

---

**The complete nucleotide sequence of the maize chlorotic mottle virus genome**

---

R.C.Nutter\*, K.Scheets, L.C.Panganiban<sup>1</sup> and S.A.Lommel<sup>2</sup>

---

Department of Plant Pathology, Oklahoma State University, Stillwater, OK 74078-0258, <sup>1</sup>Cetus Corporation, 1400 53rd Street, Emeryville, CA 94608 and <sup>2</sup>Department of Plant Pathology, Box 7616, North Carolina State University, Raleigh, NC 27695-76166, USA

---

Received December 19, 1988; Revised and Accepted March 22, 1989

EMBL accession no. X14736

**ABSTRACT**

The complete nucleotide sequence of the maize chlorotic mottle virus (MCMV) genome has been determined to be 4437 nucleotides. The viral genome has four long open reading frames (ORFs) which could encode polypeptides of 31.6, 50, 8.9 and 25.1 kd. If the termination codons for the polypeptides encoded by the 50 and 8.9 kd ORFs are suppressed, readthrough products of 111 and 32.7 kd result. The 31.6 and 50 kd ORFs overlap for nearly the entire length of the 31.6 kd ORF. Striking amino acid homology has been observed between two potential polypeptides encoded by MCMV and polypeptides encoded by carnation mottle virus (CarMV) and turnip crinkle virus (TCV). The 25.1 kd ORF most likely encodes the capsid protein. The similar genome organization and amino acid sequence homology of MCMV with CarMV and TCV suggest an evolutionary relationship with these members of the carmovirus group.

**INTRODUCTION**

Maize chlorotic mottle virus (MCMV) was first reported in the United States in 1976 when it was associated with maize dwarf mosaic virus or wheat streak mosaic virus in a severe disease in maize termed corn lethal necrosis (CLN)(1). MCMV and its association with these potyviruses in CLN is well characterized ecologically (2,3), epidemiologically (4) and physicochemically (5). However, little information exists concerning the molecular biology of the virus or the mechanism of the synergistic interaction. Previous work (5) has demonstrated a number of properties for MCMV. The virions contain a monopartite RNA genome approximately 4.4 kb in length lacking both 5' genome-linked protein (Vpg) and 3' poly(A) tail. In infected plants, MCMV produces a single, 3' co-terminal 1.3 kb sub-genomic RNA that encodes a 25 kd capsid protein. The 5' end of the virus is capped with a m<sup>7</sup>GpppA structure. Translation of the MCMV genome

by a reticulocyte system results in polypeptides of 105, 52, 44, 41, 32 and 25 kd. The 25 kd polypeptide has been shown to be capsid protein due to its specific immunoprecipitation by capsid protein antibody. This paper reports the cDNA cloning of MCMV and the complete nucleotide sequence of the viral genome. MCMV sequence analysis supports the number and sizes of translation products reported previously (5). This data permits a tentative organization of the virus-encoded polypeptides. The predicted amino acid sequences of two MCMV ORFs have striking homology to portions of carnation mottle virus (CarMV) and turnip crinkle virus (TCV).

#### MATERIALS AND METHODS

##### Construction of MCMV cDNA clones

MCMV cDNA clones were constructed essentially as outlined by Gubler and Hoffman (6). To synthesize the 3' terminal plasmid pMCM409 (Fig. 1), viral RNA was first polyadenylated (average length 10-20 bases) using poly(A) polymerase and ATP according to the supplier (BRL). First strand synthesis was performed with MMLV reverse transcriptase (BRL) using an oligo dT primer. After second strand DNA synthesis, molecules were tailed with dCTP (average length 10-20 bases), annealed to PstI-digested G-tailed pBR322 (BRL) and transformed into competent E. coli RR-1 (7). For the synthesis of the internal clone, pMCM524, a 110 bp EcoRI-PstI fragment from pMCM409 (equivalent to residues 2292-2402 in Fig. 2) was gel purified after digestion of pMCM409 with EcoRI + PstI, denatured and the strand complementary to viral RNA was used as primer for the first strand cDNA synthesis. The 5' terminal clone, pMCM602, was constructed using a synthetic deoxyoligonucleotide primer deduced from sequencing the 5' end of pMCM524 and complementary to the viral genome (5'-TGTAGTTTTCTTTGGAATTC-3'; residues 1118-1138 in Fig. 2). Bacteria containing recombinant plasmids were screened for MCMV-specific sequences using a randomly primed MCMV cDNA probe. pMCM1067 (Fig. 1) was constructed from pMCM524 and pMCM409 by ligation of the two fragments at the common EcoRI site (Fig. 1) and insertion into PstI-digested pUC18. Proper orientation of the two fragments in pMCM1067 was determined by restriction enzyme

mapping. All cDNA clones or subclones of MCMV were recloned into the appropriate sites of the Bluescript KS+ or SK+ vectors (Stratagene) for further analysis and sequencing.

