Proc. Natl. Acad. Sci. USA
Vol. 80, pp. 4055–4059, July 1983
Genetics

Nucleotide sequence of the S-2 mitochondrial DNA from the S cytoplasm of maize

(plasmid-like DNA/terminal inverted repeats/linear DNA replication/cytoplasmic male sterility)

CHARLES S. LEVINGS III AND RONALD R. SEDEROFF

Department of Genetics, North Carolina State University, Raleigh, North Carolina 27650

Communicated by C. Clark Cockerham, March 28, 1983

ABSTRACT **Mitochondria from the S male-sterile cytoplasm (*cms-S*) of maize contain two plasmid-like DNAs, S-1 and S-2, that appear to be prominently involved with the cytoplasmic male sterility trait. The complete nucleotide sequence of the S-2 DNA molecule was determined by the chain termination method. The linear S-2 DNA molecule contains 5,452 base pairs and is terminated by exact 208-base-pair inverted repetitions. Two large open reading frames were identified in the S-2 DNA, suggesting the possibility of protein-encoding genes. The nucleotide sequence of the S-2 termini are discussed with regard to models proposed for the replication of linear DNA molecules.**

In the S cytoplasm of maize, mitochondria contain plasmid-like DNAs that are distinct from the usual mtDNAs (1). These unusual DNAs, designated S-1 and S-2, are uniquely associated with the S type of cytoplasmic male sterility (*cms-S*). The S group, which includes about 20 members, is characteristically restored to pollen fertility by the nuclear gene, *Rf3*, located on chromosome 2 (2). The S-1 and S-2 DNAs are 6.4 and 5.4 kilobases (kb) long, respectively, and are commonly isolated as double-stranded linear DNA molecules with defined ends. The molecules are structurally unique in that they contain terminal inverted repeats of about 0.2 kb. Normally, S-1 and S-2 are present in equimolar quantities but are about 5-fold more abundant than the mtDNA. However, it is known that nuclear background affects the content of S-1 and S-2 (2). Although the informational content of these DNAs is unknown, it is interesting that sequences homologous with S-1 and S-2 are found integrated into the mtDNAs of all maize cytoplasms (3, 4).

Additional plasmid-like DNAs were discovered among 12 male-fertile Latin American races of maize that are distinguishable from those of *cms-S* (5). These DNA species, called R-1 and R-2, are 7.4 and 5.4 kb long. Like the S plasmids, the R plasmids are isolated as double-stranded linear DNAs that are terminated by 0.2-kb inverted repeats. The R and S plasmid-like DNAs have substantial sequence homology even though R-1 contains about 2 kb of sequence not found in S plasmids. This fact has led to speculation that S-1 may have arisen by a recombination event between R-1 and R-2 (6).

In the S male-sterile cytoplasm, spontaneous mutations to pollen fertility occur, sometimes at unusually high frequencies (2, 7). Most often the male-fertile revertants are due to cytoplasmic changes, which are maternally transmitted to subsequent generations. In these newly arisen revertant strains, free forms of S-1 and S-2 are no longer found, and rearrangements are observed that often involve sequences homologous with the S elements (8). Based upon these findings, it was suggested that S-1 or S-2 DNA or both may carry factors responsible for male fertility and behave like transposable elements. The apparent association of the S plasmids with male sterility and transpositional activity makes these molecules interesting for study.

In this report, we present the nucleotide sequence of the S-2 DNA molecule.

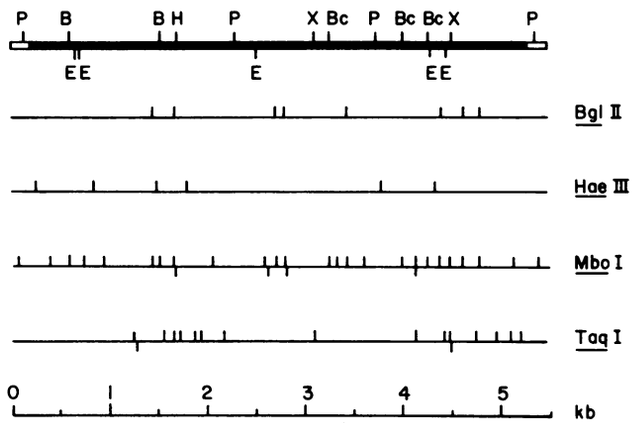

FIG. 1. Restriction map of the S-2 DNA (5,452 base pairs). Restriction sites are indicated by vertical lines: B, *Bam*HI; Bc, *Bcl* I; E, *Eco*RI; H, *Hin*dIII; P, *Pst* I; X, *Xho* I.

