The EMBO Journal vol.3 no.13 pp.3063–3068, 1984

The nucleotide sequence of maize streak virus DNA

P.M.Mullineaux, J.Donson, B.A.M.Morris-Krsinich, M.I.Boulton and J.W.Davies

John Innes Institute, Colney Lane, Norwich NR4 7UH, UK

Communicated by D.A.Hopwood

The nucleotide sequence of the DNA of maize streak virus (MSV) has been determined. The data were accommodated into one DNA circle of 2687 nucleotides, in contrast to previously characterised geminiviruses which have been shown to possess two circles of DNA. Comparison of the nucleotide sequences of the DNA of MSV with those of cassava latent virus (CLV) and tomato golden mosaic virus (TGMV) showed no detectable homology. Analysis of open reading frames revealed seven potential coding regions for proteins of mol. wt. ≥10 000, three in the viral (+) sense and four in the complementary (−) sense. The position of likely transcription signals on the MSV DNA sequence would suggest a bidirectional strategy of transcription as proposed for CLV and TGMV. Nine inverted repeat sequences which have a potential of forming hairpin structures of $\triangle G \geq -14$ kcal/mol have been detected. Three of these hairpin structures are in non-coding regions and could be involved in the regulation of transcription and/or replication.

Key words: geminivirus/nucleotide sequence/maize streak virus/genome organisation/*Zea mays*

Introduction

Maize streak virus (MSV) is a member of the geminivirus group of plant DNA viruses, members of which are characterized by their twinned, (geminate) particles and genomes of circular single-stranded (ss) DNA (Goodman, 1981). The geminivirus group comprises some members which are transmitted by whiteflies, and others which are transmitted by leafhoppers (Bock, 1982).

Three whitefly-transmitted geminiviruses have been well characterized. The genomes of cassava latent virus (CLV), tomato golden mosaic virus (TGMV) and bean golden mosaic virus (BGMV) have each been shown to consist of two ssDNA circles, and the complete nucleotide sequences of the bipartite genomes of CLV and TGMV are now known (Stanley and Gay, 1983; Hamilton *et al.*, 1984). Cloned double-stranded (ds) DNAs of CLV (Stanley, 1983) and TGMV (Hamilton *et al.*, 1983) are infectious, when mechanically inoculated onto their respective hosts.

In contrast, leafhopper-transmitted geminiviruses have not been examined at the molecular level. Furthermore, these include the only plant DNA viruses known to infect monocotyledonous hosts. We have therefore determined the nucleotide sequence of MSV DNA both from virus particles and infected tissue of *Zea mays*.

Results

Sources of DNA

Double-stranded, covalently closed circular (ccc) DNA was purified from leaf tissue of *Z. mays* L. infected with MSV. One species of cccDNA was isolated from MSV-infected plants, and restriction analysis confirmed the presence of only one class of circular molecules (data not shown). In contrast, restriction and gel electrophoresis of the cccDNA from TGMV, BGMV or CLV confirmed the presence of two classes of circular molecules (Ikegami *et al.*, 1981; Hamilton *et al.*, 1983; J.Stanley, personal communication).

Virion ssDNA was used as a template to synthesize double-stranded MSV DNA *in vitro* using the Klenow fragment of DNA polymerase I and a decameric *Bam*HI linker as primer. Only one type of circle could be discerned by restriction and gel electrophoresis in these preparations.

The nucleotide sequence of MSV DNA

The sequencing strategy for MSV DNA is shown in Figure 1. Initially, sequence data were accumulated by 'shotgun' cloning of restriction fragments into M13 generated by digestion of cccDNA with *Alu*I, *Sau*3A, *Hind*III, *Pvu*II, *Bam*HI, *Bgl*II, *Sma*I, *Xho*I, *Aha*III and *Sst*I. In addition, pMSV8, a *Bgl*II clone of cccDNA in pED891 was used as a source of specific fragments in order to complete the sequence of the circle (Figure 1) [pED891 is a pBR322-based *Bgl*II vector (Brown, 1981)].

However, small sections of the sequence were ambiguous in both senses as determined by the dideoxy chain termination method. Such regions were therefore also sequenced using the chemical degradation techniques of Maxam and Gilbert (1980). The source of DNA for the latter method was a full-length M13mp9 *Bam*HI clone derived from dsDNA synthesized *in vitro*. As a consequence of correcting ambiguous sections, the complete sequence of the dsDNA generated from MSV ssDNA *in vitro* was also obtained (see Figure 1), and thus confirmed that there were no differences in the nucleotide sequence of MSV cccDNA and that from virion DNA.

