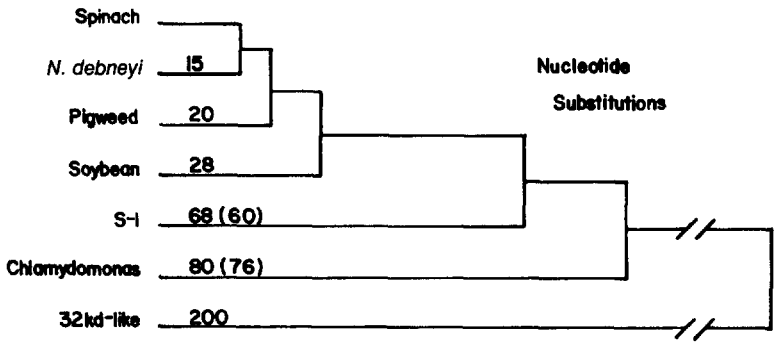

FIGURE 6.—Average linkage clusters are shown for the nucleotide sequences from the dissimilarity matrix (Table 2). All sequences were corrected for length mutation by eliminating gaps before calculating numbers of substitutions. Numbers in parentheses represent the results obtained from the alternative alignment in Figure 5.

sequence associated with transfer, integration and divergence of nonfunctional DNA. SCHWARZ-SOMMER *et al.* (1985) have argued that the sequence diversity generated by the visitation of a transposable element is an important force in evolution of new genetic functions, because small duplications may alter protein sequences in important new ways. Such duplications would have additional consequences. Small duplications would increase the size of functional genes when function was not impaired and, more frequently, would create small duplications in nonfunctional DNA. Consequently, they would increase the fraction of DNA that is classified as single-copy sequence. Mechanisms that increase DNA by duplication are ordinarily thought to contribute to the repeated fraction of DNA; however, this is not the case if the segments are less than 20 nucleotides. In the S-1 chloroplast segment, we suggest that an increase of 7.1% (28 of 392) in single-copy sequence has occurred by insertional duplication since the segment was part of a normal 32kd photogene, and we offer the speculation that increases in single-copy sequences in some eukaryotic genomes may result from activity of transposable elements.

We are intrigued by the presence of a chloroplast DNA segment in a mi-

tochondrial plasmid DNA. The linear plasmid DNAs have several features in common with certain bacterial and animal viruses (LEVINGS and SEDEROFF 1983; PAILLARD, SEDEROFF and LEVINGS 1985), suggesting an exogenous origin. No linear plasmid DNAs have been associated with chloroplasts; however, S-1 sequences have been found in the nucleus (KEMBLE *et al.* 1983). Where and how the plasmid DNAs acquired the chloroplast sequences remains a mystery.

We thank DAVID STERN and CHARLES ARNTZEN for helpful discussions and CSABA KONCZ for clone pKS6.2. Supported in part by National Institutes of Health grant GM29775 (V.W.), National Science Foundation fellowship PCM8312563 (P.B.), United States Department of Agriculture/ Science and Education grant 5901-0401-03560 (C.S.L. and R.S.) and National Science Foundation grant PCM-8010933 (C.S.L.).

LITERATURE CITED

ALT, J., J. MORRIS, P. WESTHOFF and R. G. HERRMANN, 1984 Nucleotide sequence of the clustered genes for the 44kd chlorophyll apoprotein and the "32-kd"-like protein of the photosystem II reaction center in the spinach plastid chromosome. Curr. Genet. **8:** 597–606.

BEDBROOK, J. R. and L. BOGORAD, 1976 Endonuclease recognition sites mapped on *Zea mays* chloroplast DNA. Proc. Natl. Acad. Sci. USA **73:** 4309–4313.

BROWN, W. M., E. M. PRAGER, A. WANG and A. C. WILSON, 1982 Mitochondrial DNA sequences of primates: tempo and mode of evolution. J. Mol. Evol. **18:** 225–239.

CANN, R. L. and A. D. WILSON, 1983 Length mutations in human mitochondrial DNA. Genetics **104:** 699–711.

DORING, H.-P. and P. STARLINGER, 1984 Barbara McClintock's controlling elements: now at the DNA level. Cell **39:** 253–259.

DRAKE, J. W., B. W. GLICKMAN and L. S. RIPLEY, 1983 Updating the theory of mutation. Am. Sci. **71:** 621–630.

EFSTRATIADIS, A., J. W. POSAKANY, T. MANIATIS, R. M. LAWN, C. O'CONNELL, R. A. SPRITZ, J. K. DERIEL, B. G. FOREST, S. M. WEISSMAN, J. L. SLIGHTOM, A. E. BLECHL, O. SMITHIES, F. E. BARALLE, C. C. SHOLDERS and N. J. PROUDFOOT, 1980 The structure and evolution of the human beta-globin family. Cell **21:** 653–668.

