# Nucleotide sequence of the S-1 mitochondrial DNA from the S cytoplasm of maize

# Michèle Paillard, Ronald R.Sederoff and Charles S.Levings,III

Department of Genetics, North Carolina State University, Raleigh, NC 27695-7614, USA

Communicated by C.J.Leaver

Mitochondria from the S male-sterile cytoplasm of maize contain unique DNA-protein complexes, designated S-1 and S-2. These complexes consist of double-stranded linear DNAs with proteins covalently attached to the 5' termini. To learn more about these unusual DNAs we have determined the complete nucleotide sequence of the S-1 DNA molecule (6397 bp). The sequence of S-2 has been previously determined. S-1 and S-2 are structurally similar and contain  $\sim 1.7$  kb of sequence homology. S-1 is terminated by exact 208-bp inverted repeats that are identical with the terminal inverted repeats of S-2. S-1 and S-2 also contain a 1462-bp region of nearly perfect homology, which includes one of the terminal inverted repeats. The homology between the two molecules may be maintained, in part, by homologous recombination. S-1 has three long unidentified open reading frames, URF2 (1017 bp), URF3 (2787 bp) and URF4 (768 bp). URF2 occurs in the 1462-bp region of homology and is identical in length and location in both S-1 and S-2. Based on their structural organization and their viral-like characteristics, we propose that S-1 and S-2 code for functions involved with their maintenance and replication.

Key words: plasmid-like DNAs/terminal inverted repeats/cytoplasmic male sterility/nucleotide sequence

## Introduction

Mitochondria of cms-S, a male-sterile cytoplasm of maize (Zea mays L.) contain plasmid-like DNAs in addition to and distinct from the usual mitochondrial DNAs (mtDNA) (Pring et al., 1977). Two of these unique DNAs, called S-1 and S-2 are associated with the S group of cytoplasmic male steriles. The S group, which contains >20 members, is restored to pollen fertility by the nuclear gene, Rf3, located on chromosome 2 (Laughnan et al., 1981). Customarily, S-1 and S-2 are isolated as double-stranded linear molecules with defined ends and are 6.4 and 5.4 kb long, respectively. The molecules are terminated by 0.2-kb inverted repeats (Levings and Pring, 1979; Levings and Sederoff, 1983) and have proteins covalently attached to their 5' ends that may be involved in priming DNA replication (Kemble and Thompson, 1982). Similar DNA-protein associations are also found in adenovirus and Bacillus phages where it is believed that these proteins prime for DNA synthesis (Rekosh et al., 1977; Carusi, 1977; Salas et al., 1978; Harding et al., 1978; Ito, 1978; Yehle, 1978; Yoshikawa and Ito, 1981). Interestingly, several of the Bacillus phages contain short terminal inverted repeats that have a high degree of homology with the terminal nucleotides of S-2 (Levings and Sederoff, 1983).

In most *cms-S* lines, S-1 and S-2 are present in equimolar amounts and are  $\sim$  5-fold more abundant than a typical unique © IRL Press Limited, Oxford, England.

segment of the mtDNA. The 5-fold difference suggests that mtDNA and plasmid-like DNAs replicate independently. It has been shown that nuclear background can influence the content and relative proportions of S-1 and S-2 (Laughnan *et al.*, 1981). Several studies have shown that mtDNA sequences from normal (fertile) and male-sterile cytoplasms share homology with S-1 and S-2 sequences (Spruill *et al.*, 1980, 1981; Lonsdale *et al.*, 1981). Moreover, sequences homologous to S-1 are reported to occur in the maize nuclear genome (Kemble *et al.*, 1983).

Unlike the other male-sterile cytoplasms of maize, *cms-S* is unstable and frequently mutates to the male-fertile phenotype (Laughnan *et al.*, 1981; Laughnan and Gabay-Laughnan, 1983). Most newly arisen male-fertile revertants are due to cytoplasmic changes, which are maternally transmitted through subsequent generations. Studies of cytoplasmic revertants have revealed that the change from the male-sterile to the male-fertile phenotype is accompanied by the loss of the free forms of S-1 and S-2 DNA species and by rearrangements in the mtDNA which often involve sequences homologous with the S elements (Levings *et al.*, 1980; Kemble and Mans, 1983). These results led to the speculation that S-1 and S-2 may carry determinants responsible for male sterility and behave like transposable elements. However, it is not clear that transpositional events are actually occurring and that they are responsible for the reversion to male fertility.

