Volume 16 Number 1 1988 Nucleic Acids Research

Infectivity and complete nucleotide sequence of the genome of a South African isolate of maize streak virus

Sondra G.Lazarowitz

Department of Embryology, Carnegie Institution of Washington, 115 W. University Parkway, Baltimore, MD 21210, USA

Received August 25, 1987; Revised and Accepted November 10, 1987 Accession no. Y00514

ABSTRACT

The complete infectious genome of a South African isolate of the geminivirus maize streak (MSV-S) has been cloned, characterized, and sequenced. Using an *A. tumefaciens* Ti plasmid delivery system, the cloned ~2.7 kb single circular MSV component was shown to be necessary and sufficient for infection of maize. Based on sequence analysis of the infectious clone, MSV-S is highly homologous to the previously characterized Kenyan and Nigerian isolates. While the genomic organization of MSV-S has elements in common with each of these previously characterized isolates, it is identical to neither and its analysis addresses the discrepancies between them. The result is a somewhat simplified and unified picture of the viral genome, the structural organization of which is essentially identical to that of wheat dwarf virus.

INTRODUCTION

The small single-stranded DNA-containing plant viruses known as geminiviruses are transmitted in nature either by whiteflies or by leafhoppers. Members of these two subgroups are strikingly different in at least two biological properties. Whitefly-transmitted geminiviruses in general have restricted host ranges, these restrictions not being imposed by the feeding habits of the insect vector (1,2). Cloning and sequence analyses have established that the genomes of these viruses are bipartite, consisting of two single-stranded DNA components of ~2.8 kb each (3-6). The two DNA components in any one bipartite virus are unique in sequence except for an ~200 base region of sequence identity called the Common Region. In contrast to these host range and genomic characteristics of the whitefly-transmitted geminiviruses, the leafhopper-transmitted geminiviruses have surprisingly broad host ranges and contain only a single genomic component of ~2.7-2.9 kb (7-10). The cloning and sequencing of beet curly top virus has formally demonstrated that the single genomic component is characteristic of the leafhopper-transmitted viruses and not a monocot-dicot virus difference (10).

Direct inoculation of plants with cloned beet curly top virus (BCTV) DNA (10) or with *A. tumefaciens* carrying a Ti plasmid and cloned maize streak

virus (MSV) DNA on a binary vector (11) has demonstrated that the single DNA component of the leafhopper-transmitted geminiviruses does constitute the complete viral genome. While the complete nucleotide sequence has been reported for a Kenyan (MSV-K) (7) and a Nigerian (MSV-N) (8) isolate of MSV, the biological relevance of these published sequences remains in question in view of major differences between the two sequences and insufficient data on the relationship of these sequences to infectious clones of the viral DNAs. Infectivity of the cloned MSV-K DNA (7,12) has not been demonstrated, and the reported sequence for MSV-N (8) may not be that of an infectious clone since it is a consensus derived from different clones one of which is not infectious (M.I. Boulton, personal communication).

I report here the characterization and complete nucleotide sequence of an infectious clone of the genome of MSV-S, a South African isolate of MSV (13). Although highly homologous to MSV-N and MSV-K, there are significant differences from each of the published sequences. The picture which emerges for the organization of the genome of this biologically functional clone of MSV-S DNA is simpler than previous reports would suggest and is strikingly similar to the genomic organization of wheat dwarf virus (WDV) (9).

MATERIALS AND METHODS

Preparation of nucleoprotein-containing extracts. The MSV-S isolate (form A), originally obtained from M.C. Walters, Potchefstroom, South Africa, has been maintained for several years via insect transmission by V.D. Damsteegt in containment facilities at Frederick, MD (13). Leaves from MSV-infected *Zea mays* (var. XL43 or Golden Bantam) were frozen under liquid N_2, crushed in a mortar and pestle, and powdered in a blender. Extracts containing nucleoprotein particles, which would include viral replicating complexes as well as virions, were prepared as described for tomato golden mosaic virus (TGMV) DNA components (14) with all steps being carried out at 4°C. In brief, powdered tissue was extracted for 16 h in 0.5 M KH_2PO_4, pH 7.0 - 0.75% Na_2SO_3 (2 ml/g wet wt. tissue) and clarified by filtration through Miracloth (Calbiochem) followed by centrifugation for 10 min at 12,000 x g in a Sorvall SS34 rotor. The resulting supernatant was centrifuged for 4.5 hr at 100,000 x g in a Beckman 50Ti rotor and the pelleted nucleoprotein particle fraction was resuspended in 0.02 M Tris-HCl, pH 8.0 - 0.02 M EDTA - 0.5% SDS.

Purification of Viral DNA. Nucleoprotein particle-containing extracts from infected maize tissue were deproteinized by extraction with phenol followed by two extractions with 1:1 phenol:chloroform-1.% isoamyl alcohol. DNA was

banded in ethidium bromide-containing CsCl gradients, followed by gel electrophoresis in Tris-borate-EDTA, pH 8.3 (15). Recovery of DNA was by electroelution onto DEAE-cellulose paper (16), and incubation at 68°C for 1 hr in 0.01M Tris-HCl, pH 7.4 - 1.M NaCl. The recovery was 70%-90%.

Purification of MSV Virions. Virions were purified from symptomatic leaves harvested from infected maize at 18-22 da. after inoculation with viriliferous leafhoppers (*Cicadulina mbila*). Extracts were prepared following published procedures (17), virions being banded in linear 10.% - 40.% (w/w) sucrose gradients in 0.01 M Tris-HCl, pH 7.4 - 0.1 M NaCl - 0.001 M EDTA.

Cloning of MSV DNA. Viral double-stranded DNA (dsDNA) isolated from extracts of leafhopper-inoculated maize tissue was restricted with either BamHI or SalI and cloned into pEMBL plasmids (18) using standard techniques. Host cells for cloning were *E. coli* 71/18 (18), JM101 (19), or MM294 (20).

Single-strand specific probes were synthesized from recombinant single-stranded DNA (ssDNA) clones of MSV in pEMBL plasmids. ssDNA was purified from phage particles following superinfection of cells with *f1* phage (18). Probe complementary to the inserted MSV sequences was synthesized using the standard sequencing primer located 5' to the cloned insert, and was denatured prior to hybridization. A probe specific for hybridization to the MSV sequences contained in the ssDNA clone was synthesized using a primer located 3' to the cloned insert under conditions such that synthesis would proceed partially around the circle and not into the inserted viral sequences (21). This partially single-stranded probe was used undenatured and hence hybridization was dependent on the accessible single-stranded MSV sequences.

Infectivity of cloned MSV DNA was assayed in maize using an *Agrobacterium tumefaciens* delivery system (11) with a binary vector. MSV DNA was cloned into the transintermediate vector pMON505 (22), a pMON200-derived plasmid (23) containing a wide host range origin, and was introduced into *A. tumefaciens* carrying the nopaline-type plasmid pTiT37SE (*ASE*) by the triparental mating procedure with pRK2013 as helper plasmid and appropriate antibiotic selection (23). The structure of the MSV-containing plasmid in *ASE* was verified by hybridization of blots of restriction digests with MSV and pMON505 probes.