Exonuclease III deletions A nested series of Exonuclease III deletions were generated from the original cDNA clones according to Henikoff (8). Competent *E. coli* JM109 or MV1193 (9) were used as bacterial hosts. Transformants containing plasmids from the appropriate time points were grown in 1-2 ml of selective liquid culture and plasmid DNA was isolated according to Maniatis (7). The extent of Exonuclease III deletion was determined by restriction digestion. Bacteria containing plasmids with the proper length deletions (200-250 bp for each point) were used to produce sequencing templates.

Production of ssDNA for sequencing reactions Bacteria ( $1-2 \times 10^6$ ) were inoculated into 10 ml of L-broth supplemented with 50 ug/ml ampicillin and infected with approximately  $2 \times 10^7$  M13K07 helper phage (9). ssDNA was isolated from phage particles following the procedure outlined by Viera (9).

Sequencing Reactions Dideoxy sequencing of ssDNA templates was carried out using  $^{35}\text{S}$ - $\alpha$ -S-dATP with either the Klenow fragment of *E. coli* DNA polymerase (BRL) or modified T7 DNA polymerase (Sequenase) as recommended by the supplier (US Biochemical, Cleveland, OH). Radiolabeled DNA fragments were separated on 8M urea-acrylamide gels (19:1 acrylamide:bisacrylamide) in a Tris-borate-EDTA buffer system (7). Fragments were visualized after exposure to Kodak XAR-5 film for 16-24 hrs.

Reverse Transcriptase Run-off Sequencing The sequence of the 5' terminal nucleotides was determined using the synthetic deoxynucleotide 5'-AACGCGTTGGGTCTGTTGC-3' (complementary to residues 13-32 in Fig 3) as primer. The oligonucleotide was end-labeled with polynucleotide kinase and  $\gamma$   $^{32}\text{P}$ -ATP and acrylamide gel purified. Labeled runoff cDNA was synthesized and sequenced according to the procedure described by Guilley et al. (10).

DNA Sequence Manipulation The DNA sequence was analyzed using the Pustell Sequence Analysis Program (Inter. Biotech. Inc.). Amino acid homology of MCMV polypeptides was analyzed using the FASTP protein alignment program (11) using the Genbank (NIH), EMBL and National Biomedical Research Foundation databases. The

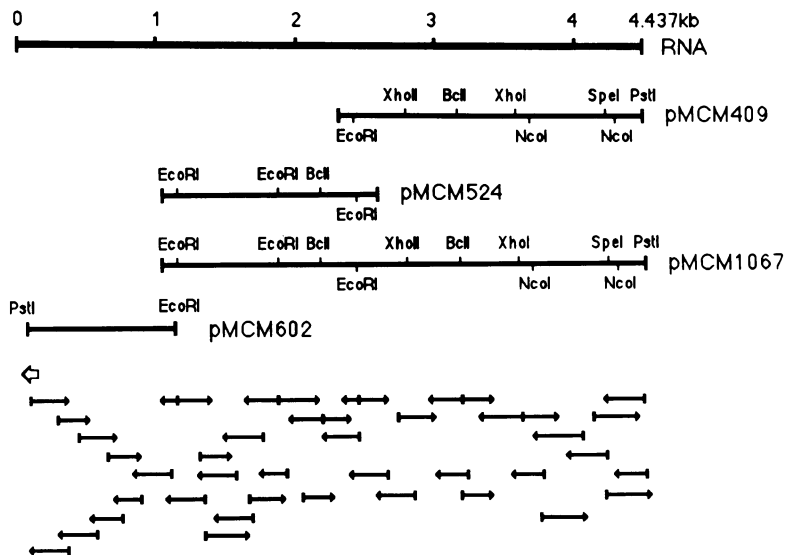


Fig 1 Schematic of cDNA clones of the MCMV genome. The major cDNA clones of MCMV used for sequencing are shown relative to their location on the MCMV genome. Arrows indicate position, direction and length of nucleotide sequence obtained from each template. The open arrow at the 5' end of pMCM602 denotes the position of the oligonucleotide used for run-off sequencing of the viral RNA.

sequence of TCV was kindly provided by Dr. T.J. Morris, U.C. Berkeley, prior to publication.