ERICKSON, J. M., M. RAHIRE and J. D. ROCHAIX, 1984 *Chlamydomonas reinhardii* gene for the 32,000 mol. wt. protein of photosystem II contains four large introns and is located entirely within the chloroplast inverted repeat. EMBO J. **3:** 2753–2762.

FARABAUGH, P. J., U. SCHMEISSNER, M. HOFER and J. H. MILLER, 1978 Genetic studies of the lac repressor. VII. On the molecular nature of spontaneous hotspots in the lacI gene of *Escherichia coli.* J. Mol. Biol. **126:** 847–857.

FARRELLY, F. and R. A. BUTOW, 1983 Rearranged mitochondrial genes in the yeast nuclear genome. Nature **301:** 296–301.

FOX, T. D. and C. J. LEAVER, 1981 The *Zea mays* mitochondrial gene coding cytochrome oxidase subunit II has an intervening sequence and does not contain TGA codons. Cell **26:** 315–323.

HIESEL, R. and A. BRENNICKE, 1983 Cytochrome oxidase subunit II gene in mitochondria of *Oenothera* has no intron. EMBO J. **2:** 2173–2178.

HIRSCHBERG, J. and L. MCINTOSH, 1983 Molecular basis of herbicide resistance in *Amaranthus hybridus.* Science **222:** 1346–1348.

JACOBS. H. T., J. W. POSAKONY, J. A. GRULA, J. W. ROBERTS, J. XIN, R. J. BRITTEN and E. H.

DAVIDSON, 1983 Mitochondrial DNA sequences in the nuclear genome of *Stronglylocentrogus purpuratus*. J. Mol. Biol. **165:** 609–632.

KEMBLE, R. J., R. J. MANS, S. GABAY-LAUGHNAN and J. R. LAUGHNAN, 1983 Sequences homologous to episomal mitochondrial DNAs in the maize nuclear genome. Nature **304:** 744–747.

KEMBLE, R. J. and R. D. THOMPSON, 1982 S1 and S2, the linear mitochondrial DNAs present in a male-sterile line of maize, possess terminally attached proteins. Nucleic Acids Res. **10:** 8181–8190.

KIM, B. D., R. J. MANS, M. E. CONDE, D. R. PRING and C. S. LEVINGS III, 1982 Physical mapping of homologous segments of mitochondrial episomes from S male-sterile maize. Plasmid **7:** 1.

LEVINGS, C. S. III, B. D. KIM, D. L. PRING, M. F. CONDE, R. J. MANS and S. J. GABAY-LAUGHNAN, 1980 Cytoplasmic reversion of cms-S in maize: association with a transpositional event. Science **209:** 1021–1023.

LEVINGS, C. S. III and R. R. SEDEROFF, 1983 Nucleotide sequence of the S-2 mitochondrial DNA from the S cytoplasm of maize. Proc. Natl. Acad. Sci. USA **80:** 4055–4059.

LEVINGS, C. S. III, R. R. SEDEROFF, W. W. L. HU and D. H. TIMOTHY, 1983 Relationships among plasmid-like DNAs of the maize mitochondria. pp. 363–371. In: *Structure and Function of Plant Genomes*, Edited by O. CIFERRI and L. DURE III. Plenum, New York.

LONSDALE, D. M., R. D. THOMPSON and T. P. HODGE, 1981 The integrated forms of the S-1 and S-2 DNA elements of maize male-sterile mitochondrial DNA are flanked by a large repeated sequence. Nucleic Acids Res. **9:** 3657–3669.

MANIATIS, T., E. F. FRITSCH and J. SAMBROOK, 1982 *Molecular Cloning*. Cold Spring Habor Laboratory, Cold Spring Harbor, New York.

MATTOO, A. K., U. PICK, H. HOFFMAN-FALK and M. EDELMAN, 1981 The rapidly metabolized 32,000-dalton polypeptide of the chloroplast is the "proteinaceous shield" relating photosystem II electron transport and mediating diuron herbicide sensitivity. Proc. Natl. Acad. Sci. USA **78:** 1572–1576.

NEEDLEMAN, S. B. and C. D. WUNCH, 1970 A general method applicable to the search for similarities in the amino acid sequence of two proteins. J. Mol. Biol. **48:** 443–452.

PAILLARD, M., R. R. SEDEROFF and C. S. LEVINGS III, 1985 Nucleotide sequence of the S-1 mitochondrial DNA from the S cytoplasm of maize. EMBO J. **4:** 1125–1128.

PALMER, J., 1983 Chloroplast DNA exists in two orientations. Nature **301:** 92–93.

