Volume 15 Number 20 1987 Nucleic Acids Research

Sequence of figwort mosaic virus DNA (caulimovirus group)

Richard D.Richins, Herman B.Scholthof and Robert J.Shepherd

Department of Plant Pathology, University of Kentucky, Lexington, KY 40546, USA

Received August 21, 1987; Revised September 9, 1987; Accepted September 18, 1987

ABSTRACT

The nucleotide sequence of an infectious clone of figwort mosaic virus (FMV) was determined using the dideoxynucleotide chain termination method. The double-stranded DNA genome (7743 base pairs) contained eight open reading frames (ORFs), seven of which corresponded approximately in size and location to the ORFs found in the genome of cauliflower mosaic virus (CaMV) and carnation etched ring virus (CERV). ORFs I and V of FMV demonstrated the highest degrees of nucleotide and amino acid sequence homology with the equivalent coding regions of CaMV and CERV. Regions II, III and IV showed somewhat less homology with the analogous regions of CaMV and CERV, and ORF VI showed homology with the corresponding gene of CaMV and CERV in only a short segment near the middle of the putative gene product. A 16 nucleotide sequence, complementary to the 3' terminus of methionine initiator tRNA ($tRNA_i^{met}$) and presumed to be the primer binding site for initiation of reverse transcription to produce minus strand DNA, was found in the FMV genome near the discontinuity in the minus strand. Sequences near the three interruptions in the plus strand of FMV DNA bear strong resemblance to similarly located sequences of 3 other caulimoviruses and are inferred to be initiation sites for second strand DNA synthesis. Additional conserved sequences in the small and large intergenic regions are pointed out including a highly conserved 35 bp sequence that occurs in the latter region.

INTRODUCTION

The caulimoviruses are a group of small double stranded DNA viruses that infect higher plants[1,2]. The type member of the group, cauliflower mosaic virus (CaMV), has a genome of about 8000 base pairs (bp) with six to eight major genes and one large and one small intergenic region[3,4,5]. Its DNA can be cloned in bacteria in an infectious form[6,7], hence, it is easy to manipulate the virus genetically. As a consequence, CaMV has become a useful model system for studying gene expression and pathogenesis in plants.

Until recently, there has been little information on other members of this group of plant viruses, including figwort mosaic virus (FMV), the subject of this communication. This virus appears to be similar to CaMV both biologically and in having a small double-stranded DNA genome[8,9,10].

However, hybridization tests have shown that FMV DNA has little apparent sequence homology with that of CaMV[10].

Recently, we have adapted a strain of FMV to solanaceous hosts which are more amenable than its original host (Scrophularia californica) to greenhouse, tissue culture and protoplast manipulations [11]. In this communication we report the sequence of an infectious clone of this adapted strain and compare it at the nucleotide and amino acid sequence level with CaMV and other sequenced caulimoviruses.

MATERIALS AND METHODS

Standard molecular biology techniques were employed throughout this study (for protocols, see Maniatis et al[12]), therefore, only novel or modified procedures will be described.

Virus source and cloning: The original virus, isolated from Scrophularia californica, was adapted to solanaceous plants (DxS strain) as described[11], then its DNA was cloned into the unique SacI site of pUC13. The resulting plasmid, designated pFMV Sc3, was found to be infectious to Nicotiana bigelovii.

DNA sequencing: For the sequencing of the FMV genome, XbaI fragments of pFMV Sc3 were subcloned into M13mp11. Deletion subclones were made for each of the XbaI subclones using the T4 DNA polymerase cutdown method of Dale et al [13]. Subclones of the regions flanking each of the XbaI sites and the unique SacI site were generated by gel isolating specific restriction fragments and cloning them into either M13mp11 or pUC119 (supplied by Jeff Viera, Waksman Institute, Piscataway, NJ).

Reagents and the apparatus (model S0) for the dideoxynucleotide sequencing procedure were purchased from Bethesda Research Laboratories (BRL, Gaithersburg, MD). ^{35}S-deoxyadenosine 5'-(α-thio)triphosphate (New England Nuclear, Boston, MA) was used in place of the ^{32}P analogue. Procedures used for the phage and template purification and sequencing reactions were from BRL's M13 Cloning/Dideoxy Sequencing Instruction Manual. Two modifications were made in the dideoxy sequencing protocol: 1) The dideoxynucleotide triphosphate concentrations were reduced to allow more bases to be read from a single subclone, and 2) Sequencing reactions were carried out at 43 C instead of 30 C.

Computer analysis: Many of the analyses were performed on an Apple IIe computer in Apple Pascal format using either the University of Minnesota or the Cornell DNA sequence analysis program. The linear amino acid homology

plots utilize a modified version of the Cornell DNA analysis program PROTHOM to give a graphical output of the data derived from the program (this program is available from the authors on request). To aid in the alignment of analogous polypeptide sequences from FMV, CaMV and CERV, the FASTP program (IBM PC version)[14] was used. Two dimensional homology plots were performed on an IBM computer using the DIAGON program[15].

RESULTS

DNA Sequence: The sequence was obtained from independent clones from both strands of the genome such that nearly every section of the genome was represented at least twice, and frequently for three or more times. Approximately 88% of the whole genomic sequence was represented by clones from both strands of the DNA. Only a small region of approximately 100 bp located near the center of ORF IV was represented by data from only one clone. This fragment was very difficult to obtain as a stable bacterial clone for unknown reasons.

The complete nucleotide sequence of the plus strand of the DxS isolate of FMV is given in Fig. 1. The genome consists of 7,743 bp, and has a GC ratio of 43%. In comparison, the genomes of the CM1841 strain of CaMV[4] (8,031 bp) and carnation etched ring virus (CERV) [17], another member of the caulimovirus group (7,932 bp), each have a GC content of 44%. We have numbered the FMV sequence beginning with the 5' end of the probable primer binding site (5'-TGGTATCAAAGCCATG-3') because essentially analogous sequences are found at the zero map positions of the CaMV and CERV sequences[3,4,5,17]. In addition, end-labelling studies performed by Hull and Donson[9] and in this lab[11] confirm that the minus strand discontinuity is located very near this region (Fig. 2).