## RESULTS

### Cloning of MCMV

The complete MCMV genome is represented by three overlapping cDNA clones. They are called pMCM409, pMCM524, and pMCM602 and have inserts 2.1, 1.4 and 1.1 kb in length, respectively (Fig. 1). A number of cDNA clones synthesized from the oligo dT primer were screened to determine which would be expected to be 3' co-terminal with viral RNA. pMCM409 was the best candidate based on extensive restriction enzyme analysis.

### Nucleotide Sequence of MCMV

The clones used to determine the sequence of MCMV are illustrated in Fig. 1. Initially, subclones of the plasmids pMCM409, pMCM1067 or pMCM602 were used as templates for

sequencing reactions. The sequence of the original subclones was verified in both directions using Exonuclease III deletions of pMCM1067 and pMCM602. The sequence of both strands has been determined for nearly the entire virus. Alternatively, independent clones of the same orientation were used when necessary.

The MCMV genome is 4437 nucleotides in length, excluding the m<sup>7</sup>G cap structure (Fig. 2). The MCMV viral genome is not polyadenylated at the 3' end, therefore, the 3' terminus of MCMV has been defined as the first adenine of the synthetic poly(A) tract. MCMV has a 5' untranslated region of 117 nucleotides and a 3' untranslated region of at least 346 nucleotides. There are two EcoRI sites (residues 1109-1114 and 1119-1124, Fig.2) separated by four nucleotides that could only be detected sequencing nested deletions of the region. If the sequence had been determined using only subclones of the cDNAs, the proper genome organization would not have been possible.

The 5' terminal sequence of MCMV was determined using a deoxyoligonucleotide primer complementary to the viral RNA. Reverse transcriptase extended the oligonucleotide to the end of the viral genome. In this manner it was determined pMCM602 lacked the 5' terminal nucleotides, 5'-AGG-3'(data not shown).

#### Analysis of the MCMV nucleotide sequence

Open reading frames (ORFs) coding for polypeptides with M<sub>r</sub> of 31,590 daltons, 50,048 daltons, 8,893 daltons and 25,147 daltons (p31.6, p50, p9 and p25) were identified upon computer analysis of the nucleic acid sequence (Fig. 3). Reading frame 1 contains p31.6 and p9, reading frame 2 contains p50 and reading frame 3 contains p25. Most of p31.6 is encoded by the same region of the viral genome that encodes p50 (137-984)(Fig. 3) but it is read from a different reading frame. Analysis of the complementary (minus-sense) strand revealed a potential ORF of 18.2 kd between nucleotides 685 (from the 5' terminus) and 1186 (data not shown).

#### Analysis of MCMV ORFs

p31.6 This ORF is the most 5' of the potential polypeptide coding regions (Table 1). The position of the initiating AUG for this ORF nearest the 5' end of the viral RNA (Fig. 3) suggests



that it is recognized by eukaryotic ribosomes (12). However, analysis of the nucleotide sequence surrounding the initiation codon (Table 1) shows it is not favored for recognition by eukaryotic ribosomes (12,13). The calculated pI of the peptide is 4.83 which is in contrast to the very basic pIs for the other peptides encoded by MCMV ORFs. An *in vitro* polypeptide of approximately 32 kd is synthesized from MCMV genomic RNA (5) which is in good agreement with the calculated size of the p31.6 polypeptide.

p50 This ORF is encoded between residues 137 and 1451 (Fig. 3). However, the single amber terminator is followed by an additional 60.7 kd open reading frame terminating at residue 3031.

Suppression of the amber termination codon would result in a polypeptide of 110.7 kd (p111). These sizes are in good agreement with polypeptides of 52 and 105 kd seen upon *in vitro* translation of MCMV RNA (5). The nucleotide sequence context surrounding the initiating codon (Table 1) does not show the preferred A at the -3 position (relative to the AUG) in eukaryotic mRNAs (12,13) but does have the preferred G at position +4 in plant mRNAs (13).

p9 p9 is located between residues 2959 and 3198 (Fig. 3). The nucleotide sequence surrounding the AUG initiation codon (Table 1) does not show the preferred A at the -3 position in eukaryotic mRNAs (12,13) but does have the preferred G at position +4 in plant mRNAs (13). After the opal termination codon (Fig. 2; Table 1), the ORF continues until residue 3832, thereby encoding a potential readthrough polypeptide of 32.7 kd (see Fig. 3, p32.7).

p25 p25 is located nearest the 3' end of the virus between residues 3384-4091 (Fig. 3). The nucleotide sequence surrounding the initiation codon (Table 1) is in excellent agreement with the consensus sequence reported for plant mRNAs (13). This suggests the polypeptide encoded by this region is efficiently translated.