The nucleotide sequence of the S-2 plasmid-like DNA has been previously completed (Levings and Sederoff, 1983). In this paper, we report the complete nucleotide sequence of the S-1 DNA molecule. S-1 contains long open reading frames, a sequence containing homology with a portion of a chloroplast gene (Ronald *et al.*, in preparation) and ~1.7 kb of highly conserved homology with S-2.

## Results

The nucleotide sequence was ascertained from S-1 DNA cleavage fragments cloned into the M13 vectors mp7-mp11. Restriction fragments from *BcII*, *BgIII*, *EcoRI*, *FnuDII*, *HaeIII*, *HindIII*, *MboI*, *XhoI*, *PstI*, *RsaI* and *TaqI* digestions were 'shotgun' cloned into the appropriate vector sites for sequence analysis. In those instances when the cloned fragments were too long for complete sequencing, double digestion was employed to generate shorter fragments, e.g., *HaeIII* and *HindIII*, *HindIII* and *SstI*, *HindIII* and *TaqI*. *Sau3*AI and *SstI*, *TaqI* and *SstI*. Both strands of S-1 were sequenced completely and the sequence was confirmed with overlapping clones (Figure 1).

After digestion with proteinase K to remove the terminal protein, S-1 DNA is purified as a linear molecule with defined ends. To sequence the ends of S-1, we have force-cloned S-1 terminal fragments, generated by *PstI* or *HindIII* digestion, into the *SmaI* and *PstI* or *HindIII* cloning sites of M13mp8 and mp9. The resulting clones have the terminus of S-1 blunt-end ligated to the *SmaI*-cut vector. Previous studies indicate that the 5' termini of the S-1 and S-2 DNAs can not be end-labeled even after proteinase K digestion and extensive phenol and chloroform extractions (Kemble and Thompson, 1982). In contrast, the 3' termini are apparently not impaired because the 3' ends can be digested