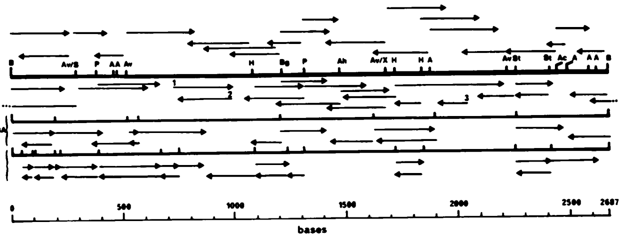

Fig. 1. DNA sequencing strategy. The thick black line represents the MSV DNA sequence. Above the line is the strategy for the sequencing by the method of Maxam and Gilbert (1980). Below the line is the strategy for sequencing for M13 clones, using the dideoxy chain termination method (see Materials and methods). In both cases the arrows denote the position, extent and direction of sequence in relation to the virion (+) sense. Restriction sites are also shown: A, *Ava*II; Ac, *Acc*I; Ah, *Aha*III; Av, *Ava*I; B, *Bam*HI; Bg, *Bgl*II; H, *Hind*III; P, *Pvu*II; S, *Sma*I; St, *Sst*I; X, *Xho*I. The restriction sites of *Alu*I and *Sau*3A are shown on separate lines. 1 and 2 are sequence data obtained from a *Hae*III clone (into M13mp9, *Sma*I site) derived from the 755 bp *Sau*3A fragment of that region. 3 is the sequence data obtained from a *Rsa*I clone (into M13mp9, *Sma*I site) derived from the 427 bp *Hind*III-*Sst*I fragment of that region.

```
GGATCCACAGAACGCCCTGTATTATCAGCCGCGGGTACCCACAGCAGCTCCGACATCCGGAGGAGTGCCGTGGAGTCGCGTAGGCGAGGTAGCTATTTTGAGCTTTGTTGCATTGATTTG
        10        20        30        40        50        60        70       C80        90       100       110       120

CTTTTACCTGCTTTACCTTTGGGTGCTGAGAGACCTTATCTTAGTTCTGAAGGCTCGACAAGGCAGATCCACGGAGGAGCTGATATTTGGTGGACAAGCTGTGGATAGGAGCAACCCTAT
       130       140       150       160       170      180C       190       200       210       220       230       240

CCCTAATATACCAGCACCACCAAGTCAGGGCAATCCCGGGCCATTTGTTCCAGGCACGGGATAAGCATTCAGCCATGTCCACGTCCAAGAGGAAGCGGGGAGATGATTCGAATTGGAGTA
       250       260       270       280       290       300       310       320     330 C       340       350       360

AGCGGGTGACTAAGAAGAAGCCTTCTTCAGCTGGGCTGAAGAGGGCTGGCAGCAAGGCCGATAGGCCATCCCTGCAAATCCAGACACTCCAGCACGCTGGGACCACCATGATAACGGTCC
       370       380       390       400       410       420       430       440       450       460       470       480

CCTCCGGAGGAGTATGTGACCTCATCAACACCTATGCCCGAGGATCTGACGAGGGCAACCGCCACACCAGCGAGACTCTGACGTACAAGATCGCCATCGACTACCACTTCGTTGCCGACG
       490       500       510       520       530       540       550       560   A   570       580       590       600

CGGCAGCCTGCCGCTACTCCAACACCGGTACCGGTGTAATGTGGCTGGTGTATGACACCACTCCCGGCGGACAAGCTCCGACCCCGCAAACTATATTTGCCTACCCTGACACGCTGAAAG
       610       620       630       640       650       660       670       680       690       700       710       720

CGTGGCCGGCCACATGGAAAGTGAGCCGGGAGCTGTGTCATCGCTTCGTGGTGAAACGGCGATGGTTGTTCAACATGGAGACCGACGGGCGCATTGGTTCGGATATTCCTCCCTCGAATG
       730       740       750       760       770       780       790       800       810       820       830       840

CAAGTTGGAAGCCTTGCAAGCGCAACATCTACTTCCACAAGTTCACGAGTGGGTTGGGAGTGAGAACGCAGTGGAAGAATGTAACGGACGGAGGAGTTGGTGCCATCCAGAGAGGAGCGC
       850       860       870       880       890       900       910       920       930       940       950       960

TGTACATGGTCATTGCCCCCGGCAATGGCCTTACTTTTACTGCCCATGGGCAGACCCGTCTGTACTTTAAGAGTGTTGGCAACCAGTAATGAATAAAACGCCGTTTTTATTATATCTGAT
       970       980       990      1000      1010      1020      1030      1040      1050      1060      1070      1080