For injection of plants, cultures from a fresh colony were grown with antibiotic selection at 28°C for 24-30 hr until near stationary phase. These were diluted 1:20, grown a further 16 hr at 28°C, and concentrated 20-fold just prior to inoculation of plants. Using a 0.3 ml insulin syringe (26 gauge), 8-to-10 day old maize seedlings were inoculated with 20 µl of the appropriate concentrated *ASE* culture by injection into the meristematic region

just above the coleoptile. Infectivity studies were done in the limited access containment greenhouse facilities of the Foreign Disease Research Laboratory at Frederick, MD. under a permit from APHIS. This is a sealed Fiberglas facility with self-contained air supply (24). Biological material and material in contact with plants was autoclaved prior to disposal. **Sequencing of MSV DNA** was accomplished through a combination of the chain termination (25) and Maxam-Gilbert (26) methods. Dideoxy products were resolved on buffer gradient gels (27). Ambiguities were resolved by formamide sequencing gels (26), dITP (28) and/or Maxam-Gilbert sequencing. The complete sequence was confirmed using synthetic oligonucleotide primers to sequence across the viral insert of two independent clones by the dideoxy method. Four independent clones of MSV were used to determine the sequence. Both strands were sequenced in their entirety averaging 6 reads/nucleotide sequenced. The database was constructed and analyzed using the programs of Staden (29,30). Synthetic oligonucleotides were synthesized on an ABI380A synthesizer (Applied BioSystems) and purified by HPLC prior to use.

RESULTS

Identification of MSV DNA in Infected Tissue. Nucleic acid-containing extracts from uninfected and MSV-infected maize were digested with ribonuclease A and analyzed on agarose gels. Two ethidium bromide staining bands (designated II and III) were specifically present in extracts from infected plants and absent from the control uninfected plants (fig. 1a). These species were sensitive to digestion with deoxyribonuclease I and resistant to digestion with S1 nuclease, demonstrating that they were dsDNA (not shown). To determine whether these were two distinct DNAs or different mobility forms of a single unique species, these DNAs were partially purified on gradients and gels, and digested with restriction endonucleases.

Digestion with SalI or BamHI (not shown) converted DNA II to a unique ~2.7 kb linear molecule comigrating with DNA III, but these two enzymes did not alter the mobility of III. Digestion with SacI converted both II and III to a unique ~2.5 kb DNA, suggesting the presence of two SacI sites very close together (not shown). Definitive evidence for the relationship of these DNAs was obtained by restriction with the multicut enzymes HindIII and TaqI (fig. 2). HindIII digestion of II and III generated the same unique ~2.0 kb and ~0.6 kb fragments. Furthermore, these two DNA fractions clearly contained a single major species with the identical TaqI restriction pattern, the fragments again adding up to ~2.7 kb. Taken together with the results

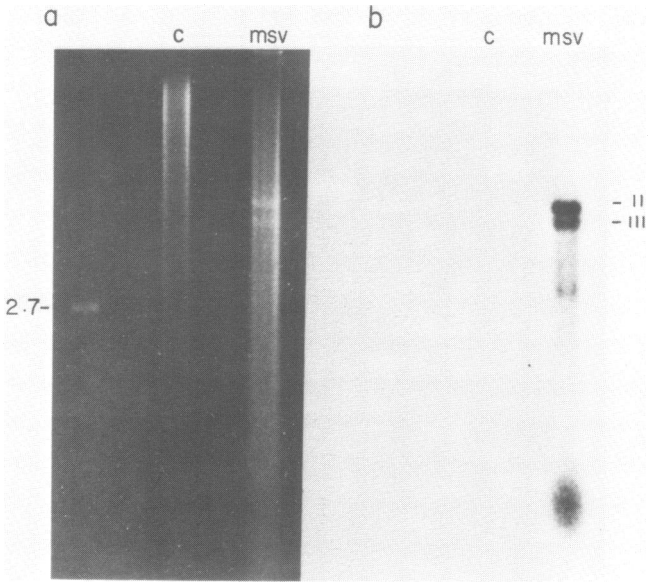

Fig. 1. Agarose gel analysis of extracts from MSV-infected (msv) and uninfected (C) maize tissue. (a) Ethidium bromide stained gel. (b) Southern blot of gel shown in (a). All samples were digested with RNase A prior to analysis. II and III mark positions of relaxed circular and linear dsDNA (see text). Marker is a supercoiled 2.7 kb plasmid DNA. Minor amounts of MSV ssDNA and supercoiled dsDNA can be seen in (b) migrating just ahead of and with the 2.7 kb marker, respectively. See fig. 5 for reference.

obtained with SalI and BamHI, II and III are thus the randomly-nicked relaxed-circular and linear forms, respectively, of the same unique ~2.7 kb DNA, an interpretation consistent with their relative mobilities. This DNA appears to be specific to MSV-infected maize (fig. 1a).

Cloning MSV DNA. Based on restriction analyses the DNA II fraction, although containing some contaminating degraded chromosomal DNA, was sufficiently purified (~50% unique DNA) to attempt to clone the presumed MSV DNA. Using either BamHI or SalI, a few hundred recombinant clones were obtained and based on restriction analyses all of those examined contained the same ~2.7 kb insert having the map deduced for the form II DNA in infected maize (see fig 7 for reference). There are unique BamHI and SalI sites, and two SacI sites ~200 bp apart. There are three HindIII sites, giving fragments of ~1950, ~630, and ~120 bp. The first two were identified in digests of the form II DNA (fig. 2). The smallest HindIII fragment was not detected due to the high background staining in the tissue extracts.

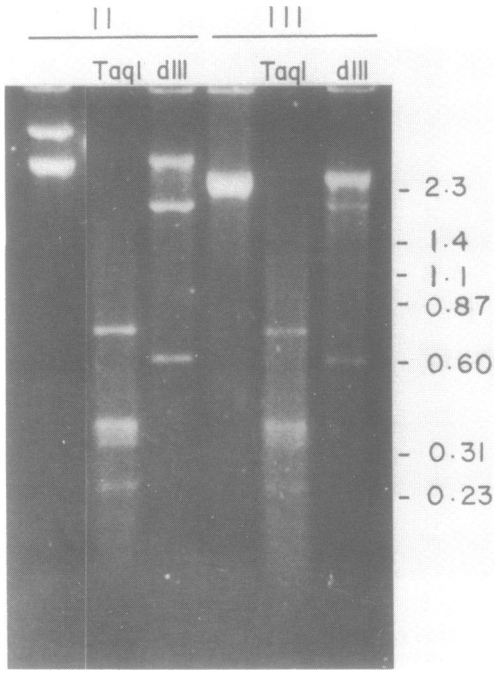

Fig. 2. Restriction analysis of DNA forms II and III purified from MSV-infected maize. Gel purified DNA (see fig. 1) was restricted with TaqI or HindIII (dIII). Undigested DNA was included for reference (unmarked lanes). The slower migrating and undigested DNA is contaminating degraded chromosomal DNA which co-purified with the viral DNA and was resolved under the conditions of electrophoresis used here.