10 20 30 40 50 60 70 80 90	
<u>AAAAGTATAČ AAGCACATČT CCAATCTAČA TAAAGATACC AACCAGGTAT CTACTTCAAA GACAGGGCGT CGGCGATCCT CTACTATTAA</u>	
GAGACAGĂTĂ ACAATGETEC CGACAGAÊĂT GGACAGAĂČŤ GCAGAGA <u>ĂTĂ CCTCTCCEGĂ GAAGTCCTTĂ CATGTCTĆĂĂ</u> <u>ACTAAATĂĂĂ</u>	
190 200 210 220 220 220 220 220 220 240 250 260 270	
ATCCTAAAGC AGGGGGGGTT ACCATJCTTG CTTCAGCATT TATGGAATTG TTAATAAATA GGGTTTACCC ATCAATCCAG GGTTCGGGGA Aso aso aso sid	
AGTTCACAČĚ CCAATATČŠA TTGAATAŤÅŠ ATGGTAAŤČČ CATTAATAŤŤ ACTITATČAŘ AAGCAATŤŘA GCTTACAŤŘĚ GCTGATGGAA	
CCCGCATTEC TAATGASTIC ATATTAAAAG AAATAATAAA TETETTGAAT AAGTACGCAG AAAACTACCA AGTTGCGAT GTTGAAGCAA	
640 650 660 670 880 690 700 710 720 Trastrias gertatica basestati tabattean trabectica atticaacta aabateanti titeanctat cteaneegag	3610 3620 3630 3640 3650 3660 3670 3680 3690
	ATTEGETEAA GAGGAAGAGT CAAGTAGTAG GTGTGGETGE TCACGGEACE CCCGGEACGG TTETTTAGGT CGATGAATEG AATETTEAA
	3700 3710 3720 3730 3740 3750 3750 3750 3750 3750 3750 3750 375
ATAAAACAGA AATGAAAAAC AAAACTCTAT TITITGTTEC TGATCTAGAA ACTCTITTAT TAAAAAGACE AEACACEGAT GICGACAAAA	3790 3860 3810 3820 3830 3840 3850 3860 3870
910 920 930 940 950 940 970 <b>980 99</b> 0	
CTCACGTÉCE CTACGCAGES GEGTATATES TEGTTEATES GAAAAAGESS GTTAACGESS ACCATATTAE GACGTELTA GCACATEAET	GGCTGATATC ATCAACCETE CTAATCITEG TATGGAGGTA ATGCACGAAC GTAATGCTCA CAACTICCT CTAGACCTAG CAGCTCITEA
1000 1000 1000 1020 1030 1030 1030 1050 1050 1050 1050 105	AGTTCCĂŢĊŢ CTTAATĢĞĂŢ AAGGTTŢŢŢĊ TCCTAAĂĊĂŢ ATAGGAĂĞĞĂ AAGCCGĞĞŢŢ CTTTTAŢĞĂĞ GAAGAAĞĊĂĞ TGCACGĂŢŢĞ
1090 1100 1110 1120 1130 1140 1150 1160 1170 Camerican centeratar tutaranti tartregit camericanta atartet citututar tamanettat amamettat	4060 4070 4080 4090 4100 4110 4120 4130 4140 Tecaccaate ceccatae ectate canadeat tctatticca ctcaaeca aaaeettac cctttacee caaetaea
	4150 4160 4170 4180 4190 4200 4210 4220 4230
1270 1280 1290 1300 1310 1320 1330 1320 136 1370	4240 $4250$ $4250$ $4260$ $4270$ $4280$ $4290$ $4300$ $4310$ $4310$
TCATGGATTC GTACCTCCTG TTGAAAGTAA AGCTCGCTGA TCTAGCCGAT AGTTTTTGCC CAGAATTGGG GGGGAAGGGA TCTTTCGATC 1360 1370 1380 1470 1440 1410 1420 1430 1440 1440	ACCEANTEC AGETATCTAS GATTETCACT ATCETCACTA TCTTCTTCAC TCCCACTACC TTCTTCTTCA CAGCCACCAC CTTCTTCTTC
ACCAGAĂTĂT CACCGUTGAT AAACTAČCTĂ GTATTAĞĞĂ AGACTCTITĂ ACTTATCTTĂ AACAAGĂTĂT TCTAATĂAČĂ GCTGCTĞITĂ	ACAGCTĂCTĂ TCCACCCCCȚ CTTTACĂĞĞC ACTACCTȚCȚ TTACTGCĂĂ CATATTCTĂȚ AGTGGCĂTŤC TCCTCTĂCĂĞ AATCTATĞŤ
