Volume 14 Number 11 1986 Nucleic Acids Research

RNA 2 of tobacco rattle virus strain TCM encodes an unexpected gene

Gerco C.Angenent, Huub J.M.Linthorst, Alex F.van Belkum, Ben J.C.Cornelissen and John F.Bol*

Department of Biochemistry, State University of Leiden, PO Box 9505, 2300 RA Leiden, The Netherlands

Received 11 March 1986; Revised and Accepted 2 May 1986

ABSTRACT

The sequence of the 3'-terminal 1210 nucleotides of RNA 1 and the complete sequence of 3389 nucleotides of RNA 2 of tobacco rattle virus (TRV) strain TCM has been deduced. The sequence of the 3'-terminal 1099 nucleotides of RNAs 1 and 2 was found to be identical. Thus the genome of this TRV strain is partially diploid, encoding a 16K protein in both RNA 1 and RNA 2. The sequence that is unique to RNA 2 contains two open reading frames: the coat protein cistron and a cistron for a 29.1K protein, which shows no homology with the RNA 1 encoded 28.8K protein. cDNA probes corresponding to these two open reading frames cross-hybridized to pea early-browning virus RNA 2, but not to RNA 2 of five other tobraviruses tested.

INTRODUCTION

The genome of tobraviruses is bipartite and consists of RNAs 1 and 2 (1). The members of this group can be divided into three clusters which mainly differ in the sequence of their RNA 1 molecules (2,3). Cluster 1 contains strains of tobacco rattle virus (TRV) serotype I-II, cluster 2 contains strains of pea early-browning virus (PEBV) and cluster 3 is represented by the CAM-strain of TRV (2). *In vitro* translation studies have shown that RNA 1 of strains PRN (cluster 1) and CAM encodes a 120K protein, and a 170K protein that is produced by read-through translation of the 120K cistron (4,5,6). Recently, we have sequenced the 3'-terminal 2077 nucleotides of RNA 1 of strain PSG (cluster 1), showing that the 170K cistron in this genome segment is followed by reading frames for a 28.8K protein and a 16K protein (3). A similar genome organization has been reported for the 3'-terminal 2kb region of RNA 1 of strain SYM (7). cDNA probes corresponding to the 3'-terminal 2077 nucleotides of PSG-RNA 1 cross-hybridized to RNA 1 of five cluster 1 strains, but not to RNA 1 of PEBV or the CAM-strain (3).

The length of RNA 2 differs from strain to strain and varies between 1800 and 4500 nucleotides (2). Recently, the complete nucleotide sequences have been reported for RNA 2 of strain CAM (1799 nucleotides)(8) and strain

© IRL Press Limited, Oxford, England.

PSG (1905 nucleotides)(3). Both genome segments contain a single open reading frame encoding the viral coat protein, and show a 3'-terminal homology with the corresponding RNA 1 molecules of 459 and 497 nucleotides, respectively. Here, we report the sequence of the 3389 nucleotides of an RNA 2 molecule of another cluster 1 virus, strain TCM, together with the sequence of the 3'-terminal 1210 nucleotides of RNA 1 of this strain. TCM-RNAs 1 and 2 share an identical sequence of 1099 nucleotides at their 3'-end. In addition to the coat protein cistron, the sequence that is unique to TCM-RNA 2 contains an open reading frame for a 29.1K protein that shows no resemblance to the RNA 1 encoded 28.8K protein. cDNA probes corresponding to different regions of TCM-RNA 2 were assayed for a possible cross-hybridization with RNA of six other tobraviruses.

MATERIALS AND METHODS

Purification of viral nucleoprotein and RNA. The sources from which the virus strains were obtained have been described by Cornelissen *et al.*(3). Virus was purified from Samsun NN tobacco by the method used for the isolation of alfalfa mosaic virus (9). RNA was extracted with phenol/chlorophorm (1:1) at 65°C, from purified virus suspensions that had been incubated in 1% SDS for 15 min at 37°C. Ethanol precipitated RNA was dissolved in 20mM Tris-HCl, pH 7.6, 0.1 mM EDTA.

Synthesis and cloning of double-stranded cDNA. DNA copies of TCM-RNA were cloned by the techniques described previously (3).

DNA sequencing. cDNA inserts or restriction fragments thereof, were sub-cloned into the mp and tg derivatives of M13 (10,11) and sequenced by the di-deoxy chain termination method (12) using (α-^{35}S)-dATP (13). If necessary breakpoints were created by unidirectional digestion with exonuclease III (14).

