Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

ELSEVIER

Virus Research 33 (1994) 145–156

Virus Research

Sequence analysis of the 3′ termini of RNA1 and RNA2 of blueberry leaf mottle virus

J.W. Bacher [a,*], D. Warkentin [a,b], D. Ramsdell [b], J.F. Hancock [a]

[a] *Department of Horticulture, Michigan State University, East Lansing, MI 48824, USA*
[b] *Department of Botany and Plant Pathology, Michigan State University, East Lansing, MI 48824, USA*

(Received 20 September 1993; revision received 10 March 1994; accepted 10 March 1994)

Abstract

The 3′ termini of RNA1 and RNA2 of blueberry leaf mottle virus (BBLMV) were cloned and the cDNA sequence of a portion of the putative polymerase gene, the complete coat protein (CP) gene, and the 3′ non-coding regions was determined. The N terminus of the coat protein gene was precisely located by comparison with the amino acid sequence determined by the Edman degradation sequencing of the purified coat protein. The coat protein gene encoded a polypeptide of 521 amino acids with a predicted M_r of 57,542. Homology to BBLMV coat protein was highest with tomato ringspot virus (TomRSV) and cherry leaf roll virus (CLRV); two other nepoviruses also belonging to a sub-group defined by the presence of large RNA2 components. The 3′ terminal 1390 nt of RNA1 and RNA2 were nearly identical and apparently non-coding. No statistically significant sequence homology was found between the 3′ non-coding regions of BBLMV and other nepoviruses. A highly conserved 3′ non-coding region of this length is unusual, but has been reported for two other related viruses, TomRSV and CLRV. The biological function of the long 3′ non-coding region and how the high level of sequence homology is maintained between RNA1 and RNA2, is unknown. Possible mechanisms for conservation of the 3′ terminus are discussed.

Key words: BBLMV; Nepovirus; 3′ Non-coding region; Coat protein; Recombination

1. Introduction

A high degree of sequence homology is usually found within the 3′ terminal region of segmented viruses (Matthews, 1991). The 3′ sequences of related viruses

* Corresponding author. Fax: +1 (608) 262-4743.

SSDI 0168-1702(94)00032-8

will often show a higher degree of sequence homology than the rest of their genomes (Meshi et al., 1983; Bouzoubaa et al., 1986). This conservation of the 3′ terminal region is well illustrated by the nepovirus group, in which nearly identical nucleotide sequences have been found between the 3′ non-coding regions of the RNA1 and RNA2 components of tomato black ring virus (TBRV) (Grief et al., 1986, 1988), grapevine chrome mosaic virus (GCMV) (Brault et al., 1989; Le Gall et al., 1989), tomato ring spot virus (TomRSV) (Rott et al., 1991a,b) and cherry leaf roll virus (CLRV) (Scott et al., 1992). In the case of CLRV and TomRSV, the homologous 3′ non-coding region represents approximately 20% of their genomes, spanning 1.4 and 1.5 kb, respectively.

The mechanism involved in maintaining this high level of sequence homology in the 3′ terminus is not clear, but this marked conservation must reflect some important functions that are selectively maintained (Matthews, 1991). A strong selection pressure for conservation of specific 3′ sequences must be present, otherwise the high mutation rates (10^{-3} to 10^{-5} per nucleotide per round of replication) present in RNA virus replication would lead to divergence between the 3′ termini (Steinhauer and Holland, 1987; Domingo and Holland, 1988).

Blueberry leaf mottle virus (BBLMV) is a member of the nepovirus group. Nepoviruses have bipartite genomes composed of positive sense ssRNA molecules which are encapsidated in 28 nm isometric particles (Harrison and Murant, 1977) and are translated as a single polyproteins which are later cleaved into functional proteins (Demangeat et al., 1990, 1991). BBLMV is classified in a subgroup with the other nepoviruses possessing large ($M_r > 1.6 \times 10^6$) RNA2 components (e.g., peach rosette mosaic virus, CLRV and TomRSV; Martelli, 1975). The relatively larger size of the RNA2 components of CLRV and TomRSV is partly due to the length of their 3′ non-coding regions. Whether BBLMV possesses a large region of homology between it's 3′ termini of RNA1 and RNA2 had not been determined previously. In order to better understand the extent, and ultimately the mechanism(s), of conservation of the 3′ termini in nepoviruses, we have cloned the 3′ termini of RNA1 and RNA2 of BBLMV and sequenced a portion of the putative polymerase gene, the complete coat protein gene, and the 3′ non-coding regions. Relationships between BBLMV and other nepoviruses were estimated by comparisons of sequence homologies and possible mechanisms for maintaining identity between RNA1 and RNA2 in nepoviruses are discussed.

2. Materials and methods

2.1. Virus purification and nucleic acid extraction

The BBLMV used in this study was originally isolated from highbush blueberry plants (*Vaccinium corymbosum* L. cv. Rubel) and subsequently propagated in *Chenopodium quinoa*. The methods of virus propagation, purification and RNA extraction were as described by Ramsdell and Stace-Smith (1981).