TGCAACGOGO TAAAGCCATT ATTTGGGAAG AGTATGGGAT TGATATCOTT AAAGTATTAA CAATATCAGO ATTGGCCCTA AAAATATTTA	4420 4430 4440 4450 4460 4470 4480 4480 4480 4480 450 4480 450 450
1940 1950 1950 1950 1970 1970 1980 1970 1980 1970 1980 1970 1970 1970 1970 1970 1970 1970 197	
1630 1640 1650 1660 1670 1680 1690 1700 1710 Gtaracear restatabe cratareges acaditati attatarege tatatater	4510 4520 4530 4540 4550 4560 4570 4580 4590
1720 1730 1740 1750 1760 1770 1760 1770 1780 1790 1790	CGTCCTTICC ACAACCGCAT CATAAGCITC TAAGATTGAT GTACTAATCA CTTCCTGTA TGGAATCCAT ACGAATAATT TACGAACTA
	GAATAAATCA ATAATCAATT CTATATCAAG ATAGTGCGGG TCCACAGATA TGAAATCTGG TGCGCTCTCA GTATAACAGT ACCACACTGT
1810 1820 1830 1840 1850 1860 1870 1880 1890	4690 4700 4710 4720 4730 4740 4750 4750 4750 4750 4760 4770
TAATTIGICE TAAGCATATT AAAAAGCTC TICTACCITA TAAAAAGGAT GATGGTACGA TAATCITCCC CACGGGGGGG TITCTCGGTG	4780 4790 4800 4810 4820 4830 4840 4850 4860
TTTATTTTTTC TEAGGAGETT AAGTATGCAG TAAGTTTGGG CTACAAAGTA TACCCGATCT GTGGATATAT CTTTGATAGA AAGGAATCAC	4870 4880 4890 4910 4920 4920 4920 4930
CATTCAAGEG TITTETETAC GACATCTACA GCAAAGEGT TGATEGTAAG GCTAAAGEGE GGAAAGETCT GGATTTGATC TATAAAATCA	CCTGITATAG AGAGACCGCA ACCCCITATA TICACTGAAA TCTGCACGAT AAATCITICI TITCAATITI AAACGCATGA TCAATACATA
2080 2090 2100 2110 2120 2130 2140 2150 2150 2160 Canteracag tectatege Ageticger Academica Actanate Itteracage Agenater Anattert	TCACTAMAGA TTTCGTITTT ATTAACCITT CGTTATAACC CTATATTITC CGGGTGTAAG TTCAGGAAGA GAAGAGCTAA TCTACTAGCT
2170 2180 2190 2200 2210 2220 2230 2240 2250 2550	SUSU SUSU SUSU SUTO SUBO 5070 5100 5110 5120 5130 Geageteaet Aeceaettec Anantanaag Atgatcaet Citciaagg Cigaetaega tatcitcisc gagetaegaet
2260 2270 2280 2290 2300 2310 2320 230 230 230	<b>5140 5150 5160 5170 5180 5190 5200 5210 5220</b>
AGCTINGTE AGACLUCULT ACTINICUE CTETTCANT ATCACCECT ETCACEGETT ATECCCECAT ANEAATECAC CCATTCATT 2350 2350 2360 2370 2380 2390 2400 2400 2400 2400 2400 2400	5230 5240 5250 5260 5270 5280 5290 5300 5310
CGAGGGATEA CTECTATIAC ACGGATACEG ACTCAETCET TETTEBAGAGE GAACTICCTE AAGAAGGET ATCACCTACC ECTITEBEIA 2440 2440 2450 2450 2450 2470 2480 2480 2490 2490 2500	ACCTCAAGGC TGCCATCCTC ATGAAGGATA AGAGCTITGT CATTGTTGTC GGCCATTAAT AACCCCATCGG ACGACCAGTT AATATCTACA 5320 5320 5320 5340 5350
AGTITAAGGA TEAGCACITI GICGAATATG GCAICITITI AGCACCIAAG ICCIAIATGC ITAAGGCATC IAGTGTGAC CAACCCATAA	CTATATĂĞĂŤ CTAGTĂCTTC TAATATĞTTĂ TCTATGŤCĂC AATCGTCGĂT TGTGAATTCG TCAGTTTGTT TACAAAGAAT AAATAGTTCA