RNA sequencing. Sequences of 5'-labeled RNA molecules were determined by the wandering spot technique as described previously (3,15). Labeling of TCM-RNA 2 at the 3'-end was done in a total volume of 30 µl, containing 40 mM Tris-HCl, pH 7.9, 0.1 mM EDTA, 50 mM NaCl, 10 mM $MgCl_2$, 10 µCi (α-^{32}P)-ATP (Amersham, 3000 Ci/mmol) and 1 µg of CTP;ATP tRNA nucleotidyl transferase. After an incubation of 16 hours at 4°C the RNA (2 µg) was phenol extracted and precipitated with ethanol. The RNA was dissolved in 10 µl 20 mM Na-citrate, pH 5.0, 9 M urea, 1 mM EDTA, 0.05% bromophenol blue and run through a 10% poly-acrylamide/8 M urea gel. Radioactive 3'-terminal fragments, located by auto-radiography, were eluted from the gel as described by Peattie and Gilbert (16). RNA sequencing was done according to Peattie (17) and Donis-Keller (18).

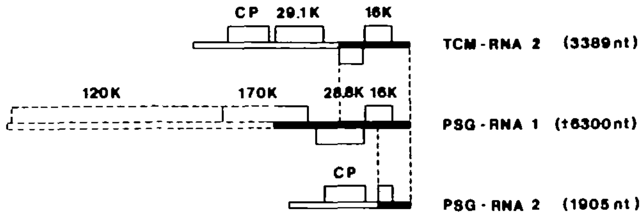


Figure 1. Schematic representation of the genetic information encoded in TCM-RNA 2 and PSG-RNA 1 and 2. Solid bars and open bars represent sequences that are unique to RNAs 1 and 2, respectively. The 3'-terminal regions that are homologous in RNAs 1 and 2 are connected by dotted lines. The location of cistrons encoding proteins with molecular weights of 120K, 170K, 28.8K, 16K, 29.1K and coat protein (CP) is indicated. The dotted region of PSG-RNA 1 has not yet been sequenced.

Digests were analysed on 15% polyacrylamide/8 M urea slabgels. The rationale for using this 3'-labeling technique will be presented elsewhere (Van Belkum *et al.*, manuscript in preparation).

Northern blotting and hybridization. This was done by the techniques described in the previous paper (3).