Fig 2 Nucleotide sequence of MCMV RNA and derived amino acid sequences of the viral polypeptides. The long open reading frames in MCMV are indicated in the left margin and the nucleic acid or amino acid residues in the right margin. The one letter amino acid codes are positioned under the first nucleotide of each codon. Terminator codons are represented by asteristics.

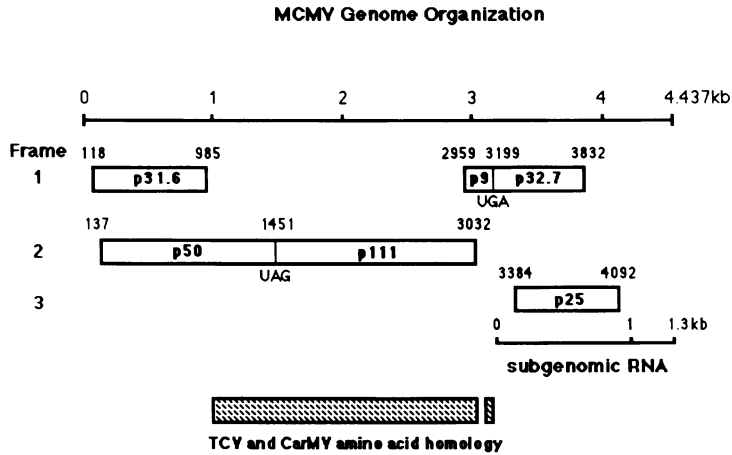


Fig 3 Major MCMV open reading frames. Open boxes represent coding regions in the MCMV genome in the appropriate reading frame. The numbers in the boxes correspond to the calculated  $m_r$  of the putative polypeptides in kilodaltons. Numbers above the boxes are the nucleotide residues of the initiation and termination codons for each viral polypeptide. Terminators that may be suppressed during translation are indicated below the open boxes. Stippled boxes indicate the regions of MCMV ORFs that are homologous to polypeptides encoded by CarMV and TCV.

The calculated pI of the polypeptide is 12.2, with 12 of the first 40 amino acids being either lysine (K) or arginine (R) (Fig. 2). This is very similar to the amino acid pattern of the CarMV and TCV capsid protein genes (10,14).

Table 1 Nucleotides surrounding the beginning and end of ORFs found in MCMV.

ORF	AUG Position	Nucleotide Context <sup>a</sup>	Termination Codon	Position
p31.6	118-120	UUUCAUGC	UAA	985-987
p50	137-139	ACUUAUGG	UAG	1450-1452
p9	2959-2961	UUGAAUGG	UGA	3199-3201
p25	3384-3386	CACAAUGG	UGA	4092-4094

<sup>a</sup> Nucleotide sequence consensus surrounding the start of translation in mammalian mRNAs (12) is CACCAUGA/G and the consensus for plant mRNAs (13) is AACAAUGG.



