The complete nucleotide sequence of cucumber green mottle mosaic virus (SH strain) genomic RNA


1 National Institute of Agrobiological Resources, 2-1-2 Kan-nondai, Tsukuba, Ibaraki 305, 2 Hokkaido Central Agricultural Experiment Station, Nagamuna, Yu-ubari, Hokkaido 069-13, 3 Miyazaki Prefectural Agricultural Experiment Station, Sadowara, Miyazaki-gun, Miyazaki 880-02 and 4 Kyushu National Agricultural Experiment Station, 2421 Suya, Nishigoushi, Kikuchi-gun, Kumamoto 861-11, Japan

The complete nucleotide sequence of the genomic RNA of cucumber green mottle mosaic virus watermelon strain SH (CGMMV-SH) was determined using cloned cDNA. This sequence is 6421 nucleotides long containing at least four open reading frames, which correspond to 186K, 129K, 29K and 17.3K proteins. The 17.3K protein is the coat protein. Sequence analysis shows that CGMMV-SH is very closely related to another watermelon strain, CGMMV-W, although three amino acid substitutions in the 29K protein were found between these strains. The sequence was also compared to those of other tobamoviruses, tobacco mosaic virus (TMV) vulgare, TMV-L (a tomato strain) and tobacco mild green mosaic virus reported by other groups. It shows 55 to 56% identity with these viruses. The size and location of the open reading frames are very similar to those of TMV but the 129K and 186K proteins are composed of 1142 and 1646 amino acids, being larger than those of TMV by 27 and 31 amino acids, respectively. The deduced amino acid sequences of these proteins are highly homologous to those of TMV, especially in the readthrough downstream region of the 186K protein.

Introduction

Cucumber green mottle mosaic virus (CGMMV) is a member of the tobamovirus group. CGMMV causes diseases in cucurbitaceae and is different in host range from tobacco mosaic virus (TMV) whose main hosts are members of the Solanaceae. Strains of CGMMV were first described as cucumber virus 3 (CV3) and cucumber virus 4 (CV4) by Ainsworth (1935). In Japan, three different strains have been described: 'cucumber strain' (Inoue et al., 1967), 'watermelon strain' (Komuro et al., 1968) and 'Yodo strain' (Kitani et al., 1970). Among them, the watermelon strain of CGMMV is the most serious disease agent, causing severe disease symptoms in infected watermelon plants, particularly the deterioration of fruit pulp which causes considerable economic losses to watermelon growers (Komuro et al., 1971). Tchihara & Komuro (1974) demonstrated that the watermelon strain has a closer relationship to an Indian bottlegourd isolate, CGMMV-C (Vasudeva et al., 1949; Vasudeva & Nariani, 1952), than to the cucumber strain, the Yodo strain or two British isolates, E1 and E2. Francki et al. (1986) pointed out that the CGMMV watermelon strain is taxonomically different from the cucumber strain (Inoue et al., 1967), shown by the lack of molecular hybridization between these two strains, and proposed a new virus name, 'kyuri green mottle mosaic virus', for the cucumber strain.

The sequence of 1071 nucleotides from the 3' end of the genomic RNA of a watermelon isolate (CGMMV-W) of the watermelon strain was determined by Meshi et al. (1983b) by analysing the cDNA of the genomic RNA. The amino acid sequence of the coat protein of the same isolate was determined by Nozu & Tsugita (1986) by the conventional protein sequencing method. Furthermore Saito et al. (1988) sequenced the region covering the 30K protein gene and compared it with the known 30K protein of other tobamoviruses.

A new type of mosaic disease was found in green-house-grown muskmelon crops in 1971, and the agent of this disease was serologically identical with the watermelon strain (Furuki & Komuro, 1973). Necrotic lesions surrounded by a water-soaked area are characteristic of
muskamelon fruit infected with the virus, and severely affected the economic value of the products (I. Furuki & T. Ohsawa, personal communication). However, the disease of muskmelon caused by CGMMV can be controlled by cross-protection, if the seedlings of disease of muskmelon caused by CGMMV can be affected by cross-protection, if the seedlings of CGMMV strain (Motoyoshi & Nishiguchi, 1988).

We have determined the complete nucleotide sequence of the genomic RNA of CGMMV-SH, a muskmelon isolate of the CGMMV watermelon strain derived from a diseased muskmelon, by preparing cDNA clones covering almost the full-length of the RNA and sequencing them with an automated fluorescent DNA sequencer, and also, partially, by directly sequencing the genomic RNA with synthetic oligonucleotides. We also discuss the difference at the level of nucleotide sequence of this isolate from other tobamoviruses.