RESULTS

Sequence of TCM-RNAs

Initially, two overlapping cDNA clones of TCM-RNA 2 were sequenced: clone 15 (nucleotides 4-1657) and clone 36 (nucleotides 1307-3301). Later on, we isolated a nearly full-length clone of RNA 2 (clone 9) that was used to sequence nucleotides 3301-3389. Sequencing of 5'-labeled RNA 2 over a length of 11 nucleotides by the wandering spot technique showed that the first three nucleotides of RNA 2 were missing in clone 15. Sequence ladders made of 3'-labeled RNA 2 showed that the 3'-terminal sequence of RNA 2 was represented in clone 9. Together, these data permitted the deduction of the complete sequence of the 3389 nucleotides of TCM-RNA 2. In addition, limited sequence data were obtained for TCM-RNA 1 and TCM-RNA 4, a putative subgenomic messenger for coat protein. The wandering spot technique revealed that the 5'-terminal sequence of TCM-RNA 1 is identical to PSG-RNA 1 (3) for at least 16 nucleotides. The 5'-terminal sequence of TCM-RNA 4 was found to be AUAAUUAUACUG---. One clone from the cDNA library (clone 7) was found to correspond to TCM-RNA 1. A comparison with PSG-RNA 1 (3) and TCM-RNA 2 (see below) showed that the sequence of the 1210 bp insert of clone 7 was co-terminal with the 3'-end of TCM-RNA 1.

```
AUAAAACAUUGCACCAAUGGUGCUGCCCUGGCUGGGGUAUGUCUUUGAACGCAGUAGAAUGUGCUAAUUGACAAGUUGGAGAACGCGGUA
        10        20        30        40        50        60        70        80        90
GAACGUACUUAUCCGACACAGCCUUUAUCCCUUUGUUGAGAGGUUUUUCUCAACUGCACCGAAAUUCUGGAUUUAGUUGGUCGCUAGAAG
       100       110       120       130       140       150       160       170       180
ACUGAAGGAAUGUAGGUGUAUUUUUAGAUAGUUAAUAGGUAUCUUCUCAAAUACCAUAAACAUGAGACGCAUCGCUUGCGAAAGUAGCAU
       190       200       210       220       230       240       250       260       270
UAAAAGCUAAACCUUACUAAAUGACAAAUUUUAGUUUGGUGGUUCAAAAACGUCACUCAAAGUCAAUUUGAGUGUGCGCAGAACAGAACG
       280       290       300       310       320       330       340       350       360
AUGUCACGUUCCUGUUCGGGGUUUCACACUAUCUUAAGAGAAAGUGUUAAAAUAGUGAGAAUAUUCUCACUAUGAGCAUAAUUAUACUGG
       370       380       390       400       410       420       430                 450
UUAUCCUCUCGCUGAUAGAGACUAUCAUUCUUAAAAUCACUUUGAAAGGUAGUCUUGUUGAUACACAGGUUGCUUUAUCACUUGCACUAC
       460       470       480       490       500       510       520       530       540
(CP) M  A  G  S  Y  G  E  T  F  D  G  K  I  L  D  D  L  S  G  A  W  V  E  K  H  N  W  S
UUACUUAUGGCAGGUAGUUAUGGGAAACGUUCGAUGGAAAAAUUCUCGACGAUCUCUCUGGGGCUUGGGUUGAGAAACAUAACUGGAGC
       550       560       570       580       590       600       610       620       630
 D  I  L  R  R  L  T  K  I  K  F  A  L  Q  A  D  R  D  M  I  P  G  I  V  E  D  L  S  T  E
GACAUCCUCCGAAGGUUGACGAAGAUCAAAUUUGCUCUCCAGGCAGACAGAGAUAUGAUCCCUGGAAUUGUUGAGGAUCUGAGUACGGAA
       640       650       660       670       680       690       700       710       720
 I  P  V  D  E  N  T  R  F  P  S  G  K  V  Y  H  L  L  T  K  E  M  L  M  A  I  E  A  I  H
AUACCAGUUGAUGAAAACACGAGGUUUCCUAGUGGGAAAGUGUAUCACCUUCUGACCAAGGAAAUGUUGAUGGCGAUCGAAGCUAUCCAU
       730       740       750       760       770       780       790       800       810
 A  A  S  S  F  K  R  R  A  E  E  K  N  N  V  N  P  R  Q  R  F  E  A  E  S  S  S  S  Q  L
GCGGCUUCGUCGUUCAAAAGAAGAGCCGAGGAGAAAAAUAAUGUCAAUCCGAGACAGAGAUUUGAAGCCGAAAGUUCGUCUUCACAGUUA
       820       830       840       850       860       870       880       890       900
 P  S  G  G  L  V  V  R  P  A  A  G  S  G  D  S  S  L  G  E  D  L  F  S  N  S  K  L  D  D
CCAUCUGGUGGUUUAGUAGUUAGACCUGCUGCUGGGAGUGGGGAUUCAUCGUUAGGUGAAGAUUUGUUUUCGAAUAGUAAAUUAGAUGAC