Coding Regions: Computer analysis of the plus strand predicts seven putative coding regions having molecular weights in excess of 10 kDa (Fig. 2). In addition, there is a 7.4 kDa protein coding region (ORF VII) immediately preceding ORF I (Fig. 2). A large intergenic region between ORFs VI and VII and a small intergenic region between ORFs V and VI, are also present. With respect to size and location, seven of these eight coding regions (ORFs I-VII) and both intergenic regions closely resemble regions found in the CaMV[3,4] and CERV[17] genomes. ORF VIII, which overlaps the C-terminal portion of ORF IV of the CaMV genome[18], is not found in the FMV and CERV genomes. The FMV genome contains an additional reading frame, designated ORF IX, which overlaps parts of regions III and IV (Fig. 2). The coordinates delineating

```
TGGTATCAAAGCCAT GTGCTCAACTAGAAA AACATCTGTAATGGA TGAAAAGGTAATTGA AAATGAAGAAATTCA TTTCCAAGAAAATTC    90
TAATGGATTTTCAGC AGATCTCACGATCCA TCAAGATAAGTTAAA ACAGATTTCAAAAAC TGGTTTAAATCTTGA AAAAGAACATATATT   180
TAATATGCCTTCTTC TCTTACAAAAGCTTT TAAAACAGCCTTTAA AAGAAAGAACGAAAT ATTTTATTGTGTTTC AACAAAAGAAATGTC   270
AGTAGATATTAAAGA TGTCTCTGGACAAGT TTATCTTCCTCTTAT AACTAAGCAGGAAAT ACAACAAAAGCTGAT GAAAATTGATCCATC   360
AGTAAGAAGTAAAAT CTCAATGATCCACCT TGGAGCAGTAAAGAT TCTACTTACAGCCCA ATTCCGGCAAGGAAT TGATACGTCAGTTAA   450
AATGGCACTTATCGA TGATAGAATCGTCAA TCGAAAAGATAGTCT GCTCGGAGCAGCAAG AGGTAATCTCGCATA TGGTAAGTTTATGTT   540
TACCGTTTACCCTAA ATTTGCATTGTCATT ACAAAGCAAGAATTT AGATAAAACATTAAG CTTTATTCAATCAATT TGAAAGAAAAGACTT   630
AATGAAAACAGGCGA TAAAGTTTTTACGGT CACATACCTAATAGG ATATGCCCTGACAAA CAGCCACCATTCTAT AGAATATAGGAAAAA   720
TTCCAATATTGAAAT AGAAGAAGTCTTTAA AGACATAGGTCAAAT AGAAGAATCACCATT CTGTGATATATCTCC TATAGACGAAAACTG   810
GACAATGGACATTGC CAGAAACAAAAAATC TATAGCCACAGGAAG TAGCTCTAGAAGAAA CTTCAGGATAGACGA AAGTTTACTTGAAAA   900
CAAAGACGAAAACCT ACTTAGGTCTATGTC TACAAAAATAGATAC TCTAGGTAAAAAACT TAGTCTTATCTAGGA CAATGAGTAAAGATC   990
ACTTAGACTTTCCTC ACGTCTACAAAAAGA ACAAAGTTCTTAAGT TAAAACCCTTAGATT TGTCAAAAAATCCTA GGAGTTATTACTTTA  1080
GCAGCCAAAATGGTA GTATTCAAGCTATTA TTAATCACTGCAATA ATATTAATCAGATAA CAGCACGAAACTGGT TAAAACTAAGCAAAG  1170
TATTATCATATTTTG GTCTAGAAAAAGACA CTAGTGATAGTATAA GCAAAAATAAAAGCC CGTTTAACAGGTTTT TAAAAGATATCTCTC  1260
ACATTTTCAGAGAGG GGGAAGGAAGTACTA AAAAAGCTTCTGAAC TAGGAGAAATCTTAG AAAAAATTAAGAATT TAGACCTAAAAATCG  1350
AAAATCTTAATAAAA GAATACCAGACAACT TGGTAACAAAAGCCT TAATTAAGGAATTAG TCAAAGACTTTGACG AAAGACTTACAGAAG  1440
TCAGAGATGACATTA AGAAAGTCATAGGAT AATGGCAGCCACATT ATCTGAAATTCGTGA ACTCATTCAAAGTTT AACCAAAACTGCTAA  1530
TGAGATCAAAGCTAT GCTTGAAAGAAATTC TGCAGGAAAACCTAC AGGTATTGAAGAAGC AGCGGCAAAAATCAT CAAAGATATTGGTGA  1620
TAAAATCGATCAGTG CGAATGCACAAAGAA AATCGAAGAAATGCT TGATCAAAAGCAAAA TGCAGACACACAGAT TGTGCCAACAAAACA  1710
AGAGTCCTCTGGACT TGTCAAATATTCGTA CCCAAACTGGAATGT TGGTAACGAAGAACT TGGAAGCTCAGGCAA CCCAAACGCGGTTAA  1800
ATGGCCACCAAGAAA ATGAGAATCTCAGAA AAACTTTGGGATAGT CTCGAAAAAGACGAG TGCAACATCGATGAA GTTGTACAACTCATG  1890
AGCCTCGACGAGGAA GAGCTCATAACCAAA TTTGCCCAAGAGGTA AGTTTAAGAATTAGG TATCTAGATGAAAAA CCTAATGAAATTTCT  1980
TTTATTGCAGAAGCA ACAGAAGACTACTCA GAACCCGAGACAGAG TCATCGGACGAGGAA ACCTATTTCCAACAA ATAAGAATGGAAAGA  2070
GGAGAAAGTAGCGAA ACAAAACGAGAACAG CAAGATCTCGGAGCT ACCAGAAAAAGGAAA ATAGAGGAAAGAAAT CCTTTCTACACTCCA  2160
CCAGTACATAAAGGT ATACCCAGTTCCACT GGAAGAGGTACCGAA ATCTCTACCCTTAAC TTAGATTGTATATCA AGTTTTGAGGAAAGA  2250
AAAGTCATGATTGAT AAATGGTTTAATGAA ATAAGCTTAATCATT CAGACAAATAAAGAA AGTTTTGACACAAGT CTAAAAGTCTTAACC  2340
TTAATGGAACATAGG ACTGAAGGAATTGCA AAATCTTTCATAAAA CAAGCAACATGGGAT ATACTCATAACCCCT GAAAAAATTATAGAA  2430
GAAGTGTTAACAGGC CAGCTCTCTTTTTAG AGAGGGGGAGTTAGG ACATTTTTGTCTTAA CAAACAGGAAATGTT ACATTTAAACGTAAC  2520
GAAGAACTTCTTATA AAATCTCAGCTTTGC AATATTTGTGAACTA GATAATTTTACTTGT TTTTACGAAAAACAA ATCAATCAACTTAAG  2610
TTTGAAGATTTTCCT AAATGGATAGAACTT TACTTAGGTAAAATA CCTATTATAGGTAAG CAGTCTAAAGAAAGA TGGGATAATGAAAAA  2700
TCTTTCACAACAAAG TATTCTTTAGCATTT GCTAAAAGGATCATT CAAGAAGAAATCGCT AAATATTGTGATTTT CAAAGAACATCTAAA  2790
AAACTAAAAAATTTT AGTAAAAAATGTTGT TCAAAAAATAGTCTT GACCCTCTTGTTTCA TTCGGGTGCAGAGAC ACAAAGAAAAAGGAT  2880
TTCAAAAAATCTTCT AAGTACAAAGCTTAT AAGAAAAAGAAAACC CTTAAAAAACTCTGG AAGAAAAAGAAAAGA AAATTACACCAGGT  2970
AAATATTTCTCTAAA AAGAAACCAGAAAAA TTCTGTCCGCAAGGA AGGAAAAAATGCAGG TGCTGGATTTGCACT GAAGAAGGACATTAT  3060
GCAAATGAGTGTCCA AATAGAAAAAGCCAC CAAGAGAAGGTAAAA ATTCTTATACATGGC ATGAACGAGGGATAC TACCCTCTAGAGGAC  3150
GCCTACACAGGTAAT CTCGAGGTATTTTCA ATGGAAATCATTGAA GAAACAACCTCAGAA GAAGAATCCACAACA GATTCTGATTCATCA  3240
TCGTCAGATGACGAG CAGCTCTCTTTTTAG AGAGGGGGAGTTAGG ACATTTTTGTCTTAA CAAACAGGAAATGTT ACATTTAAACGTAAC  3330
AAATCCAAATTCAAT TTATATTGAAGGTAA ACTATCCTTTGAAGG ATACAAATCATTTAA CATTCACTGTTATGT TGATACAGGTGCCAG  3420
TCTATGCATTGCATC AAGGTACATTATACC AGAAGAACTTTGGGA AAATTCACCAAAGGA CATTCAGGTAAAAAT TGCAAACCAGGAGCT  3510