The cloned DNA was identified as MSV by hybridization to Southern blots (31) of dsDNA in extracts from infected and uninfected plants, or of ssDNA prepared from purified virions (figs. 1b and 3). Nick-translated (32) cloned MSV dsDNA hybridized specifically to forms II and III dsDNA in infected tissue extracts, as well as to minor species with the mobilities of geminivirus ssDNA and supercoiled dsDNA (fig. 1b). It did not hybridize to any species in extracts from uninfected maize. A single strand specific probe (21) complementary to the viral insert in the particular recombinant clone used did not hybridize to the MSV virion ssDNA, although it did hybridize to dsDNA II and III in the infected maize extracts (fig. 3a), the latter being a control for differences in the specific activities of the probes used. The partially single-stranded probe made from this same ssDNA clone and specific for hybridization to the inserted viral sequences did hybridize to the MSV virion

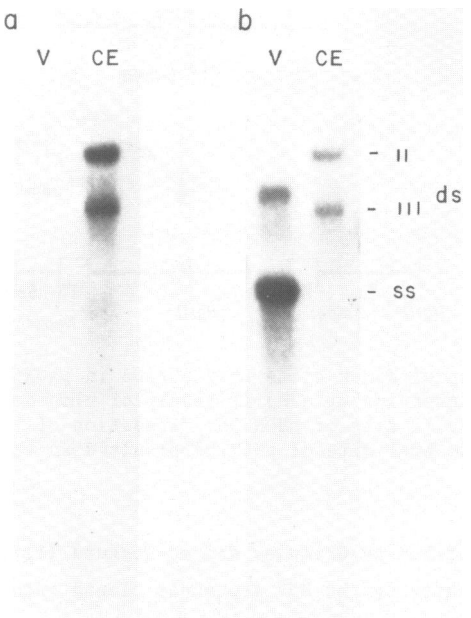

Fig. 3. Hybridization of blots of DNA isolated from MSV virions (V) or from MSV-infected maize (CE) with single strand specific probes of cloned MSV DNA. (a) Probe was synthesized from ssDNA clone pMSV7 using the standard sequencing primer located 5' to the MSV insert. (b) Probe was synthesized from pMSV7 using a primer located 3' to the MSV inserted sequences such that hybridization was dependent on the single-stranded inserted viral sequences. The identities of ssDNA and dsDNA (II and III) were determined by S1 nuclease and restriction analyses (not shown).

ssDNA, as well as to the forms II and III dsDNA (fig 3b). Thus, the cloned DNA is that of MSV. Furthermore, these analyses distinguished the cloned sequences as being the "+" strand virion DNA or the complementary "-" strand DNA. The ssDNA clone used for probes in fig. 3 contains the viral "-" strand.

These results identify the cloned DNA as that of MSV, suggesting that the viral genome is a single DNA component of ~2.7 kb. In the initial absence of infectivity data, four independent clones were selected for further study to assure that they were complete. These clones were judged to be independent based on having been cloned using two different restriction enzymes (SalI or BamHI) and isolated in both possible orientations.

Infectivity of cloned MSV DNA. Definitive evidence for the cloned ~2.7 kb DNA being the complete functional genome of MSV could be obtained only by demonstrating its infectivity. To do this, I used the agroinoculation

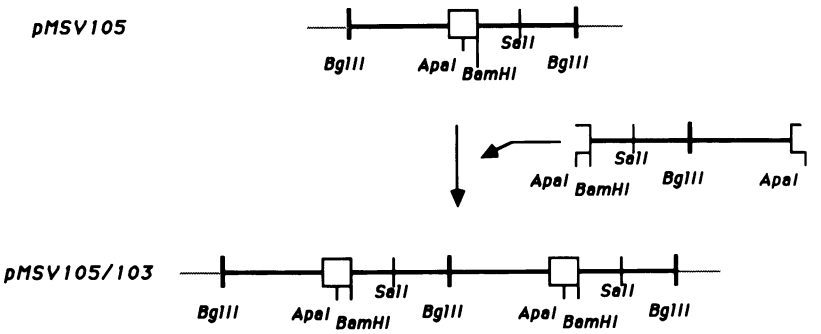

Fig. 4. Construction of dimer clone pMSV105/103 in pMON505. pMSV105 was constructed by the insertion of a BglII linear of cloned MSV-S DNA (see fig. 7) into the unique BglII site of pMON505. Insertion of the ApaI linear of MSV-S into the unique ApaI site of pMSV105 resulted in the construction of the dimer clone.

technique first described by Grimsley and co-workers (11), inoculating seedlings with *ASE* carrying the MSV sequences cloned into a binary vector.

A direct tandem repeat of two copies of MSV DNA was cloned into the

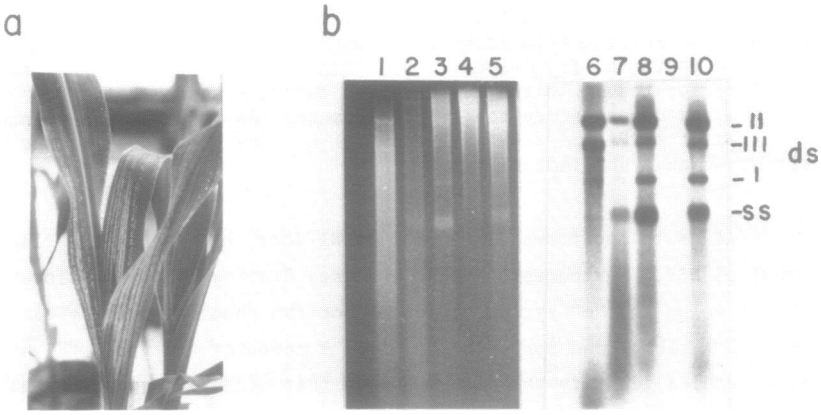

Fig. 5. Infectivity of *pMSV105/103(ASE)*. (a) *Zea mays* (var. Golden Bantam) injected with *pMSV105/103(ASE)* showing typical symptoms of maize streak disease. (b) agarose gel analysis of extracts from symptomatic and nonsymptomatic maize. (1-5) Ethidium bromide stained gel. (6-10) Southern blot of 1-5. All samples were digested with RNase A prior to analysis. (1,6) Leafhopper-inoculated symptomatic XL43. Original extract from which MSV-S was cloned (see fig. 1). (2,7) Leafhopper-inoculated symptomatic XL43. Independent extract showing presence of MSV ssDNA. (3,8;5,10) Separate symptomatic plants of Golden Bantam which were injected with *pMSV105/103(ASE)*. (4,9) Nonsymptomatic Golden Bantam. Identities of ssDNA and dsDNA (I,II,III) were based on S1 nuclease and restriction analyses (not shown).

TABLE 1
Infectivity as Assayed by the Inoculation of Maize

Inoculum	Host	Method of Inoculation	Number of Symptomatic Plants[1]
Experiment 1:			
pMSV105/103(ASE)	GB[2]	injection[3]	17/34 (50.%)
		abrasion	0/10
	XL43	injection	14/38 (37.%)
		abrasion	0/21
Experiment 2:			
pMSV105/103(ASE)	GB	injection	19/33 (57.%)
	XL43	injection	9/35 (26.%)
ASE	GB	injection	0/45
pMSV3/4[4]	GB	injection	0/40

[1]Fraction of total plants showing symptoms of maize streak disease by 18 da. postinoculation
[2]Golden Bantam
[3]8 - 10 da. old seedling were injected in the nodal region with 20µl of the *A. tumefaciens* strain or plasmid DNA as indicated
[4]Plasmid containing a tandem direct repeat of 1.8 copies of the same cloned MSV DNA used to construct pMSV105/103, cloned into pEMBL8+. Seedlings were injected with ~5 µg of purified DNA.

transintermediate plasmid pMON505 (22) as illustrated in fig. 4. A linear MSV derived from one of the two BamHI clones sequenced (pMSV5) was cloned into the unique BglII site of pMON505 (pMSV105). With no ApaI sites in pMON505 (S.G. Lazarowitz, unpublished), the dimeric insert of MSV was constructed using the unique ApaI site to clone the ApaI linear of MSV into pMSV105 (pMSV105/103). Following a triparental mating with pRK2013 for mobilization of pMSV105/103 (23), *ASE* stably carrying pMSV105/103 (*pMSV105/103(ASE)*) were selected.