       910       920       930       940       950       960       970       980       990
 A  S  T  A  F  H  K  S  L  A  T  L  K  G  S  R  P  K  A  V  V  Q  R  T  F  E  K  E  Y  S
GCCAGUACGGCCUUUCAUAAGAGUUUAGCUACUCUUAAGGGAAGUCGUCCAAAGGCGGUUGUCCAGCGCACUUUCGAAAAAGAGUAUAGU
      1000      1010      1020      1030      1040      1050      1060      1070      1080
 L  R  W  T  A  A  A  P  V  A  A  A  G  G  T  P  P  G  G  R  S  W  T  W  N  L  V  *
UUGAGGUGGACGGCUGCUGCUCCGGUCGCUGCUGCUGGUGGAACACCACCAGGUGGAAGGUCGUGGACCUGGAACCUAGUUUGAAUACUA
      1090      1100      1110      1120      1130      1140      1150      1160      1170
GAUUUCGAUAAUGACAGCAAUUGUCAUUUUAGAAACGGUUUUGCUACUUGAUUUGGUUUAAUCCUUUGACAUGGGUCGCAGGUGGCAUAG
      1180      1190      1200      1210      1220      1230      1240      1250
           (29.1K) M  G  N  G  D  W  S  S  K  W  P  N  D  H  L  F  I  D  D  F  G  K  L
UAAUUGUGAAAUUUCUUCUUUAUGGGCAACGGAGAUUGGUCUUCAAAGUGGCCUAACGAUCACUUGUUCAUAGAUGAUUUCGGAAAGUUG
      1270      1280      1290      1300      1310      1320      1330      1340      1350
 V  W  F  D  V  L  T  D  I  V  K  I  S  H  F  V  S  Q  V  P  T  D  L  S  P  I  P  S  S  Y
GUUUGGUUCGAUGUACUGACUGACAUAGUUAAGAUCUCUCAUUUCGUUUCACAGGUACCAACCGAUUUGUCUCCAAUUCCGAGCUCCUAU
      1360      1370      1380      1390      1400      1410      1420      1430      1440
 I  S  F  I  D  G  R  I  P  M  C  I  N  H  L  G  W  V  Y  I  R  V  K  F  E  S  E  E  V  F
AUUAGCUUCAUCGAUGGUCGAAUUCCCAUGUGUAUCAAUCACCUAGGUUGGGUGUAUAUACGUGUUAAAUUUGAGUCUGAGGAAGUAUUU
      1450      1460      1470      1480      1490      1500      1510      1520      1530
 Y  Q  K  F  G  E  V  D  V  S  R  F  G  E  S  E  L  P  P  D  F  E  V  T  F  S  K  V  T  T
UAUCAGAAGUUCGGUGAGGUAGAUGUUAGUAGGUUCGGCGAGUCGGAACUUCCUCCGGAUUUCGAAGUCACUUUCAGUAAGGUGACUACG
      1540      1550      1560      1570      1580      1590      1600      1610      1620
 L  V  N  K  S  L  V  R  K  S  E  L  L  E  K  M  N  N  E  L  K  Q  E  L  T  N  K  L  D  S
CUGGUGAAUAAGUCCUUGGUGCGAAAAAGCGAGCUUCUCGAAAAGAUGAACAAUGAACUCAAACAGGAAUUAACUAACAAGCUUGAUUCG
      1630      1640      1650      1660      1670      1680      1690      1700      1710
 L  E  K  V  N  V  Q  L  K  K  E  L  S  Q  A  Q  Q  S  N  F  T  E  L  R  D  G  L  I  L  N
CUGGAAAAGGUAAAUGUUCAGUUAAAGAAGGAACUUAGUCAAGCGCAACAAAGUAAUUUCACUGAAUUGCGCGACGGUUGAUCUUAAAC
      1720      1730      1740      1750      1760      1770      1780      1790      1800
 F  S  K  V  G  G  R  I  H  R  M  V  V  R  S  I  Q  N  Q  L  K  L  V  S  E  I  N  N  D  G
UUUUCAAAAGUUGGCGGGAGAAUCCAUCGUAUGGUUGUCAGAUCCAUCCAAAAUCAGCUUAAGCUUGUUUCGGAAAUCAAUAAUGAUGGU
      1810      1820      1830      1840      1850      1860      1870      1880      1890
 D  R  W  A  T  M  G  A  T  V  I  L  K  E  G  A  Q  Y  L  G  F  V  V  V  K  N  D  G  K  I
GACAGGUGGGCUACAAUGGGAGCUACGGUAAUUCUAAAGGAAGGAGCUCAGUAUUUAGGGUUUGUUGUUGUAAAGAAUGAUGGAAAAAUC
      1900      1910      1920      1930      1940      1950      1960      1970      1980
 G  V  K  F  D  L  T  R  L  N  S  L  D  N  Q  T  L  L  A  S  V  V  *
GGUGUCAAAUUCGAUUUAACCAGGUUAAAUUCUUUGGAUAACCAAACUUUGCUGGCGUCUGUGGUUUAAUAUGUUGUAGGUUUAAACUCC
      1990      2000      2010      2020      2030      2040      2050      2060      2070
CUUUUGAGCGGGGCGUGAAAGUCGUCCGGUGGAUCGUACCUUAUCGAUUUCUGUUGUAGGAAGCAGAUAAAAUAAUUCCUGUAGCUACGU
      2080      2090      2100      2110      2120      2130      2140      2150      2160
UCACGUAGGCUGGUUGUUACCAUAAACAAUAAUGUCGUUGGUUGCAUUUAAAAUAAUUCCAUUGACCGAUAUUCUAUCAUGGAAACGUC
      2170      2180      2190      2200      2210      2220      2230      2240      2250
```