GATAAAGATCACTAA GGTATGCAAAAACCT CAAGGTAAAATTCGC AGGTAAATCCTTCGA AATACCTACAGTATA CCAGCAGGAAACAGG  3600
TATTGATTTCTTAAT CGGTAACAACTTCTG CAGATTATACAACCC GTTCATACAATGGGA AGATAGAATAGCTTT TCACCTAAAAAATGA  3690
AATGGTTCTAATCAA AAAAGTTACTAAAGC CTTTTCGGTCAGTAA CCCATCATTTTTAGA AAACATGAAAAAAGA TTCTAAAACTGAACA  3780
AATTCCGGGAACAAA TATTTCAAAGAATAT TATCAACCCAGAAGA AAGGTATTTTCTAAT TACAGAAAAATATCA AAAAATAGAACAATT  3870
ACTAGACAAAGTATG TTCAGAAAATCCTAT AGACCCTATTAAGTC TAAACAATGGATGAA AGCATCAATTAAATT AATTGATCCGCTAAA  3960
AGTTATTAGGGTAAA ACCTATGAGCTACAG TCCACAGGACAGGGA AGGATTTGCCAAACA AATTAAGGAACTCCT TGATTTAGGTCTAAT  4050
TATACCCAGTAAATC CCAACATATGTCACC AGCCTTCCTAGTTGA AAATGAAGCAGAAAG ACGTAGGGGTAAAAA GAGAATGGTTGTAAA  4140
TTATAAAGCAATCAA TCAGGCTACTATTGG AGATAGTCACAATCT ACCTAACATGCAGGA ATTACTTACATTACT GCGAGGTAAAAGCAT  4230
CTTTTCCAGTTTTGA CTGTAAAAGTGGATT TTGGCAGGTAGTTCT TGACGAAGAAAGCCA AAAGCTTACAGCTTT CACGTGCCCACAAGG  4320
ACATTTCCAGTGGAA AGTGGTCCCATTCGG CCTAAAGCAAGCACC AAGCATATTTCAAAG ACATATGCAGACAGC CTTAAATGGCGCAGA  4410
TAAGTTTTGTATGGT TTACGTTGATGATAT CATCGTTTTCAGTAA TTCAGAACTTGATCA CTATAACCATGTTTA TGCAGTTTTAAAAAT  4500
TGTAGAAAAATATGG AATAATCCTTTCAAA AAAGAAAGCTAATCT TTTTAAGGAAAAAAT CAATTTTCTAGGACT AGAAATTGACAAAGG  4590
GACACATTGTCCACA AAATCATATTTTGGA AAACATTCATAAATT TCCAGATAGATTAGA AGATAAAAAACATCT TCAGAGATTTTTAGG  4680
AGTACTTACATATGC CGAGACATATATTCC TAAACTCGCAGAAAT TAGGAAACCTTTACA GGTTAAACTAAAAAA GGATGTAACCTGGAA  4770
TTGGACACAATCTGA TTCAGATTACGTAAA AAAGATAAAGAAAAA CCTAGGATCATTTCC TAAGCTTTACTTACC TAAACCAGAAGATCA  4860
TCTCATTATTGAAAC AGATGCATCTGATAG TTTTTGGGGCGGTGT CCTAAAAGCAAGAGC GCTTGACGGTGTTGA ATTAATTTGCAGGTA  4950
TAGTTCAGGAAGCTT CAAACAAGCAGAAAA GAATTATCACAGCAA TGATAAAGAATTACT AGCAGTCAAGCAAGT AATCACCAAGTTTTC  5040
AGCTTACCTAACACC AGTCAGGTTTACAGT AAGAACTGATAACAA AAATTTTACTTATTT CCTTAGAATTAATCT TAAAGGTGATAGTAA  5130
ACAAGGACGATTAGT CCGTTGGCAAAATTG GTTCAGCAAGTATCA ATTTGATGTCGAACA TCTTGAAGGTGTAAA AAACGTTTTAGCAGA  5220
TTGCCTCACGAGAGA TTTTAATGCTTAAAA ACGTAAGCGCTGACG TATGATTTCAAAAAA CGCAGCTATAAAAGA AGCCCTCCAGCTTCA  5310
AAGTTTTCATCAACA CAAATTCTAAAAACA AAATTTTTTAGAGAG GGGGAGTGATGGAAG AATTAAAGGCCTTGC GCCTAAAAGAAAAAA  5400
TTCTAGAAATAGAAT TAAATTCTGTAAAAA TGCAGATTCATGCTT ATGAGGAAAGCCTCA AAGCCACAACTGTTA ATTCAGTACAGGAAG  5490
GAGAAATCCTACAAA CTGAATCTATTCCTG AGTGTCCGGCTCAAG GTAAAGAAACACCAA ATCCGGTGAAAGCTG ATAGTTTGCTAAAAA  5580
CTATTCTTGGTAATG AACGCCAGAATCCTC TTGAAGGAAAAAGTT CAAAGTTGGTTAATT TAACACCAAAATCCG ACAAGGATAAAGTTA  5670
AATCTAGTCCGGTTG CAAACGGCTCCGGTA AAGACTCAACAAATC CGTTGAACCCTGTTG CTTTGGGAAAAAGCA AAATGACTATTCTCG  5760
GTCAAAAACAAGCTG ATGAAGAAGAGTTCA AGCCTGATTATCTCA GAGCAGCCTCTAATG GTCAGTCATGGTTTG CAGTATATAAAGGCC  5850
CAAACAAGGAGTTTT TCACTGAATGGGAAA TTGTTGCTGATATTT GCAAGAAACGACAAA AGTCTAAAAGATTCC GATCAAAGGAGCAAG  5940
CTGAAGTGTCTATCA GTCTTTACAACAAAG ACATCCAAGACCCAG TAAATTTTCTAAGAC CAGTCAAACTAGTCA AAGAAGAAAGAGCAG  6030
CACAGCCATTGAAGT TCAAGGCCATTGCAG CAGAACAAACCATAC AATTTGATGAGTTTC GTCAGATATGGGAAA AATCAAGACTTTCAG  6120
ATCTTGAAGATGGTG TACAAGAAAAGTTCT ACACCAACGATTCAG CAAGCAAAAGTACAT ACACTTTTGTTGAAA ACGCAGAACCTTATC  6210
TGGTTCACACAGCAT TCCGAGCAGGATTAG CCAAGGTCATCTATC CGAGTCCTAACCTAC AGGAACTAAAATGGT TCCCAGAAGGAATTG  6300
TTAAAGCAATTAAAA ACTTTCGAAAAAAGG TTTTAAATGCAAAAG ATGCAGCTATATTCA TCAAAATATTTAGCA GCATTCCAGATTGGG  6390
TTCAATCAACAAGGT ACGAGCCATATCACT TTATTCAAATTGGTA TCGCCAAAACCAAGA AGGAACTCCCATCCT CAAAGGTTTGTAAGG  6480
AAGAATTCTCAGTCC AAAGCCTCAACAAGG TCAGGGTACAGAGTC TCCAAACCATTAGCC AAAAGCTACAGGAGA TCAATGAAGAATCTT  6570
CAATCAAAGTAAACT ACTGTTCCAGCACAT GCATCATGGTCAGTA AGTTTCAGAAAAAGA CATCCACCGAAGAGT TAAAGTTAGTGGGCA  6660
TCTTTGAAAGTAATC TTGTCAACATCGAGC AGCTGGCTTGTGGGG ACCAGACAAAAAAGG AATGGTGCAGAATTG TTAGGCGCACCTACC  6750
AAAAGCATCTTTGCC TTTATTGCAAAGATA AAGCAGATTCCTCTA GTACAAGTGGGGAAC AAAATAACGTGGAAA AGAGCTGTCCTGACA  6840
GCCCACTCACTAATG CGTATGACGAACGCA GTGACGACCACAAAA GAATTCCCTCTATAT AAGAAGGCATTCATT CCCATTTGAAGGATC  6930
ATCAGATACTGAACC AATATTTCTCACTCT AAGAAATTAAGAGCT TTGTATTCTTCAATG AGAGGCTAAGACCCT AAAGAGTTTCGAAAG  7020
AGAAATGTAGTATAG TAAGAGTCCTCCCAG TCCGGGAGATTGTAA TAAAGAGATCTTGTA ATGGATCCAAGTGTC TGTAATTTTTGGAAA  7110
AATTGATCTATAAAA TATTCAATCTTTCTT TAAGCTTATTCAAAG AACAAACATACTATC TATCATCCAAATCCA CAGAGTGACAGAGAG  7200
AAAATGGTCTGTGTT GTGTGGATCTGAAGT ACCGCCGAGGCAGGA GGCCGTTAGGGAAAA AGGGACTGTTTTGAC CGTCAAAGTATCAGG  7290
CTGGCTCTAGGAAGG AAGATGAAGATATCA GGTATTGGTTTATGT TCTAAAAAATAAGTA ATAAAGAAAAAGTT TATTAAAAAGAAAAT  7380
TTTATCAAGAGCAAA TTACATGTCTAGAGG ATACCTAGATCTATA TTACAATAATCTTAC TTACATGTTTTATTT CGTGACTCTAAATTA  7470
AAAAATTGTTTAATT GTTTATTCAAAACAA TGCCAGGACTAACCC TCCAGCAAGAGTATA TACTCTTAGCACACC TTATTCTTCAGGTAC  7560
TCGAAGAAGTCAAGC AGGTACAACTGCATT CAGGAGACTTCCAGT CTCTCAGAAGTCTAT ATGCTAGGCTTAACG GGCTTCGGTCACACC  7650
AAGCTCATCTCCAAG CGAGAATTTCAGCTG TTTCTCAACACGGCA ATTGAAGCGTCATGA ACTTCAAGAAAATCT CGGATCCCGTCCGCA  7740
TTC  7743
```