2.2. cDNA cloning and nucleotide sequencing

All standard recombinant DNA techniques were performed according to Sambrook et al. (1991) unless indicated otherwise. Total viral RNA was extracted from purified virus particles and fractionated from an ultracentrifuged linear log sucrose density gradient made with RNase-free sucrose in 1 × SSC buffer, pH 7, containing 6 μg/ml purified bentonite according to the methods of Fraenkel-Conrat et al. (1961). Ethanol precipitated RNA1 and RNA2 mixtures were resuspended in TE (Tris 10 mM, EDTA 1 mM, pH 8.0) and used as templates for oligo(dT) primed cDNA synthesis essentially as described by Gubler and Hoffman (1983). Size fractionated cDNA was blunt end ligated into SmaI cut Bluescript KS + vector (Stratagene), then used to transform XL1-Blue *Escherichia coli* competent cells (Stratagene). Clones were separated into two classes (presumably RNA1 or RNA2) based on restriction enzyme analysis, and the largest representative from each class, clone 24 (3.1 Kb) and clone 34 (3.3 Kb), were selected for sequencing. The viral origin of the clones was tested by probing a northern blot of purified BBLMV RNA with ^{32}P-labeled cDNA clones. Subclones used to determine the cDNA sequence were obtained by Exonuclease III generated nested deletions and by restriction enzyme digestion. Both strands of the 3′ portion of clone 24 and the entire length of clone 34 were sequenced by the dideoxynucleotide chain termination method (Sanger, 1981) using single stranded templates prepared according to Viera and Messing (1987).

2.3. N terminal amino acid sequencing of the viral coat protein

To identify which clone encoded the coat protein gene, the N terminus of the viral coat protein subunit was sequenced. The intact virions were denatured by combining equal volumes of the purified virus (10 μl of a 1-mg/ml suspension) and dissociation buffer [1% (v/v) 2-mercaptoethanol, 4 M Urea, and 1% (w/v) SDS in 0.1 M sodium phosphate buffer, pH 7.2] in a microcentrifuge tube and heating the mixture in a 100°C water bath for 90 s (Ramsdell and Stace-Smith, 1981). The coat protein was purified on a 12% SDS-PAGE gel, stained with Coomassie blue, electroblotted onto a polyvinylidene fluoride membrane (Scheicher and Schuell) and directly sequenced using an Applied Biosystems model 477A protein sequencer at the Macromolecular Structural Facility at Michigan State University, East Lansing, Michigan (Hunkapiller et al., 1983).