**Methods**

**Virus purification and RNA isolation.** CGMMV-SH is an isolate from a muskmelon leaf that was supplied by Dr I. Furuki, Shizuoka Agricultural Experiment Station. The virus sample for this study was obtained after five passages through local lesions in the leaves of *Gomphrena globosa* and propagated in muskmelon plants (*Cucumis melo* L. cv. Earl's Favorite Natsuke-i-4).

Plants inoculated with the virus were grown in the greenhouse. About 2 weeks after inoculation, the leaves were harvested and stored at -60 °C. Virus was purified according to the method reported by Nozu et al. (1971). Viral RNA was isolated from the purified virus preparation using a phenol-SDS method (Gierer & Schramm, 1956; Fraenkel-Conrat et al., 1957). The RNA was resuspended in H2O and kept at -70 °C until use.

**Polyadenylation of virus RNA.** Virus RNA was polyadenylated at the 3' end using poly(A) polymerase by the method of Mishi et al. (1982) with a slight modification. Two-hundred μl of the reaction mixture containing 50 mM-Tris-HCl pH 8.0, 0.2 M-NaCl, 10 mM-MgCl2, 0.4 mM-EDTA, 1 mM-DTT, 30 μg RNA, 8.1 mM-[γ-32P]ATP (Amersham, 50 Ci/mol) and poly(A) polymerase (Bethesda Research Laboratories, 20 units) was incubated at 37 °C for 13 min. The resulting poly(A)-tailed RNA was passed through a Sephadex G-50 column and collected at -70 °C until use.

**Preparation of double-stranded cDNA to polyadenylated CGMMV RNA, and construction and transformation of hybrid plasmids.** Using the cDNA synthesis system (Amersham) based on the method of Gubler & Hoffman (1983), cDNA was synthesized by the manufacturer's protocol. Larger sized cDNA was recovered by fractionation through a Sephadex G-50 column. Addition of oligo(dC) to the double-stranded cDNA at the 3' end was carried out using terminal deoxynucleotidyl transferase (Takara Shuzo Company). The tailed cDNA (about 100 ng) was annealed with 660 ng of the vector, oligo(dG)-tailed pUC9 (Pharmacia). The hybrid plasmid RNA was used to transform competent cells of *Escherichia coli* JM101 prepared by the modified RbCl/CaCl2-mediated method (Hanahan, 1985). White colonies on the plates of selection medium were picked and used for colony hybridization.

**Preparation of probes and colony hybridization.** Partially degraded CGMMV-SH RNA or the Ω fragment of TMV-OM RNA was used as a probe after labelling at the 5' end with [γ-32P]ATP by the method described by Mendiola et al. (1982). The Ω fragment of TMV-OM RNA was prepared as described by Mandeleis (1968). Colony hybridization was performed at 40 °C overnight using a probe of approximately 105 c.p.m. per filter (nitrocellulose filter BA85, Schleicher & Schüll) as described by Ohshima (1981). For the SH RNA probe, the filter was washed once with 4 × SSC, twice with 2 × SSC for 15 min each and then treated with RNase A at 20 μg/ml for 30 min following by washing with 2 × SSC twice for 15 min. When the TMV-OM RNA Ω fragment was used as a probe, the filter was washed only twice with 4 × SSC.

**cDNA sequencing.** Sequencing of cDNA was performed by the dideoxynucleotide method (Sanger et al., 1977). The cDNA was cloned into the pBluescript vector (Stratagene) and was used to make a series of deletion clones as reported by Henikoff (1984). The deletion clones were sequenced using an automated fluorescent DNA sequencer (370A, Applied Biosystems) as described by Ugaki et al. (1988). The DNA was sequenced in both directions.

**RNA sequencing.** The sequence of the coat protein-encoding and 3' non-coding region of SH RNA was further confirmed by direct sequencing of SH RNA using reverse transcriptase (Seikagaku Kogyo) as described by Mishi et al. (1983a). The primer was a 5'-labelled cDNA fragment which had been digested with the appropriate restriction enzymes. The 5'-terminal region of the genome RNA was determined as follows: (i) SH-RNA was annealed to a synthetic primer (GGCGTACGGTGGTTGATT, positions 81 to 102) which had been labelled with [γ-32P]ATP and extended with reverse transcriptase at the 5' end and (ii) the base at the 5' end was identified by the methods described by Moss (1977) and Konarska (1984). The 3' end of SH-RNA was determined as described by Miura et al. (1974) after labelling of the RNA with T4 RNA ligase (England et al., 1980).