Maize seedlings were inoculated with *pMSV105/103(ASE)* or the parental *ASE* lacking MSV sequences. Plants inoculated with *pMSV105/103(ASE)* developed typical symptoms of maize streak disease beginning 5-7 da. postinoculation, the maximum number of symptomatic plants occurring between 12-18 da. (fig. 5a and Table 1). *C. mbila* fed on symptomatic injected plants efficiently transmitted maize streak disease to new seedlings (V.D. Damsteegt and S.G. Lazarowitz, unpublished). The host variety of maize had some affect on the efficiency of infection. While 50.% - 60.% of Golden Bantam seedlings were symptomatic by 18 da., ~25.% -~35.% of XL43 seedlings inoculated with the same

cultures on the same day developed symptoms (Table 1). In repeated attempts, Golden Bantam consistently gave 1.5-2 fold higher efficiencies of infection than XL43. None of the seedlings injected with the parental *ASE* developed symptoms of maize streak disease. This was also true of seedlings inoculated with *pMSV105/103(ASE)* by leaf abrasion using Celite, or by injection with purified cloned MSV DNA (Table 1).

DNA-containing extracts were prepared from symptomatic and nonsymptomatic injected plants and analyzed on Southern blots for the presence of MSV DNA (fig. 5b). In contrast to the previous study with MSV-N (11), high amounts of monomer size MSV ssDNA as well as the MSV dsDNA forms I (covalently closed supercoiled circles), II and III were evident in the extracts from symptomatic plants inoculated with *pMSV105/103(ASE)* (fig. 5b, lanes 3,5,8 and 10). These forms of MSV DNA precisely comigrated with those found in leafhopper-inoculated MSV-infected plants (fig. 5b, lanes 1,2,6 and 7). Thus, as originally reported for tandem direct repeats of TGMV in transgenic plants (23) as well as for MSV (11), it appears that following introduction into the plant cells the MSV DNA excises from the bacterial plasmid to form unit length viral DNA circles which initiate the systemic infection. The lack of MSV ssDNA in the study of MSV-N (11) may result from differences in the preparation of tissue extracts. Hybridization of the extracts shown in fig. 5 with the vector pMON505 did not detect any homologous sequences (not shown). Thus, it appears that the vector sequences have been lost. No viral DNA was detected in nonsymptomatic plants (fig. 5b, lanes 4 and 9). These results clearly demonstrate that the ~2.7 kb DNA which I have cloned and sequenced (see below) is the complete, infectious genome of MSV-S.

Nucleotide Sequence of the MSV Genome. As mentioned above, four independent MSV clones were initially chosen for sequence analysis in an attempt to assure the functional relevance of these studies. Restriction mapping and sequence analyses showed each clone to form a unique circle. The infectious clone (pMSV5) was sequenced in its entirety. 95.% of the sequence was determined for at least 2 of the independent clones, 55.% for at least 3 of the clones, and 21.% for all four clones. No differences were found in the sequences of the four clones with the exception of two transversions, an A/C at position 2682 and a C/T at position 1019 in one clone not tested for infectivity. The former is in an intergenic region. The latter is a third position change in an open reading frame which does not alter the predicted amino acid.

The sequence of the virion strand of the infectious genome of MSV-S is shown in fig. 6, and the structural features are diagrammed in fig. 7 and

```
        10         20         30         40         50         60         70         80         90        100
GGATCCACAG AACGCCCTGT ATTATCAGCC GCGGGTACCC ACTGCAGCTC CGACATCCGG AGGAGTGCCG TGGAGTCGCG TAGGCGAGGT AGCTATTTTG

       110        120        130        140        150        160        170        180        190        200
AGCTTTGTTG CATTGATTTG CTTTTACCTG CTTTACCTTT GGGTGCTGAG AGACCTTATC TTAGTTCTGA AGGCTCGACA AGGCAGATCC ACGGAGGAGC

       210        220        230        240        250        260        270        280        290        300
TGATATTTGG TGGACAAGCT GTGGATAGGA GCAACCCTAT CCCTAATCTA CCAGCACCAC CAAGTCAGGG CAATCCCGGG CCATTTGTTC CAGGCACGGG

       310        320        330        340        350        360        370        380        390        400
ATAAGCAATC AGCCATGTCC ACGTCCAAGA GGAAGCGGGG AGATGATGCG AATTGGAGTA AGCGGGTGCC TAAGAAGAAG CCTTCTTCAG CTGGGCTGAA

       410        420        430        440        450        460        470        480        490        500
GAGGGCTGGA AGCAAGGCCG ATAGGCCATC CCTCCAAATC CAGACACTCC AGCATGCTGG GACCACCATG ATAACTGTCC CATCCGGAGG AGTATGTGAC

       510        520        530        540        550        560        570        580        590        600
CTCATCAACA CCTATGCCCG AGGATCTGAC GAGGGCAACC GCCACACCAG CGAGACTCTG ACGTACAAGA TCGCCGTCGA CTACCACTTC GTTGCCGACG

       610        620        630        640        650        660        670        680        690        700
CGGCTGCCTG CCGCTACTCC AACACCGGAA CCGGTGTAAT GTGGCTGGTG TATGACACCA CTCCCGGCGG ACAAGCTCCG ACCCCGCAAA CTATATTTGC

       710        720        730        740        750        760        770        780        790        800
CTACCCTGAC ACGCTAAAAG CGTGGCCGGC CACATGGAAA GTGAGCCGGG AGCTGTGTCA TCGCTTCGTG GTGAAACGGC GATGGTTGTT CAACATGGAG

       810        820        830        840        850        860        870        880        890        900
ACCGACGGTC GGATTGGTTC GGATATCCCT CCCTCGAATA CAAGTTGGAA GCCTTGCAAG CGCAACATCT ACTTCCACAA GTTCACGAGT GGGTTGGGAG

       910        920        930        940        950        960        970        980        990       1000
TGAGAACGCA GTGGAAGAAT GTAACGGACG GAGGAGTTGG TGCCATCCAG AGAGGAGCTT TGTACATGGT CATTGCCCCC GGCAATGGCC TTACATTTAC

      1010       1020       1030       1040       1050       1060       1070       1080       1090       1100
TGCCCATGGG CAGACCCGCC TGTACTTTAA GAGTGTTGGC AACCAGTAAT GAATAAAAAC GCCCGTTTTA TTATATCTGA TGAATGCTGA AAGCTTACAT

      1110       1120       1130       1140       1150       1160       1170       1180       1190       1200
TAATATGTCG TGCGATGGCA CGAAAAAACA CACGCAATCA ATACAGGGGG GTAGTCGGCG GGCGGCTAAG GGTGGTGCTC GGCGGGCAAA ACATCGAAAA

      1210       1220       1230       1240       1250       1260       1270       1280       1290       1300
ATCAAGATCT ATATGAATTA CACTTCCTCC GTAGGAGGAA GCACAGGGGG AGAATACCAC TTCTCCCCCG GCGACATAAT GTAAATGATG CAGTTTGCCT

      1310       1320       1330       1340       1350       1360       1370       1380       1390       1400
CGAAATACTC CAGCTGCCCT GGAGTCATTT CCTTCATCCA ATCTTCATCC GAGTTGGCGA GGATTATTGT AGGCTTAGAC TTCTTCTGCA CCTTTTTCTT

      1410       1420       1430       1440       1450       1460       1470       1480       1490       1500
CTTACCATAC TTGGGGTTTA CAATGAAATC CCTCTGACAG CCAACTAACT GTTTCCAACA AGGACAGAAT TTAAACGGAA TATCATCTAC GATGTTGTAG

      1510       1520       1530       1540       1550       1560       1570       1580       1590       1600
ATTGCGTCTT CGTTGTATGA AGACCAATCA ACATTATTTT GCCAGTAATT ATGAACCCCT AGGCTTCTGG CCCAAGTAGA TTTTCCGGTT CTTGTTGGGC

      1610       1620       1630       1640       1650       1660       1670       1680       1690       1700
CGACGATGTA GAGGCTCTGC TTTCTTGATC TTTCATCTGA TGACTGGATA CAGAATCCAT CCATTGGAGG TCAGAGATTG CATCCTCGAG GGTATAACAG

      1710       1720       1730       1740       1750       1760       1770       1780       1790       1800
GTAGGTTGAA GGAGCATGTA AGCTTCGGGA CTAACCTGGA AGATGTTAGG CTGGAGCCAA TCATTGATTG ACTCATTACA AAGTAAATCA GGTGAGGAGG

      1810       1820       1830       1840       1850       1860       1870       1880       1890       1900
GTGGATGAGG ATTGGTGAAC TCTTCCTGAA TCTCAGGAAA AAGCTTATTT GCAGAGTATT CAAAATACTG CAATTTTGTG GACCAATCAA AGGGAAGCTC

      1910       1920       1930       1940       1950       1960       1970       1980       1990       2000
TTTCTGGATC ATGGAGAGGT ACTCTGCTTT GGAAGTAGCG TGTGAAATAA TGTCTCGCAT TATTTCATCT TTAGAAGGCT TTTTTTCCTT TACCTCTGAA

      2010       2020       2030       2040       2050       2060       2070       2080       2090       2100
TCAGATTTTC CTAGGAAGGG GGACTTCCTA GGAATGAAAG TACCTCTCTC AAACACAGCC AGAGGTTCCT TGAGAATGTA ATCCCTCACT CTGTTAACTG

      2110       2120       2130       2140       2150       2160       2170       2180       2190       2200
ACTTGGCACT CTGAATATTT GGGTGAAACC CATTTATATC AAAGAACCTT GAGTCAGATA TCCTTACCGG CTTCTCTGTC TGAAGCAATG CATGTAAATG

      2210       2220       2230       2240       2250       2260       2270       2280       2290       2300
CAAACTTCCA TCTTTATGTG CCTCTCGGGC ACATAGAATA TACTTGGGAA TCCAACGAAC GACGAGCTCC CAGATCATCT GACAGGCGAT TTCAGGATTT

      2310       2320       2330       2340       2350       2360       2370       2380       2390       2400
TCTGGACACT TTGGATAGGT TAGGAACGTG TTAGCGTTCC TGTGTGAGAA CTGACGGTTG GATGAGGAGG AGGCCATAGC AGACGACGGA GGCTGAGGCT

      2410       2420       2430       2440       2450       2460       2470       2480       2490       2500
GAGGGATGGC AGACTGGGAG CTCCAAACTC TATAGTATAC CCGTGCGCCT TCGAAATCCG CCGCTCCCTT GTCTTATAGT GGTTGTAAAT GGGCCGGACC

      2510       2520       2530       2540       2550       2560       2570       2580       2590       2600
GGTCCGGCCC AGCAGGAAAA GAAGGCGCGC ACTAATATTA CCGCGCCTTC TTTTCCTGCG AGGGCCCGGT AGGGACCGAG CGCTTTGATT TAAAGCCTGG

      2610       2620       2630       2640       2650       2660       2670       2680       2690
TTCTGCTTTG TATGATTTAT CTAAAGCAGC CCAATCTAAA GAAACCGGTC CCGGGCACTA TAAATTGCCT AACAAGTGCG ATTCATTCAT
```