MATERIALS AND METHODS

S-2 DNA was obtained from maize strains carrying the S (U.S. Department of Agriculture) maize cytoplasm, designated *cms-S*. mtDNA was isolated from dark-grown seedlings as described (9). S-2 DNA was fractionated by electrophoresis on 0.9% agarose gels and purified by electroelution (10).

Cloning was carried out by using M13 bacteriophage vectors mp7, mp8, and mp9 (11). Double-stranded replicative form was cleaved at the appropriate restriction sites and ligated to DNA preparations of S-2 digested with *Bam*HI, *Bcl* I, *Bgl* II, *Eco*RI, *Hae* III, *Mbo* I, *Pst* I, *Taq* I, and *Xho* I. Ligation and transformation procedures generally followed protocols provided by New England BioLabs (Fig. 1). In some cases, recloning was done to invert a cloned fragment or to subclone an internal fragment from an existing clone. To do this, double-stranded preparations were made from 1-ml cultures by a plasmid preparation technique that included LiCl precipitation to remove single-stranded DNA before recloning (12).

The DNA nucleotide sequence was determined by the chain termination method of Sanger *et al.* (13) with a universal primer furnished by New England BioLabs or P-L Biochemicals. Sequencing gels were either 6% or 8% polyacrylamide and 0.4 mm thick. The sequence was analyzed by the computer programs of Intelligenetics.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviation: kb, kilobase(s).

```
        10         20         30         40         50         60         70         80         90        100
AAAAGTATAC AAGCACATGT CCAATCTACA TAAAGATACC AACCAGGTAT CTACTTCAAA GACAGGGCGT CGGCGATCCT CTACTATTAA GAGACAGATA
       110        120        130        140        150        160        170        180        190        200
ACAATGGTGC CGACAGAGAT GGACAGAACT GCAGAGAATA CCTCTCCGGA GAAGTCCTTA CATGTCTCAA ACTAAATAAA TCCAACCTGC AAGAAGACAC
       210        220        230        240        250        260        270        280        290        300
ACAAAAAAGA AAAATATGAA GTATCCTCCA CTGACAGCAA AAAAATTGGC CGAAGTGAAA AGACTCTTGA AAAAGAGTCA GATTCCTCAA TTGAAATATA
       310        320        330        340        350        360        370        380        390        400
GAGAATACGA GTATGAGCTT AGATATGAGA AAGAAATGCG TGAATTAGAA TGGCAGAAAG ATAAAATTAA GAGCGATTTC TTTAACTGTA ATTATAACCG
       410        420        430        440        450        460        470        480        490        500
ATCAGGGAGT CAGTTAGGGT ATCAACGGAT AGAGACCCTG ATTTAGAGGA TGAAAAAAGA GAGCAGCTAG GAGAGTCTAT GCAGACTGAA TTGGAGAGAC
       510        520        530        540        550        560        570        580        590        600
TAACGCGGGG TGTAGAAGTC GGTACTTCAG AAGATATCAA TGTTGACACA GTCAAACGTT GGGGTTTGCA AAATAATAAG TACAACGCTG AACACTGGAT
       610        620        630        640        650        660        670        680        690        700
CCCTCCAGGG GGTCAGCGCA CAGCTGAGGA TATGGGAAAA CTCTACCAAT TCTGGGATGA ATTCATTGAA ACATATGAAA ATGAGGTTAT GATGAATTCG
       710        720        730        740        750        760        770        780        790        800
GGGTACAGGG AACCCTTAAA TAAAGATAAA GTGTTTAATA CACTACTAGA TCATTCTAAT AATAAGACCA ATCAGACGGC TGAGGAATTA AAAGGCATTC
       810        820        830        840        850        860        870        880        890        900
AAACTAAAAT AGAACGTATG ACTATGTATT TTTATGAGGC CAATGTCTAT GAATCATTAG GTCAGTTGAA GAATAAGTTC ATCAATATAG ATGAGCAAAC