GAATGCTGAAAGCTTACATTAATATGTCGTGCGATGGCACGAAAAACACACACAATCAATACAGGGGGGTAGTCGGCGGGCGGCTAAGGGTGGTGCTCGGCGGGCAGAACATCGAAAAAT
      1090      1100      1110      1120      1130      1140      1150      1160      1170      1180      1190      1200

CAAGATCTATCTGAATGTACTGCCTCCGTAGGAGGCAGCTCAGGGGGAGAATACCATTTCTCCCCCGGCGACATAATGTAAATGATGCAGTTTGCCTCGAAATACTCCAGCTGCCCTGGA
      1210      1220      1230      1240      1250      1260      1270      1280      1290      1300      1310      1320

GTCATTTCCTTCATCCAATCTTCATCCGAGTTGGCGAGGATTATTGTAGGCTTAGACTTCTTCTGCACCTTTTTCTTTTTACCATACTTGGGGTTTACAATGAAATCCCTCTGACAGCCA
      1330      1340      1350      1360      1370      1380      1390      1400      1410      1420      1430      1440

ACTAACTGTTTCCAACAAGGACAGAATTTAAACGGAATATCATCTACGATGTTATAGATTGCGTCTTCGTTGTATGAAGACCAATCAACATTATTTTGCCAGTAATTATGAACCCCTAGG
      1450      1460      1470      1480      1490      1500      1510      1520      1530      1540      1550      1560

CTTCTGGCCCAAGTAGATTTTCCGGTTCTTGTTGGGCCGACGATGTAGAGGCTCTGCTTTCTTGATCTTTCATCTGATGACTGGATACAGAATCCATCCATTGGAGGTCAGAAATTGCAT
      1570      1580      1590      1600      1610      1620      1630      1640      1650      1660      1670      1680

CCTCGAGGGTATAACAGGTAGGTTGAAGGAGCATGTAAGCTTCGGGACTAACCTGGAAGATGTTAGGCTGGAGCCAATCATTGATTGACTCATTACAAGTAAATCAGGTGATGAGGGTG
      1690      1700      1710      1720      1730      1740      1750      1760      1770      1780      1790  G   1800

GATGAGGATTGGTGAACTCTTCCTGAATCTCAGGAAAAAGCTTATTTGCAGAGTATTCAAAATACTGCAATTTTGTGGACCAATCAAAGGGAAGCTCTTTCTGGATCATGGAGAGGTACT
      1810      1820      1830      1840      1850      1860      1870      1880      1890      1900      1910      1920

CTTCTTTGGAAGTAGCGTGTGAAATAATGTCTCGCATTATTTCATCTTTAGAAGGCTTTTTTTCCTTTACCTCTGAATCAGATTTTCCTAGGAAGGGGGACTTCCTAGGAATGAAAGTAC
      1930      1940      1950      1960      1970      1980      1990      2000      2010      2020      2030      2040

CTCTCTCAAACACAGCCAGAGGTTCCTTGAGAATGTAATCCCTCACCCTGTTTACTGACTTGGCACTCTGAATATTTGGGTGAAACCCATTTATATCAAAGAACCTTGAGTCAGATATCC
      2050      2060      2070      2080      2090      2100      2110      2120      2130      2140      2150      2160

TTACCGGCTTCTCTGTCTGAAGCAATGCATGTAAATGCAAACTTCCATCTTTATGTGCCTCTCGGGCACATAGAATGTATTTGGGAATCCAACGAACAACGAGCTCCCAGATCATCTGAC
      2170      2180      2190      2200      2210      2220      2230      2240      2250      2260      2270      2280

AGGCGATTTCAGGATTTTCTGGACACTTTGGATAGGTTAGGAACGTGTTAGCGTTCCGGTGTGAGAACTGACGGTTGGATGAGGAGGAGGCCATTGCCGACGACGGAGGTTGAGGCTGAG
      2290      2300      2310      2320      2330      2340      2350      2360      2370      2380      2390      2400

GGATGGCAGACTGGGAGCTCCAAACTCTATAGTATACCCGTGCGCCTTCGAAATCCGCCGCTCCCTTGTCTTATAGTGGTTGCAAATGGGCCGGACCGGGCCGGCCCAGCAGGAAAAGAA
      2410      2420      2430      2440      2450      2460      2470      2480      2490      2500      2510      2520

GGCGCGCACTAATATTACCGCGCCTTCTTTTCCTGCGAGGGCCCGGTAGGGCCCGAGCGATTTGATGTAAAGTTTGGTCCTGCTTTGTATGATTTATCTAAAGCAGCCCATTCTAAAGAA
      2530      2540      2550      2560      2570      2580      2590      2600      2610      2620      2630      2640

TCCGGTCCCGGTCACTATAAATTGCCTAACAAGTGCGATTCATTCAT

      2650      2660      2670      2680   2687
```

