The Complete Nucleotide Sequence of Tobacco Rattle Virus RNA-1

By W. D. O. HAMILTON, M. BOCCARA, D. J. ROBINSON and D. C. BAULCOMBE

Molecular Genetics Department, Plant Breeding Institute, Maris Lane, Cambridge CB2 2LQ and Scottish Crop Research Institute, Invergowrie, Dundee DD2 5DA, U.K.

(Accepted 6 July 1987)

SUMMARY

The complete nucleotide sequence of tobacco rattle virus (TRV) strain SYM RNA-1 was determined from a series of overlapping cDNA clones. cDNA prepared by primer extension was used to determine the exact 5' terminus. The RNA sequence was 6791 nucleotides in length and contained four open reading frames (ORFs). The ORF nearest the 5' terminus coded for a polypeptide of predicted mol. wt. 134000 (134K) and terminated at an opal (UGA) stop codon. Readthrough of this stop codon would result in the production of a protein of 194K. The gene for a 29K polypeptide started one base beyond the 194K stop codon and, in turn, was followed by the gene for a 16K protein at the 3' end of RNA-1. Amino acid comparisons of the 194K protein with the putative replicase of tobacco mosaic virus showed three regions of strong homology, suggesting that the 134K and 194K proteins were similarly involved in virus replication. The 5' terminal sequences of both genome RNA species of TRV strains ORY, N5 and PRN together with that of SYM RNA-2 were also determined. Alignments of these sequences showed that there was a 22 base repeated sequence close to the 5' terminus in all these RNA species. It was also shown that the 5' terminus of RNA-1 was homologous with the same region in RNA-2.

INTRODUCTION

Tobraviruses are single-stranded RNA viruses possessing a bipartite genome. They have the widest host range of any plant virus (Harrison & Robinson, 1986) and are naturally transmitted by soil-inhabiting nematodes. The two genome RNA species (RNA-1 and RNA-2) are of positive polarity, are capped at their 5' termini, and have a 3' structure with tRNA-like features (Abou Haidar & Hirth, 1977; Pelham, 1979; Harrison & Robinson, 1986; van Belkum et al., 1987). The group comprises three serologically distinct viruses (Robinson & Harrison, 1985): tobacco rattle virus (TRV), pepper ringspot virus (PRV, previously known as the CAM strain of TRV; Harrison & Woods, 1966) and pea early-browning virus (PEBV). All three viruses have RNA-1 molecules that are approximately 6-5 kilobases (kb) in length whereas the RNA-2 molecules vary from 1.8 kb for PRV (Bergh et al., 1985) to 3.9 kb for the SYM strain of TRV (Robinson et al., 1983).

Translation studies on TRV strain PRN in vitro have shown that RNA-1 codes for two polypeptides of approximate mol. wt. 140000 and 170000 (140K and 170K, Mayo et al., 1976; Fritsch et al., 1977) and that the larger protein may arise by the readthrough of a 'leaky' termination codon (Pelham, 1979). Similar studies using PEBV indicate that proteins of 134K and 165K are produced in the same manner (Hughes et al., 1986). A third protein of mol. wt. 29K is also coded by RNA-1 in TRV but is translated from a 1.5 kb subgenomic RNA (Robinson et al., 1983). A further gene product, a protein of mol. wt. 16K, is encoded at the 3' end of TRV RNA-1 and is translated from a subgenomic RNA of 0.7 kb (Boccara et al., 1986; Cornelissen et al., 1986). The gene for the 16K protein had not been detected previously by

† Present address: Laboratoire de Pathologie Vegetale I.N.A., 16 rue Claude Bernard, 75005 Paris, France.
The virion coat protein is encoded by RNA-2, although in the case of strain SYM it is translated from a subgenomic RNA species (Robinson et al., 1983). The nucleotide sequences of the RNA-2 species from PRV and TRV strain PSG show only one major open reading frame (ORF) (Bergh et al., 1985; Cornelissen et al., 1986) whereas the sequence determined for TRV strain TCM has, in addition to the coat protein gene, an ORF coding for a 29.1K protein and a 16K protein (Angenent et al., 1986). The predicted 16K protein sequences encoded by SYM RNA-1, PSG RNA-1 and RNA-2 are almost identical and reflect the large degree of sequence conservation at the 3' ends of RNA-1 and RNA-2 in these viruses.