**Sequence analysis.** Sequence data were analysed using DNAMAN (Hitachi).

**Results and Discussion**

**Cloning**

For cDNA synthesis, virus RNA was extracted from purified virus particles and subjected to agarose gel electrophoresis. Two main RNA bands were detected (data not shown). The main larger one corresponded to full-length genomic RNA. The smaller RNA band was presumed to be derived from short virus particles as shown by Okada et al. (1980) and Fukuda et al. (1981).

Some of the cDNA clones are shown in Fig. 1. All of these were selected using fragmented SH-RNA probes. The largest cDNA clone, pSH-K-3, covers almost the whole length of the genomic RNA, but lacks only 33 nucleotides at the 5' end, proved by sequencing using a cDNA primer. This and two other clones (pSH-G-6 and pSH-J-44) were found to hybridize with the probe of TMV-OM RNA Ω fragments. Clones of pSH-G-6, pSH-J-44 and pSH-P-11 were thought to be derived from the fragments which lost large 3' portions of SH-RNA prior to the addition of poly(A).
Genome sequence of CGMMV

Sequencing was performed using a series of deletion clones from pSH-K-3 in both directions. The region not covered by the clone was directly sequenced for SH-RNA as described above. The genome organization of CGMMV-SH was compared to three other sequenced tobamoviruses, TMV vulgaris (Goelet et al., 1982), TMV-L (Ohno et al., 1984) and tobacco mild green mosaic virus (TMGMV) (Sols & Garcia-Arenal, 1990) (Fig. 2). The genome of CGMMV-SH is 6421 nucleotides long. At least four open reading frames (ORFs) were found in the positive strand, encoding putative proteins of 186K, 129K, 29K and 17.3K. The proteins encoded correspond to the 180K, 130K, 30K and 17.5K proteins of TMV vulgaris or TMV-L. Each ORF of the CGMMV-SH genome is slightly larger than the corresponding ORF of the other tobamoviruses in the number of nucleotides except for the ORF encoding the 30K protein of TMV vulgaris and TMV-L. Some ORF-like structures, at the largest 324 nucleotides long, were found in the negative strand (data not shown). It is apparent, however, that they are rather short to encode proteins although we have not yet tested this by in vitro or in vivo translation assays.

The complete nucleotide sequence of CGMMV-SH is shown in Fig. 3. The first AUG initiation codon was found at residues 61 to 63, which starts an ORF encoding a protein composed of 1142 amino acids (129K). This ORF terminates at an amber codon, UAG, positioned at residues 3490 to 3492. The ORF encoding the read-through protein composed of 1646 amino acids (186K) terminates at residues 5002 to 5004. The terminal 14 bases of the ORF encoding the 186K protein overlap the ORF encoding the 29K protein, which terminates with an amber codon. The coat protein gene initiates at 25 bases upstream from the terminal nucleotide of the amber codon for the 29K protein. The number of nucleotides encoding the 29K protein was exactly the same as that of another CGMMV isolate (CGMMV-W) (Meshi et al., 1983b; Saito et al., 1988). The coat protein gene of our CGMMV isolate was also found to be composed of the same number of nucleotides as that found in CGMMV-W (Meshi et al., 1983b). In total, 27 nucleotide substitutions (six in the 186K protein, 14 in the 29K protein and seven in the coat protein) were found between CGMMV-SH and CGMMV-W, when the sequenced region (1878 nucleotides from the 3' end) of CGMMV-W was compared (Fig. 4). They include 25 transitions (15 between T and C, 10 between A and G) and two transversions (one between T and A, one between A and C). Twenty-five are in the third codon position and only two are in the first. Three resulted in amino acid substitutions which were all found in the 29K protein. The nucleotide substitutions at 5375 (CCT→TCT), 5675 (GGT→AGT) and 5758 (GAA→GAC) correspond to P in CGMMV-SH for S in CGMMV-W (at position 128), G for S (228) and E for D (255), respectively. The number of T residues in a cluster (6401 to 6404) is one nucleotide less in CGMMV-SH than that in the corresponding T cluster of CGMMV-W.