Fig. 2. Nucleotide sequence of MSV DNA. The viral strand (+) sequence is shown starting with the nucleotide at the 5′ end of the unique *Bam*HI site. This starting point was chosen so that the reader subsequently would find it easier to compare various aspects of the analysis of the sequence with those of CLV and TGMV.

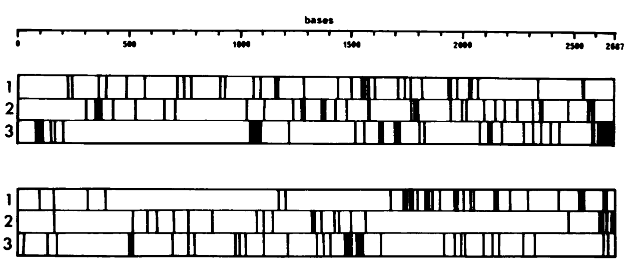

Fig. 3. Open reading regions contained within the MSV DNA. The open reading regions found for MSV DNA in the viral (+) sense and its complement (−) are shown. The nucleotide numbering is directly related to that of Figure 2, open reading frames 1, 2 and 3 begin at nucleotides 1, 2 and 3 of the sequence, respectively. Each reading frame was divided up into groups of 10 bases and shaded if it contained the second base of a stop codon.

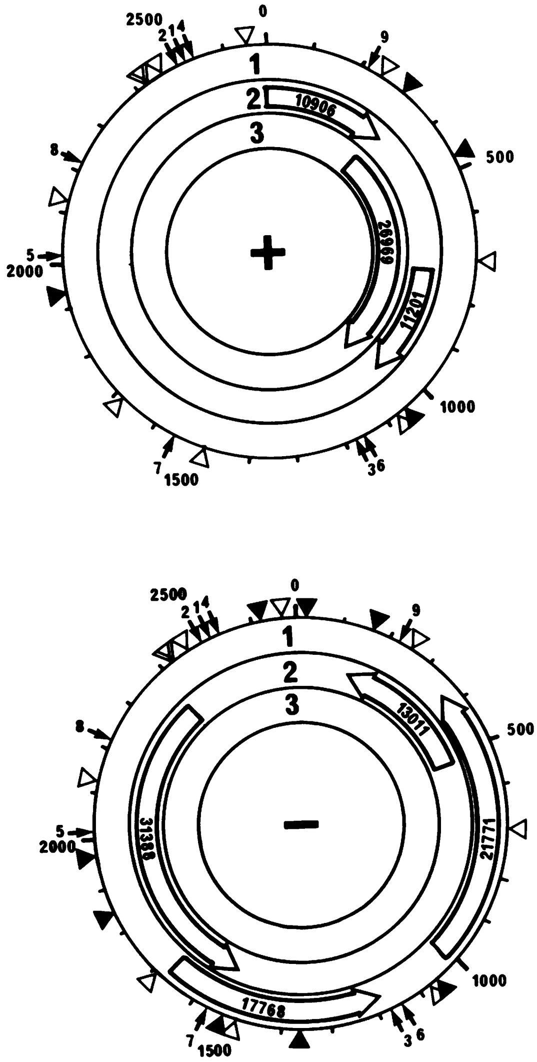

Fig. 4. Potential protein coding regions within MSV DNA in both the virion DNA sense (+) and its complement (−). Assuming that the first in-phase ATG triplet of each open reading frame of Figure 3 initiates protein synthesis, those regions with a coding capacity of mol. wt. ≥10 000 are given. Open triangles (△) indicate the position of a TATA box, and solid triangles (▲) the sequence $^{A}_{G}$ATAA. The numbered arrows (→) indicate the positions of the inverted repeat sequences which are potentially capable of forming hairpin structures with a $\triangle G \geq -14$ kcal/mol. The numbering system for the inverted repeats is that of Figure 5.

Table I. Positions of potential genes

Product mol. wt.	Reading frame	Start	Stop	TATA	$^{A}_{G}$ATAA
10 906	2+	2686	302	2656	301
26 969	3+	315	1047	247	1052
11 201	2+	734	1028	692	1052
31 388	2−	2374	1558	2431	1535
17 768	1−	1661	1202	1693	1071
21 771	1−	1007	389	1074	160
13 011	2−	469	163	695	160

Finally, some sequencing by the methods of Maxam and Gilbert (1980) was carried out directly on the *Hae*III digested single-stranded virion DNA to determine the polarity of the consensus sequence. During the cloning experiments, no viral cloned DNA was found which did not correspond to the known viral DNA species described above. In addition, all derived sequence data could be accommodated into one circle.

The nucleotide sequence of MSV DNA is shown in Figure 2. The sequence is presented in the virion (+) sense and contains 2687 nucleotides. Approximately 98% of the sequence was determined in both orientations. Although the entire sequence was determined from one preparation each of cccDNA and single-stranded virion DNA, a number of nucleotide variations were encountered, and are shown in Figure 2 as an alternative nucleotide below the sequence. None of these sequence variations had any effect on our interpretation of the sequence in terms of open reading frames. MSV has no known local lesion host through which we could passage the isolate, thus a slightly mixed population of molecules might be expected.

Potential genes

To investigate the potential coding capacity of the MSV DNA, the sequence was screened in all three reading frames for potential genes. The results of this, both for the viral (+) nucleotide sequence (Figure 2) and its complement (−), are presented in Figure 3. Based on these data, and assuming there is no post-transcriptional or post-translational processing, a proposal for a number of virus-specific proteins is suggested in Figure 4. The precise location of these potential genes on the sequence is given in Table I. It was assumed that the first ATG triplet in each potential gene would initiate protein synthesis, and only those regions with a potential coding capacity of mol. wt. ≥10 000 are included in this figure.