Fig. 6. Complete nucleotide sequence of the virion strand of the infectious clone of MSV-S. The unique BamHI site was taken as position 1.

summarized in Table 2. The unique BamHI site has been defined as position 1 to conform with published sequences for MSV-K and MSV-N. The DNA is 2690 nucleotides in length, three nucleotides longer than the published sequence of

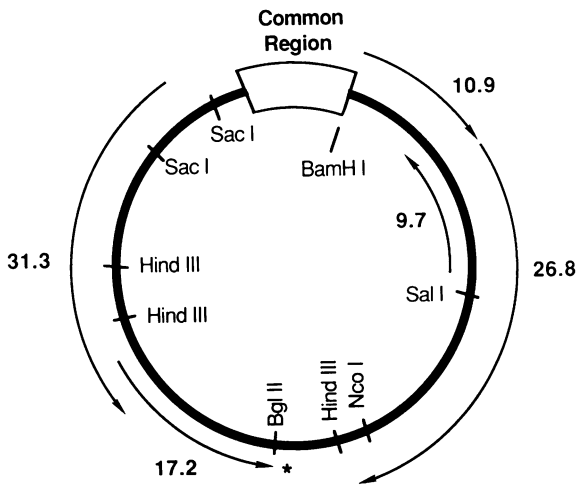

Fig. 7. Structural organization of the genome of MSV-S. Arrows denote potential ORFs. The equivalent of the geminivirus Common Region and the intergenic region containing the DNA primer (*) are marked. See text for details.

MSV-N (8) and nine nucleotides longer than the published sequence of MSV-K (7). There are five potential open reading frames (ORFs) in MSV-S, two on the virion "+" strand and three on the complementary "-" strand, and two apparent intergenic regions located on opposite sides of the viral genome.

The two ORFs on the virion "+" strand potentially encoding polypeptides of 10.9 kd (position 2689-301) and 26.8 kd (position 315-1046) correspond to those reported for MSV-N and MSV-K both by position and sequence homology (not shown). Unlike the analogous ORFs in WDV (9) and BCTV (10), these two ORFs do not overlap but are in different reading frames separated by 13 bases (Table 2). By sequence homology to MSV-N (8) the ORF for the 26.8 kd polypeptide is the gene for the virion capsid subunit (33). As reported for MSV-N, a third ORF on the virion strand between positions 734-1027 is completely included within the ORF for the capsid subunit and would potentially encode an 11.2 kd polypeptide. Codon usage and base preference analyses (30) for this ORF are unfavorable suggesting that it is unlikely to exist, and it has not been diagrammed in fig. 7. Unlike MSV-N, the only likely TATA box consensus (AC<u>TATAAA</u>) for both "+" strand ORFs is located in the intergenic region at position 2659 (fig. 8). This is consistent with the two overlapping transcripts which have been identified for the viral "+" strand (33), the longer of which has its cap site at nucleotide 2685, and suggests the

TABLE 2
Predicted Open Reading Frames of MSV-S

Translation Frame	TATA[1]	Initiation Codon[2]	Termination Codon	$^{A}_{G}$ATAA[3]	Protein (daltons)
2+	2659	2689	302[4]	(300)[6]	10,893
3+	2659	315	1047[4]	1052	26,778
2-	2480	2377	1561[5]	1538 1368	31,294
3-	2140	1664	1220[4]	1072	17,194
2-	1075	505	250[5]	160	9,685

[1]Most likely TATA box matching the consensus sequence $T^{C}_{G}TATA^{T}_{A}A_{1-3}$ (42)
[2]AUG
[3]Most likely polyadenylation signal (42)
[4]UAA
[5]UAG
[6]A large transcript which would contain this ORF has been identified for MSV-N. This transcript has its cap site at position 2682 and terminates at position 1114 (equivalent to positions 2686 and 1115 in MSV-S).

possibility of transcript processing from this region of the viral DNA. Transcription mapping suggests that the ORF for the 10.9 kd polypeptide, if it exists, is contained on a large transcript from nucleotide 2685 to 1115 (33).

There are two ORFs on the viral "-" strand which overlap by 102 bases and potentially encode a 31.3 kd (positions 2377-1562) and a 17.2 kd (positions 1664-1221) polypeptide (fig. 7 and Table 2). This overlapping arrangement of ORFs is identical to that reported for the analogous region of MSV-N (8) and, as in MSV-N, the predicted polypeptides are homologous to NH_2- and COOH-terminal domains respectively in the predicted product of ORF AL1 in the dicot geminiviruses (10,34). The analogous ORFs in WDV also overlap (9). While the ORF for the 17.2 kd polypeptide in MSV-S and MSV-N start with an ATG codon, the corresponding ORF in WDV starts with GUG. Of these two ORFs, only the upstream one has a good candidate TATA consensus relatively nearby at position 2480 (AC<u>TATAA</u>). Combined with the overlapping nature of the two ORFs, this again suggests the possibility of transcript processing in this region.