CGMMV-SH was isolated originally in Shizuoka Prefecture which is more than 200 km distant from Chiba Prefecture where CGMMV-W was isolated from watermelon. The number of nucleotide substitutions between these two strains was, however, very small, suggesting that CGMMV-SH is a variant of the CGMMV watermelon strain. It is likely that the three amino acid substitutions found in the 29K protein are responsible for adaptation of these strains to different host species. The 30K protein of tobamoviruses is known to be necessary for the movement of virus from cell to cell (Deom et al., 1987; Meshi et al., 1987). However, it remains to be determined whether CGMMV-SH moves from cell to cell more efficiently in muskmelon plants than does CGMMV-W.
M. Ugaki and others

> I29K. 186K

~ A N I N E Q I N N O R D A A A S G R N
GTTTTAATTTTTATAATTAAACAAACAACA
ACAACAACAACAAACAATTTTAAAACAACA
ATGGCAAACATTAATGAACAAATCAACAACAACGTGACGCCGCGGCTAGCGGGAGAAAC 120

NLVSQLASKRVYDEAVRSLDHODRRPKMNFSRVVSTEHTR
AATCTCGTTAGCCAATTGGCGTCAAAAAGG
GTGTATGACGAGGCTGTTCGCTCGTTGGATCATCAAGACAGACGCCCGAAAATGAATTTT
TCTCGTGTGGTCAGCACAGAGCACACCAGG 240

LVTDAYPEFSISFTATKNSV HSLAGGLRLLELEYMMMQVP
CTTOTAACTGACGCGTATCCGGAGTTTTCG
ATTAGCTTTACCGCCACCAAGAACTCTGTA
~ CTCCC~GCGGG%GG~C~GAGGC~C~GAA~TGGAA~A~ATGATGATGCAGGTGCCC 360

YOSPCYDIGGNYTQHLFKGRSYVHCCNPCLDLKDVARNVM
TACGGCTCACCTTGTTATGACATCGGCGGT
AACTATACGCAGCACTTGTTCAAAGGTAGA
TCATATGTGCATTGCTGCAATCCGTGCCTA
GATCTTAAGGATGTTOCGAGGAATGTGATG 480

YNDMITQHVQ RHKGSCGCHP LFTFQIDAFRRYDSSPCAVT
TACAACGATATTOATTACGCAACATGTACAG
AGGCACAAGGGATCTTGCGGGTGCAGACCT
CTTCCAACTTTCCAGATAGA~GCATTCAGG
AGGTACGATAGTTCTCCCTGTGCGGTCACC 600

CSDVFQECSYDFGSGRDNHA VSLHSIYDIPYSSIGPALHR
TGTTCAGACGTTTTCCAAGAGTGTTCCTAT
GATTTTGGGAGTGGTAGGGATAATCATGCA
GTCTCGTTGCATTCAATCTACGATATCCCT
TATTCTTCGATCGGACCTGCTCTTCATAGG 720

KNVRVCYAAFHFSEALLLGS PVGNLNSIGAQFRVDGDDVH
AAGAATGTGCGAGTTTGTTATGCAGCCTTT
CATTTCTCGGAGGCA~TGCTTTTAGGTTCG
CCTGTAGGTAATTTAAATAGTATTGGCGCT
CAGTTTAGGGTCGATGGTGATGATGTGCAT 840

FLFSEESTLHYTHSLENIKLIVMRTYFPADDRFVYIKEFM
TTTCTTTTTAGTGAAGAGTCTACTTTGCAT
TATACTCATAGTTTAGAAAATATCAAGTTA
ATCGTGATGCGTACTTACTTTCCTGCTGAT
GATAGGTTTGTATATATTAAGGAGTTCATG 960

VKRVDTFFFRLVRADTHMLHKSVGHYSKSK SEYFALNTPP
GTTAAGCGTGTGGATACTTTTTTCTTTAGG
TTGGTCAGAGCAGATACACACATGCTTCAT
AAATC~GTGGGGCACTATTUGAAATGGAAG~C¢GAGTACTTCGCGCTGAATACCCCTCCO 1080

IFQDKATFSVWFPEAKKVLIPKFELSRFLSGNVKISRMLV
ATCTTCCAAGATAAAGCCACGTTTTCTGTG
TGGTTTCCTGAAGCGAAGAAGGTGTTGATA
CCCAAGTTTGAACTTTCGAGATTCCTTTCT
GGGAATGTGAAAATCTCTAGGATGCTTGTC 1200