Figure 4 shows that, when read in the virion (+) sense, the sequence of MSV DNA could encode three proteins of mol. wts. 26 969, 11 201 and 10 906. When read in the complementary (−) sense, 75% of the sequence could be involved in protein coding, to give products of mol. wts. 31 388, 21 771, 17 768 and 13 011. The region encoding a 26 969 mol. wt. product on the viral sense sequence (see Figure 4) may be the coat protein gene. Firstly, the coat protein of MSV has an estimated size of 28 000 mol. wt. (Bock *et al.*, 1977), and secondly the calculated amino acid contents from the sequence data and the experimentally determined amino acid contents closely agreed (M.Short, personal communication). No other potential gene product was in agreement with the experimental data described above.

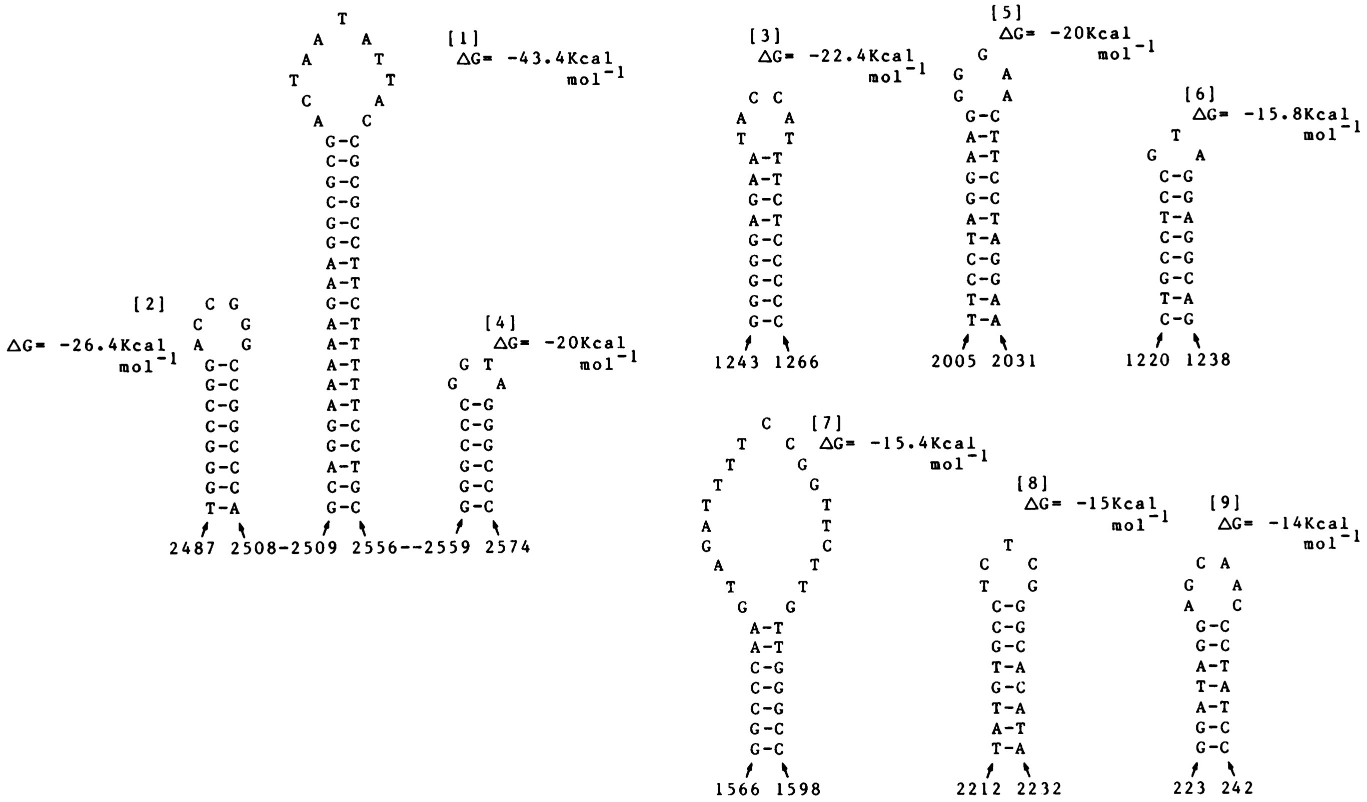
[1] ΔG= -43.4Kcal mol^{-1}
[2] ΔG= -26.4Kcal mol^{-1}
[3] ΔG= -22.4Kcal mol^{-1}
[4] ΔG= -20Kcal mol^{-1}
[5] ΔG= -20Kcal mol^{-1}
[6] ΔG= -15.8Kcal mol^{-1}
[7] ΔG= -15.4Kcal mol^{-1}
[8] ΔG= -15Kcal mol^{-1}
[9] ΔG= -14Kcal mol^{-1}
2487 2508-2509 2556--2559 2574
1243 1266
2005 2031
1220 1238
1566 1598
2212 2232
223 242