       910        920        930        940        950        960        970        980        990       1000
AGCAAAGGAT TACCTTCTTG ATAAGTTAGA GAAACCAGAC GATCTAGATA TCGTTAGAGC TATGGGCACC TATACACTGG AATGTATAGT TGTATTTGTC


      1010       1020       1030       1040       1050       1060       1070       1080       1090       1100
TTAAGCAAGC AGTTTAATGT GTTTGCTTTG GATAAGGCTA GTGTTCAAGT TGCTACATTG GTGAGTGAAC TAGATTCCGC AGCTAAGATT GAGTATAATC
      1110       1120       1130       1140       1150       1160       1170       1180       1190       1200
GTATAATCGC TGAGGATAGA AAGCGCAAAA AAGCTAAGCG GGGTGATAAG CAGATAAATG AAGATGAAAA CATCTTATTG CAAGATAAAG GTGTGGATAC
      1210       1220       1230       1240       1250       1260       1270       1280       1290       1300
CTTTAATGGT GATGTACATT CCGAGCTGGA TAGAAAGAAA AAGAGCATAT CAAAACAAAC GAATCGAAAA AGAGTTCGAA TCAGTAAACA TAAGGAGCCC
      1310       1320       1330       1340       1350       1360       1370       1380       1390       1400
GGTATAGGGG AAGCCCTGTT TGGCTGGCTG GCTAGCAGGA AATTGATAGA AGTCAAAAAA CCTGTTTTTC TCTTTGATAA GAAGAATAAG AAGCCAAAAA
      1410       1420       1430       1440       1450       1460       1470       1480       1490       1500
ATGTAATATA CCCGGGCTAT GTAGATTGTC TATTTGACAT TAAAGATCTA CCCTTCTGTT CAACCTTACC TATGGTATAC CCTCCAGCCG ATTGGGAATT
      1510       1520       1530       1540       1550       1560       1570       1580       1590       1600
ACGGCCCACC GCAACCGCGG AAGTCTCGGA TCCATATATA ACGAATCTGA CCCCTCTATC GAGTTATAGA GGTGGTTACC TCACATCCCT ACAAAGGGAG
      1610       1620       1630       1640       1650       1660       1670       1680       1690       1700
AGTGGAGATA GTCCGACTCT CTTAAGTGAA AAAGATTATG GTGTATTTGA CATACATATC GATCGCGAGA GATCTCAACC AGTGTTGAGT GCTGTGAAGA
      1710       1720       1730       1740       1750       1760       1770       1780       1790       1800
AGCTTCAGCG GCAGCCCTAT CGAATTAATA AATTAGTCTA TGATTTTATA CAAAAACATT GGAGTGTATT AGTGTCCGTG GGGCTTCTCA GCCCGAAGAA
      1810       1820       1830       1840       1850       1860       1870       1880       1890       1900
TCTAGCTTTA TTTAAACGAA AGGAGGCGCT CAGGCTACTA TCTAGCCTTT TGTTTAAACA CGAGGAGCTT TCAACGATTT ATCGATATAG TGAGTTCAAA
      1910       1920       1930       1940       1950       1960       1970       1980       1990       2000
TCTGTATTGT TAAAAAATAT ACACGCGTCA ACCTTCGAAC TATATACTAT GAAAATAGCT GAGGCTTATC TAGATTATAA AATCTATTTT CCAATCTTTC


      2010       2020       2030       2040       2050       2060       2070       2080       2090       2100
TGGACTTCAG GGGGCGAAAT TACCGCCATG GACCCTTCCA TTTCCACGAA CGTGATTTAG TGAGATCACT CATCATATTT GATGAAAGTG ATGACTCAGC
      2110       2120       2130       2140       2150       2160       2170       2180       2190       2200
AGCACATACT ATAAATAGTG ATGTTGGGGA TAGAATCCTC CATAATTTCC TTATATCAGC GGCATACCAT AAATCGAAAT TTGGTGTATA CCGAGAGGCT
      2210       2220       2230       2240       2250       2260       2270       2280       2290       2300
CTGGAGTTTA TCTATAATAA AATAGAAGAT ATGCAAAGTA AACCTACATT CTTTGAAAAA GATATCTTTG TGGATACACT GTGCTGCAGG CACCCATTTC
      2310       2320       2330       2340       2350       2360       2370       2380       2390       2400
AGTATATATC ATCATGTATA AGTTTGAAGA CGTACGCGGA TACAAAAGAC TTGTCTGTGC TGCGATACAC ACCAGTATTC CAAGATGCCT CAGCAAGTGC
      2410       2420       2430       2440       2450       2460       2470       2480       2490       2500