MCMV		MATLPSIHA	9
MCMV	FWKLVWPTFSEERKMTVVQALRNGLTRKLFQITQIAQQILRWPPYI PERMTSSLGQSLVK		69
MCMV	TFQTLGLLRKALSRLCLTRNMVNDIPTMLLIPFKFLSQCISQIVASANFVIQTFRHGLSR		129
MCMV	GLCLARCRNPETMRSGWSKHTLQMNCSSTSPLMTPFPQIARDLSNPHGLWMTSRNKLT		189
CarMV	MCLPFLVEGVIGCTLVGGLVAVGSAALAVRATIGVVEFNREQVGRARRVSSGGCLWV		60
TCV	MLLHLTNTALAVGILGARYYPEVQTFILGLPDYVGHMKNVVSUFQSGSLMVSSSDTVGV		60
MCMV	RTGCSAIQQTSTRYVRLVRFPSLLAAGAVYISCVLARKALKKEVYRWELKSDQELPTREQSQ		249
CarMV	CSFQVNPNGCLINGEDEEIDNVEESTPVVEITFLIEVKAIVDGD--KRVVVSQKVVVNHILR		118
TCV	RGTYSNRQISSSICILAVPDSGADIEIILLNLVGTETETATSCLVENAVGSLADUPFERVR		120
MCMV	ISQVEQAMEQNGAMVMVKLNGPTAPRLLSLQCLTKHVLVPEIVDEGGTVVQEQSTFLIR		309
CarMV	Q--RFVVSIAIEARVHFGGDISPKNMVLVSKFTLQCKKERHVVBARHFDVSSAMVLM		176
TCV	QNGRFAMHAAVNAAKLHFGVPKPTBANKLAVSKMLVQYCKERHVVDSHRTIUNTALEPRV		180
MCMV	EFSPAVRNILVTLAKLIFCGIPKRIITANLIVVWFVLRKCEHANNMPTDSRTAITSALVVM		369
CarMV	FTPPVHEH--EMMAGTASDAAYGIKTAAPSILNRKQCKMIMVPLDRAMWENKCVVNGE		235
TCV	FTPPAADI--QVVLILISVRAHDEHNAIAEAGKVRKVMVNLAMETIGRSASRAWTKLCAI		239
MCMV	EMECRTLVGASIFLRDESIEICQYRAQFVEETP-LKRVFNPLTKKARRWVHSLGLI		428
CarMV	DSNKPVIFPK*GGLIFVINGVEYKIRRGGHSVIEVDGQCIFRERKLYVONAITTGYEYVH		294
TCV	EDDAISEVH*GCLRELVGEEEDQISRGENPAMRVVFLANEEHVRRIIFHICGNGCLDFGV		298
MCMV	DDPALFQELK*GCLRELVGVSRRITAEHFKLBSHFKDRHRTKRVVFDALGLNITVEFGV		488
CarMV	HNSAVNLRRLGILERVFYVER-NKELVSCPQEPGSH--RNGYVRRRFRVAGGNHTRISA		352
TCV	HNSLNLRLRGLMERVYVEDACQIKPAPQIPGIFGH--LSDIIRRIIVLALGNHTPVPR		357
MCMV	HNSAVNLRGLMERVYVVDKGLVSCFPPFEGITVWVLRGPNSTVHFVHGVVSN		548
CarMV	NDLVDCMGRKRTIYENAAASILDRAIERKDKLTKTKAERKVNIRSDPAPRVIQPRS		412
TCV	EKYPSFKGKRATIVKCALDLYMTCRISRKDAELKTFVKAEKINPANNDPAPRVIQPRD		417
MCMV	ETFLAVYISGRITVYEKAVNSIVHEPVYDDANLTKTFVKAEKINLNSADDFPVIQPRD		608
CarMV	PRYNVELGRYLKVEYEHNAVVALDKINGGPTVHKGYTTERVAQHISAWNCFQTPVAIGFD		472
TCV	PRYNVEGKYLKPYEYHNAVRAIDALMCGGPTVIRGYLVGELNIMSILMDFRKTCAIGFD		477
MCMV	PRYNVELGRYLKVEYEHNAVVAIDKINGGPTVHKGYVVEQIGHIDNAFRSITDPAIGFD		668
CarMV	MSRFDQHSVAATEPEHSQVLAQFEGDAHLANLTKQQLVHSGVGPASNSLITVGGCRM		532
TCV	MSRFDQHSVDAALRWEHSVYNGENCPQLAQLLWQLINAGVGRASDGFINYQVDGCRM		536
MCMV	MSRFDQHSVVALRWEHSVYSRIYGYPELITOLLVQVIRHSGTAVASDGFANYQVDGCRM		728
CarMV	SGDMNTALGNCLLACIT--EMLMR--IRSRLLNNGDDCVLICETDIDIVVSNITTCQSR		589
TCV	SGDMNTALGNCLLACIT--EMLMR--GIKCELLNNGDDCVLFEADEVDRVREIRLHH--MID		593
MCMV	SGDMNISLGNCLLATALTHDFVTLNIGLIPARLINNGDNLVLIQFAVEYGRVNOQLVRLN		788
CarMV	FGFNCAIAEEPVEEERKIFCOMAPVFDGAGWLMVRDHLVSMKDSHSLVHWNNEINAKQK		649
TCV	FGFNCAIAEEHVEYLEKVEFCOMSHFDGEGWVVRVPRVLSKDSYSITLQANNEKIDARW		653
MCMV	YGFVEVISEEPVYILEVEFCOMREVFDDGQYTHVDRPRTNSKDAVAVPNTHTDARW		848
CarMV	LKSVGCGILKAGGVVAVGESH---QRYVETAGVRENTWITEKSSSGFFMADRANG		705
TCV	LAPDCEGILKAGGVVILGYSYSLKRNFGPIAGDYK--RSDVSPDSGFFHLSRNGHRG		712
MCMV	MRAVGCGLISLTCGTPVQGEYH----TALVKHGLDFPNTIKKIDSGITMLSLNSK		908
CarMV	YSAVSVCARFSFYVAFGUTPDCIALGCEINSLIMTVGFL--QCEADSIWILNRKYQ		762
TCV	SKVSDARFSFYRGVYTPDECEALEEYVNDLLECEVPTGYNEELSDRWIILNEPFTL		773
MCMV	WQVQESAVSFWLAFGYTPDECEALEEYFNSMPTFMSSTIGLILGIFECLLILKHNLPPT		965