All tobravirus strains give rise to so called ‘NM’ isolates (Cadman & Harrison, 1959) which are characterized by the loss of RNA-2 (Harrison & Robinson, 1978). Although no virions are produced, RNA replication, intercellular movement of viral RNA, symptom production and cross-protection between TRV strains all occur. As all these properties are therefore coded by RNA-1 we have directed our efforts towards determining the primary structure of this RNA species. We report here the complete nucleotide sequence of TRV strain SYM RNA-1 together with the 5' terminal sequences from a number of other TRV strains. These data not only give further insights into tobravirus evolution but also provide the basis for future work towards the elucidation of viral pathogenesis at the molecular level.

**METHODS**

Preparation of viral RNA. RNA from TRV strains SYM (Kurppa et al., 1981), PRN (Cadman & Harrison, 1959), ORY (Lister & Bracker, 1969) and N5 (Harrison et al., 1983) were propagated in *Nicotiana clevelandii* as described by Robinson & Harrison (1985) and RNA was extracted from purified virions as described by Robinson (1983). Strain ORY was held under licence from the Department of Agriculture and Fisheries for Scotland.

Construction of cDNA clones. Construction of an overlapping set of TRV SYM RNA-1 cDNA clones in GC-tailed pBR322 or pUC19 has already been described (Boccara et al., 1986). Clone 543B was found to extend furthest in the 3' direction. Based on comparisons of this sequence with the 3' terminal sequences of other tobravirus RNA species, we estimated that the 543B sequence was short of the 3' end of SYM RNA-1 by less than 100 nucleotides. In order to obtain clones extending to the extreme 3' end, viral RNA was polyadenylated *in vitro* using poly(A) polymerase (a gift from Dr Phillip Goelet) and cDNA prepared by the RNase H procedure using an Amersham cDNA Synthesis System Kit. The cDNA was cleaved with *PstI* and fractionated by size on a 5% polyacrylamide gel. DNA of 0.8 to 2 kb was isolated, ligated into *PstI-SmaI*-digested pUC19, and transformed into calcium chloride-treated *Escherichia coli* strain MC1022 (Casadaban & Cohen, 1980). Clones containing the 3' *PstI* fragment were identified by colony hybridization (Maniatis et al., 1982) using a probe synthesized by random priming (Feinberg & Vogelstein, 1984) on a 688 base *AvaiI-Xhol* fragment from cDNA of 543B. The identification was confirmed by DNA sequence analysis using the plasmid DNA as template. The plasmid DNA was prepared as described by Holmes & Quigley (1981) and the DNA sequencing method was as described by Chen & Seeburg (1985), except that the supercoiled DNA was treated with 0.2 volumes of 1 M-NaOH, 1 mM-EDTA for 5 min at room temperature and then desalted by centrifugation through Sepharose CL-6B (Pharmacia) after the RNase digestion (G. J. P. Murphy, personal communication).

Nucleotide sequencing. Restriction fragments from the cDNA were cloned into M13 mp18 or mp19 phage vectors (Messing, 1983) and the nucleotide sequence derived by the dideoxy chain termination method of Sanger et al. (1977). Most of the sequence was determined by the random cloning of cDNA fragments produced by digestion with restriction enzymes that have four base recognition sites. Subsequent cloning of specific restriction fragments was carried out in order to complete sequence determination in both directions.