TTATCAGATT ATGAGTTACT TTTGTTGGA TATTGATTAT GGGATACATA CAAATCTTTT AAAAAAAACA AATACGGATG GCAGATATAT CAGGGATATA
      2510       2520       2530       2540       2550       2560       2570       2580       2590       2600
TATGAATTCA TGCGCGGTTG TTTAATAAAG TATCTCATCG CTGAAGAAAA GATAGAACTA GCAATTAAAC TACTTACGCC TAATGAAAAA GATCAAGAAA
      2610       2620       2630       2640       2650       2660       2670       2680       2690       2700
GTGTATTAGC TAAGATTGTT AGTATATTTG ATCGCAATGT GGTAAAGAAA ATGTTTATGC CCATGATGTA TGGGAAAACC GATTATACCT TAAAAAAGGA
      2710       2720       2730       2740       2750       2760       2770       2780       2790       2800
TGTTGAAGAT CTATTAAAGG GTAAGTCGGA TTCCGAGGGA ATAAACCTAA TTTCAAAACA TATTTCTACA TATTGGAAAG TGAATTTTGG AAAAATGAAA
      2810       2820       2830       2840       2850       2860       2870       2880       2890       2900
GATCTTATGG ATCTTATCAA TTATGTTAGC TGGTTTGGAG CAGGACAGGA TAAACCGGTG GTATATAGTA CGCCGTACTG GGTAACGCTG CAAACATATA
      2910       2920       2930       2940       2950       2960       2970       2980       2990       3000
AGCGGCGTAA ACGCGTAAAA ATGAAAATAC AGTATGAAAC AACCAAAAAT AACGAAAAGG AGGTGAAAAC AACATCGGCT AAAATGCTTA TACCGTTAAA


      3010       3020       3030       3040       3050       3060       3070       3080       3090       3100
CGATAACGAT ATTAGAAAGA GCTCAACCTC AACCTTTGCC AACTTCATCC ATCAAAAGGA TGCATTTACT GCTATCCAGC TTGTTGACTT TATCAATAAA
      3110       3120       3130       3140       3150       3160       3170       3180       3190       3200
CTCGAGAATG CTTCCTCAAT TCCTATATAC GCAGTACATG ATAATTTTAT AACTATGCCT GAATATGCTA GCATTTTGCC GACCCTTTAT AGGGATTCAA
      3210       3220       3230       3240       3250       3260       3270       3280       3290       3300
TCTTTCGTAT GGGGCACCCA CTCATCATAA TAAACAAATT TTTATTTGAT CATATACTTA TACCTGCAAT ACAAAACGAA CATCCTCAAA ATAAACACTT
      3310       3320       3330       3340       3350       3360       3370       3380       3390       3400
ATTCTCCGTG GAAGAGCGCT CTATGTTAGA TCGTATGATG ATTGATTTAC AGAATCCATT GATTCCCGAT TTTGGAAGTG TTGATATTAC TAAGGCTAGA
      3410       3420       3430       3440       3450       3460       3470       3480       3490       3500
ATCAAATCTA TAGTGATTCC GAAAGATCTT CTTCTTAAGT GCTTTTCATG TTTATGGATG AGTAAAACTA AAAAAATATC TTTAGTTAGA TGGGAATCTT
      3510       3520       3530       3540       3550       3560       3570       3580       3590       3600
GTCGTGATAA AATCATCAAG GTATACATGA GGTATACTGA TGATATCTCT TCAGATGAAG GGGTTAGTAG ATGGTTGGAA TACAAGAATA ATCTTGAGTT
      3610       3620       3630       3640       3650       3660       3670       3680       3690       3700
TGCTAGTGAT CCTGTATGGA GTAGTGATAA TACTAATGGT ACTCAGGCGG ATTCGCTTGA TAAAGGTGAG GATGACTACT GCATCCATTA TTAAATGTTG
      3710       3720       3730       3740       3750       3760       3770       3780       3790       3800
TTTTACTCAA CTGATTCCCT GCAGCTTTCT CAGCATAAAA CATATTTGAT ATCCCGGTTT AGTAGGTATA TACAAATACC GAGGCCACCA CTACCAACTA
      3810       3820       3830       3840       3850       3860       3870       3880       3890       3900
CTTGGTAGCC GTGTGGGAAA GAAGTGTGGG AAAGTGGGCT TCTTTCGCTG TGAATACAGA TGTTTCTCCC CCCTTGAGAC AGGGAAAACA TTCAAGAGCT
      3910       3920       3930       3940       3950       3960       3970       3980       3990       4000
TATTTCATTA TTTCCATTTC TATTTAGTGA TGTATAAAGC GTATAGCGTA GATTTAGCGT TTGATTATTT CAAGTTGAAA TGATATTTGT CATGATCAAT
```