Fig. 1. The Nucleotide Sequence of FMV DNA (DxS strain). The plus strand is presented here with the numbering beginning as described in the text and in Figure 5.

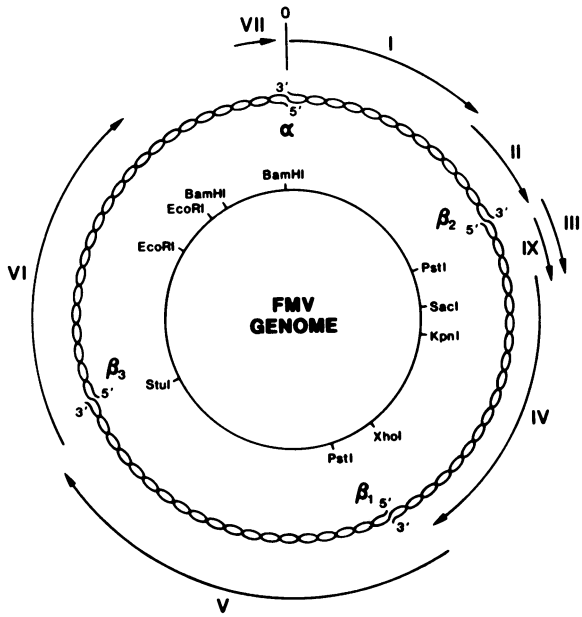

Fig. 2. Organization of the FMV Genome. The inner circle shows the positions of selected restriction sites found in the 7743 bp circular genome. Single-stranded discontinuities are shown in the middle (chained) circle. We have numbered the three single-stranded discontinuities found in the plus strand such that the two discontinuities β_1 and β_2 at positions 1250 and 3300 are numbered the same as those found at these approximate positions in the CaMV genome. The additional discontinuity (at position 5350) is denoted β_3. The peripheral arrows depict the positions of the major open reading frames and are positioned to show the reading frames in which they occur.

each of these ORFs, together with the molecular weights (MW) and percent direct homologies for the putative translation products of FMV, CaMV and CERV are given in Table 1.

Only two open reading frames of significant size are observed in the minus strand. The predicted MWs of these two translation products are 10,840 and 10,399 Da. However, no sequences associated with transcriptional promoters or terminators were found flanking these ORFs so it seems unlikely that these regions are expressed.

Comparisons of each of the putative FMV and CaMV translation products using the matrix analysis program of Staden[15] are shown in Fig. 3.

ORF I is one of the most highly conserved coding regions in the FMV genome (second only to region V) with nearly 54% of the predicted amino acids being exactly conserved compared to the region I protein of CaMV (Fig 3A). A

Table 1. Protein Coding Regions of the FMV, CaMV and CERV Genomes

Open Region	Starting Nucleotide[a]	Ending Nucleotide[a]	Protein Molecular Weights (Daltons)[b] and Amino Acid Homologies[c]		
	FMV	FMV	FMV	CaMV (%)	CERV (%)
I	14	985	36,957	36,825 (54%)	36,484 (49%)
II	978	1472	18,934	17,878 (42%)	18,774 (38%)
III	1472	1819	12,659	14,136 (37%)	14,214 (31%)
IV	1801	3270	57,346	56,614 (38%)	58,175 (33%)
V	3248	5248	76,938	78,698 (64%)	76,433 (64%)
VI	5364	6902	58,133	57,833 (26%)	56,350 (21%)
VII	7499	7700	7,447	11,398[d]	10,301[d]
VIII	----	----	----	12,430	----
IX	1529	1840	11,924	----	----

a) Assuming that translation begins with the first in frame ATG and ends with the first in frame stop codon within each open reading frame.

b) The molecular weights of the translation products of CERV[16] and CaMV[4] are presented for comparison.

c) The numbers in parentheses after each molecular weight is the percent direct homology of the amino acid sequences in each putative coding region.

d) The homologies between region VII were so low that a reliable degree of homology could not be determined. This was also the case when CaMV ORF VII was compared with CERV ORF VII.

function has not yet been identified for this gene product. However, it is expressed in CaMV infected plants[19,20].