TAAAGTTTAA AGGAGCGGGT AAAGATGAGG CTGATGAAGA ATGGTTCATC AACCABCTAG CTGACCCTAG GGCTAAAAAA GTGATTTCAT	
ZOSO ZOSO ZOSO ZOSO ZOSO ZOSO ZOSO ZOSO	5410 5420 5430 5440 5450 5460 5470 5480 5490
	5500 5510 5520 5520 5530 5540 5550 5560 5570 5580
2710 2720 2730 2740 2750 2760 2770 2780 2790	GITATITICAC CTATAGGTGT TATCCCCGTG GAGGGGAAAA AGCTGTGATT ATGAAAGTCA TAGATAGGGA GATCTATAAT CCGTTCAACC 5590 5640 5650 5610 5620 5620
2800 2810 2820 2830 2840 2840 2840 2850 2860 2870 2880	AGATTGTAAT ATAGTCCATA TTGTTTTGTA AATTTCTGGG ATGTTAATAT ACGCGAAGTA GGTATATAAT CCATACAATC ACTCAAATCA
ATCEGATAAT TATGAATITE TTAGAGGAGA ATGAAGATCT CCGAAATGAA TTITCTAATT CGGAAATGAT GATTGCTAAT CGGGAGATTA 2890 2900 2910 2920 2920 2930 2940 2950 2950 2950 2950	ATACTARGA CITTARGAC ATTITCTTGT CGATTCARGA ATGCGTARTG CATGGCTTAT AGATCTITGA ATTTAGAAAT GTTAGGCTGA
AAGCTGATGC TECTAAAAAG AGGAAAAGCT CTAAGCTAAG AGGAAAGCT CTAAGABGBG GAGACCCC TTCTCATATT GAAGAATAGG	5770 5780 5790 5800 5810 5820 5820 5830 Gettintia ciettigita Astastita Gatascort centres testores controls and 5830
CATEGRATEC ATTOCCOASC TOTATACTAT AGACCATTEC TEAAACAAGA TATEAATEAA CAATTACAAC CECTTOTCAA ECTAATACCA	5860 5870 5880 5890 5900 5910 5920 5930 5940
AACAAAGCCC GAAACCACTC AAGTAAAGCA GAAGCTGAAA CCAACGAACC TSTGCAGGCT AGGCACCTT TACTAGATCAG CTACAAAACA	SPS0 SPS0 SPS0 SPS0 SPS0 SPS0 SPS0 SPS0
3160 3170 3180 3190 3200 3210 3220 3230 3230 3230 3230	TTAATATTTA TITGATAATA AACAACGGGA CCAAAACGGT ATTTACCTGT TAAGTTCAAT TGATTTATGA AATCATTAAT CATGAAAGGC
3250 3260 3270 3280 3280 3380 3380 3380 3380 3380 338	AGCACAÇETE GAACCETIÇĂ AGAATCEȚIĂ TAGATAGATE GTACITETAA AACCAGETIC IGEITIGE AATTATITE GEGAATATE
3340 3350 3360 3360 3360 3360 3340 3420	6130 6140 6150 6160 6170 6180 6190 6100 600 600 600 600 600 600 600 600 6
ACGAAGAICA AIACAICGGA TICCTICTIG TCTGAITIGGA AGGGAIGGAI TGIGCTIIGG CICAIGIACI CIAIAAGGGA TAGTAAGCA 4430 - 3440 - 3440 - 3450 - 3450 - 3450 - 3450 - 3450 - 3550	6220 6210 6220 6220 6250 6250 6260 6270 6290 6270 6290 6290
GGGAATGECT TTGATTATET TITTTAETTI CTATAAGGET CITGTTÄTET CCTITGTETE TCTCTCTETT CTATAAGGET CTTGTTTÄGT	SILO SILO ALANDIA ALANDIA GALATIZICE GAGAGAGIAI ICICIGCAGI ICIGICAIC ICIGICAGE CALIGUIA
GGTATAČTČT TTGATCČTTA CTCTATTTAČ TACGCAĂATĂ AATTGTTČGT GGCAGGGAAA CCGGTGAAAA ATGAAGAATC TGGGTTČGCC	CTETCTETTA ATAGTAGAGG ATEGEEGACG CECTGTETTT GAAGTAGATA CETGGTTGGT ATETTATGT AGATTGGACA TGTGETTGTA
	TACTIT