2.4. Nucleotide and amino acid sequence analysis

Sequence data were analyzed using the Genetics Computer Group Sequence Analysis Software Package for the VAX, Version 7.1 (Devereux et al., 1984). Alignments of the amino acid and nucleotide sequences were obtained using the GAP, BESTFIT, PILEUP and PRETTY algorithms. The statistical significance of

```
1     CCGAATATGACTGGACGTGCATCTATTCCTGTTCAGACTAACATTAGGAATAGCCCCAGA      60
1      P  N  M  T  G  R  A  S  I  P  V  Q  T  N  I  R  N  S  P  R       20

61    ATTGTAGATGGGGAAGAAATAACGCCTCCTCGGTTTACAACGTGCAATAGTGGTCTGATA     120
21     I  V  D  G  E  E  I  T  P  P  R  F  T  T  C  N /S  G  L  I       40

121   GCAGACACTAGTATAGCGCATGTTGTGCAGGGGTGGGTGCCCAAAGACGCAACAAAAGGC     180
41     A  D  T  S  I  A  H  V  V  Q  G  W  V  P  K  D  A  T  K  G       60

181   CGTGTTTTAGAGGCCATTAATTTGCGTGAAGATATCGCTACTAGTGATAATTTAGTTAAA     240
61     R  V  L  E  A  I  N  L  R  E  D  I  A  T  S  D  N  L  V  K       80

241   TATGAGTGGCTTGCCAAAGGACTCATACATCCAGATTTGAAGTTGCGTATGACTGTCGGC     300
81     Y  E  W  L  A  K  G  L  I  H  P  D  L  K  L  R  M  T  V  G      100

301   CAAAATCCCTTTGTTGGGATATCGATAGGCATATGCTGCGATTACTTTGGGCGCCTTAGT     360
101    Q  N  P  F  V  G  I  S  I  G  I  C  C  D  Y  F  G  R  L  S      120

361   AAATATTATGAGGGCGACACTGCTTTACCTATAGAGGTATGCAATCAATTGCCCAATTTT     420
121    K  Y  Y  E  G  D  T  A  L  P  I  E  V  C  N  Q  L  P  N  F      140

421   GTTTGCCCAATATCGGAGAAGAGTGTCTTCGAGTTTGATTTAGATATGTCCCTCGCGGGA     480
141    V  C  P  I  S  E  K  S  V  F  E  F  D  L  D  M  S  L  A  G      160

481   TATAACCTTTTTCAAACTTCTAAAGGCTTCGCTGATCCGGTATTATTAGTGTACATAATA     540
161    Y  N  L  F  Q  T  S  K  G  F  A  D  P  V  L  L  V  Y  I  I      180

541   GATACTAATTCTTTACCCGCCAGTGATGAATGGGTTTACACATGTGAGGTTTGTATAAAA     600
181    D  T  N  S  L  P  A  S  D  E  W  V  Y  T  C  E  V  C  I  K      200

601   TCTGCCTTGCATGCCACTTCTGTGGCAAATAAACCCATTCTATCGCTACCACATTTTTTT     660
201    S  A  L  H  A  T  S  V  A  N  K  P  I  L  S  L  P  H  F  F      220

661   GACGGTCGTCTCCCACTTGACTTGTGGAGGGGACCTTTTTCTTTTGAGTTAGGTAGAAGT     720
221    D  G  R  L  P  L  D  L  W  R  G  P  F  S  F  E  L  G  R  S      240

721   TCCAAAAGGGAGAATCACATCGGCATCAATTTTGGTAGTGCTCGTGTTGTCTCTGGGACC     780
241    S  K  R  E  N  H  I  G  I  N  F  G  S  A  R  V  V  S  G  T      260

781   AATACCTTTTATTCTTTTCCTGCTGCCTATACTCAGCTTTTACAGAGTGTAGGTGGTATT     840
261    N  T  F  Y  S  F  P  A  A  Y  T  Q  L  L  Q  S  V  G  G  I      280

841   TTACATGGTACTGTCGTTCAAACTGGCAGTAAGGCTATATCTTGTGAGATGTTTCTTATC     900
281    L  H  G  T  V  V  Q  T  G  S  K  A  I  S  C  E  M  F  L  I      300

901   CTTCAACCGGATAAGACCGCCCACAATTTAGAGCAGGCTCTCCGCCTTCCTGGTTGTCGT     960
301    L  Q  P  D  K  T  A  H  N  L  E  Q  A  L  R  L  P  G  C  R      320

961   ATACCAACTGGGGGTGGACCATTTTCTATTCGTATACAGACTCCCTTCCAGCGAGAGCAA    1020
321    I  P  T  G  G  G  P  F  S  I  R  I  Q  T  P  F  Q  R  E  Q      340

1021  ATTTTTAATACCGGCGTTCAGCTGGTAATCTATGCTGTTGGGGGTCCTATGGGAGCACAA    1080
341    I  F  N  T  G  V  Q  L  V  I  Y  A  V  G  G  P  M  G  A  Q      360

1081  GCTATATCTGCACCATATCAATATATGGTGCATTTCTCCCATATACAAGAAGAGGGGGAT    1140