ORFs II and III are also conserved in the FMV genome, but with less homology than was observed for ORFs I and V (Fig 3B,C). Gene II of CaMV has been shown to encode an aphid transmission component.[21,22,23] The function of the ORF III protein has not been determined, although this gene of CaMV is expressed in infected plants[24]. With only 115 amino acid residues, this region of FMV is the smallest conserved gene.

The ORF IV translation product of FMV shows approximately the same overall degree of homology (vs similar regions of CERV and CaMV) as do ORFs II and III. Near the amino terminus, the ORF IVs of FMV, CERV[17] and CaMV[3,4] encode proteins which are rich in glutamic and aspartic acids. Except for the high proportion of these acidic amino acids throughout this region, there is little additional amino acid homology. The middle third of the FMV region IV protein shows an increased degree of homology with the equivalent region of the CaMV and CERV capsid protein with over 40% of the amino acid sequence

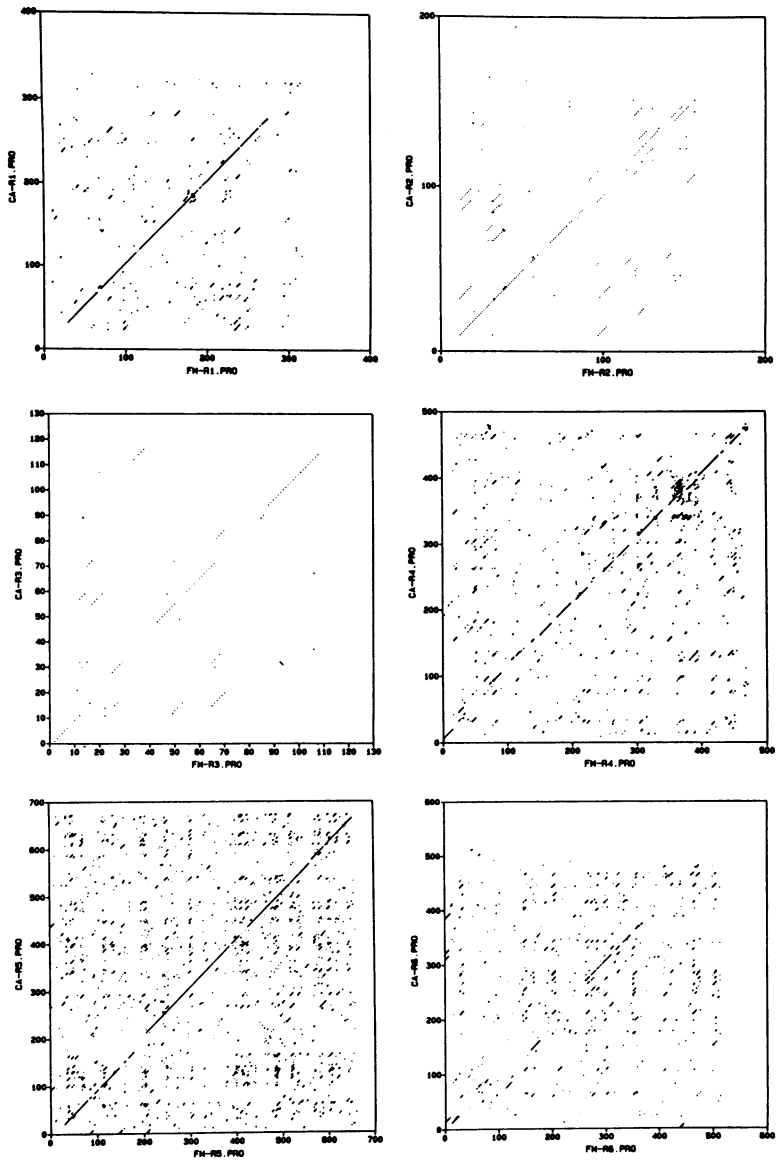

Fig. 3. Matrix analyses[15] comparing six of the major putative translation products of FMV with homologous translation products of the CaMV genome.[4] The translation products of both FMV and CaMV begin from the first methionine in each reading frame and end at the first termination codon (TAA, TAG or TGA). For each analysis, a 'window' of eight amino acids was used with positive matches being given for scores exceeding 96. The comparisons begin with ORFs I in the upper left hand corner and end with ORFs VI in the lower right hand corner.

being conserved. A highly basic domain containing over 40% lysine residues begins at amino acid 316 and ends at amino acid 396. Similar lysine-rich regions are found in CaMV[4] and CERV[17] region IV proteins.

Immediately following this highly basic domain in the putative coat protein gene is a region with a highly conserved array of cysteine residues with the arrangement of: $CysX_2CysX_4HisX_4Cys$. As pointed out by Covey[25], this array is conserved in the gag gene of all replication competent retroviruses and the Drosophila retrotransposon copia. Furthermore, it also appears in the coat protein genes of CaMV and CERV. Beyond this Cys-motif is the C-terminal domain that is rich in glutamic and aspartic acids and serine.

The putative region V protein of FMV shows a particularly interesting homology with its CaMV counterpart. When the inferred amino acid sequences of FMV and CaMV are aligned, the three distinct domains pointed out by Toh et al[26] for reverse transcriptases can be distinguished. The first domain is comprised of amino acids 26 through 143 of FMV (18 through 135 of CaMV), has high homology with the avian retroviral gag-specific protease p15[27]. The next major domain within the putative region V protein, beginning at amino acid 211 (203 of CaMV) and ending at amino acid 564 (556 of CaMV), together with the next domain (consisting of amino acids 574 through 666 of FMV or 563 through 679 of CaMV), may be the reverse transcriptase[26]. With retroviruses, it is associated with both RNase H and RNA dependent DNA polymerase activities. The homology between FMV and CaMV within these two domains is about 73% and 68%, respectively. Hence, this is the most highly conserved protein coding regions of these viral genomes (Fig. 3E).

The least conserved coding region of the major genes of FMV and CaMV and CERV is that of region VI. Except for a small conserved tract near the center of the protein, little homology is apparent in this ORF for the three viruses. Although Hull et al[17] also have found gene VI to have the least homology between CERV and CaMV, they found considerably more homology than is the case for the FMV and CaMV genes.

A graphical plot (Fig. 4) showing the homologies of the deduced gene VI translation products of the three viruses emphasizes the relationships of these putative proteins. Although short tracts of homologous sequences occur near the N-terminus for each of the ORF VI polypeptides, the most extensive region is near the center of the gene. Within this conserved central domain there is a run of 47 amino acid residues which shows 47% homology between FMV and either CaMV or CERV.