The third ORF on the "-" strand of MSV-S is at position 505 - 251, overlapping the ORF for the 10.9 kd polypeptide and the capsid gene on the "+" strand (fig. 7 and Table 2). This ORF would encode a 9.7 kd polypeptide and coincides with the 5'-end of an ORF for a 13. kd polypeptide in MSV-N. While

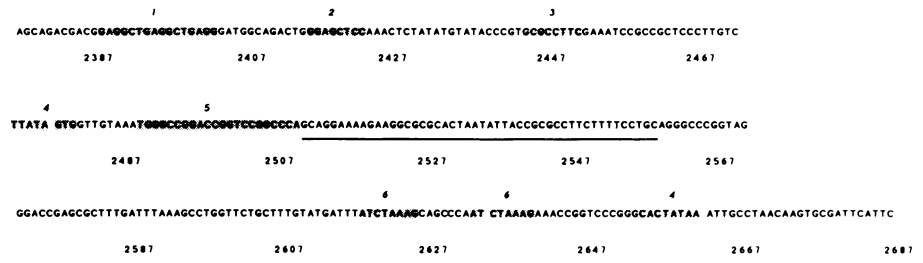

Fig. 8. Structural features of the intergenic Common Region of MSV-S. The conserved geminivirus sequence element (2510 - 2559) is underlined. Potential hairpin is formed by the 18 base inverted repeats at the ends of the 46 base sequence. Hatching indicates direct and inverted repeat elements. (1) Overlapping 10 base direct repeats of GAGGCTGAGG. (2) 8 base perfect palindromic sequence. (3) 8 base sequence repeated in direct (2547 - 2554) and inverted (2525 - 2532) orientation at the top of the stem on either side of the loop in the potential hairpin structure of the conserved sequence element. (4) 8 base inverted repeats of CACTATAA containing potential TATAA consensus sequences. (5) 22 base perfect palindromic sequence located immediately adjacent to the conserved sequence element. (6) 8 base direct repeats of ATCTAAAG, the second of which contains part of the potential consensus CCAAT sequence.

codon usage and base preference analyses do not help in deciding whether this ORF could exist, it does not have a reasonable TATA consensus nearby and a strain difference in MSV-S resulting in a conservative amino acid substitution on the "+" strand introduces a stop codon on the "-" strand and thus truncates ~25.% of the potential protein. This suggests that this ORF may not exist.

Finally, there are two apparent intergenic regions in MSV-S located on opposite sides of the genome (fig. 7). Both have been reported for MSV-N and MSV-K (7,8,12). One of these regions (position 2378-2688) contains the conserved sequence element found in the Common Regions of the bipartite geminiviruses and in potential intergenic regions of the single component geminiviruses (for review, see 6) (fig. 7, 8). This sequence element is defined by its potential to form a hairpin structure and the presence in the loop of the conserved sequence TAATATTAC in all geminiviruses. Thus, this sequence element locates the equivalent of the intergenic Common Region for the single component geminiviruses, although the precise limits of this region in MSV have yet to be defined functionally. The 46 base sequence element in MSV-S is identical to that reported for MSV-N (8). This intergenic region in MSV-S also contains motifs of direct and inverted repeats of short nucleotide sequences (fig. 8) as is found in the Common Regions of the bipartite geminiviruses (6). The potential TATA consensus sequences for the "+" and "-"

strands are each located in an 8 base inverted repeat element within this region, oriented as expected for transcription (fig. 8). There is also a potential CCAAT consensus sequence located within this region as part of one copy of two short direct repeats (fig. 8, see below).

The second apparent intergenic region in MSV-S has also been reported for MSV-N and MSV-K and is located at position 1050-1217 (fig. 7). As first shown for MSV-N (35), this intergenic region is the location of an 80 base DNA "self-primer" found hybridized to the virion DNA. Incubation of virion ssDNA *in vitro* with the Klenow fragment of *E. coli* polymerase I and all four deoxyribonucleotide triphosphates directed the synthesis of a complementary viral DNA copy without the addition of exogenous primer (not shown) as previously reported (7,12,35). As determined by dideoxy-sequence analysis combined with ribonuclease or alkali treatment, the DNA primer in MSV-S is ~80 bases in size with a nonuniform length of ribonucleotides at its 5'-end, and is at position 1203-1124. The presence of a primer DNA hybridized to the virion DNA has yet to be demonstrated for any other geminivirus.

DISCUSSION

Until recently, the fact that successful mechanical inoculation of maize could not be demonstrated using either virion (36) or cloned DNA (8; S.G. Lazarowitz, unpublished) has been a major obstacle in the biological analysis of MSV. With the elegant demonstration that *A. tumefaciens* carrying a Ti plasmid and a tandem repeat of cloned MSV DNA in a binary vector can transmit maize streak disease (11), the biological relevance of published sequences for MSV can now be assessed. The infectious clone of MSV-N (11) was derived from viral dsDNA found in infected tissue. The published sequence for MSV-N (8) was derived from a consensus of both viral dsDNA and virion ssDNA, not from a single clone. One of the clones constructed from virion DNA and used to assemble this consensus is not infectious. While there are three nucleotide variations reported in the published sequence of MSV-N (8), the authors are unable to state the relationship of the published sequence to that of an infectious clone (M.I. Boulton, personal communication). Furthermore, there are several discrepancies in the published sequences for MSV-N and MSV-K, with the infectivity of the latter yet to be demonstrated. Thus, I have reported here the complete nucleotide sequence of MSV-S, analyzing a clone which I have shown to be infectious.

Inoculation of 8 - 10 da. old seedlings by direct injection into the nodal region of *pMSV105/103(ASE)*, which contains a tandem repeat of the cloned

MSV-S sequences, resulted in the development of typical symptoms of maize streak disease with the concomitant accumulation of monomer-sized circular forms of MSV ssDNA and dsDNA (fig. 5). Thus, as first shown for tandem inserts of TGMV DNA in transgenic plants (23) and previously reported using this approach for MSV-N (11), the MSV-S DNA once introduced into plant cells appears to excise by homologous recombination and replicate as monomer-sized circles. Whether this excision is replication-mediated remains unknown. The bacterial plasmid sequences introduced with the MSV DNA cannot be detected in the infected plants. Furthermore, no symptoms were observed in plants inoculated with the parental *ASE* (Table 1). Thus, the cloned ~2.7 kb genomic component of MSV-S is infectious for maize, this single component being necessary and sufficient for the transmission of maize streak disease. Leafhoppers fed on symptomatic injected plants successfully transmitted maize streak disease to new maize seedlings (V.D. Damsteegt and S.G. Lazarowitz, unpublished), thus formally fulfilling Koch's postulates.

Plants inoculated by leaf abrasion with *pMSV105/103(ASE)* (Table 1) or by injection into regions of the stem higher than ~1. cm above the node (S.G. Lazarowitz and A. Pinder, unpublished) did not develop maize streak disease. Injection in the meristematic region appears to be essential for the success of this technique. Why injection of high concentrations of MSV DNA directly into this region of the seedlings never works (Table 1) remains unexplained. *A. tumefaciens* would appear to provide a required function. Whether this is simply protective or more complex in nature requires further investigation. Our efficiency of infection in Golden Bantam (50.%-60.%) was somewhat lower than that reported by Grimsley and co-workers (11). Not surprisingly, we have observed greenhouse temperatures to affect the efficiency of agroinoculation, with temperatures above ~30°C being severely deleterious. This is likely the result of curing of the Ti plasmid and/or decreased viability of the bacteria at higher temperatures. Thus, we attribute our slightly lower efficiency to a notably hot Maryland summer and suspect that better controlled greenhouse temperatures in cooler weather will improve this situation.