Primer extension. Two oligonucleotides were synthesized (A. Northrop, Microchemical facility, I.A.P., Babraham, U.K.); a 17-mer 5'-AAGTACGTCTACCGGC-G-3' and a 22-mer 5'-CCCCACTCGGTAC-TAAGCTGG-Y. Five hundred ng of end-labelled oligomer was annealed with 5 µg of viral RNA in 15 µl 50 mM-Tris-HCl pH 7.5, 75 mM-KCl, 3 mM-MgCl₂, 10 mM-dithiothreitol at 65 °C for 2 min. The reaction was allowed to cool at room temperature for at least 30 min. While keeping the buffer concentration the same, the reaction volume was increased to 35 µl by the addition of 250 units of reverse transcriptase (M-MLV, Bethesda Research Laboratories), 45 units of RNasin (Anglian Biotechnology, Colchester, Essex, U.K.), 5 µg of bovine serum albumin and 300 µM with respect to all four dNTPs. Following incubation for 1 h at 37 °C, the reaction mix was fractionated by electrophoresis on a 6% urea–polyacrylamide sequencing gel. After autoradiography, the cDNA fragments were excised, electroeluted, extracted with phenol/chloroform and precipitated with ethanol. The cDNA was resuspended, and sequenced by the chemical cleavage method of Maxam & Gilbert (1980) as modified by Barker et al. (1983).
**Nucleotide sequence of TRV RNA-1**

The nucleotide sequence of 2 kb from the 3' terminal region of RNA-1 included within cDNA clones 25B and 543B was described previously (Boccara et al., 1986). These clones and others forming an overlapping set extending over most of RNA-1 are shown aligned with the viral RNA in Fig. 1. The nucleotide sequences of these additional cDNA clones were determined using the dideoxy chain termination method on M13 templates generated by subcloning specific restriction fragments.

To determine whether the extreme 5' end of the RNA had been cloned and sequenced, primer extension studies were carried out using a synthetic 17-base oligomer. This was designed to be complementary to a position 14 bases away from the 5' end of cDNA 24B (equivalent to nucleotides 59 to 75, see Fig. 2). Using viral RNA as a template, it was shown that the genuine 5' end was not represented in cDNA 24B and that approximately 50 bases were missing. Subsequently, extension products primed with end-labelled 17-mer were isolated from preparative polyacrylamide gels and sequenced by the Maxam & Gilbert procedure (see Methods) to show that 24B was 45 bases short of the genuine 5' end of the viral RNA. A second oligomer of 22 nucleotides complementary to the sequence 60 bases to the 3' side of the 17-mer (nucleotides 120 to 141) was used in the same way to confirm the 5' terminal sequence beyond the end of clone 24B.

Comparison of the SYM RNA-1 3' terminal sequence determined previously (Boccara et al., 1986) with the RNA-2 sequences of TRV strains PSG (Cornelissen et al., 1986) and TCM (Angenent et al., 1986) showed considerable homology. Alignment of the sequences indicated that the SYM sequence was 79 bases shorter, relative to the 3' ends of both PSG and TCM RNA-2 and TCM RNA-1. In order to obtain the sequence from this missing region, cDNA was

---

**Sequence analysis.** Computer programs of Staden (1980) were used for data management and collation of the sequence. University of Wisconsin (UWGCG, Devereaux et al., 1984) programs were used for nucleotide and protein sequence analysis. Values given for the direct amino acid homologies in Table 1 were determined using the BESTFIT program. TRV sequences covering the three conserved domains (amino acids 50 to 450, 875 to 1275 and 1350 to 1650) were aligned with the sequences for tobacco mosaic virus (TMV), alfalfa mosaic virus (AMV), brome mosaic virus (BMV) and cucumber mosaic virus (CMV) using the default parameters. Gaps were introduced to optimize the alignments. The amino acid homologies given in Table 1 correspond to the region of the TRV protein sequence that was found to align with all four viruses (i.e. amino acids 95 to 405, 897 to 1155 and 1361 to 1590) and thus give a minimal estimate for the relative homologies in each domain.

---

**RESULTS**

**Sequencing TRV RNA-1**

The nucleotide sequence of 2 kb from the 3' terminal region of RNA-1 included within cDNA clones 25B and 543B was described previously (Boccara et al., 1986). These clones and others forming an overlapping set extending over most of RNA-1 are shown aligned with the viral RNA in Fig. 1. The nucleotide sequences of these additional cDNA clones were determined using the dideoxy chain termination method on M13 templates generated by subcloning specific restriction fragments.