OADFVHTIINHISTYDNKALVWKNVQSFVESIRSRVIVNG
GATGCTGATTTCGTCCATACCATTATTAAT
CACATTAGCACGTATGATAACAAGGCCTTA
GTGTGGAAGAATGTTCAGTCCTTTGTGGAA
TCCATACGTTCAAGAGTAATTGTAAACGGA 1320

VSVKSEWNVPVDQLTDISFS IFPLVKVRKVQIELMSDKVV
GTTTCCGTGAAATCTGAGTGGAACGTACCG
GT TGATCAGCTCACTGATATCTCGTTCTCG
ATATTCCCTCTCGTGAAGGTTAGGAAGGTA
O~OA~CGAGTTAATGTCTGATAAAGTTGTA 1440

IEARGLLRRFADSLKSAVEG LCDCVYDALVQTGWFDTSSD
ATCGAGGCGAGGGGTTTGCTTCGGAGGTTC
GCAGACAGTCTTAAATCTGCCGTAGAAGGA
CTAGGTGATTGCGTCTATGATGCTCTAGTT
CAAACCGGCTGGTTTGACACCTCTAGCGAC 1560

ELKVLLPEPFMTFSDYLEGMYEADAKIEHESVSELLASGD
GAACTGAAAGTATTGCTACCTGAACCGTTT
ATGACCTTTTCGGATTATCTTGAAGGGATG
TACGAGCCAGATGCAAAGATCGAGAGAGAG
GTGTCTCTGAGTTGCTCGCTTCCGGTGAT 1680

DLFKKIDEIRNNYSGVEFDV EKFQEFCKELNVNPMLIGHV
GATTTGTTCAAGAAAATCGATGAGATAAGA
AACAATTACAGTGGAGTCGAETTTGATGTA
GAGAAATTCCAAGAA~TTGCAAGGAACTG
AATGTTAATCCTATGCTAATTGGCCATGTC 1800

IEAIFSQKAGVTVTGLGTLSPEMGASVAL$STSVDTCEDM
ATCGAAGCTATTTTTTCGCAGAAGGCTGGG
GTAACAGTAACGGGTCTGGGCACGCTCTCT
CCTGAGATGGGCGCTTCTGTTGCGTTATCC
AGTACCTCTGTAGATACATGTGAAGATATG 1920

DVTEDMEDIVLMADKSHSYMSpEMARWADVKYGNNKGALV
GATGTAACTGAAGATATGGAGGATATAGTG
TTqATGGCGGACAAGAGTCATTCTTACATG
TCCCCTGAAATGGCGAGATGGGCTGATGTT
AAATATGGCAAOAATAAAGGGGCTCTAGTC 2040

IEAIFSQKAGVTVTGLGTLSPEMGASVAL$STSVDTCEDM
ATCGAAGCTATTTTTTCGCAGAAGGCTGGG
GTAACAGTAACGGGTCTGGGCACGCTCTCT
CCTGAGATGGGCGCTTCTGTTGCGTTATCC
AGTACCTCTGTAGATACATGTGAAGATATG 2160

EYKVGTSMTLPATWAEKVKAVLPLSGICVRKPQFSKPLDE
GAGTACAAAGTCGGAACCTCGATGACTTTA
CCTGCCACCTGGGCAGAGAAAGTTAAGGCT
GTCTTACCGTTGTCGGGGATCTGTGTGAGG
AAACCCCAATTTTCGAAQCCGCTTGATGAG 2280

EDDLRLSNMNFFKVSDLKLK KTITPVVYTG TIRERQMKNY
GAAGATGACTTGAGGTTATCAAACATGAAT
TTCTTTAAGGTGAGCGATCTAAAGTTG~AG
AAGACTATCACTCCAGTCGTTTACACTGGG
AOCATTCGAGAGAGGCAAATGAAGAATTAT 2280

IDYLSASLGSTLGNLERIVRSDWNGTEESMQTFGLYDCEK
ATTGATTACTTATCGGCCTCTCTTGGTTCC
ACGCTGGGTAATCTGGAGAGAATCGTGCGG
AGTGATTGGAATGGTACTGAGGAGAGTATG
CAAACGTTCGGGTTGTATGACTGCGAAAAG
IDRHTKAMVYYTVVFDAVTS IIADVEKVDO SILTMFATTV 2400