(*Fig. 2 continues on next page.*)

```
      4010       4020       4030       4040       4050       4060       4070       4080       4090       4100
ACATATCACT AAAGATTGAG TTTTTATTAA CCTTTCGTTA TAACCCTATA TTTTCCGGGT GTAAGTTCAG GAAGAGAAGA GCTAATCTAC TAGCTGGAGG
      4110       4120       4130       4140       4150       4160       4170       4180       4190       4200
TGAGTAGCGA GTTGCAAAAT AAAAGATCGA TCAGTCCTCT AAGAGCCGAG TAGCATCATG GTAGATATCT TCTGCGAGGT AGGCTTTCCT ATCTGCCGAG
      4210       4220       4230       4240       4250       4260       4270       4280       4290       4300
TAGACCTATC TAGCTTTACC TTCGCGGGTA TTCTTACAAT AGGGATACCA TGATCAAATG TCTTCTGAAT TCCACACCTC AAGGCTGCCA TCCTCATGAA
      4310       4320       4330       4340       4350       4360       4370       4380       4390       4400
GGATAAGAGC TTTGTCATTG TTGTCGGCCA TTAATAACCC ATCGGACGAC CAGTTAATAT CTACACTATA TAGATCTAGT ACTTCTAATA TGTTATCTAT
      4410       4420       4430       4440       4450       4460       4470       4480       4490       4500
GTCACAATCG TCGATTGTGA ATTCGTCAGT TTGTTTACAA AGAATAAATA GTTCATGACC GTATCGACTA TAGGTGATCC CCGGATACCG AGACTCGAGC
      4510       4520       4530       4540       4550       4560       4570       4580       4590       4600
AAGGAATCTA CAGTATTCTG ATAGAAATTA TGCAATATAA CGTGTGTTAT TTCACCTATA GGTGTTATCC CCGTGGAGGG GAAAAAGCTG TGATTATGAA
      4610       4620       4630       4640       4650       4660       4670       4680       4690       4700
AGTCATAGAT AGGGAGATCT ATAATCCGTT CAACCAGATT GTAATATAGT CCATATTGTT TTGTAAATTT CTGGGATGTT AATATACGCG AAGTAGGTAT
      4710       4720       4730       4740       4750       4760       4770       4780       4790       4800
ATAATCCATA CAATCACTCA AATCAATACT AACGACTTTA ACGACATTTT CTTGTCGATT CAAGAATGCG TAATGCATGG CTTTCAGATC TTTGAATTTA
      4810       4820       4830       4840       4850       4860       4870       4880       4890       4900
GAAATGTTAG GCTGAGCTTT ATTAACTGTT GTTAAAATAA CATTAGATAA CCCCTTCAAT ACAAGTAGGT CCATAGGAAC TCTACACTTA TACACTCCTA
      4910       4920       4930       4940       4950       4960       4970       4980       4990       5000
TAAGACTATG AGTATCCAGA TTCCTTTCAT CCAACCATAA ATAGCCCTTC GATGCTTCCA TACGAATCAC ATCATAGATT AAAGGCGGAA TCTCCTTAAT

      5010       5020       5030       5040       5050       5060       5070       5080       5090       5100
ATTTATTTGA TAATAAACAA CGGGACCAAA ACGGTATTTA CCTGTTAAGT TCAATTGATT TATGAAATCA TTAATCATGA AAGGCAGCAC ACCTCGAACC
      5110       5120       5130       5140       5150       5160       5170       5180       5190       5200
CTCCAAGAAT CGTTATAGAT AGATCGTACT TCTAAAACCA GCTTCTGCTT TTCGCAATTA ATTTGGGGAA TATGTGCAAT TTTACGAGGA GATATAAACA
      5210       5220       5230       5240       5250       5260       5270       5280       5290       5300
TCGAATGCCC GGTACCCTTA GTAGTATGTT TTGCCATATT TGTATTTTTT GTGTGTCTTC TTGCAGGTTG GATTTATTTA GTTTGAGACA TGTAAGGACT
      5310       5320       5330       5340       5350       5360       5370       5380       5390       5400
TCTCCGGAGA GGTATTCTCT GCAGTTCTGT CCATCTCTGT CGGCACCATT GTTATCTGTC TCTTAATAGT AGAGGATCGC CGACGCCCTG TCTTTGAAGT
      5410       5420       5430       5440       5450
AGATACCTGG TTGGTATCTT TATGTAGATT GGACATGTGC TTGTATACTT TT
```