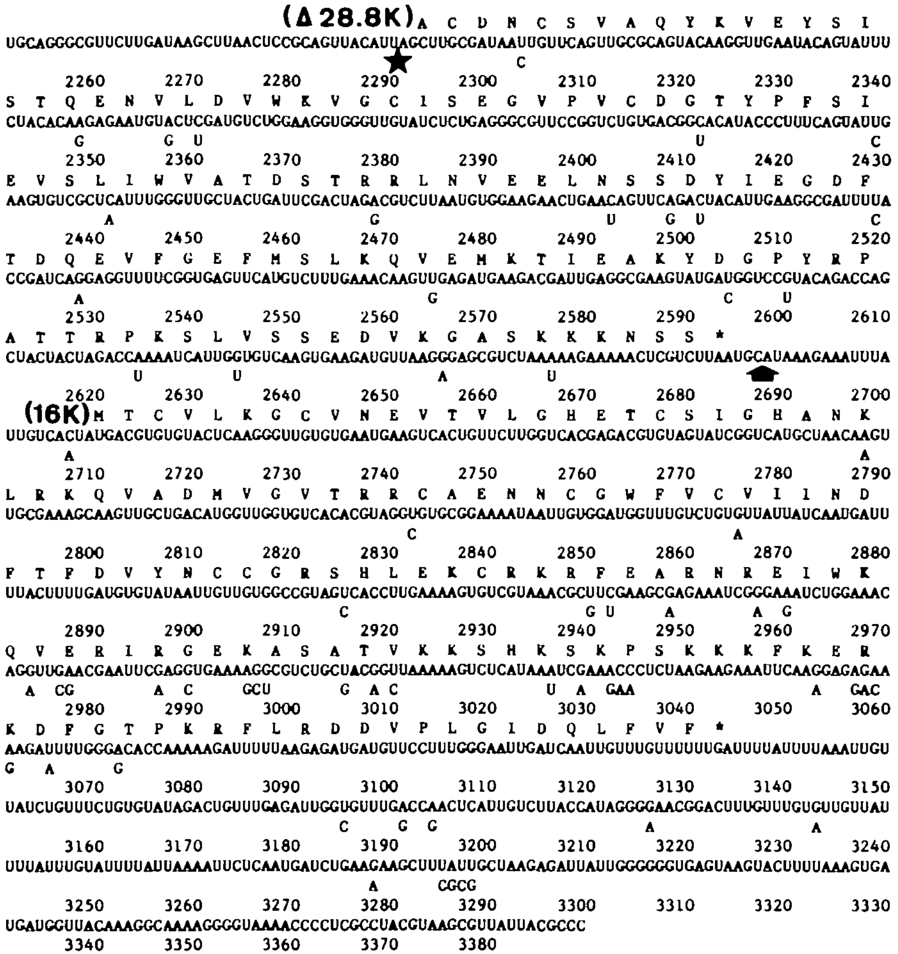


Figure 2. Nucleotide sequence of TCM-RNA 2. The amino acid sequences deduced from the open reading frames for the coat protein (CP), 29.1K protein, the C-terminal region of the 28.8K protein (Δ28.8K) and the 16K protein are given. The arrows at positions 438, 1257 and 2688 indicate the 5'-terminus of subgenomic RNA 4 and the putative 5'-termini of subgenomic RNAs 6 and 5, respectively. The asterisk marks the beginning of the 3'-terminal sequence of 1099 nucleotides that is homologous to TCM-RNA 1; differences in this region with the sequence of PSG-RNA 1 are indicated below the TCM-RNA 2 sequence.

Figure 1 shows a schematic representation of the open reading frames in TCM-RNA 2; also for comparison the genome structure of strain PSG is given (3). A complete listing of the nucleotide sequence of TCM-RNA 2, together with the amino acid sequences corresponding to the open reading frames is given in

Figure 2. The sequence of the 3'-terminal 1099 nucleotides of TCM-RNA 2 is homologous to the 3'-termini of TCM-RNA 1 and PSG-RNA 1 for 100% and 94%, respectively. Differences between TCM-RNA 2 and PSG-RNA 1 in this region are listed in Figure 2. The insert in clone 7 contained 111 nucleotides of TCM-RNA 1 upstream of the sequence that is identical to TCM-RNA 2. Four out of these 111 nucleotides differed from the corresponding sequence of PSG-RNA 1 (data not shown). The region of TCM-RNA 2 that is homologous to RNA 1 contains the reading frames for the C-terminal portion of the RNA 1 encoded 28.8K protein and the 16K protein. This means that the TCM genome is partially diploid, encoding the 16K protein in both RNAs 1 and 2.

The sequence that is unique to TCM-RNA 2 contains two open reading frames. The 5'-proximal reading frame is the coat protein cistron because the encoded protein (molecular weight 22,405) shows extensive homology to the coat proteins of strains PSG and CAM (see below). The second open reading frame encodes a protein of molecular weight 29,147 (29.1K). A "dot plot" comparison generated by a computer programme supplied by the University of Wisconsin Genetics Computer Group (UWGCG) did not reveal any homology between the amino acid sequences of the 29.1K protein and the 28.8K protein or any other known TRV protein.