SVKAIKVSGAGILRPELTKIKGKIITFTOSDKQSLIKSGY
TC~AAGGCARTTAA~TCAG~GCC GGTATTCTGAGGCCTGAGTTDACAAAGATC
AAAGGAAAGATAATAACGTTTACTCAATCT
GATAAGCAGTCCTTGATCAAGAGTGGGTAC 2520

NDVNTVHEIO GET ETT ATV RAPTPTGLALT ANDISAVLA 2640

DGVPGCGKTAEIIARVNWKTDLVLTPGREAAAMIRRRACA
GACGGAGTGCCGGGTTGTGGAAAGACCGCC
GAGKTTATAGCGAGGGTCAATTGGAAAACT
GATCTAGTATTGACTCCCGGAAGGGAGGCA
GCTGCTATGATTAGGCGGAGAGCCTGCGCC 2760

LHKSPVATNDNVRTFDSFVMNRKIFKFDAVYVDEGLMVHT
CTOCACAAGTCACCTGTGGCAACCAATGAC
AACGTCAGAACTTTCGATTCTTTTGTGATG
AATAGGAAAATCTTCAAGTTTGACGCTGTG
TATGTTGACGAGGGTCTGATGGTCCATACG 2880

GLLNFALKISGCKKAFVFGDAKQIPFINRV MNFDYPKELR
GGATTACTT&~TTTTGCGTT~K~TCTCA GGT~GTA~AAGCCTTCGTCTT~GGTGAT
GCTAAGCAAATCCCGTTTATAAACAOAGTC
ATGAATTTCGATTATCCTAAGGAGTTAAGA 2900

TLIVDNVERRYVTHRCPRDV TSFLNTIYKAAVATTSPVVH
ACTTTAATAGTCGATAATGTAGAGCGTAGG
TATGTCACCCATAGGTGTCCTAGAGATGTC
ACTAGTTTTCTTAATACTATCTATAAAGCC
GCTGTCGCTACTACTAGTCCGOTTGTACAT 3120

SVKAIKVSGAGILRPELTKIKGKIITFTOSDKQSLIKSGY
TC~AAGGCARTTAA~TCAG~GCC GGTATTCTGAGGCCTGAGTTDACAAAGATC
AAAGGAAAGATAATAACGTTTACTCAATCT
GATAAGCAGTCCTTGATCAAGAGTGGGTAC 3240

NDVNTVHEIQGETFEETAVV RATPTPIGLI ARDSPHVLVA
ARDSPHVLVA
AA~A~AATACTG~CA~AAATTCAG GGAGAAACCTT~AGGAGACG~CAGT~TG CGTGCCACCCCGACTCCAATAGGTTTGATT GCCCGTGATTCACCACATGTACTAGTGGCC 3360

LTRHTKAMVYYTVVFDAVTS IIADVEKVDO SILTMFATTV 3480
Fig. 3. The complete nucleotide sequence of the CGMMV-SH genome RNA and the deduced amino acid sequences of ORFs. The numbering refers to the nucleotide sequence. The amino acid sequences are given in the one-letter code. The nucleotide sequence within 1878 nucleotides from the 3' end determined by Meshi et al. (1983b) and Saito et al. (1988), and the deduced amino acid sequences of the CGMMV-W genome RNA are shown only in the positions where they are different from those of CGMMV-SH. The wavy line marks a T cluster where the number of Ts is three in CGMMV-W.
The identity of the entire nucleotide sequences is approximately 55 to 56% between CGMMV-SH and TMV *vulgare*, between CGMMV-SH and TMV-L or between CGMMV-SH and TMGMV.

5' and 3' Non-coding regions

The 5' non-coding region of the ordinary TMV (such as *vulgare*) is referred to as Ω and is characteristically free of G residues (Mandeles, 1968). The 5' non-coding regions of several tobamoviruses are compared in Fig. 4. The deletion of the 5'-terminal eight nucleotides (GUAAUUUU) of TMV-L caused the complete loss of infectivity and the other sequence of the region seemed not to be crucial although a large deletion caused considerable effects (Takamatsu et al., 1991). The role of the sequence of eight nucleotides is however not known. This region is well conserved among five tobamoviruses. The ribosome-binding site is thought to be AUU (Tyc et al., 1984), being present in all the five viruses. The three nucleotides (ACA) upstream and the three (GCA) downstream to the initiation codon are also well conserved. The 5' non-coding region of CGMMV-SH has 60 nucleotides and is 67 to 68% similar to that of TMV *vulgare*, TMV-L and a Spanish isolate of pepper mild mottle virus (PMMV-S; Avila-Rincon *et al.*, 1989) or TMGMV. The first nucleotide of the CGMMV-SH genome was found to be G, as are those of the other three (TMV *vulgare*, TMV-L and TMGMV).