Direct comparison of the three sequenced MSV clones shows them to be highly homologous and equally related to each other. MSV-S and MSV-N are identical at 2628 positions. MSV-S and MSV-K are identical at 2626 positions. Of the 60 nucleotides where MSV-S and MSV-K differ, 33 are identical in MSV-S and MSV-N, and 27 are identical in MSV-K and MSV-N. With four exceptions, all of the differences among the three clones are single base substitutions. In MSV-S, 18 single base substitutions which are different from both MSV-N and

MSV-K occur in ORFs. Of these, 11 are third position substitutions which do not alter the predicted amino acid. The 3 second position changes and 4 first position changes all result in conservative amino acid substitutions.

Despite this extensive sequence identity in these three MSV clones, there are several differences in genomic organization predicted from the published sequences of MSV-N and MSV-K. The overall organization of the genome of MSV-S most resembles that of MSV-N, although it is identical to neither. While there is general agreement on the locations of the intergenic regions and the arrangement of the two ORFs on the virion "+" strand for the 10.9 kd and capsid polypeptides (fig. 7; 7,8,12), the arrangement of ORFs on the complementary "-" strand is different in the three isolates. Both MSV-S and MSV-N have an identical overlapping arrangement of ORFs for a 31.3 kd and 17.2 kd polypeptide on the "-" strand, the same arrangement as that reported for the analogous ORFs in WDV (9). This same region of the "-" strand of MSV-K has been reported to have two nonoverlapping ORFs, separated by 24 nucleotides and in the same frame (7). The first of these is identical to the ORF for the 31.3 kd polypeptide in MSV-S and MSV-N. The difference results from a single inserted G residue in MSV-K at position 1502 within the beginning of the ORF for the 17.2 kd polypeptide which alters its starting position. This has the effect of eliminating the first 44 amino acids and completely changing predicted amino acids 45 - 52 in the potential product of this ORF in MSV-K.

MSV-S, MSV-N, and MSV-K also differ in the position of the end of this ORF for the 17.2 kd polypeptide. This results from a single base deletion in MSV-N and MSV-K at the position of the UAA stop codon for this ORF in MSV-S, the relevant sequences being:

```
                P   P   T   E   E   V   *           *
    MSV-S  ...CCT CCT ACG GAG GAA GTG TAAT TCA TAT AGA TCT TGA TT ...
           ::: ::: ::: ::: : : ::    :: :::  :: ::: ::: ::: ::
MSV-N,MSV-K ...CCT CCT ACG GAG GCA GTA -CAT TCA GAT AGA TCT TGA TT ...
                P   P   T   E   A   V   H   S   D   R   S   *
```

Thus, the predicted product of this ORF would have five extra amino acids at its carboxy terminus in MSV-N and MSV-K as compared to MSV-S. All four independent clones of MSV-S unambiguously have this same sequence as shown.

Finally, there is a large ORF on the viral "-" strand in MSV-N between positions 1007 - 389 which overlaps the entire capsid coding sequence and part of the ORF for the 10.9 kd polypeptide on the "+" strand and would encode a 21.7 kd polypeptide (8). This ORF is not present in MSV-S because of a strain difference at position 716 which introduces a UAG stop codon on the "-"

strand. This same base substitution is found in MSV-K. This is a third position codon change in the capsid coding sequence which does not alter the amino acid present at that position. However, it creates two ORFs on the viral "-" strand where MSV-N has the single large one. The upstream of these two ORFs is unlikely to exist based on unfavorable codon usage and base preference. The downstream ORF is predicted to encode a short polypeptide of only 22 amino acids. Thus, the only ORF which may exist in this region of the viral "-" strand in MSV-S is that for the 9.7 kd polypeptide (fig. 7).

The last region of major difference among the three MSV isolates is within the intergenic Common Region. In MSV-K there is an 8 base deletion of the sequence 5'-AGGGCCCG-3' adjacent to the 3'-end of the conserved viral sequence element (position 2561 -2568). This sequence is present in both MSV-S and MSV-N. While the conserved sequence element is identical in both MSV-S and MSV-N, there is a single base change at the fourth nucleotide of the element in MSV-K which would alter the base pairing stability of the lower part of the potential stem structure.

Thus, the picture of the organization of the genome of MSV that emerges from the analysis of the infectious clone of MSV-S has some elements in common with each of the previous reports for MSV-N and MSV-K, being to some extent a composite of these (fig. 7) and essentially identical in organization to WDV (9). There are five ORFs, two on the virion strand and three on the complementary strand, although it seems likely that the ORF for the 9.7 kd polypeptide does not exist. There are two apparent intergenic regions located on opposite sides of the genome. One, located between the starts of the ORFs for the 31.3 kd ("-" strand) and 10.9 kd ("+" strand) polypeptides, is in the equivalent position to the bipartite viral Common Region and contains the conserved sequence element which is its hallmark. The other intergenic region is found between the ends of the ORFs for the 17.2 kd ("-" strand) and capsid ("+" strand) polypeptides and is the location of the DNA primer in the virion.

In overall organization, the genome of the infectious clone of MSV-S is very similar to that reported for MSV-N. In the absence of direct data on the relation between the reported sequence for MSV-N and that of the noninfectious clone, one can only speculate on the importance of the few differences between this published sequence and that reported here for MSV-S. Two of the three base variations reported in MSV-N (8) are identical in the virion consensus of MSV-N and in MSV-S. At position 1793 in MSV-N, 132 bases upstream from the AUG for the 17.2 kd polypeptide, MSV-S has the alternate base reported. If there is RNA processing and this is within an intron, the sequence as found in

MSV-S matches the conserved site associated with lariat formation in mammalian systems (37) and this base change could be functionally important. However, the only good consensus donor splice junction (38) in MSV, occurs downstream from this at position 1738-1730. If this base change occurs within the coding sequence of the 31.3 kd polypeptide, it is a third position change not affecting the protein. All of the other single base substitutions in ORFs in MSV-N and MSV-S are either third position changes or result in conservative amino acid substitutions.

This leaves three differences of potential relevance to biological function in the two viral DNAs: (1) deletion of a base in MSV-N altering the end of the ORF for the 17.2 kd polypeptide; (2) two single base deletions in the intergenic DNA primer region of MSV-N; and (3) a single base substitution in the intergenic Common Region of MSV-N which alters a CCAAT consensus sequence. The first of these, while it has the two distinct effects of adding five amino acids to the carboxy-terminal end of the 17.2 kd polypeptide and positioning the primer DNA so it precisely overlaps the UGA stop codon for this ORF, both of which could be functionally significant, occurs in both MSV-N and MSV-K. Although the infectivity of cloned MSV-K has not been shown, this coincidence in sequence makes it less likely that this difference could explain lack of infectivity in MSV-N. The two single A residues deleted in MSV-N as compared to MSV-S and MSV-K are in short runs of A's in the primer DNA region -- one at position 1059 in MSV-S, 10 nucleotides downstream of the termination codon for the capsid subunit gene and adjacent to the potential polyadenylation signal (fig. 7); and the second at position 1128 in MSV-S at the 3'-end of the primer DNA. The first deletion leaves a polyadenylation signal (AATAA) and would not obviously affect mRNA stability through basepairing interactions. The latter still leaves 4 A residues at the 3'-end of the primer. It is not apparent why either of these differences would destroy viral infectivity.