PSG coat protein is probably translated from a subgenomic RNA (RNA 4) that lacks the 5'-terminal 474 nucleotides of PSG-RNA 2 (3). Subgenomic messengers involved in the expression of the PSG-RNA 1 encoded 28.8K protein (RNA 3) and the 16K protein (RNA 5) have been identified (3). We have suggested that the PSG subgenomic RNAs all have the 5'-terminal sequence AUA and that in the respective genomic RNAs this 5'-end is preceded by the sequence GC and a nearby AUG triplet (3). The TCM-RNA 2 sequence around position 438, which is the 5'-terminal nucleotide of RNA 4 (arrow in Figure 2), is in agreement with this prediction. By analogy it is proposed that a possible subgenomic messenger (RNA 6) involved in the expression of the 29.1K protein may start at nucleotide 1257 of TCM-RNA 2 (arrow in Figure 2). Such a messenger has not yet been identified. It may be noted that the 5'-terminal sequence of RNA 4 and the putative 5'-terminus of CAM-RNA 4 (3) are identical for the first twelve nucleotides.

Homology between TRV-RNAs.

The RNA 2 sequences of TCM, PSG and CAM were compared by "dot plot" analysis using the UWGCG programme. Significant homologies were found between the coat protein cistrons and the 3'-terminal 44 nucleotides of the three RNA

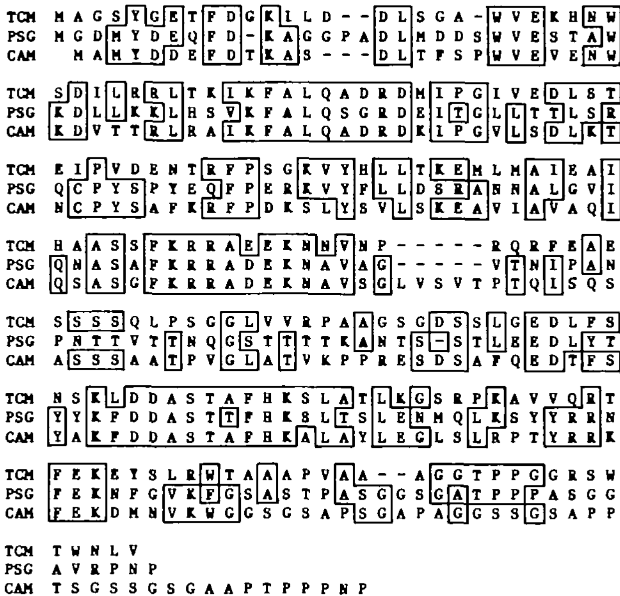
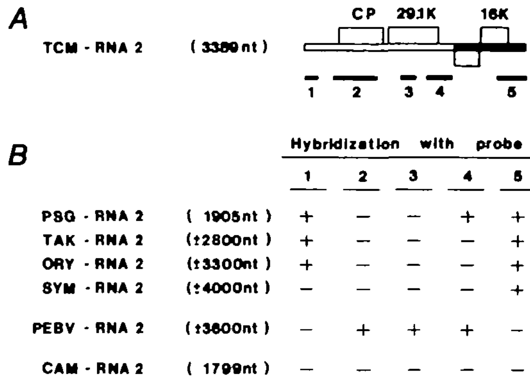


Figure 3. Alignment of the amino acid sequences of the coat proteins of strains TCM (205 amino acids), PSG (209 amino acids) and CAM (223 amino acids). *Identical residues are boxed.*

molecules. The homology between the coat protein cistrons is also reflected at the level of the encoded proteins. Figure 3 shows an alignment of the amino acid sequences of the three coat proteins. The following homologies can be observed: TCM/PSG, 39% homology (80 amino acids in common); TCM/CAM, 40% homology (83 amino acids in common); PSG/CAM, 43% homology (90 amino acids in common).

A comparison of TCM-RNA 2 and PSG-RNA 2 reveals several homologies that are absent in CAM-RNA 2. At the 5'-end there is a 90% homology between the first 142 nucleotides. The 3'-terminal PSG-RNA 2 sequence of 497 nucleotides that is homologous to RNA 1, is also present in TCM-RNA 2 with a few substitutions. However, in TCM-RNA 2 the homology with RNA 1 continues for another 602 nucleotides. Just upstream of the RNA 1 derived sequences, there is a 70% homology between a sequence of about 80 nucleotides in TCM-RNA 2 (nucleotides 2144-2230) and PSG-RNA 2 (nucleotides 1328-1409).