FIG. 2. Nucleotide sequence of the linear S-2 DNA molecule from the mitochondria of the S cytoplasm of maize. The 208-base-pair exact terminal inverted repeats are underlined. The sequence is presented in the 5′→3′ direction.

RESULTS AND DISCUSSION

The DNA sequence was determined from S-2 restriction fragments cloned into the M13 vectors mp7, mp8, and mp9. *Bam*HI, *Bcl* I, *Bgl* II, *Eco*RI, *Hae* III, *Mbo* I, *Pst* I, *Taq* I, and *Xho* I restriction fragments were "shotgun" cloned into the appropriate vector sites and subjected to sequence analysis. When cloned fragments were too long for sequence analysis, double digestion was used to prepare shorter fragments—e.g., with *Pst* I and *Taq* I, *Bam*HI and *Taq* I, or *Mbo* I and *Pst* I. The locations of these restriction sites are shown on the map (Fig. 1). Sequences of both strands were determined from positions 1 through 5,378. The remaining sequence, 5,379–5,452, was determined from the same strand of several independent clones. In most instances, the sequence was further verified by overlapping clones.

The S-2 DNA molecule contains 5,452 base pairs (Fig. 2) and the strand shown has a base composition of 33.2% adenine, 17.3% cytosine, 20.2% guanine, and 29.3% thymidine. The molar G+C content of S-2 is 37.5%, which is substantially lower than that of mtDNA of maize, 47% (14).

S-2 DNA is isolated as a linear molecule with defined ends. It is terminated by exact 208-base-pair inverted repetitions (Fig. 2, underlined sequences). These repeats are responsible for the stem-loop (panhandle) configurations observed by electron microscopy after denaturation and hybridization of S-2 DNA at low concentration (15). The S-1, R-1, and R-2 plasmid-like DNAs are also terminated by similar repeats as judged by hybridization or heteroduplexing studies (ref. 5; unpublished data). The occurrence of these repeats among the various plasmid-like DNAs may suggest a common origin. The function of the inverted repeats is unknown. It is possible that they play a role in replication, rearrangement, or transpositional activities. Sequences homologous with the inverted repeats have been confirmed in high molecular weight mtDNA by nucleotide sequence determination (data not shown).

Two large open reading frames were identified by computer analysis using the universal code (Fig. 3). A 3,294-nucleotide-long unidentified reading frame (1,098 amino acids) begins at position 398 and ends at 3,691. On the other strand, a 1,017-nucleotide-long reading frame (339 amino acids) starts at position 5,273 and ends at 4,257. Although genes have not yet been assigned to the S-2 DNA molecule, the occurrence of long reading frames suggests the possibility of protein-encoding genes. Codon usage by plant mitochondria is not well established. Analysis of the cytochrome oxidase subunit II gene *mox1* in maize has indicated two possible departures from the universal code (16): the UGA codon, which in mitochondria of mammals and fungi codes for tryptophan, may not be read in plant mitochondria, and the CGG codon may code for tryptophan rather than arginine.

Kemble and Thompson (17) recently reported that the 5′ termini of S-1 and S-2 are covalently linked to proteins, which they suggest may be involved in priming replication of the DNAs. Similar DNA–protein associations have been demonstrated in adenovirus (18, 19) and in *Bacillus* phages (20–24); these DNA terminal proteins are thought to play a role in DNA replication. Both of these viral DNAs initiate replication at or close to either DNA end and proceed by a mechanism of strand displacement (25–29). In adenovirus and ϕ 29 it has been proposed that the protein linked to 5′ termini of the linear DNA strand may serve as a primer for DNA synthesis (18, 25, 28–30).

Adenovirus DNA contains terminally inverted repeated sequences that are approximately 100 nucleotides long (31–33). Short terminal inverted repeats have been found in *Bacillus*

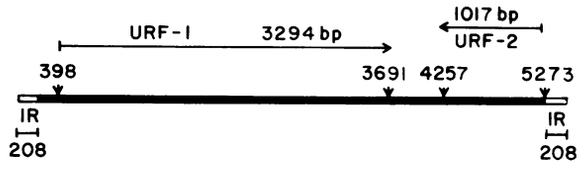

FIG. 3. Schematic map of S-2 DNA showing the location of two large unidentified reading frames (URF). Codon usage was that of the universal code. On one strand, URF-1 begins at nucleotide 398 and ends at 3,691. On the other strand, URF-2 begins at nucleotide 5,273 and ends at 4,257. IR, position of terminal inverted repeat; bp, base pairs.