This leaves 13 base substitutions in MSV-N in the intergenic Common Region. Of these, one of potential significance occurs in a candidate CCAAT consensus sequence for "+" strand transcription at position 2634 in MSV-S (fig. 8). A transcript starting 4 bases upstream from the ATG for the 10.9 kd polypeptide has been identified (33). There is a TATA consensus sequence (AC<u>TATAAA</u>) -27 bases from this cap site and the CCAAT consensus sequence (GC<u>CCAAT</u>C) -27 from this potential TATA box in MSV-S. The reported sequence for MSV-N has a single base substitution changing the CCAAT consensus sequence to GC<u>CCATT</u>C, a substitution which could potentially impair transcriptional

activity (39-41). Saturation mutagenesis of the β-globin promoter has shown that this substitution significantly decreases transcription of the β-globin gene (41). This change could also be important in altering the first position of the second 8 base repeat element. Current analyses should determine the importance of this and the TATA box sequence in viral gene expression.

The sequence of the infectious cloned DNA of MSV-S provides a unified and somewhat simplified picture of the genomic organization of MSV, one that is in overall organization almost identical to that reported for WDV (9). With the complete sequence of the infectious clone and the demonstration of leafhopper-transmission of the virus from plants inoculated with this cloned viral DNA, we can now analyze the effects of specific mutations on infectivity and transmission in an attempt to translate this structural map into a relevant functional map of the viral genome.

ACKNOWLEDGMENTS

I thank Allison Pinder and Betty Conde for their excellent technical assistance during different phases of this work. I am also grateful to Vern Damsteegt and Anna Hewings of the Foreign Disease-Weed Science Research Unit, Frederick, MD, for generously providing leafhopper-inoculated maize streak-infected tissue, laboratory space and support services, as well as for their encouragement and invaluable advice; to W.F. Dowler for making the facilities of his department at Frederick available to me; to Mike Goff of the Md. Dept. of Agriculture for his help in obtaining permits for this work; and to Steve Rogers for his enthusiastic support and helpful discussions. This work was supported in part by NSF grant PCM8216268 to S.G.L., and by funds from the Monsanto Company.

REFERENCES

1. Goodman, R.M. (1981) J. Gen. Virol. *54*, 9-21.
2. Goodman, R.M. (1981) "Handbook of plant virus infection and comparative diagnosis" (E. Kurstak, ed.), 879-910.
3. Stanley, J. and M.R. Gay (1982) Nature *301*, 260-262.
4. Hamilton, W.D.O., V.E. Stein, R.H.A. Coutts and K.W. Buck (1984) EMBO J. *3*, 2197-2205.
5. Howarth, A.J., J. Caton, M. Bossert and R.M. Goodman (1985) Proc. Natl. Acad. Sci. USA *82*, 3572-3576.
6. Lazarowitz, S.G. (1986) Plant Molec. Biol. Rep. *4*, 177-192.
7. Howell, S.H. (1984) Nucl. Acids Res. *12*, 7359-7375.
8. Mullineaux, P.M., J. Donson, B.A.M. Morris-Krsinich, M.I. Boulton and J.W. Davies (1984) EMBO J. *3*, 3063-3068.
9. MacDowell, S.W., H. Macdonald, W.D.O. Hamilton, R.H.A. Coutts and K.W. Buck (1985) EMBO J. *4*, 2173-2180.
10. Stanley, J., P.G. Markham, R.J. Callis and M.S. Pinner (1986) EMBO J. *5*, 1761-1768.
11. Grimsley, N., T. Hohn and B. Hohn (1987) Nature *325*, 177-179.
12. Howell, S.H. (1985) Nucl. Acids Res. *13*, 3018-3019.
13. Damsteegt, V.D. (1983) Plant Dis. *67*, 734-737.
14. Hamilton, W.D.O., D.M. Bisaro, R.H.A. Coutts and K.W. Buck (1983) Nucl. Acids Res. *11*, 7387-7396.

15. Maniatis, R., E.F. Fritsch and J. Sambrook (1982) "Molecular cloning", Cold Spring Harbor Laboratory, New York.
16. Dretzen, G.M., M. Bellard, P. Sassone-Corsi and P. Chambon (1981) Anal. Biochem. *112*, 295-298.
17. Hamilton, W.D.O., D.M. Bisaro and K.W. Buck (1982) Nucl. Acids Res. *10*, 4901-4912.
18. Dente, L., G. Cesareni and R. Cortese (1983) Nucl. Acids Res. *11*, 1645-1655.
19. Norrander, J., T. Kempe and J. Messing (1983) Gene *26*, 101-106.
20. Boyer, H.W. and D. Roulland-Doussoix (1969) J. Mol. Biol. *41*, 459-472.
21. Karess, R.E. and G.M. Rubin (1984) Cell *38*, 135-146.
22. Horsch, R.B. and H.J. Klee (1986) Proc. Natl. Acad. Sci. *83*, 4428-4432.
23. Rogers, S.G., R.B. Horsch and R.T. Fraley (1986) Meth. Enzymol. *118*, 627-640.
24. Melching. J.S., K.R. Bromfield and C.H. Kingsolver (1983) Plant Dis. *67*, 717-722.
25. Sanger, F., S. Nicklen and A.R. Coulson (1977) Proc. Natl. Acad. Sci. *74*, 5463-5467.
26. Maxam, A.M. and W. Gilbert (1980) Methods Enzymol. *65*, 499-518.
27. Biggin, M.D., T.J. Gibson and G.F. Hong (1983) Proc. Natl. Acad. Sci. *80*, 3963-3965.
28. Mills, D.R. and F.R. Kramer (1979) Proc. Natl. Acad. Sci. *76*, 2232-2236.
29. Staden, R. (1979) Nucl. Acids Res. *6*, 2601-2610.
30. Staden, R. (1985) Genetic Engineering *7*, 67-114.
31. Southern, E. (1975) J. Mol. Biol. *98*, 503-517.
32. Rigby, P.W.J., M Dieckmann, C. Rhodes and P. Berg (1977) J. Mol. Biol. *113*, 237-251.
33. Morris-Krsinich, B.A.M., P.M. Mullineaux, J. Donson, M.I. Boulton, P.G. Markham, M.N. Short and J.W. Davies (1985) Nucl. Acids Res. *13*, 7237-7256.
34. Mullineaux, P.M., J. Donson, J. Stanley, M.I. Boulton, B.A.M. Morris-Krsinich, P.G. Markham and J.W. Davies (1985) Plant Mol. Biol. *5*, 125-131.
35. Donson, J., B.A.M. Morris-Krsinich, P.M. Mullineaux, M.I. Boulton and J.W. Davies (1984) EMBO J. *3*, 3069-3073.
36. Bock, K.R. (1982) Plant Dis *66*, 266-270.
37. Reed, R. and T. Maniatis (1985) Cell *41*, 95-105.
38. Mount, S.M. (1982) Nucl. Acids Res. *10*, 459-472.
39. Graves, B.J., P.F. Johnson and S.L. McKnight (1986) Cell *44*, 565-576.
40. McKnight, S. and R. Tjian (1986) Cell *46*, 795-805.
41. Meyers, R.M., K. Tilly and T. Maniatis (1986) Science *232*, 613-618.
42. Messing, J., D. Geraghty, G. Heidecker, N. Hu, J. Kridl and I. Rubenstein (1983) "Genetic Engineering of Plants", (T. Kasuge, C.P. Meredith and A. Hollaender, eds.), Plenum Press, NY, pp. 211-227.