361    A  I  S  A  P  Y  Q  Y  M  V  H  F  S  H  I  Q  E  E  G  D      380

1141  CCTCCGCCTCGTCCTATTGGCAATGTGTTGCTCTTTAATTGGGCCACCATTTCTGAAATG    1200
381    P  P  P  R  P  I  G  N  V  L  L  F  N  W  A  T  I  S  E  M      400

1201  ACGAATCTTACCCGGTTTCAGATTCCGGCGCGATTAAGTGATCTCGTGTTGCCAGGTCAA    1260
401    T  N  L  T  R  F  Q  I  P  A  R  L  S  D  L  V  L  P  G  Q      420
```

```
1261  ACTGTCACCATGAGGCGAAATGCTTTAGCGAATCTGATAAGGTCTTGTGGGTTCTTCCGG  1320
421    T  V  T  M  R  R  N  A  L  A  N  L  I  R  S  C  G  F  F  R   440

1321  GGCCGTGTTACATTTGTGTTCCAATGGACATTGAATGTAGCACATATTGTACCAACTGCT  1380
441    G  R  V  T  F  V  F  Q  W  T  L  N  V  A  H  I  V  P  T  A   460

1381  ACAATGCAAATTTTAACGGCAGTTGGGCGCGTTGGCAATGCGGAGACTAATGGTTCACAA  1440
461    T  M  Q  I  L  T  A  V  G  R  V  G  N  A  E  T  N  G  S  Q   480

1441  ATCCTACAAAGTTGGATTGTGCCCGTTAGTCAGGTCTTTGAGAAAGAGGTTGAAATGGAT  1500
481    I  L  Q  S  W  I  V  P  V  S  Q  V  F  E  K  E  V  E  M  D   500

1501  CTCACCGATTATCCCGGTTTTAATACATCTGGGGGAATTGGTGCTGACCATGATCAGCCT  1560
501    L  T  D  Y  P  G  F  N  T  S  G  G  I  G  A  D  H  D  Q  P   520

1561  TACATTGACATTGCTTGTGGTAATTTTCCACAAATATTCTATATGAATATCAATGTGCGT  1620
521    Y  I  D  I  A  C  G  N  F  P  Q  I  F  Y  M  N  I  N  V  R   540

1621  GTACACCCCGGGTTTGAGCTCTATGGTAGGAGTATCACACCCCTACGCATTTAGTATGCT  1680
541    V  H  P  G  F  E  L  Y  G  R  S  I  T  P  L  R  I  *

1681  AAGTGTGTTTATAGGAACTAGAACCCTTAAGTTCTAGGAGTTGGTCTGTCCTCTCTGACA  1740
1741  GGCCTTCAAAGGATAGAGAATAGCTCGAACTCTCTGTAATACGAGAGGTCCGGACCTGTA  1800
1801  GGTCTTCCTGGCATATACCCAGGTTTTGAGATAGTAGTAAACTACTCTTCGATGTAGCGA  1860
1861  ATCGTCGTAAATAGGACACCCTCCTAAAACGAAGCCTTAAATAGGAACTTGAAAAAGTTT  1920
1921  CCTTTCCACTTTGTGGAGGATAGTATAAGGGACGGTGGTGCCAGCTTGATGACTGCTTAA  1980
1981  GAGCAGGAGGTTGCTCGTTAACCTTACACGAGCTAGGACGTTCTAGTAGAGATGAGACAT  2040
2041  CTACCTCGAAAAACGTCAGAATTACTATATGATTCAAAAGCGTGGTTTTTCCAACGTTAA  2100
2101  CCAATGGAAACCAGGTGCACATAGGTTAGTTGTGCTGATTTGCTACCTTTTAAGAAAGGA  2160
2161  GATTATTCTGGTGAAATTCCAGATCTATCTTAGTTTTGTTGTTTCAGTTTGATTGCAATA  2220
2221  AACCCACATAAACTGTCGTCATTAGGACGGCATACCATTGAGCTCTTTAGGGCGCCTCTG  2280
2281  GTTCCGTGAAATCGGTATACGTGTGAAGATTAGGGTTTGCTCGAACCATAGAGAGCTAGG  2340
2341  TTGTTGGAGCCGAACTGAGTCCAACCGCATTTGTCAGTTTTAGATATAACTGTCCAAGGT  2400
2401  CTACTGCTTCCGAGCCTGAAAAAATCTTAAAGCGCCCAGGCGTCCGTGACTTCACGGCAC  2460
2461  TCGGGGACAGAGTTTAGGGAAACTCTAGAAAAATTCCCTCGCCTTTTAGTTGTGTGGCCG  2520
2521  TGATGGACACAACTCTCTCTTCTTTCTGAGAGTGTACCGCTGTTTTAGTATCTGGTGATG  2580
2581  ATGTAGTTTTGAAACTACCAGAGATGTCTCAGTGGAGAAGCGTCTTGCCAAACGATATTG  2640
2641  GCTTAAGGTCTATGTGACGATAATTTGCTAGTGTACTCTAGAGAATGTGGGGTGGCACCC  2700
2701  ACTTCTTGGATGAGGGCCGGAGATGAAAACCGGGGAGTAATAAACTCCAGCTAGCGGCAT  2760
2761  AGGCCGACCACCGTGAGGGAGCTCACGGCGCAATTTGGACCATTTTTAGACATAAATGGC  2820
2821  CATGTTAGTGTAGCGCTTTGCGCATGTTGAATGATAATGAACCATGCGTTGCAGCGCATG  2880
2881  CCTTTCGAGATCGGATGTGATTACCGTGAGAAAGGGGAAACAATGCCAACATGTTCAATT  2940
2941  CGTTGTACTATGTTTTCTTTCTTTTTGTAGACTCCTGTGAGGATTATCCAACAGCAGGTT  3000
3001  GTGCCTTCAGTAAGCACACAAAAAGATTTCGCATTTTTCTTTGTGTTAGATAGTTTTATA  3060
3061  TCTATAATGTCTTTATTTCAC - poly(A)...
```