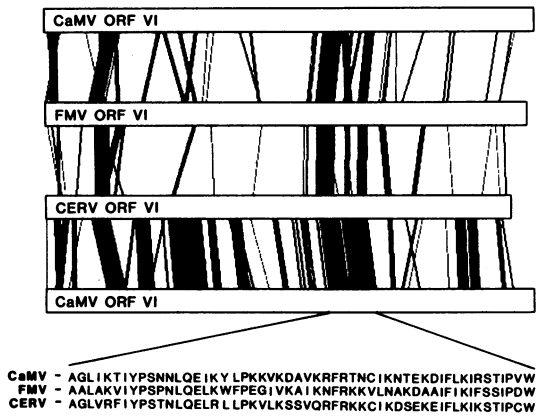

Fig. 4. Graphical plot of the amino acid homologies of ORF VI of FMV with region VI of CaMV and CERV. Homologous regions between polypeptide chains are displayed as dark lines. The plot utilizes a modified version of the Cornell DNA Sequence Analysis Program (PROTHOM) to generate comparisons. A window of 8 amino acids was used for this comparison with homologies exceeding 62% being scored positive. Homologous regions offset more than 25 amino acid residues in either direction along the polypeptide chain were not recorded in the data plot.

Regions VII, VIII (CaMV) and IX (FMV) were also compared, but no significant levels of homology were noted (data not shown).

Noncoding Regions: As noted earlier, two intergenic regions are present in the genome of FMV: a large intergenic region located between ORFs VI and VII consisting of 577 bp, and a small intergenic region located between ORFs V and VI consisting of 117 bp.

Within the small intergenic region there is a TATA-like box implicated in promoter function (TATAAAA) about 70 bp upstream from the ATG start of ORF VI. CG rich regions flank this sequence. Although the virus-specific RNA species occurring in FMV infected plants have not been characterized, it seems by analogy with CaMV that this region serves as a transcriptional promoter for ORF VI.

The large intergenic region contains several short runs of nucleotides homologous with the corresponding regions of CaMV and CERV. A graphical representation of the sequence homologies between the large intergenic regions of FMV, CaMV and CERV is presented in Fig. 5. A sequence similar to the TATA or Goldberg-Hogness box is found in each of the three viruses (see Fig. 5). In the FMV genome the probable TATA box occurs just inside the 3'-

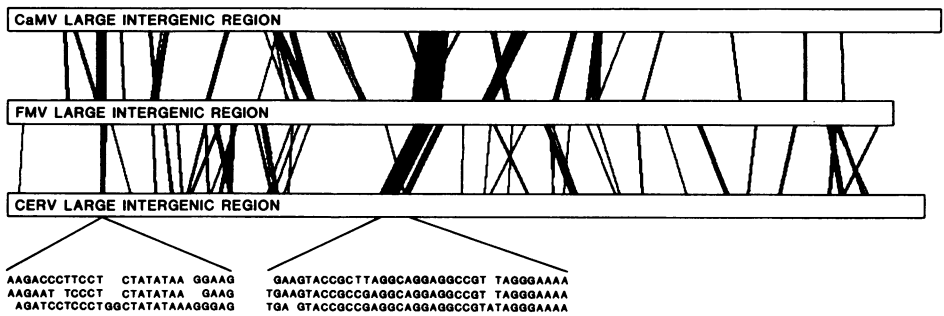

Fig. 5. Graphical homology comparisons of the large intergenic regions of three caulimoviruses (CaMV-top, FMV-middle, CERV-bottom). The comparison starts 100 bp upstream from the proposed TATA box of the genomic-length promoter and continues to the start codon of region I. The program used for this plot is identical to the one described in Fig. 4, except that a window of 10 bp was used and homologies exceeding 80% were scored positive. Regions of homology offset more than 40 bp in either direction were not recorded. The sequence of regions with a high degree of homology are given.

end of ORF VI (at nucleotide 6893). An identical sequence is found in the large intergenic regions of both CaMV and CERV (TATATAA). This sequence is reported to be essential for high level expression of eucaryotic genes including those of plants[28]. Messing et al[29] have made comparisons among the sequences of plant 5'-controlling regions and remarked on the frequent occurrence of the dinucleotide TC preceding the TATA box. This TC dinucleotide is found preceding the TATA boxes of both the FMV and CaMV large intergenic regions. However, the dinucleotide GC precedes the TATA box of the large intergenic region of CERV as well as the TATA box found in the small intergenic region of FMV.

At position 7064 of FMV a sequence associated with polyadenylation of mRNA (AATAAA) occurs. This sequence lies 162 bp downstream from the inferred FMV TATA box. For comparison, the possible polyadenylation sequence found in the intergenic region of CaMV is positioned about 197 bp downstream from the TATA box. Termination of both major RNA transcripts of CaMV occurs just 11 bp downstream from the initial nucleotide of this sequence[30,31]. However, a consensus sequence for termination in plants has not been defined. Even the function of AATAAA as a polyadenylation signal in plants is uncertain, since a recent compilation of termination sequences of 26 well characterized plant genes has pointed out that only 39% contain the conventional AATAAA signal[32].

The most extensive and striking region of homology within the large intergenic region of the three viral DNAs occurs downstream of the tentative

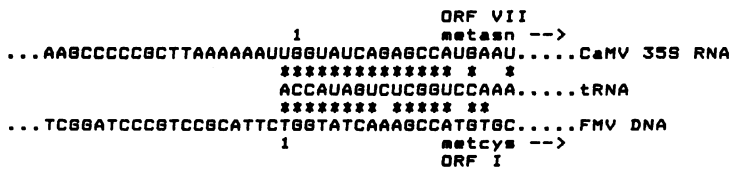

Fig. 6. Comparative sequences at the $tRNA_i^{met}$ binding sites of FMV and CaMV. Regions which best complemented the 3' terminus of bean cytoplasmic $tRNA_i^{met}$ [35,36] are shown. Asterisks denote complementarity with the $tRNA_i^{met}$ 3' terminus. The complete sequence of FMV given in Fig. 1 begins with the T residue designated "1" in this figure. The CaMV sequence presented is that of strain CM1841[4].

polyadenylation signal. This consists of a 35 bp sequence which is virtually identical for the three viral DNAs (Fig. 5). This highly conserved region occurs at position 7228-7262 in the FMV sequence, at 7740-7774 in that of CaMV, and at 7343-7378 of the CERV genome. The conservation of such an extensive tract in the large intergenic region suggests it is of some biological significance.

Probable Primer Binding Sites: Because of similarities in the sizes and positions of coding and noncoding regions of the DNA genomes of FMV and CaMV and evidence for replication of the latter in a reverse transcriptional mode[16,33,34], we sought a tRNA binding site on the FMV plus strand under the assumption that a full length RNA transcript would serve as a template for an RNA dependent DNA polymerase. This search yielded a sequence near the unique minus strand discontinuity (see Fig. 2,6) with complementarity to the 3' end of bean and wheat $tRNA_i^{met}$[35,36]. Fifteen of the first 17 nucleotides were found to be complementary to this sequence on the plus DNA strand in the vicinity of the minus strand discontinuity. This sequence is very similar to that postulated to be the minus strand primer for CaMV[16,33,34,37].