To determine whether the extreme 5' end of the RNA had been cloned and sequenced, primer extension studies were carried out using a synthetic 17-base oligomer. This was designed to be complementary to a position 14 bases away from the 5' end of cDNA 24B (equivalent to nucleotides 59 to 75, see Fig. 2). Using viral RNA as a template, it was shown that the genuine 5' end was not represented in cDNA 24B and that approximately 50 bases were missing. Subsequently, extension products primed with end-labelled 17-mer were isolated from preparative polyacrylamide gels and sequenced by the Maxam & Gilbert procedure (see Methods) to show that 24B was 45 bases short of the genuine 5' end of the viral RNA. A second oligomer of 22 nucleotides complementary to the sequence 60 bases to the 3' side of the 17-mer (nucleotides 120 to 141) was used in the same way to confirm the 5' terminal sequence beyond the end of clone 24B.

Comparison of the SYM RNA-1 3' terminal sequence determined previously (Boccara et al., 1986) with the RNA-2 sequences of TRV strains PSG (Cornelissen et al., 1986) and TCM (Angenent et al., 1986) showed considerable homology. Alignment of the sequences indicated that the SYM sequence was 79 bases shorter, relative to the 3' ends of both PSG and TCM RNA-2 and TCM RNA-1. In order to obtain the sequence from this missing region, cDNA was...
The complete nucleotide sequence of TRV SYM RNA-1, shown as DNA. Amino acid sequences of the coding regions shown in Fig. 1 are given. Termination codons and the opal readthrough codon are indicated by asterisks.

TRV SYM RNA-1 was found to code for four proteins of predicted mol. wt. 134,287 (134K), 194,220 (194K), 28,826 (29K) and 16,337 (16K) by analysis of the sequence data. The first AUG was found at base 203, the start of the first ORF which coded for 1187 amino acids (134K protein). The ORF continued after an UGA (opal) stop codon for a further 519 amino acid residues (194K protein). Fig. 3 illustrates a series of dot matrix comparisons of the TRV 194K protein and the putative replicase genes of TMV (Goelet et al., 1982), A1MV (Cornelissen et al., 1983a, b), BMV (Ahlquist et al., 1984) and CMV (Rezaian et al., 1984, 1985). Three regions of homology between TMV, A1MV and BMV have been described previously (Cornelissen & Bol, 1984; Haseloff et al., 1984). The comparisons in Fig. 3 show that these three domains were also present in the TRV 194K protein. In addition, it is apparent that the homology between TRV and TMV was more extensive than between TRV and A1MV, BMV or CMV. Using the UWGCG BESTFIT program (see Methods) the three domains within the TRV sequence which showed homology to regions in all four viruses were determined. The number of direct homologies found after alignment of these domains is summarized in Table 1. Even though these

synthesized on polyadenylated viral RNA, cloned, and about twenty possible 3' end clones were identified as described in Methods. When the isolated plasmids were sequenced directly, one (MA61) was shown to contain cDNA that extended the sequence by the missing 79 bases. It was believed to represent the correct 3' end based on homology with the 3' terminal sequences of other tobraviruses (Cornelissen et al., 1986; Angenent et al., 1986; Bergh et al., 1985).

The complete TRV SYM RNA-1 nucleotide sequence comprised 6791 bases and had a composition of 29.3% A, 28.5% U, 16.6% C, 25.6% G. This was almost identical to that reported for PRV RNA-1 (Minson & Darby, 1973). The complete sequence is shown in Fig. 2.
Nucleotide sequence of TRV RNA-1

Fig. 3. Comparison of the TRV 194K protein with the putative replicases of TMV, AIMV, BMV and CMV. The dot-plots were made using a window of 31 and a stringency of 0.8. For TRV and TMV, the lines across the plots indicate the position of the readthrough codons. For AIMV, BMV and CMV there are two replicate proteins derived from separate RNA species. The amino acid sequences have been joined together to aid comparison and the positions of the join are indicated by lines across the plots. References for the sequences used are given in the text.