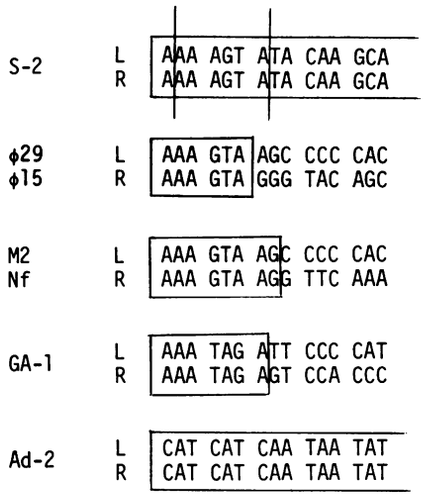

FIG. 4. Alignment of nucleotides at the termini of S-2 with five different *Bacillus* phages (ϕ, 29, ϕ 15, M2, Nf, and GA-1) and adenovirus 2 (Ad-2) DNAs. Only 5′ end sequences are shown. L and R, left end and right end, respectively. Terminal inverted repeats are indicated in boxes; S-2 and Ad-2 inverted repeats are longer than the 15 nucleotides shown. Vertical lines indicate the S-2 sequence common to the *Bacillus* phages.

phages ϕ 29, ϕ 15, Nf, M2Y, and GA-1 (34–36). Alignment of the terminal nucleotides of S-2 and the five phages indicates a high degree of homology (Fig. 4). ϕ 29 contains a six-base-pair inverted repeat sequence, A-A-A-G-T-A, which is found in the inverted repeat of S-2 beginning at the second nucleotide from the 5′ ends.

The terminal sequences of ϕ 29, adenovirus, and S-2 DNAs all are rich in A·T pairs. A+T-rich regions are needed at DNA sites where local melting of DNA is required; origins of replications of *Escherichia coli*, λ, and G4 DNAs contain such A+T rich regions (37–39).

The terminal sequences of S-2 DNA, like those of adenovirus and ϕ 29 DNAs, do not contain extensive self-complementary regions that could generate perfect hairpin loops (31–33, 35). Therefore, it seems unlikely that S-2 DNA would support a mechanism for initiation of synthesis that requires the formation of hairpin loops (40).

When adenovirus DNA replicates by the strand-displacement mechanism, daughter duplex DNA and parental single-stranded molecules are generated. The parental single-stranded DNA could hybridize to the self-complementary terminal sequences to form a "panhandle"-shaped intermediate (25). These panhandle-shaped single-stranded DNAs could initiate DNA synthesis by the same mechanism as occurs at the ends of double-stranded DNA. Because S-2 DNA contains a long terminal inverted repeat, it could form the panhandle intermediates as suggested for adenovirus DNA. Collectively, the chemical and structural similarities of S-2 termini to adenovirus and *Bacillus* phages strongly suggest that they may replicate their DNAs in an analogous fashion.

To determine the sequence of the termini, we have force-cloned S-2 terminal fragments, derived from *Pst* I digestion, into the *Sma* I and *Pst* I sites of M13mp8 and -mp9. By this procedure, clones were obtained in which the end of S-2 is blunt-end ligated to the blunt end of *Sma* I-cut vector. It is not known if the blunt-end ligation occurred *in vitro* before transformation or if ligation took place after transformation inside the bacterial cell after repair. It was reported that, even after proteinase K treatment followed by phenol and chloroform extractions, the 5′ termini of S-2 DNA could not be end labeled (17). Apparently, the 3′ ends are not modified and are not sterically impaired by the 5′ attached protein because the 3′ ends are digested with exonuclease III and are labeled with terminal transferase. This is further indicated by the fact that full-length, linear S-2 DNA has been cloned by homopolymeric tailing (41). If termini lacking a nucleotide or so are preferentially cloned, then our terminal sequence could be incomplete. In any event, we have consistently obtained sequences ending in the same nucleotide order. Additional studies will be needed to determine the chemical structure of the ends.

The expert technical assistance of Jane Suddith, Annmarie Tuttle, and C. Farrell Lanier is gratefully acknowledged. We also thank Lois Miller for her assistance in manuscript preparation. This work was supported by grants from the National Science Foundation (PCM-8010933), U.S. Department of Agriculture, Competitive Research Grants Office (8200429), and Agrigenetics. This is paper no. 8798 of the Journal Series of the North Carolina Agricultural Research Service, Raleigh.