Fig. 1. The cDNA sequence and the predicted amino acid sequence of the 3′ terminus of BBLMV RNA2. The N terminus (underlined) and proteolytic cleavage site (N/S) of the coat protein gene were determined by Edman degradation sequencing of the purified coat protein. The putative 3′ non-coding region of RNA2 is 1410 nucleotides and nearly identical to the 3′ non-coding region of RNA1.

the sequence comparisons was determined by comparing percent similarity of each pair of sequences against that expected by chance (using GAP/RAN algorithm). For each pair of sequences, 25 randomized comparisons were made by repeatedly shuffling one of the nucleotide sequences, then aligning it with the other non-randomized sequence. The similarity was judged to be significant if it exceeded the mean randomized similarity plus three standard deviations (Doolittle, 1981). Viral sequences were obtained from GenBank or EMBL databases.

3. Results and discussion

The nucleotide sequences of the 3′ portion of BBLMV RNA1 and RNA2 were determined (Figs. 1 and 2). Both sequences contained a single long ORF, nearly identical 1.4 kb 3′ non-coding regions and poly(A) sequences on their 3′ termini.

```
1     ATGCTTGATATTCCGGTGGAGAAAACTAAGCTCACCAGGGCAATGGCAAACCAGGGTGAG     60
1      M  L  D  I  P  V  E  K  T  K  L  T  R  A  M  A  N  Q  G  E     20

61    GCAGATTATAAATGCCAGGTGGCTGAACGCATTTTTGTGTGCGGCCCCAAAAGCAATTTG    120
21     A  D  Y  K  C  Q  V  A  E  R  I  F  V  C  G  P  K  S  N  L     40

121   AGCCAGGAGCCAGCTCATTTTGTCGTTTCTCACACTGGTTCTCTTAAAAGAGGTGAAACC    180
41     S  Q  E  P  A  H  F  V  V  S  H  T  G  S  L  K  R  G  E  T     60

181   GGTATTGTTGCTCCTGTCGACTTTGTTAGTGAAGGCCAGGGTCGGCTGCCAACCCAGTTG    240
61     G  I  V  A  P  V  D  F  V  S  E  G  Q  G  R  L  P  T  Q  L     80

241   TGGGTTAAAAAATTTCGCTCTGAATCACACACATTCAGGGTCATGATTAACGATGCCTAC    300
81     W  V  K  K  F  R  S  E  S  H  T  F  R  V  M  I  N  D  A  Y    100

301   ACAAGGGGACATTCCATTTATTTTAGGAGTGATCCCCCTTACATTACTAATTGGCTTAGT    360
101    T  R  G  H  S  I  Y  F  R  S  D  P  P  Y  I  T  N  W  L  S    120

361   GCCACCTCCTTTGCTCTTGGCAAGGGGATGGACTATAAAGCCATTTTGGGGTTGTACCAT    420
121    A  T  S  F  A  L  G  K  G  M  D  Y  K  A  I  L  G  L  Y  H    140

421   AACGTGTGCACTCCAGACGCGCAATGTTTAGATGAATATTTTGTTAGTGCGCGTTTTAGA    480
141    N  V  C  T  P  D  A  Q  C  L  D  E  Y  F  V  S  A  R  F  R    160

481   AGAGCTGGATGCCTATCGGCCTCCACATATCACCAGTAGGAACTAGAACCTTTAAGTTCT    540
161    R  A  G  C  L  S  A  S  T  Y  H  Q  *

541   AGGAGTTGGTCTGTCCTCTCTGACAGGCCTTCAAAGGATAGAGAATAGCTCGAACTCTCT    600
601   GTAATACGAGAGGTCCGGACCTGTATGTCTTCCTGGCATATACCCAGGTTTTGAGATAGT    660
661   AGTAAACTACTCTTCGATGTAGCGAATCGTCGTAAATAGGACACCCTCCTAAAACGAAGC    720
721   CTTAAATAGGAACTTGAAAAAGTTTCCTTTCCACTTTGTGGAGGATAGTATAAGGGACGG    780
781   TGGTGCCAGCTTGATGACTGCTTAAGAGCAGGAGGTTGCTCGTTAACCTTACACGAGCTA    840
841   GGACGTTCTAGTAGAGATGAGACATCTACCTCGAAAAACGTCAGAATCACTATATGATTC    900
901   AAAAGCGTGGTTTTTCCAACGTTAACCAATGGAAACCAGGTGCACATAGGTTAGTTGTGC    960
961   TGATTTGCTACCTTTTAAGAAAGGAGATTATTCTGGTGAAATTCCAGATCTATCTTAGTT   1020
1021  TTGTTGTTTTAGTTTGATTGCAATAAACCCACATAAACTGTCGTCATTAGGACGGTATAC   1080
1081  CATTGAGCTCTTTAGGGCGCCTCTGGTTCCGTGAAATCGGTATACGTGTGAAGATTAGGG   1140
1141  TTTGCTCGAACCATAGAGAGCTAGGTTGTTGGAGCCGAACTGAGTCCAACCGCATTTGTC   1200
1201  AGTTTTAGATATAACTGTCCAAGGTCTACTGCTTCCGAGCCTGAAAAAAATCTTAAAGCG   1260
1261  CCCAGGCGTCCGTGACTTCACGGCACTCGGGGACAGAGTTTAGGGAAACTCTAGAAAAAT   1320
1321  TCCCTCGCCTTTTAGTTGTGTGGCCGTGATGGACACAACTCTCTCTTCTTTCTGAGAGTG   1380
1381  TACCGCTGTTTTAGTATCTGGTGATGATGTAGTTTTGAAACTACCAGAGATGTCTCAGTG   1440
1441  GAGAAGCGTCTTGCCAAACGATATTGGCTTAAGGTCTATGTGACGATAATTTGCTAGTGT   1500
1501  ACTCTAGAGAATGTGGGGTGGCACCCACTTCTTGGATGAGGGCCGGAGATGAAAACCGGG   1560
1561  GAGTAATAAACTCCAGCTAGCGGCATAGGCCAACCACCGTGAGGGAGCTCACGGCGCAAT   1620
1621  TTGGACCATTTTTAGACATAAATGGCCATGTTAGTGTAGCGCTTTGCGCATGTTGAATGA   1680
1681  TAATGAACCATGCGTTGCAGCGCATGCCTTTCGAGATCGGATGTGATTACCGTGAGAAAG   1740
1741  GGGAAACAATGCCAACATGTTCAATTCGTTGTACTATGTTTTCTTTCTTTTTGTAGACTC   1800
1801  CTGTGAGGATTATCCAATAGCAGGTTGTGCCTTCAGTAAGCACACAAAAAGATTTCGCAT   1860
1861  TTTTCTTTGTGTTAGATAGTTTTATATCTATAATGTCTTTATTTCAC - poly(A)...
```