Table 1. Direct amino acid homologies found in the three conserved replicase domains of TMV, AIMV, BMV and CMV relative to TRV

<table>
<thead>
<tr>
<th>Domain</th>
<th>TRV sequence (95–405)*</th>
<th>TMV 76t (81–421)</th>
<th>AIMV 68 (99–389)</th>
<th>BMV 61 (80–362)</th>
<th>CMV 74 (84–395)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td>97 (826–1084)</td>
<td>91 (831–1093)</td>
<td>78 (678–946)</td>
<td>75 (706–974)</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>90 (1290–1520)</td>
<td>60 (1361–1590)</td>
<td>69 (1337–1559)</td>
<td>71 (1415–1638)</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* Amino acid residues are given in parentheses. For TRV the numbers indicate the limit of the protein sequence that could be aligned with all four viruses.
† Number of directly homologous amino acids.

Data represent the minimum degree of homology between TRV and each of the other viruses in these domains, they demonstrate that the homology between TRV and TMV is consistently greater than that found with the other viruses.

Amino acid comparisons between putative RNA-dependent polymerases for plant, animal and bacterial viruses have shown the conservation of a number of amino acid residues (Kamer & Argos, 1984). The most notable of these is a pair of aspartic acid residues which are flanked on
TRV p194 1511-1556 QKKSQGADTYANSBHTLCA -11 aa- MVTVGDDSSLIAFPR
TMV p183 1442-1487 QKKSQGVTTFIGMTVIYAAC -11 aa- KGAFCGDDSSLYYPK
ALMV p90 586-633 QRTGDAQLYGNTIVTAC -13 aa- FYVASSGDDSLGTVE
BMV p94 521-566 QRTGDAFTYFGNTVYMAK -11 aa- CAFSGDDSLISKVL
CMV p94 570-615 QRTGDAFTYFGNTIVTMAK -11 aa- RLLFSGDDSLAFLK
CarMV 526-569 QRTGDAFTYFGNTIVTMAK -11 aa- RLLFSGDDSLAFLK
SNBV nsP4 op+ 490-537 MMKSGMFLTLFVNTVLNVVI -13 aa- CAAFIHDNIIHGYV
MIDV nsP4 op+ 490-537 MMKSGMFLTLFVNTVMNTI -13 aa- CAAFIHDNIIHGYK
SFV nsP4 485-532 MMKSGMFLTLFVNTVLNI -13 aa- CAAFIHDNIIHGYK
FMDV 295-346 GMPSGCSATSIINTILNNIY -17 aa- TMISYGDVVASD
EMCV 2120-2171 GLPSGCATSMNINMNII -17 aa- KVLSYGDVVATNY
POLIO 285-336 GMPSGCSATSIINFNI -17 aa- KIAAYGDVSAYP
COX 286-337 GMPSGCSATSIINTI -17 aa- KIAAYGDVSAYP
RHINO14 2003-2054 GMPSGCSATSIINTI -17 aa- KIAAYGDVSAYP
RHINO 2 1974-2025 GVPSGCSATSIINTI -17 aa- KIAAYGDVSAYP
BBV p102 646-700 GVKSGSSTTPHMTQYNGCV -20 aa- IGPKCDGGDSLRAI
HEP A 2037-2091 SMPSGCSSTTPHMTQYNGCV -20 aa- RILCYGDVVFSR
CpMV 1494-1554 GIPSGFPMTVIVNSIFNEIL -26 aa- GPPATGDVSAI
TEV 2586-2634 GNSQGSPTyVNDTLMVIIA -15 aa- VYLYNDDDLIDAIHP
TVMV 2531-2584 GNSQGSPTyVNDTLMVLA -19 aa- KPDANDDDDILIAKP