The homology of TCM-RNA 2 to six other tobravirus RNAs was also investigated using restriction fragments of TCM-cDNA as probes in the hybridization to Northern blots. Figure 4A shows a schematic representation of the five probes that were used; Figure 4B lists the results that were obtained. Probe



		Hybridization with probe				
		1	2	3	4	5
PSG - RNA 2	(1905nt)	+	−	−	+	+
TAK - RNA 2	(±2800nt)	+	−	−	−	+
ORY - RNA 2	(±3300nt)	+	−	−	−	+
SYM - RNA 2	(±4000nt)	−	−	−	−	+
PEBV - RNA 2	(±3600nt)	−	+	+	+	−
CAM - RNA 2	(1799nt)	−	−	−	−	−

Figure 4. Homology between RNA 2 of strain TCM and six tobraviruses. (A) Five restriction fragments derived from cDNA of TCM-RNA 2 were used as hybridization probes: 1, nucleotides 4-198; 2, nucleotides 445-1103; 3, nucleotides 1462-1701; 4, nucleotides 1864-2270; 5, nucleotides 2948-3389. (B) The five probes were hybridized to Northern blots loaded with RNAs of six tobraviruses. Hybridization of the probes to RNA 2 of the repective strains is indicated by a plus-sign; the absence of hybridization is indicated by a minus-sign.

1 hybridizes to PSG-RNA 2 because of the homology in the 5'-terminal 142 nucleotides; a similar sequence is probably also present at the 5'-termini of TAK-RNA 2 and ORY-RNA 2. Probe 2 does not hybridize to PSG-RNA 2 or CAM-RNA 2. Apparently, the homology between the coat protein cistrons revealed by "dot-plot" analysis, is to low to permit detectable cross-hybridization at the stringency used. However, probe 2 strongly hybridized to PEBV-RNA 2, indicating that the coat protein cistrons of TCM and PEBV are highly homologous. Similarly, the results with probes 3 and 4 indicate that PEBV-RNA 2 contains a reading frame that is homologous to the 29.1K cistron in TCM-RNA 2. The hybridization of probe 4 with PSG-RNA 2 is probably due to the homologous sequence of 80 bases occurring just upstream of the RNA 1 derived sequence. Probe 5 recognizes the RNA 1 derived sequences in RNA 2 of the cluster 1 viruses (PSG, TAK, ORY, SYM), but does not hybridize to RNA 2 of PEBV or CAM. Probes 1 to 4 did not hybridize to RNA 1 of any strain; probe 5 hybridized to RNA 1 of the cluster 1 viruses.

DISCUSSION

The results reported in this paper indicate that the variability in length of TRV-RNA 2 is caused by two factors: (1) the presence of a second gene in the RNA 2 specific sequences of relatively long RNA 2 molecules, and

(2) a variation in the length of the 3'-terminal RNA 1 derived sequences. TRV-RNA 1 is able to replicate systemically in host plants in the absence of RNA 2 (1). The homology with gene products of tobacco mosaic virus suggests that the 120K and 170K proteins are involved in the replication of TRV-RNA, whereas the 28.8K protein may be involved in cell-to cell transport (3,7). The RNA 1 induced replicase is probably able to replicate RNA 2 which in its turn provides the coat protein that encapsidates both TRV-RNAs. We do not yet know whether the 29.1K cistron in TCM-RNA 2 is expressed into a functional protein. If this is the case, one may wonder what function this protein performs in the replication of TCM. One possibility is that the 29.1K protein is functionally equivalent to the RNA 1 encoded 28.8K protein, and that its gene in RNA 2 is in fact redundant just as the 16K cistron in TCM-RNA 2 is redundant. There are only 15 amino acid substitutions in the 16K proteins of TCM and PSG but there is no obvious homology between the 29.1K and 28.8K proteins. It is possible that the 16K cistron in TCM-RNA 2 is derived from TCM-RNA 1 and that the 29.1K cistron is derived from a yet unknown non-homologous RNA 1 molecule during an earlier symbiotic period of TCM-RNA 2. Alternatively, it is possible that the 29.1K gene provides TCM with a unique function that is not involved in the replication of strain PSG.