Fig. 2. The cDNA sequence and the predicted amino acid sequence of the 3′ terminus of BBLMV RNA1. The region of identity between RNA1 and RNA2 begins with the TAG termination codon (*) of the RNA1 encoded ORF. The seven nucleotides that differed between the 3′ non-coding regions of RNA1 and RNA2 are underlined.

3.1. Identification of coat protein gene

The location of the coat protein gene and the putative protease cleavage site were determined from the amino acid sequencing of the N-terminus of the purified coat protein. The sequence of the N-terminal 15 amino acids of the BBLMV coat protein was found to be S-G-L-I-A-D-T-S-I-A-H-V-V-Q-G by automated Edman degradation sequencing. This amino acid sequence was identical to the predicted amino acid sequence of cDNA clone 34 (Fig. 1). The presence of the coat protein gene on clone 34 indicated that the source of the clone was RNA2, since in all other nepoviruses sequenced to date the coat protein has been found on the 3′ end of the RNA2 component (Bertioli et al., 1991; Blok et al., 1992; Brault et al., 1989; Grief et al., 1986; Meyer et al., 1986; Rott et al., 1991b; Serghini et al., 1990).

The coat protein gene spanned 1562 nucleotides and terminated at a TAG stop codon, followed closely by two additional in-frame stop codons (Fig. 1). The coat protein gene encodes a polypeptide containing 521 amino acids with a predicted M_r of 57,542. This is slightly larger than the estimated M_r of 54,000, previously determined by SDS polyacrylamide gel electrophoresis (Ramsdell and Stace-Smith, 1981). Providing no processing of the N terminus occurs after cleavage, the proteolytic cleavage site of the BBLMV coat protein would be (N/S) at position 108/109. This cleavage site is unique for the nepovirus group; however, there does not seem to be much conservation of cleavage sites among nepoviruses. The known cleavage sites for the release of other nepovirus coat proteins from polyprotein precursors are: K/A for TBRV (Demangeat et al., 1991), R/A for GCMV (Brault et al., 1989), C/A for raspberry ringspot virus (RRV) (Blok et al., 1992) and R/G for arabis mosaic virus (AMV) (Bertioli et al., 1991) and grapevine fanleaf virus (GFLV) (Serghini et al., 1990).

3.2. Identification of polymerase gene

The GDD motif, found in RNA-dependent RNA polymerases (Kamer and Argos, 1984), was located in the 5′ end of clone 24 (Fig. 3). The GDD motif and

```
BBLMV       lTvllNSifn elliRyvwkt 20 livYGDDnli svhpeflpyf
GFLV        lTvvlNSifk ellmRycfkk 20 litYGDDnvf tvaqsvmqyf
GCMV        lTvvmNSifn eiliRyaykt 20 llvYGDDnli svspavaswf
TBRV        lTvvvNSvfn eiliRyaykk 20 llvYGDDnli svspsiaswf
CPMV        mTvivNSifn eiliRyhykk 20 lvtYGDDnli svnavvtpyf
POLIO       gTsifNSmin nliiRtlllk 20 miaYGDDvia syphevdasl
Consensus   -T---NS--- ----R----- 20 ---YGDD--- ----------
```

Fig. 3. Alignment of a portion of the putatitive RNA-dependent RNA polymerase sequence of BBLMV (derived from the 5′ end of clone 24) with poliovirus and RNA1 of GFLV, GCMV, TBRV, and CPMV. Capital letters indicate a consensus between the BBLMV sequence and all other viral sequences. The **GDD** motif is common to RNA-dependent RNA polymerases (Kamer and Argos, 1984).

the surrounding sequences shared homology with RNA1 of GFLV, GCMV, TBRV, cowpea mosaic virus (CPMV), and poliovirus, therefore, the origin of clone 24 was assumed to be RNA1 of BBLMV.

3.3. The 3′ non-coding regions of RNA1 and RNA2

The 3′ terminal sequences of BBLMV RNA1 and RNA2 were compared and found to be nearly identical for 1390 nucleotides, with differences at only seven positions (Figs. 1 and 2). The region of homology began with the first stop codon of the RNA1 ORF and 18 nt downstream of the first stop codon of the RNA2 ORF. There was no significant sequence homology between RNA1 and RNA2 upstream of these stop codons. Downstream of these stop codons, within the putative non-coding regions, were additional stop codons in all three reading frames (37, 26 and 31 in RNA1; and 24, 42 and 28 in RNA2). All ORFs in this 3′ region were less than 141 nucleotides and are believed to be untranslated, since only a single long ORF is thought to be translated from each RNA component in nepoviruses (Matthews, 1991). A 3′ non-coding region of this length is unusual in plant RNA viruses, but has been reported for two other nepoviruses, TomRSV and CLRV (Rott et al., 1991a; Scott et al., 1992).