Fig. 4. Conserved amino acids (aa) found in viral replicases. The amino acids coded for by nucleotide 4733 to 4870 on TRV SYM RNA-1 are aligned with equivalent regions from a number of other putative viral RNA polymerases. Numbers of the amino acids from each protein are given. Asterisks indicate conserved residues (Schwartz & Dayhoff, 1978; residues in bold lettering are directly conserved in all the sequences. Sindbis and Middleburg viruses (SNBV and MIDV; Strauss et al., 1983); Semliki Forest virus (SFV), foot-and-mouth disease virus (FMDV), encephalomyocarditis virus (EMCV), poliovirus (POLIO), coxsackievirus (COX), rhinoviruses 2 and 14 (RHINO2 and RHINO14), black beetle virus (BBV p102) and hepatitis A virus (HEP A) are cited in Hodgman & Zimmern (1987); cowpea mosaic virus (CpMV, Lomonossof & Shanks, 1983); other viruses are from sources cited in the text.

either side by six hydrophobic residues. This feature was present in all the sequences examined, including the TRV protein (Fig. 4) where the two aspartic acid residues were encoded from nucleotide 4844 in the domain of the 194K protein. Glycine, threonine and asparagine residues, situated about 24 to 32 amino acids on the N-terminal side of this region, were also conserved in all the RNA-dependent polymerases. These conserved residues are also present in the nuclear inclusion proteins of two potyviruses, tobacco etch virus (TEV, Allison et al., 1986) and tobacco vein mottling virus (TVMV, Domier et al., 1986; Fig. 4). Although the precise function of this domain is not known, it has been proposed that the -GDD- motif may be a common nucleic acid recognition site and/or an active processing region (Kamer & Argos, 1984).

A third homologous domain in the putative replicase genes of plant viruses was also found in the TRV protein (encoded between nucleotides 2906 and 2944). Based on homologies with ATPases and GTP binding domains in various non-viral proteins, it has been proposed that this region also has a nucleotide binding function (see Hodgman & Zimmern, 1987).

The readthrough codon into the C-terminal region of the 194K protein is an opal codon. This is consistent with the result of Pelham (1979) who used a translation system in vitro with yeast suppressor tRNAs and found that readthrough of the 134K protein did not involve amber or ochre termination codons. The use of opal readthrough codons is a feature of mammalian viruses, for example Sindbis and Middleburg viruses (Strauss et al., 1983) although not so far found in other plant viruses. TMV (Goelet et al., 1982), beet necrotic yellow vein virus RNA-2 (BNYVV; Bouzoubaa et al., 1986), carnation mottle virus (CarMV; Guilley et al., 1985) and turnip yellow mosaic virus (TYMV; Morch et al., 1982) all possess amber readthrough codons. Comparisons do not reveal any obvious features of primary or secondary sequence which may have a role in allowing translational readthrough.

Termination of the 194K protein at nucleotide 5326 resulted in an intergenic region between the 194K and 29K proteins of one base. This was a correction of previous data (Boccara et al., 1986) due to an extra base inserted at position 5056. The sequence of PSG RNA-1 reported by
Nucleotide sequence of TRV RNA-1

Cornelissen et al. (1986) shows an intergenic region of 68 residues resulting from a single nucleotide (C) insertion at base 5237, relative to the SYM RNA-1 sequence (Fig. 2). This insertion causes the reading frame of the PSG 194K protein to terminate early. The sequence in this region of SYM RNA-1, as shown in Fig. 2, has been confirmed by analysis of a second, independent cDNA clone. The C-terminal domain of the 194K protein including the region which was out of frame in the PSG sequence showed significant homology with the readthrough region of the TMV 183K protein.

Comparison of the 3' terminal 2077 nucleotides of PSG RNA-1 (Cornelissen et al., 1986) with the equivalent region of SYM RNA-1 showed them to be greater than 95% homologous. Within the 29K coding region there were 27 nucleotide changes, all in the third bases of codons, with the result that the PSG and SYM 29K proteins were identical in amino acid sequence. In the region coding for the 16K protein, 23 base differences occurred of which only 10 were in third base positions. This resulted in 12 amino acid differences (out of a total of 141) between the 16K genes of PSG RNA-1 and SYM RNA-1. The 16K gene that is duplicated on TRV strain TCM RNA-2 (Angenent et al., 1986) had 19 amino acid differences compared with the SYM RNA-1 16K gene.