Fig 4 Homology between MCMV p111 CarMV p87 and TCV p87. The one letter amino acid codes for each polypeptide have been aligned using the FASTP protein homology program. The origin of the amino acid sequence is in the left margin and the amino acid residues of each viral polypeptide is shown in the right margin. Homology is designated by open boxes. Stars represent the amber terminators that presumably can be suppressed during translation. Amino acids found in the 'GDD' motif of viral RNA replicases (Kamer and Argos) have solid bars over them.

MCMV Homology to Carmoviruses

No significant nucleic acid homology was detected between MCMV and any plant viruses on the Bionet system. The

predicted amino acid sequence of MCMV-specific polypeptides show two regions of striking homology to polypeptides encoded by two other plant RNA viruses. Figure 4 shows the amino acid homology between MCMV p111, CarMV p87 (10) and TCV p87 (15). MCMV p111 has 46% homology to the TCV p87 and 41% homology to the CarMV p87. The homology is even greater within the region beyond the amber terminator (amino acid residue 439, 246 and 250 in MCMV, TCV and CarMV, respectively). All three polypeptides are found in the same region of the viral genome, result from the suppression of an amber termination codon and the terminators are located in the same position when the greatest amino acid homology is aligned (Fig. 4). The highly conserved 'GDD' motif found in a large number of RNA replicase genes from a wide variety of RNA viruses (10,15,16) is found within the readthrough portion of all three polypeptides (Fig. 4).

Significant amino acid homology also occurs within the carboxyl terminus of p9 and the corresponding regions of CarMV p7 and TCV p8 (Fig. 5). Eight of the last twenty amino acids in the MCMV p9 are identical to amino acids in either one or both of the other peptides. Of particular note is the FNF (Phe-Asn-Phe) tri-peptide sequence that is conserved in all three polypeptides.

#### DISCUSSION

The monopartite MCMV genome is 4437 nucleotides in length, which is in good agreement with size estimations from sucrose gradients (17) and denaturing agarose gels (5). Since the MCMV genome is not polyadenylated (5), the 3' terminus is currently defined to be the first adenine in the synthetic poly(A) tract. The possibility exists additional nucleotides are present at the 3' end of the viral RNA. Computer analysis of the nucleotide sequence of MCMV reveals at least four ORFs encoding polypeptides of  $M_r$  in excess of 8000 daltons. Readthrough of single termination codons would result in two additional polypeptides. Analysis of the sequence of the negative strand of MCMV RNA indicates an ORF that could encode a polypeptide of 18.2 kd. It is possible the 18.2 kd polypeptide is translated from some replicative intermediate of the virus in the infected cell. At this time however, there is no data showing the synthesis of such

```

CarMV p7      M I E S R P V V G K N L I G N R G K Q K T R R S V A K D A I R K P A S D S T N G C M V N V A D R L E V H I H F N F
TCV p8       M D P E R I P Y N S L S D A T G R K R K G G E K S A K K L V A S H A A S S V L N M R N E G S A S H C C M V I V A D K V E S I I P N F
MCMV p9      M V H N G Y F G R N S R M S S S Q T S P P T D N E R Q O T G T I V R S E D N H R N Q L E N I A V G L I T S E G A P A Q O N V I I N K R V I N N H F N F

```

Fig 5 Homology between MCMV p9 CarMV p7 and TCV p8. Amino acids have been aligned for greatest homology and open boxes represent identical amino acids.

a polypeptide in infected plants. The calculated sizes of the polypeptides encoded by four of the MCMV ORFs; p111, p31.6, p50 and p25 closely correspond to the observed molecular weights of the 105, 32, 52 and 25.4 kd polypeptides synthesized in vitro from MCMV RNA (5). It is not known if a in vitro polypeptide corresponding to p9 exists because proteins in the translation system have the same mobilities in acrylamide gels (Lommel and Nutter, unpublished). The data reported in this paper are consistent with p25 being the capsid protein. The single capsid protein for MCMV has been estimated to have a molecular weight of 25.4 kd and is synthesized from a 1.3 kb subgenomic RNA. A 25.4 kd polypeptide synthesized in vitro is specifically immunoprecipitated by antisera to the viral capsid protein (5). p25 is in the region of the viral genome expected for capsid protein, has biochemical properties similar to other capsid proteins (10,14) and is very close to the calculated size of MCMV capsid protein. The portion of the MCMV genome encoding the 41 and 44 kd polypeptides synthesized in vitro is not known at this time.