Fig. 1. Nucleotide sequence of the linear S-1 DNA molecule from the mitochondria of the S cytoplasm of maize. The 208-bp terminal inverted repeats are underlined. The sequence is presented in the 5'-3' direction.

with exonuclease III and end-labeled or tailed with terminal transferase (Kemble and Thompson, 1982; Meints *et al.*, in preparation). Nevertheless, our terminal sequence could be incomplete if termini missing a nucleotide or so are preferentially cloned. At any rate, we have consistently obtained terminal sequences ending in the same order for both S-1 and S-2 (Levings and Sederoff, 1983).

The S-1 DNA molecule is 6397 bp long while S-2 is 5453 bp in length (Levings and Sederoff, 1983). We have recently discovered an error in the published S-2 sequence which increases its length by 1 bp to 5453. The molar G+C content of S-1 is 39.4%, which is similar to S-2 (37.5%) but different from the maize mtDNA (47%). The maize chloroplast DNA is 40.8% and the maize nucleus is 42.9% (Shah and Levings, 1974; Levings and Sederoff, 1983).

The S-1 DNA molecule is terminated by exact 208-bp inverted repetitions (Figure 1, underlined sequence); S-2 DNA is also terminated by the same 208-bp inverted repeats as S-1. A short region of nearly perfect homology (15/16) between S-1 and S-2 begins at the end of the inverted repeat in S-1 and S-2. This 16-bp homology, which starts at nucleotide 209 and ends at 224 in S-1 and S-2, contains a single mismatch at position 221. When this homology is considered with the inverted repeat, a 224-bp homology with one mismatch is observed.

A 1462 nucleotide stretch of S-1, beginning at position 4936 and running through 6397 is found highly conserved in S-2 (Levings and Sederoff, 1983). This homology, which includes the inverted repeat, would be perfectly conserved if it were not for a 2-bp mismatch occurring at positions 4963 and 4964 in S-1

1126

and at 4019 and 4020 in S-2. The difference is illustrated by comparing this region in S-1, 5' TC 3', with the same region in S-2, 5' GA 3'. The S-1 sequence in this region was determined by sequencing 10 independent clones. In one exceptional case, the sequence of S-1 was 5' GA 3' at positions 4963 and 4964, respectively, which is the same as in S-2. It was definitely established that this clone was from S-1 by sequencing beyond the homologous regions. Since this exception occurred only once, it was not included in the final sequence. Nevertheless, it appears that two forms of S-1, differing at nucleotides 4963 and 4964, can exist.

Three long open reading frames were found in S-1 by computer analysis (Figure 2). They include a 2787-bp unidentified reading frame URF3 (929 amino acids) starting at position 180 and terminating at 2966. The other strand contains two long open reading frames: the first, URF2, has a 1017 nucleotide-long open reading frame (339 amino acids) that begins at position 6218 and ends at 5202; the second, URF4, has a 768 nucleotide open reading frame (256 amino acids) that begins at position 4950 and ends at 4183. Since codon usage in plant mitochondria is not established, the universal code was employed in the computergenerated translation. Two possible exceptions to the universal code have been suggested based on an investigation of the cytochrome oxidase subunit II gene of plant mitochondria (Fox and Leaver, 1981; Hiesel and Brennicke, 1983). The UGA codon, which codes for tryptophan in animal and fungi mitochondria, may not code for an amino acid in plant mitochondria, and the codon CGG may code for tryptophan instead of arginine. Although protein synthesis probably initiates at the



Fig. 2. Schematic maps of S-1 and S-2 DNAs showing the location of the long unidentified open reading frames (URF). Codon usage was that of the universal code. Numbers at the beginning and end of the arrows indicate the start and stop of the respective open reading frames. IR indicates the terminal inverted repeats. The 1462-bp region of homology is indicated by a bold line.

methionine codon, AUG, this has not been unequivocally established in plant mitochondria.

#### Discussion

The long open reading frames of the plasmid-like DNAs may code for functions involved with their maintenance or replication. Other linear DNAs that replicate by a strand displacement mechanism require additional functions to replicate, i.e., a terminal protein, a DNA-binding protein and a polymerase (Stillman, 1983). In adenovirus and several *Bacillus* phage these polypeptides are coded by the viral genome, therefore, these functions may also be coded in the plasmid-like DNAs. Since the plasmid-like DNAs of maize are characteristically found in pairs, these molecules may share essential replication functions.

The sequence of S-1 shows that the terminal inverted repeats of the left and right ends of the molecule are identical to each other and to those of S-2. This identity may be due, partly, to the requirement for a specific sequence at the ends of the molecules related to replication and for possible promoter sites suggested by the location of open reading frames. Alternatively, homologous recombination could generate homogeneity between repeats. Schardl et al. (1984) have reported that S-1 and S-2 recombine with homologous sequences present on the complete circular mitochondrial chromosome. Recombination can also explain the exceptional S-1 derived clone that contains a GA sequence at positions 4963 and 4964. This sequence is characteristic of the 1462-bp homology region of S-2. The occasional appearance of this sequence in an S-1 molecule is expected if recombination occurs between homologous regions of S-1 and S-2.