Non-coding regions

The 5' leader sequence of SYM RNA-1 was 202 bases long and contained no AUG start codons. This compares with a leader sequence of 570 bases for PSG RNA-2 (Cornelissen et al., 1986), 546 bases for TCM RNA-2 (Angenent et al., 1986) and 573 bases for PRV RNA-2 (Bergh et al., 1985), all of which contain multiple AUG codons prior to the start of the coat protein genes. PRV RNA-2 was also found to contain two regions of 119 and 76 nucleotides that were almost directly repeated within the leader sequence. No comparable repeats could be found in SYM RNA-1, although a shorter sequence was repeated in the 5' leader sequence (see below).

The 3' termini of SYM RNA-1, PSG RNA-2 and TCM RNA-2 were almost identical, having 3' non-coding regions of 255 bases. However PRV RNA-2 had a 3' non-coding region of 554 bases of which only the 3' terminal 45 nucleotides showed any homology (78%) to the SYM, PSG and TCM RNA sequences.

Comparison of the 5' termini of different TRV RNA species

Robinson et al. (1987) compared the RNA species of various tobravirus strains by hybridization methods and detected significant homology between the 5' terminal regions of RNA-1 and RNA-2. Similarly, the sequence of SYM RNA-1 was found to be homologous with the 5' terminus of RNA-2 from strains PSG and TCM (Angenent et al., 1986). In order to investigate in detail the homology between RNA-1 and RNA-2, the 17-mer oligonucleotide which was included in these homologous regions was used to prime cDNA synthesis on RNA-1 and RNA-2 of different TRV strains. The nucleotide sequence of the cDNA was then determined by the chemical cleavage method. RNA from TRV strains PRN (Cadman & Harrison, 1959), ORY (Oregon Y; Lister & Bracker, 1969) and N5 (Harrison et al., 1983) was investigated in this manner. In each instance, as with SYM, at least two cDNA products of differing lengths were obtained with each sample of unfractionated viral RNA. These were assigned as either RNA-1 or RNA-2 on the basis of their homology to SYM RNA-1 (Fig. 2) or PSG or TCM RNA-2 (Cornelissen et al., 1986; Angenent et al., 1986). Using the 17-mer as a primer, some of the RNA preparations gave rise to three or four extension products due to repetition of a sequence containing the priming site in the 5' region of individual RNA molecules. The sequences are shown aligned in Fig. 5. A second primer (22-mer) was used to confirm these data except for RNA-2 of N5 and ORY. In these instances the sequence of the 22-mer was apparently outside the zone of homology with SYM RNA-1. The 5' termini of all the RNA species started with the motif AUAAAACA- except for N5 RNA-2 which started AUAAAUAU- . An alignment of the sequences showed two major features (Fig. 5): a 22 base imperfect repeat was found to be conserved in all the TRV sequences and the RNA-1 sequences contained two deletions, either side of the repeats, with respect to the RNA-2 sequences. The RNA-1 sequences were very homologous and most of the differences that occurred
Fig. 5. Alignment of the 5' terminal nucleotide sequences from different tobraviruses. Sequences are aligned with the 5' terminal sequences of PSG RNA-2 (Cornelissen et al., 1986), TCM RNA-2 (Angenent et al., 1986) and PRV RNA-2 (Bergh et al., 1985). Gaps, indicated by dots, have been introduced to obtain optimal alignments. Oligonucleotides used for primer extension studies are written above the aligned sequences. Boxes define the conserved 22 base imperfect repeats. Asterisks show the limit of the derived sequences.

distinguished the SYM and ORY sequences from the PRN and N5 sequences. SYM RNA-1 could thus be regarded as being more related to ORY RNA-1, and N5 RNA-1 to PRN RNA-1. Although the RNA-2 sequences were also very homologous to one another, a number of base changes were evident but grouping of the strains on their apparent relatedness of their RNA-2 species was not possible.