Fig. 5. Potential hairpin structures on MSV DNA. The inverted repeat sequences shown may be capable of forming hairpin structures with a free energy ($\triangle G$) of ≥ -14 kcal/mol, as calculated by the rules of Tinoco *et al.* (1973). The structures are numbered in order of potential stability, and their calculated free energies are shown adjacent to them. The coordinates refer to the sequence of Figure 2, and the positions of these potential hairpin structures are shown in Figure 4.

Non-coding regions

There are two regions of the MSV sequence, from nucleotides 1047 to 1202 and 2374 to 2686, which are not part of any putative gene in either sense.

Taking a free energy of $\triangle G$ as ≥ -14 kcal/mol (the free energy of the primosome assembly site of ϕX174; Arai and Kornberg, 1981) as a base line, nine stem-loop structures were identified on the sequence with a greater value of $\triangle G$ (Figure 5). The $\triangle G$ of these potential hairpin structures was calculated according to the rules of Tinoco *et al.* (1973). Three of these potential hairpin structures are located in the two non-coding regions described above (hairpin numbers: 1, 2, and 4; Figures 4 and 5), and could be involved in functions such as replication, regulation of transcription and sites for the assembly of coat protein. Some or all of the hairpins in Figure 5 (plus a number of others with lower potential free energy) may have consequences for the conformation of MSV ssDNA.

Possible transcriptional control signals

The best characterized DNA sequence of eukaryotic promoters is the TATA box of Goldberg and Hogness (see Proudfoot, 1979), which is associated with transcription both *in vitro* and *in vivo* (Breathnach and Chambon, 1981; Messing *et al.*, 1983). The complete consensus sequence for promoters from animal genes is TATA$^{A}_{T}$A$^{A}_{T}$ (Breathnach and Chambon, 1981) and from plant genes is TT$^{C}_{G}$TATA$^{T}_{A}$A$_{1-3}$ A, based primarily on sequences from dicotyledonous species, but including information from the zein multigene family (Messing *et al.*, 1983). Nevertheless, information on the transcriptional signals from monocotyledonous species is so limited that we were constrained to search only for the 'core' TATA box in the MSV DNA sequence. The results of this search are shown in Figure 4 and Table I. With the exception of the region encoding for the 13 011 mol. wt. peptide on the complementary sense, all the remaining potential genes possess TATA sequences within 100 bp of the beginning of the first ATG. Also, AT-rich centres, rather than the TATA box, could be involved in promoter sequences. For example, in the DNA sequence of the maize transposable element, Mu 1, there are no core TATA sequences, although an AT-rich centre has been cited as containing possible promoter sequences (Barker *et al.*, 1984).

Consensus polyadenylation signals of plant genes, $^{A}_{G}$ATAA (Messing *et al.*, 1983) are also shown in Figure 4. All the potential gene products, except that coding for the 21 771 mol. wt. protein on the complementary sense, have a polyadenylation signal located within 200 bp of the termination codon.

Comparison with CLV and TGMV

A comparison of the nucleotide sequence of MSV DNA with those of circles 1 and 2 of CLV (Stanley and Gay, 1983) and circles A and B of TGMV (Hamilton *et al.*, 1984) revealed no detectable homology. The parameters were a block size of 11, and a score of 7 using the DIAGON program (Staden, 1982). These parameters allowed the programme to display the homologies reported between the sequences of CLV and TGMV (Hamilton *et al.*, 1984).

The number of potentially stable hairpin structures (see Figures 4 and 5) compared with one each reported for CLV 1 (Stanley and Gay, 1983) and TGMV A (Hamilton *et al.*, 1984) does not allow the assignment to the MSV DNA sequence of a segment of DNA that may be involved in replication, such as has been suggested for the common regions of circles 1 and 2 of CLV (Stanley and Gay, 1983) and circles A and B of TGMV (Hamilton *et al.*, 1984). However, the organization of the potential genes and their associated promoter signals (Figure 4) suggests a similar bidirectional strategy for transcription as proposed for CLV and TGMV (Stanley and Gay, 1983; Hamilton *et al.*, 1984).

Discussion

The determination of the MSV DNA sequence reveals several novel features. The identification of only one circle of MSV DNA in virus particles and infected tissue of *Z. mays* L., is the most striking difference between MSV and the other characterized geminiviruses, which have two circles of DNA (Stanley and Gay, 1983; Hamilton *et al.*, 1984). Furthermore, there is no detectable sequence homology between MSV and DNA circles 1 and 2 of CLV and circles A and B of TGMV. In contrast there is 60% homology between CLV 1 and TGMV A, and 39.6% homology between CLV 2 and TGMV B (Hamilton *et al.*, 1984). This is also reflected by the serological relationships of the virus particles (Roberts *et al.*, 1984).