1. Pring, D. R., Levings, C. S., III, Hu, W. W. L. & Timothy, D. H. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 2904–2908.
2. Laughnan, J. R., Gabay-Laughnan, S. & Carlson, J. E. (1981) *Stadler Genet. Symp.* **13**, 93–114.
3. Thompson, R. D., Kemble, R. J. & Flavell, R. B. (1980) *Nucleic Acids Res.* **8**, 1999–2008.
4. Spruill, W. M., Levings, C. S., III, & Sederoff, R. R. (1980) *Dev. Genet.* **1**, 363–378.
5. Weissinger, A. K., Timothy, D. H., Levings, C. S., III, & Goodman, M. M. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 1–5.
6. Levings, C. S., III, Sederoff, R. R., Hu, W. W. L. & Timothy, D. H. (1982) in *Structure and Function of Plant Genomes*, eds. Ciferri, O. & Dure, L. (Plenum, New York), pp. 363–372.
7. Laughnan, J. R. & Gabay, S. J. (1978) in *Maize Breeding and Genetics*, ed. Walden, D. B. (Wiley, New York), pp. 427–446.
8. Levings, C. S., III, Kim, B. D., Pring, D. R., Conde, M. F., Mans, R. J., Laughnan, J. R. & Gabay-Laughnan, S. J. (1980) *Science* **209**, 1021–1023.
9. Pring, D. R. & Levings, C. S., III (1978) *Genetics* **89**, 121–136.
10. Smith, H. O. (1981) *Methods Enzymol.* **65**, 371–380.
11. Messing, J. (1982) in *Genetic Engineering—Principles and Methods*, eds. Setlow, J. K. & Hollaender, A. (Plenum, New York), pp. 19–35.
12. Birnboim, H. C. (1982) *Methods Enzymol.* **100**, 243–255.
13. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
14. Shah, D. M. & Levings, C. S., III (1974) *Crop Sci.* **14**, 852–853.
15. Levings, C. S., III, & Pring, D. R. (1979) in *Physiological Genetics*, ed. Scandalios, J. G. (Academic, New York), pp. 171–193.
16. Fox, T. D. & Leaver, C. J. (1981) *Cell* **26**, 315–323.
17. Kemble, R. J. & Thompson, R. D. (1982) *Nucleic Acids Res.* **10**, 8181–8190.
18. Rekosh, D. M. K., Russell, W. C. & Bellet, A. J. D. (1977) *Cell* **11**, 283–295.
19. Carusi, E. A. (1977) *Virology* **76**, 390–394.
20. Salas, M., Mellado, R. P., Vinuela, E. & Sogo, J. M. (1978) *J. Mol. Biol.* **119**, 269–291.
21. Harding, N., Ito, J. & David, G. S. (1978) *Virology* **84**, 279–292.
22. Ito, J. (1978) *J. Virol.* **28**, 895–904.
23. Yehle, C. D. (1978) *J. Virol.* **27**, 776–783.
24. Yoshikawa, H. & Ito, J. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 2596–2600.
25. Lechner, R. L. & Kelly, T. J., Jr. (1977) *Cell* **12**, 1007–1020.
26. Ariga, H. & Shimojo, H. (1977) *Virology* **78**, 415–424.
27. Sussenbach, J. S. & Kuijk, M. G. (1977) *Virology* **77**, 140–157.
28. Inciarte, M. R., Salas, M. & Sogo, J. M. (1980) *J. Virol.* **34**, 187–199.
29. Harding, N. E. & Ito, J. (1980) *Virology* **104**, 323–338.
30. Winnacker, E. L. (1978) *Cell* **14**, 761–773.
31. Steenbergh, P. H., Maat, J., Van Ormondt, H. & Sussenbach, J. S. (1977) *Nucleic Acids Res.* **4**, 4371–4389.
32. Arrand, J. R. & Roberts, R. J. (1979) *J. Mol. Biol.* **128**, 577–594.
33. Shinagama, M. & Padmanabhan, R. (1979) *Biochem. Biophys. Res. Commun.* **87**, 671–678.
34. Yoshikawa, H., Friedmann, T. & Ito, J. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 1336–1340.

35. Escarmis, C. & Salas, M. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 1446–1450.
36. Yoshikawa, H. & Ito, J. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 2596–2600.
37. Messer, W., Meijer, M., Bergmans, H. E. N., Hansen, F. G., von Meyenburg, K., Beck, E. & Schaller, H. (1978) *Cold Spring Harbor Symp. Quant. Biol.* **43**, 139–145.
38. Moore, D. D., Denniston-Thompson, K., Kruger, K. E., Furth, M. E., Williams, B. G., Daniels, D. L. & Blattner, F. R. (1978) *Cold Spring Harbor Symp. Quant. Biol.* **43**, 155–163.
39. Hobom, G., Grosschell, R., Lusky, M., Scherer, G., Schwarz, E. & Kossel, H. (1978) *Cold Spring Harbor Symp. Quant. Biol.* **43**, 165–178.
40. Cavalier-Smith, T. (1974) *Nature (London)* **250**, 467–470.
41. Köncz, C., Sumegi, J., Udvardy, A., Racsmany, M. & Dudits, D. (1981) *Mol. Gen. Genet.* **183**, 449–458.