PRING, D. R. and C. S. LEVINGS III, 1978 Heterogeneity of maize cytoplasmic genomes among male-sterile cytoplasms. Genetics **89:** 121–136.

PRING, D. R., C. S. LEVINGS III, W. W. L. HU and D. H. TIMOTHY, 1977 Unique DNA associated with mitochondria in the "S" type cytoplasm of male-sterile maize. Proc. Natl. Acad. Sci. USA **74:** 2904–2908.

RAO, J. K. M., P. A. HARGRAVE and P. ARGOS, 1983 Will the seven-helix bundle be a common structure for integral membrane proteins? FEBS Lett. **156:** 165–169.

RENGER, G., 1976 Studies on the structural and functional organization of system II of photosynthesis. The use of trypsin as a structurally selective inhibitor at the outer surface of the thylakoid membrane. Biochem. Biophys. Acta **440:** 287–300.

SAEDLER, H. and P. NEVERS, 1985 Transposition in plants: a molecular model. EMBO J. **4:** 585–590.

SANGER, A. R., A. R. COULSON, G. F. HONG, D. F. HILL and G. B. PETERSON, 1982 Nucleotide sequence of bacteriophage lambda DNA. J. Mol. Biol. **162:** 729–773.

SCHARDL, C. L., D. M. LONSDALE, D. R. PRING and K. R. ROSE, 1984 Linearization of maize mitochondrial chromosomes by recombination with linear episomes. Nature **310:** 292–296.

SCHWARZ-SOMMER, Z., A. GIERL, H. CUYPERS, P. A. PETERSON and H. SAEDLER, 1985 Plant transposable elements generate the DNA sequence diversity needed in evolution. EMBO J. **4:** 591–597.

SNEATH, H. A. and R. R. SOKOL, 1973 *Numerical Taxonomy.* pp. 228–229. W. H. Freeman, San Francisco.

SPEILMANN, A. and E. STUTZ, 1983 Nucleotide sequence of soybean chloroplast DNA regions which contain the *psb*A and *trn*H genes and cover the ends of the large single copy region and one end of the inverted repeats. Nucleic Acids Res. **11:** 7157–7167.

SPRUILL, W. M., JR., C. S. LEVINGS III and R. R. SEDEROFF, 1980 Recombinant DNA analysis indicates that the multiple chromosomes of maize mitochondria contain different sequences. Dev. Genet. **1:** 363–378.

SPRUILL, W. M., JR., C. S. LEVINGS III and R. R. SEDEROFF, 1981 Organization of mitochondrial DNA in normal and Texas male-sterile cytoplasms of maize. Dev. Genet. **2:** 319–336.

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

STERN, D. B. and J. D. PALMER, 1984 Extensive and widespread homologies between mitochondrial DNA and chloroplast DNA in plants. Proc. Natl. Acad. Sci. USA **81:** 1946–1950.

STREISINGER, G., Y. OKADA, J. EMRICH, J. NEWTON, A. TSUGITA, E. TERZAGHI and M. INOUYE, 1966 Frameshift mutations and the genetic code. Cold Spring Harbor Symp. Quant. Biol. **31:** 77–84.

TAKAIWA, F. and M. SUGIURA, 1982 Nucleotide sequence of the 16S-23S spacer region in an rRNA gene cluster from tobacco chloroplast DNA. Nucleic Acids Res. **8:** 2665–2676.

TIMMIS, J. N. and N. S. SCOTT, 1983 Sequence homology between spinach nuclear and chloroplast genomes. Nature **305:** 65–67.

VERMAAS, W. F. J., C. F. ARNTZEN, L-Q GU and G-A YU, 1983 Interactions of herbicides and azidoquinones at a photosystem II binding site in the thylakoid membrane. Biochem. Biophys. Acta **723:** 266–275.

WEISSINGER, A., D. H. TIMOTHY, C. S. LEVINGS III, W. W. L. HU and M. M. GOODMAN, 1982 Unique plasmid-like mitochondrial DNAs from indigenous maize races of Latin America. Proc. Natl. Acad. Sci. USA **79:** 1–5.

ZURAWSKI, G., H. J. BOHNERT, P. R. WHITFELD and W. BOTTOMLEY, 1982 Nucleotide sequence of the gene for the 32,000-M_r thylakoid membrane protein from *Spinacia oleracea* and *Nicotiana debney*: predicts a totally conserved primary translation product of M_r 38,950. Proc. Natl. Acac. Sci. USA **79:** 7699–7703.

ZURAWSKI, G., M. T. CLEGG and A. H. D. BROWN, 1984 The nature of nucleotide sequence divergence between barley and maize chloroplast DNA. Genetics **106:** 735–749.

Communicating editor: M. R. HANSON