More inverted repeat sequences, capable of forming stable hairpin structures, are found in the MSV sequence compared with either CLV or TGMV, which may reflect differences in replication, regulation of transcription and conformation of the virion DNAs. This may be a consequence of the specificities of CLV and TGMV for their dicotyledonous hosts and MSV for its monocotyledonous hosts, or adaption to whitefly and leafhopper insect vectors, respectively.

The positions of transcriptional signals would suggest that like CLV and TGMV, MSV has bidirectional transcription, although the mechanisms of transcriptional regulation could still be different. Once more is known about the products of the potential coding regions it will be interesting to reflect on how the organization of the potential genes and putative transcriptional signals relates to the lack of sequence homology or serological relationship between MSV and CLV or TGMV. The construction of infective double-stranded clones of TGMV and CLV have allowed the nucleotide sequence of the genomes of these two viruses to be defined (Stanley, 1983; Stanley and Davies, 1984; Hamilton *et al.*, 1983, 1984). We find that MSV, like most other leafhopper-transmitted geminiviruses (perhaps excluding beet curly top virus), is not mechanically transmissable, either as whole virus or viral DNA. Intra-haemocoelic injection of virus into *Cicadulina mbila* (Naudé), a leafhopper vector for MSV, results in systemic infection. However we have not been able to demonstrate infectivity with virion or cloned DNA (results not shown). We believe, however, that we have identified and exhaustively sequenced all the components of MSV DNA, with all sequence obtained fitting into the one circle, and that the lack of infectivity is a failure to deliver intact viral DNA to the site where infection begins, rather than the absence of a minor and as yet undiscovered component. In view of dissimilarities with CLV and TGMV it has not escaped our notice that MSV being the type member of a group containing these two viruses (Matthews, 1982) may be misleading.

Determination of the nucleotide sequence of MSV DNA is a prerequisite for a more detailed understanding of gene organization and function of a DNA virus which is restricted to monocotyledonous plants, and thus may be useful in gaining some insight into aspects of the molecular biology of these hosts. In addition, provided that the problem of infectivity can be overcome, the sequence may be of use in the development of MSV as a vector for the introduction of chimaeric DNA into some important monocotyledonous crop plants.

Materials and methods

Materials

MSV (Nigerian isolate) was propagated in *Z. mays* L. var. Golden Cross Bantam, following transmission by *C. mbila* (Naudé). *Escherichia coli* DNA polymerase I (large fragment) and calf intestinal phosphatase were purchased from Boehringer, and polynucleotide kinase from PL Biochemicals. Restriction enzymes were obtained from New England Biolabs or Bethesda Research Laboratories. Radiochemicals were from Amersham International or New England Nuclear.

Isolation of MSV DNA

Virus particles and virion DNA were isolated according to the methods of Harrison *et al.* (1977). Supercoiled dsDNA was isolated from infected leaf tissue of *Z. mays* L. by the method of Sunter *et al.* (1984). The production of *in vitro* dsMSV DNA from virion ssDNA was by the method of Stanley and Gay (1983) using a synthetic decameric *Bam*HI linker (BRL) as a primer. Restriction fragments from RF DNA or *in vitro* synthesized dsDNA were ligated into the appropriately linearized M13mp vectors of Messing and Vieira (1982). Recombinant phages were identified by the *lac* complementation assay of Benton and Davies (1977). Bacteriophage isolation and DNA extraction were carried out as described by Sanger *et al.* (1980). Sequencing of this cloned DNA was performed using the dideoxy chain termination method of Sanger *et al.* (1977) with [α-^{32}P]dATP (800 Ci/mmol) using the 17-mer M13 primer of Duckworth *et al.* (1981). The products were subjected to electrophoresis on 6% (w/v) polyacrylamide gels (Sanger and Coulson, 1978) or 6% (w/v) buffer gradient gels (Biggin *et al.*, 1983). The gels were then fixed, dried and subjected to autoradiography (Biggin *et al.*, 1983).

Another method used was to sequence the products of restriction digests labelled at the 5′ termini by the chemical degradation techniques of Maxam and Gilbert (1980). The products of these sequencing reactions were subjected to electrophoresis in 6% (w/v), 8% (w/v) or 20% (w/v) polyacrylamide gels (Maxam and Gilbert, 1980) and treated as described above. The source of restriction fragments for this sequencing method was a full length *Bam*HI clone of *in vitro* dsDNA into the *Bam*HI site of M13mp9; or ss viral DNA fragments generated by restriction with *Hae*III or *Taq*I. Sequence information derived from the above methods was stored, assembled and analysed using the computing methods of Staden (1980).