Two MCMV polypeptides, p111 and p32.7 may result from the occasional suppression of single translational termination codons. Suppression of termination codons has been seen in a number of single-stranded RNA viruses (10,15,18,19,20). In some instances, the resulting polypeptide is the replicase gene (10,15,18), but the QB capsid protein has been reported to result from the same mechanism (21). Suppression of translational termination codons has been observed both in vitro (19) and in vivo (22,23) and therefore has a central role in the expression of viral genomes.

The location of the p9 in the viral genome raises the possibility it is initiated from an internal AUG. This is unusual in that the genomes of nearly all single stranded RNA

plant viruses are considered to be monocistronic or functionally monocistronic (24). In some cases, internally encoded polypeptides are presumed to be translated from additional subgenomic RNAs (14,25,26). Most notable are the reports that CarMV and TCV, which shows considerable genome structure and amino acid sequence similarity, produces a second subgenomic RNA that is thought to encode the p7 and p8 polypeptides (14,26). Northern analysis of double stranded RNA isolated from MCMV-infected corn plants has detected a single subgenomic RNA that is the proper length for encoding the viral capsid protein (5). It is possible a subgenomic MCMV RNA for the p9 is synthesized only at low levels early in virus infection. This would be analogous to the I<sub>2</sub> subgenomic RNA responsible for synthesis of the 30 kd protein in TMV (25). The I<sub>2</sub> RNA has been detected as long as twelve hours post-infection.

The coding regions for p31.6 and p50 overlap for nearly the entire length of p31.6, indicating highly efficient genome organization. This has also been reported recently with the barley yellow dwarf virus (BYDV) capsid protein and a 17.1 kd ORF (20). The presence of overlapping polypeptides in MCMV and BYDV, and the possible internal initiation of the MCMV p9 suggests a modification of the scanning ribosome model for initiation of translation (12) may be necessary to account for translation of some polypeptides in these viruses.

There is no significant homology at the nucleic acid level between MCMV and other viruses contained in the indicated databases. Likewise, we found no amino acid homology between the MCMV capsid protein and other viral capsid proteins, although as noted above, the amino portion of the p25 of MCMV has a large number of basic amino acids like other viral capsid proteins (10,14). Significant amino acid homology does exist between two putative MCMV polypeptides and polypeptides encoded by CarMV and TCV. The similarity between the MCMV p111 and the putative CarMV and TCV replicase proteins is most striking. When the identical amino acids are aligned, very few gaps are needed to preserve alignment. Even the position of the termination codon presumed to be occasionally suppressed during translation is conserved.

The greatest amino acid homology is seen in the region of the 'GDD motif' (16). The presence of this amino acid sequence has been demonstrated in the putative RNA replicase of a wide variety of single stranded RNA viruses (10,15,20,27,28). This data argues the p111 is the MCMV-specific RNA replicase. The homology between the MCMV p8.9 and proteins of similar sizes in CarMV and TCV, while less pronounced, are significant. Conservation of these amino acids argues a similar polypeptide function.

Earlier reports have suggested classification of MCMV with the sobemovirus group based on the size of the viral genome, size of the viral particles and probable beetle vectors (29,30,31). The sequence of the cowpea strain of southern bean mosaic virus (SBMV-C) has been recently published (28) allowing comparison of the genome organization and nucleic acid and amino acid sequences of SBMV-C to MCMV. There is very limited amino acid homology between MCMV and SBMV in the immediate 'GDD' motif (16) of the putative polymerase genes. However, there are major differences in 5' cap structure, genome organization and no significant nucleic acid or amino acid homologies are found. These comparisons do not support the criteria used for the original classification of MCMV in the sobemovirus group.

There is considerable amino acid homology between potential polypeptides encoded by MCMV and polypeptides made by CarMV and TCV. Additionally, there are similarities in overall genome organization between MCMV, CarMV and TCV. The use of these criteria indicate that MCMV is more closely related to these members of the carmovirus group (32). As more detailed information is accumulated it may be possible to develop clearer taxonomic relationships between MCMV and other single-stranded RNA plant viruses.

ACKNOWLEDGEMENTS This is publication number 5509 from the School of Agriculture, Oklahoma State University. The research was funded in part by a NSF/EPSCOR grant to Oklahoma State University and a grant from Stauffer Chemical Company to SAL.