Although the nature of the cap was not determined, we assume it to be m7Gppp. The 3' non-coding region of CGMMV-SH genome is composed of 176 nucleotides. The 3'-terminal sequence of CGMMV-SH fits the pseudoknot structure model which was proposed for tobamovirus genomes by Van Belkum *et al.* (1985) (data not shown).

Open reading frames

No CGMMV-encoded proteins except for the coat protein have yet been identified *in vivo* (Okada, 1986). However by analogy with other tobamoviruses, at least four ORFs can be identified on the genome sequence.

The putative 129K and 186K proteins correspond to the 130K and 180K proteins of TMV *vulgare* and TMV-L. The former protein of CGMMV-SH has 1142 amino acids which is 27 amino acids more than those of TMV *vulgare* and TMV-L. The 186K readthrough product of
CGMMV-SH is longer by 31 amino acids than those of TMV vulgar and TMV-L. A homology plot of the amino acids of the 180K (130K) protein is shown in Fig. 5. Amino acid sequence identity of the CGMMV-SH 186K protein is approximately 48% with the corresponding 180K protein of TMV vulgar.

The N-terminal one-third, the C-terminal one-third and the remainder of the CGMMV-SH 129K protein have identities of 48%, 47% and 33% with those of TMV vulgar respectively. These values are the same for TMV-L except for that of the C-terminal one-third of the 130K protein (49% identity). The level of similarity of the amino acid sequence of the middle one-third portion of the CGMMV-SH 129K protein is not as high as those of the other portions. The readthrough part of CGMMV-SH 186K has 58% identity with the corresponding regions of TMV vulgar and TMV-L, which is higher than those of any other regions of the protein.

Rozanov et al. (1990) indicated that the N-terminal portions of large putative NTPases of 'Sindbis-like' plant viruses including tobamoviruses might be methyltransferases. It is of interest that the 130K protein of TMV-U1 (vulgar) was found to have guanylyltransferase-like activity (Dunigan & Zaitlin, 1990). Although the capping mechanism of plant virus RNA is not known, it might be possible that motifs for capping-related enzyme activities are located in the N-terminal one-third portion of the tobamovirus 130K protein.

Habili & Symons (1989) showed that positive-strand viruses could be grouped based on amino acid sequence motifs of nucleic acid helicases and RNA polymerases, and proposed a new luteovirus supergroup. The amino acid sequences were compared in the presumed motifs in Fig. 6. These motifs in the 186K protein of CGMMV-SH are characteristic for tobamoviruses. The amino acid sequence identities in the helicase motifs region (a) were 56% and 53% with TMV-L and TMV vulgar and those in the polymerase motifs region (b) were 70% and 67%, respectively.

A homology plot of amino acids of the 30K protein between CGMMV-SH and TMV vulgar is also shown in Fig. 5. Saito et al. (1988) showed local high homology in the 30K protein among tobamoviruses and a tobravirus. The 30K protein is thought to have at least two functions, binding to single-stranded nucleic acid (Citovsky et al., 1990) and interacting with plasmodesmata of host cells.

Fig. 6. Comparison of partial amino acid sequences of the 186K protein of CGMMV-SH with those of the 180K protein of TMV vulgar (Goelet et al., 1982). These sequences contain motifs indicated by Habili & Symons (1989). (a) The region contains nucleic acid helicase motifs; (b) the region contains RNA polymerase motifs. Upper and lower lines are the amino acid sequences of CGMMV-SH and TMV vulgar, respectively. Identical amino acids are written in white letters. Motifs are overlined. Roman letters represent numbers of motifs used by Habili & Symons (1989). Gaps have been introduced to obtain the closest match.
(Tomenius et al., 1987). A closely similar region exhibited in Fig. 5 may be important in the binding of the 30K protein to single-stranded nucleic acid, which was indicated by Citovsky et al. (1990).

The authors are very grateful to Drs I. Furuki and T. Ohsawa for providing them with the sample of CGMMV-SH and the muskmelon seeds.

**References**


(Received 29 November 1990; Accepted 25 March 1991)