The observation that the 3'-termini of CAM-RNA 2, PSG-RNA 2 and TCM-RNA 2 show a 100% homology with the corresponding RNA 1 molecules over a length of 459, 497 and 1099 nucleotides, respectively, suggests that in some way sequence differences are corrected in this region. Inspection of the junction between the RNA 2 and RNA 1 specific sequences in PSG-RNA 2 (3) and TCM-RNA 2 (asterisk in Figure 2) shows the occurrence of the sequence AUAAUUGUU just downstream both junctions. This sequence resembles the 5'-termini of genomic and subgenomic TRV-RNAs (3) and may reflect a possible internal initiation site for the viral replicase in a minus-strand template. If a replicase molecule, attached to a nascent RNA 2 chain, leaves the minus-strand RNA 2 template, and reinitiates at an internal site of a minus strand RNA 1 template, a chimaeric transcript is made with a 3'-end that is identical to RNA 1. Such a transcript may be replicated by the RNA 1 encoded replicase in preference over transcripts with 3'-termini that are not identical to the RNA 1 sequence.

Tobraviruses of cluster 1 (serotype I-II) and cluster 3 (strain CAM, serotype III) are considered to be serologically distinct groups (2). The comparison made in Figure 3 shows, however, that at the amino acid sequence level there is no clear distinction between the coat protein of the two serotype I-II viruses (TCM and PSG) on the one hand and that of the CAM strain

on the other hand. A possible serological relationship between the coat proteins of the three strains is being investigated. The observation that TCM-cDNA probes cross-hybridize with PEBV-RNA 2 (Figure 4) correlates well with an earlier report that a Dutch isolate of PEBV is serologically related to a TRV serotype I-II isolate (19). The results shown in Figure 4 indicate that, in their association with TCM-RNA 1 and PEBV-RNA 1, the coding sequences of TCM-RNA 2 are accomodated with different 5'- and 3'-terminal sequences. It is possible that the mechanism of copy-choice, as discussed in the previous paragraph, permits the attachment of different termini to the TCM-RNA 2 coding sequence, thus enabling the replication of this sequence by the replicases encoded by either TCM-RNA 1 or PEBV-RNA 1. For a more detailed insight in this matter, sequence studies on the PEBV genome are required.

ACKNOWLEDGMENTS

Thanks are due to Mr. P. van Zuidam and Mr. F.Th. Brederode for assistence in the sequence studies. This work was supported in part by the Netherlands Foundation for Chemical Researh (S.O.N.) with financial aid from the Netherlands Organization for the Advancement of Pure Research (Z.W.O.).

*To whom correspondence should be addressed

REFERENCES

1. Harrison, B.D. and Robinson, D.J. (1978) Adv. Virus Res. 23, 25-77.
2. Robinson, D.J. and Harrison, B.D. (1985) J. gen Virol. 66, 171-176.
3. Cornelissen, B.J.C., Linthorst, H.J.M., Brederode, F.Th. and Bol, J.F. (1986) Nucleic Acids Res. 14, 2157-2169.
4. Mayo, M.A., Fritsch, C. and Hirth, L. (1976) Virology 69, 408-415.
5. Fritsch, C., Mayo, M.A. and Hirth, L. (1977) Virology 77, 722-732.
6. Pelham, H.R.B. (1979) Virology 97, 256-265.
7. Boccara, M., Hamilton, W.D.O. and Baulcombe, D.C. (1086) EMBO J. 5, 223-229.
8. Bergh, S.T., Koziel, M.G., Huang, s.-C., Thomas, R.A., Gilley, D.P. and Siegel, A. (1985) Nucleic Acids Res. 13, 8507-8518.
9. Bol, J.F., Brederode, F.Th., Janze, G.C. and Rauh, D.C. (1976) Virology 65, 1-15.
10. Messing, J. (1983) Methods Enzymol. 101, 20-78.
11. Kieny, M.P., Lathe, R. and Lecocq, J.P. (1983) Gene 26, 91-99.
12. Sanger,F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463-5467.
13. Biggin, M.D., Gibson, T.J. and Hong, G.F. (1983) Proc. Natl. Acad. Sci. USA 80, 3963-3965.
14. Hennikoff, S. (1984) Gene 28, 351-359.
15. Koper-Zwarthoff, E.C., Lockard, R.E., Alzner-De Weerd, B, RajBhandary, U.L. and Bol, J.F. (1977) Proc. Natl. Acad. Sci. USA 74, 5504-5508.
16. Peattie. D. and Gilbert, W. (1980) Proc. Natl. Acad. Sci. USA 77, 4679-4682.
17. Peattie, D. (1979) Proc. Natl. Acad. Sci. USA 76, 1760-1764.
18. Donis-Keller, H. (1977). Nucleic Acids Res. 4, 2527-2538.
19. Maat, D.Z. (1963) Neth. J. Plant Pathol. 69. 287-239.