We have previously suggested that S-1 arose from related plasmid-like DNAs found in exotic races of maize (Levings *et al.*, 1983). Among 12 male-fertile Latin American races of maize, plasmid-like DNAs are found that are distinct from those of *cms-S* (Weissinger *et al.*, 1982). These DNA species, designated R-1 and R-2, are 7.4 and 5.4 kb long, respectively, and like the S elements, are isolated as double-stranded linear DNAs terminated by 0.2-kb inverted repeats. The R and S plasmids share considerable sequence homology (Levings *et al.*, 1983; Weissinger *et al.*, 1982) although R-1 contains  $\sim 2.6$  kb of sequence not found in the S plasmids. Based on sequence homology, we proposed that S-1 arose by a recombination event between R-1 and R-2 (Levings *et al.*, 1983). This proposal affords an explanation for the 1462 bp of sequence homology between S-1 and S-2.

A short segment of chloroplast DNA sequence has been found in the S-1 molecule (Ronald *et al.*, in preparation). The chloroplast segment contains a sequence from the gene coding for the 32-kd quinone-binding protein (psbA) of the thylakoid membrane.

Several characteristics suggest that the plasmid-like DNAs have a viral origin. These include the structure of the ends of S-1 and S-2, the presence of long open reading frames, and the occurrence of autonomous replication. However, no higher plant viruses are known that have the same physical properties (Shepard, 1979). All higher plant DNA viruses are circular, including cauliflower mosaic virus, which is primarily doublestranded (Hull and Shepard, 1977; Volovitch *et al.*, 1978) and gemini viruses, which are single-stranded (Goodman, 1977). If the plasmid-like DNAs do have a viral origin, they may be derived from an evolutionarily distant biological system.

#### Materials and methods

S-1 DNA was obtained from maize strains carrying the S (US Department of Agriculture) maize cytoplasm, designated *cms-S*. The mtDNA was isolated from etiolated seedlings as previously reported (Pring and Levings, 1978). S-1 DNA was purified by electrophoresis on 0.9% agarose gels and electroelution (Smith, 1981).

S-1 DNA preparations cleaved with *BcII*, *BgIII*, *EcoRI*, *HaeIII*, *MboI*, *PstI*, *TaqI*, *FnuDII*, *HindIII*, *RsaI* or *XhoI* were cloned into the M13 bacteriophage vectors mp7-mp11 (Messing, 1982). Digestion, ligation and transformation procedures followed the protocols furnished by Bethesda Research Laboratories or New England Biolabs. In some instances, recloning was carried out to invert a cloned fragment or to subclone an internal fragment from an existing clone.

The DNA nucleotide sequence was resolved by the dideoxy-chain termination method (Sanger *et al.*, 1977) with a 17-nucleotide, universal primer provided by P-L Biochemicals. Sequencing gels were 0.4 mm thick and either 6% or 8% polyacrylamide. Sequence analysis was performed with the computer programs of Intelligenetics.

#### Acknowledgements

We wish to thank Jane Suddith, Annmarie Tuttle and Maria Patroni for excellent techical assistance. This work was supported by grants from the National Science Foundation (PCM-8010933), the competitive grants program of the US Department of Agriculture (USDA/SEA 82-CRCR-1-1085), and Agrigenetics. M.Paillard received post-doctoral support from Le Ministère de la Recherche et de la Technologie and La Societe Nationale ELF-AQUITANE. This article is contribution No. 9548 of the Journal Series of the North Carolina Agricultural Research Service, Raleigh.