3.4. Comparison of coat protein and 3′ non-coding sequences with other nepoviruses

Coat protein amino acid sequence comparisons (Table 1) and 3′ non-coding nucleotide sequence comparisons (Table 2) were used to generate dendrograms showing the clustering relationships between BBLMV and other nepoviruses (Fig. 4).

Table 1
Percent amino acid sequence identity between nepovirus coat proteins (all nepovirus sequence comparisons were statistically significant except RRV by CLRV)

	GCMV	AMV	GFLV	RRV	BBLMV	CLRV	TomRSV
TBRV	57	22	23	23	24	24	22
GCMV		23	24	23	23	20	24
AMV			69	23	23	20	25
GFLV				23	22	21	25
RRV					21	21 [a]	24
BBLMV						29	29
CLRV							22
TomRSV							–
TEV [b]	16	15	16	14	16	14	15
Random [c]	16	15	16	16	18	15	18

[a] Not statistically significant.
[b] TEV (tobacco etch virus) was used for a comparison to a nonrelated virus.
[c] Randomized sequence of BBLMV was used for comparison to nonrelated sequences with similar amino acid composition.

Table 2
Sequence comparison of the 3′ non-coding regions showing the percent nucleotide sequence similarity between the different nepoviruses

	GCMV	AMV	GFLV	RRV	BBLMV	CLRV	TomRSV
TBRV [a]	68 *	56 *	60 *	41	42	38	40
GCMV [a]		57 *	58 *	36	43	39	44
AMV [a]			74 *	46	44	47	43
GFLV [a]				49	42	44	41
RRV [b]					43	39	37
BBLMV [c]						37	37
CLRV [c]							39
TomRSV [c]							–
Random [d]	43	37	42	35	37	37	39

* Indicates a statistically significant nucleotide sequence similarity.
[a] Nepoviruses with short (< 300nt) 3′ non-coding regions.
[b] RRV has 397 nt in its 3′ non-coding region.
[c] Nepoviruses with long (> 1389nt) 3′ non-coding regions.
[d] Randomized sequence of BBLMV was used for comparison to nonrelated sequences with similar base composition.

Low, but statistically significant coat protein amino acid sequence homology was found between BBLMV and the other nepoviruses (Table 1). Higher amino acid sequence homology between BBLMV, CLRV and TomRSV (all possessing large