Acknowledgements

We would like to thank Dr H.W.Rossell for the gift of the MSV (Nigerian) isolate; Drs John Stanley, George Lomonossoff and Rod Townsend for interesting discussion throughout the course of this work; Dr Angus Hepburn for advice on some aspects of computing, Dr Margaret Short for the amino acid analysis of the coat protein and Dr P.Markham for leafhopper transmission expertise. This work was funded by Agrigenetics Research Corporation, and carried out under MAFF licence number 49/152.

References

Arai,K. and Kornberg,A. (1981) *Proc. Natl. Acad. Sci. USA*, **78**, 69-73.

Barker,R.F., Thompson,D.V., Talbot,D.R., Swanson,J. and Bennetzen,J.L. (1984) *Nucleic Acids Res.*, **12**, 5955-5967.

Benton,W.D. and Davis,R.W. (1977) *Science (Wash.)*, **196**, 180-182.

Biggin,M.D., Gibson,T.J. and Hong,G.F. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 3963-3965.

Bock,K.R. (1982) *Plant Dis.*, **66**, 266-270.

Bock,K.R., Guthrie,E.J., Meredith,G. and Barker,H. (1977) *Ann. Appl. Biol.*, **85**, 305-308.

Breathnach,R. and Chambon,P. (1981) *Annu. Rev. Biochem.*, **50**, 349-383.

Brown,A. (1981) Ph.D. Thesis, University of Edinburgh.

Duckworth,M.L., Gait,M.J., Goelet,P., Hong,G.F., Singh,M. and Titmas, R.C. (1981) *Nucleic Acids Res.*, **9**, 1691-1706.

Goodman,R.M. (1981) in Kirkstak,E. (ed.), *Handbook of Plant Virus Infections and Comparative Diagnosis*, Elsevier, Amsterdam, Holland, pp. 879-910.

Hamilton,W.D.O., Bisaro,D.M., Coutts,R.H.A. and Buck,K.W. (1983) *Nucleic Acids Res.*, **11**, 7387-7396.

Hamilton,W.D.O., Stein,V.E., Coutts,R.H.A. and Buck,K.W. (1984) *EMBO J.*, **3**, 2197-2205.

Harrison,B.D., Barker,H., Bock,K.R., Guthrie,E.J., Meredith,G. and Atkinson,M. (1977) *Nature*, **270**, 760-762.

Ikegami,M., Haber,S. and Goodman,R.M. (1981) *Proc. Natl. Acad. Sci.*

USA, **78**, 4102-4106.

Matthews,R.E.F. (1982) *Intervirology*, **17**, No. 1-3.

Maxam,A.M. and Gilbert,W. (1980) *Methods Enzymol.*, **65**, 499-560.

Messing,J. and Vieira,J. (1982) *Gene*, **19**, 269-276.

Messing,J., Geraghty,D., Heidecker,G., Hu,N.T., Kridl,J. and Rubenstein, I. (1983) in Kasuge,T., Meredith,C.P. and Hollaender,A. (eds.), *Genetic Engineering of Plants*, Plenum Press, NY, pp. 211-227.

Proudfoot,N.J. (1979) *Nature*, **279**, 376.

Roberts,I.M., Robinson,D.J. and Harrison,B.D. (1984) *J. Gen. Virol.*, in press.

Sanger,F. and Coulson,A.R. (1978) *FEBS Lett.*, **87**, 107-110.

Sanger,F., Nicklen,S. and Coulson,A.R. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5463-5467.

Sanger,F., Coulson,A.R., Barrell,B.G., Smith,A.J.H. and Roe,B.A. (1980) *J. Mol. Biol.*, **143**, 161-178.

Staden,R. (1980) *Nucleic Acids Res.*, **8**, 3673-3694.

Staden,R. (1982) *Nucleic Acids Res.*, **10**, 2951-2961.

Stanley,J. (1983) *Nature*, **305**, 643-645.

Stanley,J. and Davies,J.W. (1984) in Davies,J.W. (ed.), *Molecular Plant Virology*, C.R.C. Press, Boca Raton, FL, USA.

Stanley,J. and Gay,M.R. (1983) *Nature*, **301**, 260-262.

Sunter,G., Coutts,R.H.A. and Buck,K.W. (1984) *Biochem. Biophys. Res. Commun.*, **118**, 747-752.

Tinoco,I., Borer,P.N., Dengler,B., Levine,M.D., Uhlenbeck,O.C., Crothers, D.M. and Gralla,J. (1973) *Nature, New Biol.*, **246**, 40-41.

Received on 17 September 1984

Note added in proof

Since this paper was accepted, the complete sequence of a Kenyan isolate of MSV has been published. [Howell, (1984), *Nucleic Acids Res.*, **12**, 7359-7375.] It is curious that the virion (+) sense sequence of this isolate shows >99% homology with the complementary (−) sense of the nucleotide sequence of MSV (Nigerian isolate) shown in this paper.