\*To whom correspondence should be addressed

REFERENCES

1. Niblett, C.L. and Claflin, L.E. (1978) *Plant Dis. Rep.* 62, 15-19.
2. Bockelman, D.L., Claflin, L.E. and Uyemoto, J.K. (1982) *Plant Dis.* 66, 216-218.
3. Uyemoto, J. (1983) *Plant Dis.* 67, 1-10.
4. Nault, LR, Styer, WE, Coffey, ME, Gordon, DT, Negi, LS and Niblett, CL (1978) *Phytopathology* 68, 1071-1074.
5. Lommel, S.A., Kendall, T. and Nutter, R.C. (1988) *Journal of General Virology* Submitted.
6. Gübler, U. and Hoffman, B.J. (1983) *Gene* 25, 263-269.
7. Maniatis, T, Frisch, EF and Sambrook, J (1982) *Cold Springs Harbor Press, New York.*
8. Henikoff, S (1987) *Methods in Enzymology* 155, 156-165.
9. Viera, J. and Messing, J. (1987) *Methods in Enzymology* 153, 3-11.
10. Güilley, H., Carrington, J.C., Balazs, E., Jonard, G., Richards, K and Morris, T.J. (1985) *Nucleic Acids Res.* 13, 6663-6677.
11. Lipman, D.J. and Pearson, W.R. (1985) *Science* 227, 1435-1441.
12. Kozak, M. (1980) *Nucleic Acids Res.* 9, 5233-5252.
13. Lutcke, H.A., Chow, K.C., Mickel, F.S., Moss, K.A., Kern, H.F. and Scheele, G.A. (1987) *EMBO J.* 6, 43-48.
14. Carrington, J.C., Morris, T.J., Stockley, P.G. and Harrison, S.G. (1987) *J. Mol. Biol.* 194, 265-276.
15. Carrington, J.C., Heaton, L.A., Zuidema, D., Hillman, B.I., and Morris, T.J. (1989) *Virology* in press.
16. Kamer, C. and Argos, P. (1984) *Nucleic Acids Res.* 12, 7269-7282.
17. Ferguson, M.W., Shahin, E.A. and Uyemoto, J.K. (1978) *Phytopathol. News* 12, 87.
18. Goelet, P, Lomonosoff, G.P., Butler, P.J.G., Akam, M.E. Gait, M.J. and Karn, J. (1982) *Proc. Nat. Acad. Sci. (USA)* 79, 5818-5822.
19. Morch, M.D. and Benicourt, C. (1980) *Eur. J. Biochem.* 105, 445-451.
20. Miller, W.A., Waterhouse, P.M. and Gerlach, W.L. (1988) *Nucleic Acids Res.* 16:6097-6111.
21. Weiner, A.M. and Weber, K. (1971) *Nature New Biology* 234, 206-209.
22. Beier, H., Barciszewska, M., Krupp, G., Mitnacht, R. and Gross, H.J. (1984) *EMBO J.* 3, 351-356.
23. Beier, H., Barciszewska, M. and Sickinger, H.D. (1984) *EMBO J.* 3, 1091-1096.
24. Joshi, S (1984) *FEBS Letters* 177, 163-174.
25. Watanabe, Y., Emori, Y., Ooshika, I., Meshi, T., Ohno, T. and Okada, Y. (1984) *Virology* 133, 18-24.
26. Carrington, J.C. and Morris, T.J. (1985) *Virology* 144, 1-10.
27. Domier, L.L., Shaw, J.G. and Rhoads, R.E. 1987 *Virology* 158, 20-27.
28. Wu, S-X, Rinehart, CA and Kaesberg, P (1987) *Virology* 161, 73-80.
29. Tremaine, J.H. and Hamilton, R.I. (1983) In: *Description of plant viruses.* Commonwealth Mycological Institute/

- Association of Applied Biologists. #274 Kew, Surrey, England.
30. Gordon, D.T., Bradfute, O.E., Gingery, R.E., Nault, L.R. and Uyemoto, J.K. (1984) In: Descriptions of plant viruses. Commonwealth Mycological Institute/Association of Applied Biologists. #284 Kew, Surrey, England.
  31. Hull, R. (1988) In: R. Koenig, (ed.), The Plant Viruses. Plenum Publishing Corp. Vol. 3. pp.113-146.
  32. Morris, T.J. and Carrington, J.C. (1988) In: R. Koenig, (ed.), The Plant Viruses. Plenum Publishing Corp. Vol. 3 pp. 73-112.