In a search of the FMV sequence for plus strand primer binding sites we have used the data of Hull and Donson[9] as well as our own results[11] from mapping of the single stranded breaks. These occur at the approximate nucleotide positions of 1250, 3300 and 5350. Purine-rich regions were found near each of these coordinates. In Fig. 7, we list these sequences and propose a consensus second strand primer binding sequence based on the compilation of 9 such sequences from 4 different caulimoviruses. A notable feature of these tracts is that a pyrimidine-rich region occurs immediately upstream from each run of purine nucleotides.

Gap	Virus	Sequence
B1 (G2)	CERV	TCTTTTT.AGAAGAGGGGGGGAA...TCATGTAGA
	CaMV	TTCTTTC....AGAGGGGAGGAGG..TTATCAGA
	FMV	TCTTTTT..AGAGAGGGGGAG.....TTAGGACATTT
	SoyCMV	TCATTTC..GAGGAGGGAGGAAG...TTACAAAAGGAA
B2 (G3)	CERV	AATCTTT..AAGGAGGGGGGAGGAAATGGTGA
	CaMV	CATTTTTAAGAGTAGGGGGG.....TTGATTACTCGA
	FMV	CATTTTC..AGAGAGGGGGAAGGAAGTACTAA
	SoyCMV	AATTGTT..GAGGAGGGTGAGGA...TAAGGGAAAAA
B3 (G4)	FMV	ATTTTTT..AGAGAGGGGGAG.....TGATGGAAGAA
Consensus		YTTTYR$_{6-4}$RGAGGGGR$_{1-2}$...T

Fig. 7. Purine-rich regions observed near the single stranded discontinuities of plus strand DNA of CERV, CaMV, soybean chlorotic mottle virus[53] and FMV. A consensus priming sequence is presented (R = any purine; Y = any pyrimidine).

DISCUSSION

The FMV genome is remarkably similar to that of CaMV and CERV with respect to its size and organization. Its genome consists of 7743 bp (compared to 8031 bp for CaMV DNA[4] and 7932 bp for CERV DNA[17]) organized into seven major open reading frames and two intergenic regions.

The sequence reveals that only one strand of the DNA genome (the plus strand) has long open coding regions. Although the complementary strand contains two ORFs which if transcribed would yield proteins in excess of 10 kDa, neither of these regions are conserved in the complementary strand of CaMV or CERV. Furthermore, these ORFs are not preceded by sequences associated with promoter activity. Hence, the transcription of FMV is probably asymmetric like that of the CaMV genome[38,39,40].

A unique ORF (ORF IX) of 313 nucleotides overlaps portions of regions III and IV of the FMV genome (Fig. 2). However, this region is not found in either the CaMV or CERV genomes, and therefore like ORF VIII of CaMV, it may be fortuitous in nature.

The five closely spaced ORFs of the FMV genome, and similar genes of CaMV and CERV raise the question of how these genes are expressed. From knowledge of CaMV it seems likely that genes I-V are transcribed via a genomic length RNA molecule and that no subgenomic versions of these genes will be found (with the possible exception of gene V[41]).

If a genomic length molecule serves as a mRNA for translation of genes I through V, initiation probably occurs on internal AUGs as proposed by others for CaMV[42]. However, this suggestion conflicts with the scanning hypothesis of Kozak[43] for translation in eucaryotes in which only AUGs of 5'-proximal cistrons serve as initiator codons. However, there is now good evidence with mammalian cells that reinitiation occurs if translation is terminated in the

vicinity of internal AUGs[44,45]. In fact, translation of caulimoviral genomic length RNA may be one of numerous exceptions to the 5' proximal rule.

Another important difference in the initiation of translation between plants and certain other eucaryotes is that plants probably do not require a particular context of sequence surrounding the initiation codon. A recent compilation of the sequences around the initiation codon of approximately 50 plant genes has revealed that no strongly favored consensus exists[46].

The presence of short overlaps between the stop of one translational reading region and the beginning of the next may not be an obstacle for efficient internal initiation. An examination of the translation of downstream coding regions of artificially constructed multicistronic transcription units in an SV40 gene vector in mammalian cells has revealed that reach back reinitiation occurs with good efficiency[44]. However, the efficiency varies with the distance between the terminator and initiator AUG. This suggests that similar reach back reinitiation may account for translation of FMV ORFs II, IV and V which in each case overlap the 3' end of their upstream cistrons. The initiator AUG of ORF IV, for example, lies 18 nucleotides upstream of the ORF III termination codon. ORF V overlaps ORF IV by 22 nucleotides.

Similarities in the FMV and CaMV genomes extend to the putative primer binding sites of the two viral genomes. A sequence complementary to the 3' end of $tRNA_i^{met}$ occurs in the FMV sequence near or at the single interruption in the minus DNA strand. This suggests an RNA priming event for the initiation of DNA replication like that believed to occur during CaMV DNA replication[16,33,34]. The occurrence of this RNA priming site plus the presence of gene V which has much resemblance to known reverse transcriptases[26], provides a strong indication that FMV also replicates by reverse transcription.

The highly conserved array of cysteine residues with the $CysX_2CysX_4HisX_4Cys$ motif (Cys-motif) that is present in the nucleic acid binding portion of the gag gene product of retroviruses[25] is also observed in the putative gene products of ORF IV in FMV, CaMV[3,4] and CERV[17]. The Cys-motif seems to be a general characteristic in coat protein (precursors) of viruses that are conjectured to use $tRNA_i^{met}$ for the initiation of reverse transcription[25]. This suggests a dual function of the Cys-motif in encapsidation of genomic RNA and initiation of reverse transcription in immature virions of caulimoviruses[47]. It is notable, for example, that the Cys-motif resembles the metal binding regions in nucleic acid binding

"finger" proteins[48,49,50], several of which appear to be important in regulation of transcription[51]. Perhaps the Cys-motif has a related function in replication and encapsidation of caulimovirus DNA.

Only a small segment of the FMV ORF VI translation product showed a high degree of homology with the gene VI product of CaMV or CERV (Fig. 4). Except for a short tract of similar amino acid sequence near the N-terminal and a more extensive region near the middle of the gene, there was a very low level of homology in these proteins. Conservation of such a short segment of the amino acid sequence of this gene among caulimoviruses suggests that this region of the protein may be responsible for its molecular function. The more variable regions to either side of this conserved region may interact with the host cell in some manner to establish a compatible host-virus relationship. Gene VI of CaMV has been shown to specify the host reaction during infection and to contain a major host range determinant[52].

Our alignment of the large intergenic sequences of FMV with the other two sequenced caulimoviral DNAs was made by a subjective choice of sequence which best corresponded to the TATA box of CaMV. Other TATA-like sequences are observed in this region including the sequence TCTATAAAATA which starts at nucleotide 7116. The chosen alignment, shown graphically in Fig. 5, revealed a conserved 35 bp intergenic sequence in a nearby downstream region. An almost exact sequence occurs in the same position in CaMV and CERV DNA. The conserved 35 bp sequence in the large intergenic region of the caulimoviruses occurs downstream of the termination site of the 35S and 19S RNAs of CaMV (at nucleotide 7615). Since downstream sequences are known to be involved in transcriptional termination in eucaryotes[32], this suggests a possible function for the 35 bp sequence.