#### References

- Carusi, E.A. (1977) Virology, 76, 380-394.
- Fox, T.D. and Leaver, C.J. (1981) Cell, 26, 315-323.
- Goodman, R.M. (1977) Virology, 83, 171-179.
- Harding, N.E., Ito, J. and David, G.S. (1978) Virology, 84, 279-292.
- Hiesel, R. and Brennicke, A. (1983) EMBO J., 2, 2173-2178.
- Hull, R. and Shepard, R.J. (1977) Virology, 79, 216-230.
- Ito, J. (1978) Virology, 28, 895-904.
- Kemble, R.J. and Mans, R.J. (1983) J. Mol. Appl. Genet., 2, 161-171.
- Kemble, R.J. and Thompson, R.D. (1982) Nucleic Acids Res., 10, 8181-8190.
- Kemble, R.J., Mans, R.J., Gabay-Laughnan, S. and Laughnan, J.R. (1983) *Nature*, **304**, 744-747.
- Laughnan, J.R. and Gabay-Laughnan, S.J. (1983) Annu. Rev. Genet., 17, 27-48.
  Laughnan, J.R., Gabay-Laughnan, S.J. and Carlson, J.E. (1981) Stadler Genet.
  Symp., 13, 93-114.
- Levings, C.S., III and Pring, D.R. (1979) in Scandalios, J.G. (ed.), Physiological Genetics, Vol. 5, Academic Press, NY, pp. 171-192.
- Levings, C.S., III and Sederoff, R.R. (1983) Proc. Natl. Acad. Sci. USA, 80, 4055-4059.
- Levings, C.S., III, Kim, B.D., Pring, D.L., Conde, M.F., Mans, R.J., Laughnan, J.R. and Gabay-Laughnan, S.J. (1980) Science (Wash.), 209, 1021-1023.
- Levings, C.S., III, Sederoff, R.R., Hu, W.W.L. and Timothy, D.H. (1983) in Ciferri, O. and Dure, L., III (eds.), Structure and Function of Plant Genomes, Plenum Publishing Corp., pp. 363-371.
- Lonsdale, D.M., Thompson, R.D. and Hodge, T.P. (1981) Nucleic Acids Res., 9, 3657-3668.
- Messing, J. (1982) in Setlow, J.K. and Hollander, A. (eds.), Genetic Engineering, Principles and Methods, Plenum Press, NY, pp. 19-35.
- Pring, D.R. and Levings, C.S., III (1978) Genetics, 89, 121-136.
- Pring, D.R., Levings, C.S., III, Hu, W.W.L. and Timothy, D.H. (1977) Proc. Natl. Acad. Sci. USA, 74, 2904-2908.
- Rekosh, D.M.K., Russell, W.C., Bellet, A.J.D. and Robinson, A.J. (1977) Cell, 11, 283-295.
- Salas, M., Mellado, R.P. and Vinuela, E. (1978) J. Mol. Biol., 119, 269-291.
- Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA, 74, 5463-5467.
- Schardl, C.L., Lonsdale, D.M., Pring, D.R. and Rose, K.R. (1984) Nature, 310, 292-296.
- Shah, D.M. and Levings, C.S., III (1974) Crop Sci., 14, 852-853.
- Shepard, R.J. (1979) Annu. Rev. Plant Physiol., 30, 405-423.
- Smith, H.O. (1981) Methods Enzymol., 65, 371-380.
- Spruill, W.M., Jr., Levings, C.S., III and Sederoff, R.R. (1980) Dev. Genet., 1, 363-378.
- Spruill, W.M., Jr., Levings, C.S., III and Sederoff, R.R. (1981) Dev. Genet., 2, 319-336.
- Stillman, B.W. (1983) Cell, 35, 7-9.
- Volovitch, M., Drugeon, G. and Yot, P. (1978) Nucleic Acids Res., 5, 2913-2925.
- Weissinger, A.K., Timothy, D.H., Levings, C.S., III, Hu, W.W.L. and Good-
- man, M.M. (1982) Proc. Natl. Acad. Sci. USA, 79, 1-5.
- Yehle, C.O. (1978) Virology, 27, 776-783.
- Yoshikawa, H. and Ito, J. (1981) Proc. Natl. Acad. Sci. USA, 78, 2596-2600.

Received on 19 December 1984; revised on 21 February 1985