DISCUSSION

Analysis of the nucleotide sequence of TRV strain SYM RNA-1 has identified four ORFs. Natural TRV isolates have been found that contain only RNA-1 and thus expression of these proteins is sufficient for pathogenesis. We have described the 29K and 16K proteins.

The RNA-2 molecules of TRV, strains PSG and TCM, and of PRV all possess AUG initiation codons 5' to the start of the coat protein gene. The possibility that the 134K gene on SYM RNA-1 does not start at the first AUG cannot be ruled out but is unlikely as the nucleotide context of the first AUG is more similar to that proposed as the consensus motif for plant genes (AACAAUGGC; Lutcke et al., 1987), than are those of the following eight in-frame AUG codons. The TMV 126K protein also starts at the first AUG codon (Goelet et al., 1986). The new results presented here show homology between the TRV 134K/194K proteins and putative replicase proteins of other viruses, and therefore suggest that the TRV proteins have a similar function.
Nucleotide sequence of TRV RNA-1

recombinants. Hybridization studies show that both have TRV-like RNA-1 species but their coat proteins are antigenically related to PEBV. The 3' terminal homology between SYM RNA-1 and RNA-2 is less than 100 nucleotides (Robinson et al., 1987). It is possible, therefore, that the essential regions for minus strand replication are contained in this short 3' sequence. The 5' end homologies between RNA-1 and RNA-2 molecules break down just on the 3' side of the repeat sequence motif (Fig. 5). Thus it is likely that the origin for positive strand replication lies between this point and the 5' terminus. The deletions in RNA-1 molecules may be concerned with the regulation of RNA-1 replication, giving rise to differences between the amounts of RNA-1 and RNA-2 found in infected cells. It will thus be interesting to compare the TRV terminal sequences with those of PEBV when they become available.

We have already suggested that TRV and TMV share a common viral ancestor based on their similar viral morphology, gene organization, translational strategy and amino acid homology in the 29K/30K proteins (Boccara et al., 1986). This proposal is further supported by analysis of the 194K and 183K proteins in TRV and TMV respectively where the homology is much higher than between either of these viruses and the equivalent proteins of other plant viruses. Furthermore, the recent work of van Belkum et al. (1987), which shows tRNA-like features of the 3' terminus of TRV RNA-1, has indicated an additional similarity that was not obvious previously due to the inability of the TRV RNA to be aminoacylated.

The major difference between the tobamo- and tobraviruses remains their respective monopartite and bipartite genome organization. Current understanding of plant viruses does not allow an easy explanation of how this difference may have arisen. One possibility is that a primitive monopartite genome gave rise, by degeneration, to an RNA-2 molecule which contained only the terminal sequences required for replication and the coat protein gene. This would have then allowed the coat protein to be lost, by degeneration and deletion from the RNA-1, and the whole process may have been driven by the selective advantages associated with multipartite plant viral genomes (Matthews, 1981).

An alternative hypothesis proposes evolution from a primitive bipartite genome via a flow of genes between RNA-1 and RNA-2. The mechanism of such a process might involve RNA recombination in which the terminal homologies (Fig. 5 and Robinson et al., 1987) help align the two molecules. RNA recombination has been shown in an experimental situation with BMV (Bujarski et al., 1986) and is the probable explanation for the evolution of tobraviral RNAs, such as N5, which are hybrid between PEBV and TRV.

Irrespective of the evolutionary relationship between TRV and TMV, it is clear from the extent of protein sequence homologies and from features of virus structure that there are functional similarities between these viruses. These various similarities will serve as a useful guide for direct analysis of the TRV gene functions. Such analysis is now greatly facilitated by the ability to produce infectious RNA transcripts from full length TRV and TMV clones (W. D. O. Hamilton & D. C. Baulcombe, unpublished results; Dawson et al., 1986; Meshi et al., 1986).

This work was supported in part by the Rockefeller Foundation. M.B. held an award from the Royal Society while on sabbatical leave at the Plant Breeding Institute.

REFERENCES


Nucleotide sequence of TRV RNA-1


(Received 7 May 1987)