ACKNOWLEDGEMENTS

The authors wish to thank Susan Ballard for preparing many of the sequencing templates and Roger Hull for supplying the sequence of carnation etched ring virus DNA before its publication. We would also like to express our appreciation to Dr. Les Domier for his assistance in doing the computer assisted amino acid alignments. This research was supported by the U.S. Department of Agriculture CGO grants 85-CRCR-1-1743 and 86-CRCR-1-2261. The authors are also grateful for the support and facilities provided by Kentucky State THRI project no. 5-41086.

REFERENCES

1. Shepherd, R. J., Wakeman, R. J. and Romanko, R. (1968). Virology 36, 150-152.
2. Harrison, B. D., Finch, J. T., Gibbs, A. J., Hollings, M., Shepherd, R. J., Valenta, V. and Wetter, C. (1971). Virology 45, 356-363.
3. Franck, A., Guilley, H., Jonard, G., Richards, K. and Hirth, L. (1980). Cell 21, 285-294.
4. Gardner, R. C., Howarth, A. J., Hahn, P., Brown-Luedi M., Shepherd, R. J. and Messing, J. (1980). Nucleic Acids Research 9, 2871-2888.
5. Balazs, E., Guilley, H., Jonard, G. and Richards, K. (1982). Gene 19, 239-249.
6. Howell, S. H., Walker, L. L. and Dudley, R. K. (1980). Science 208, 1265-1267.
7. Lebeurier, G., Hirth, L., Hohn, T. and Hohn, B. (1980). Gene 12, 139-146.
8. Shepherd, R. J. (1979). Ann. Rev. Pl. Physiol. 30, 405-423.
9. Hull, R. and Donson, J. (1982). Journal of General Virology 60, 125-134.
10. Richins, R. D. and Shepherd, R. J. (1983). Virology 124, 208-214.
11. Shepherd, R. J., Richins R. D., Duffus, J. E. and Handley, M. K. (1987). Phytopathology (in press).
12. Maniatis, R., Fritsch, E. F. and Sambrook, J. (1982). "Molecular Cloning." Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
13. Dale, R. M. K., McClure, B. A. and Houchins, J. P. (1985). Plasmid 13, 31-40.
14. Lipmann, D. J. and Pearson, W. R. (1985). Science 227, 1435-1441.
15. Staden, R. (1982). Nucleic Acids Research 10, 2951-2961.
16. Pfeiffer, P. and Hohn, T. (1983). Cell 33, 781-789.
17. Hull, R., Sadler, J. and Longstaff, M. (1986). EMBO Jour. 5, 3083-3090.
18. Hohn, T., Richards, K. and Lebeurier, G. (1982). Curr. Topics in Microbiol. Immunol. 96, 194-236.
19. Young, M. J., Daubert, S. and Shepherd, R. J. (1987). Virology 158, 444-446.
20. Harker, C. L., Mullineaux, P. M., Bryant, J. A. and Maule, A. J. (1987). Plant Mol. Biol. 8, 275-287.
21. Armour, S. L., Melcher, U., Pirone, T. P., Lyttle, D. J. and Essenberg, R. C. (1983). Virology 129, 25-30.
22. Daubert, S., Shepherd, R. J. and Gardner, R. (1983). Gene 25, 201-208.
23. Woolsten, C., Covey, S., Penswick, J. and Davies, J. (1983). Gene 23, 15-23.
24. Xiong, C., Lebeurier, G. and Hirth, L. (1984). Proc. Natl. Acad. Sci. (USA) 81, 6608-6612.
25. Covey, S. N. (1986). Nucleic Acids Res. 14, 623-632.
26. Toh, H., Kikuno, R., Hayashida, H., Miyata, T. Kugimiya, W., Inouye, S., Yaki, S. and Saigo, K. (1985). EMBO Jour. 4, 1267-1272.
27. Dickson, C., Eisenman, R. and Fan, H., Hunter, E. and Teich, N. (1985). In RNA Tumor Viruses 2nd Edition. Weiss, R., Teich, N., Varmus, H. and Coffin, J. (editors). Cold Spring Harbor Laboratory, N.Y. pp. 513-648.
28. An, G., Ebert, P. R., Yi, B-Y and Choi, C-H. (1986). Mol. Gen. Genet. 203, 245-250.
29. Messing, J., Geraghty, D., Heidecker, G., Hu, N-T., Kridl, J. and Rubenstein, I. (1983). In Genetic Engineering of Plants. T. Kosuge, C. Meredith and A. Hollaender, editors. Plenum Press, N.Y. pp. 211-227.
30. Covey, S. N., Lomonossoff, G. P. and Hull, R. (1981). Nucleic Acids Res. 9, 6735-6747.

31. Guilley, H., Dudley, R. K., Jonard, G., Balazs, E. and Richards, K. E. (1982). Cell 30, 763-773.
32. Hunt, A. G., Chu, N. M., Odell, J. T., Nagy, F. and Chua, N-H. (1987). Plant Mol. Biol. 8, 23-35.
33. Guilley, H., Richards, K. E. and Jonard, G. (1983). EMBO Jour. 2, 277-282.
34. Hull, R. and Covey, S. N. (1983). Sci. Progr. Oxf. 68, 403-422.
35. Canaday, J., Guillemaut, P. and Weil, J.-H. (1980). Nucleic Acids Res. 8, 999-1008.
36. Ghosh, H. P., Ghosh, K., Simsek, M. and Raybhandary, U. L. (1982). Nucleic Acids Res. 10, 3241-3247.
37. Turner, D. S. and Covey, S. N. (1984). FEBS Letters 165, 285-289.
38. Howell, S. H. and Hull, R. (1978). Virology 86, 468-481.
39. Hull, R., Covey, S. N., Stanley, J. and Davies, J. W. (1979). Nucleic Acids Res. 7, 669-677.
40. Guilfoyle, R. J. (1980). Virology 107, 71-80.
41. Plant, A. L., Covey, S. N. and Grierson, D. (1985). Nucleic Acids Res. 13, 8305-8321.
42. Dixon, L. and Hohn, T. (1984). EMBO Jour. 3, 2731-2736.
43. Kozak, M. (1978). Cell 15, 1109-1123.
44. Peabody, D. S. and Berg, P. (1986). Mol. Cell. Biol. 6, 2695-2703.
45. Peabody, D. S., Subramani, S. and Berg, P. (1986). Mol. Cell. Biol. 6, 2704-2711.
46. Heidecker, G. and Messing, J. (1986). Ann. Rev. Pl. Physiol. 37, 439-466.
47. Marsh, L., Kuzj, A. and Guilfoyle, T. (1985). Virology 143, 212-223.
48. Miller, J., McLachlan, A. D. and Klug, A. (1985). EMBO Jour. 1609-1614.
49. Vincent, A. (1986). Nucl. Acids Res. 14, 4385-4391.
50. Berg, P. M. (1986). Science 232, 484-487.
51. Maniatis, T., Goodbourn, S. and Fischer, J. A. (1987). Science 236, 1237-1245.
52. Schoelz, J. E., Shepherd, R. J. and Daubert, S. (1986). Mol. and Cell. Biol. 6, 2632-2637.
53. Verver, J., Schijns, P., Hibi, T. and Goldbach, R. (1987). J. Gen. Virol. 68, 159-167.