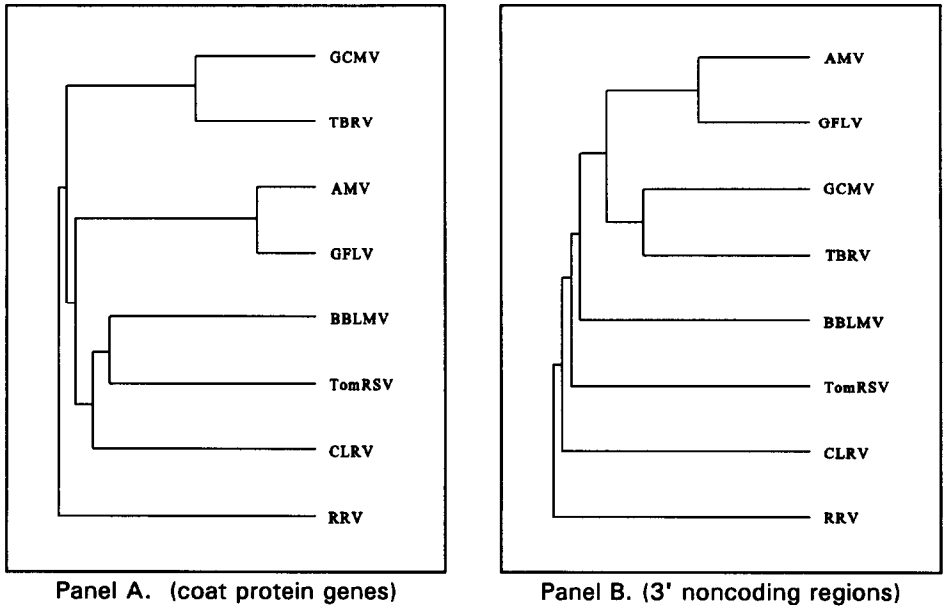

Fig. 4. Dendrograms showing the clustering relationships of different nepoviruses based on amino acid sequence comparisons of coat proteins (panel A) or nucleotide sequence comparisons of 3′ non-coding regions (panel B) Generated by Genetics Computer Group program PILEUP.

```
BBLMV GGaTTaTCCa acagCaggTt gTGCcTTcag taaGcACACA AAAAGATTTc gcaTTTTTct
RRV   GtTgctcCCT cTaagaggTc gTGCcTTtag caaGcACACA AAAAtATgca tTtgTTTTTg
CLRV  .GTgTtTCaa aattCgctTa tTGtaTgagt GTcGGACtCA ggcAGtgTTT aggTTTTaTt
TRSV  aGggTtTtgT TggtCCgtTt gTGttTcaaa acgctgCttt gcAAttTTcT tTtTTgTTTt
AMV   tGTTTgTCCT TTggaCacac tTGCcTT... GTtGGACgCA AAAAGATTTT aTtTTcTTTt
GFLV  tGTTTgTCCT TTggaCacac tTGCcTa... GTtGGACgCA AAAAGATTTT tccTTTcTTt
GCMV  GGTTTgTCCT TTtctCatgt tTGCtTT... GTtGGACACA AAAAGATTTT aTaTTTcTTa
TBRV  GGTTTgTCCT TTtcCCtgTg gTGCtaT... GTtGGACACA AAAAGATTTT cTcTTTTgTa
CPMV  GagcTcctgT TTagCaggTc gTcCcTTcag caaGGACACA AAAAGATTTT aatTTTaTT.
Con.  GGTTT-TCCT TT--CC--T- -TGC-TT--- GT-GGACACA AAAAGATTTT -T-TTTTT-
```

Fig. 5. Multiple sequence alignment (using Genetics Computer Group PRETTY program) showing a conserved region found in the 3′ terminal non-coding regions of nine nepoviruses and cowpea mosaic virus (CPMV). BBLMV sequences correspond to positions 2981 to 3041 of RNA2 (Fig. 1). A consensus sequence (Con) is given when a plurality of 5 out of 9 sequences exists.

RNA2 components with long 3′ non-coding regions), resulted in their clustering in the dendrogram generated by GCC PILEUP program (Fig. 4). This clustering supports Martelli's (1975) earlier grouping of nepoviruses with large RNA2 components. However, the nepoviruses with larger RNA2 components, and correspondingly longer 3′ non-coding regions, did not share significant sequence similarity between their 3′ non-coding regions (Table 2). Contrasting results between comparisons based on coat protein and 3′ non-coding sequences may reflect the general trend for greater conservation within coding vs. non-coding regions of a genome.

Despite low overall sequence homology between the different nepoviruses, a region of localized homology was found between BBLMV and the other nepoviruses about 40 nt upstream from the poly(A) sequence (Fig. 5). This conserved region in BBLMV shared 15 out of the 17 nucleotides of a 3′ consensus sequence (GGACACAAAAAGATTTT) previously identified in GFLV, GCMV, TBRV and CPMV by Serghini et al. (1990).

3.5. Maintenance of identity of the 3′ non-coding region

The marked conservation of sequences in the 3′ terminal region of each nepovirus may reflect an important function or functions that are selectively maintained. If the highly conserved 3′ termini contain an optimal sequence for a particular process, such as a function involved in replication, then changes in this sequence may reduce the efficiency of the process and result in a virus that is at a replicative disadvantage, or is completely inactivated. Variant genotypes may replicate at reduced levels compared to the optimal or wild type sequence and competition could eventually eliminate them from the viral population. Selection

for these functions could maintain the optimal sequence on both RNA components and, thereby, maintain identity.

However, to maintain identity in the 3′ terminus of BBLMV, selection would have to act over a high number of individual nucleotides (i.e., 1.4kb), all of which would need to have some effect on the fitness of the virus in order to be selected for or against. Because of the excessive length of the conserved 3′ non-coding region, it is likely some secondary and/or tertiary structure is involved. Any modifications in the 3-dimensional structure of a viral RNA molecule that reduced its rate of replication would be selected against. Strong selection pressure for back mutations and compensatory mutations that restored an optimal secondary and/or tertiary structure could, however, help maintain a specific 3′ structure while allowing sequence divergence.

Whether or not selection, by itself, can maintain identity in the 3′ non-coding region of BBLMV is still uncertain. High frequency RNA recombination has been suggested as a mechanism for maintaining identity of the 3′ terminus of two other nepoviruses, TomRSV (Rott et al., 1991a) and CLRV (Scott et al., 1992). Rott et al. (1991a) proposed that replication always begins on the same RNA and that template switching occurs at or near the junction between the non-coding and coding regions, thus maintaining identical sequences in the 3′ non-coding regions of both RNAs. Such a mechanism would be analogous to the leader-primed generation of subgenomic RNAs proposed for coronaviruses (Lai et al., 1990).

Recombination may have been responsible for the original duplication of the 3′ terminus of BBLMV and other nepoviruses. However, the role that recombination plays in the continued maintenance of identity in the 3′ terminus is unknown. Experiments are now in progress to estimate recombination frequencies in the 3′ non-coding regions of BBLMV.

References

Bertioli, D.J., Harris, R.D., Edwards, M.L., Cooper, J.I. and Hawes, W.S. (1991) Transgenic plants and insect cells expressing the coat protein of arabis mosaic virus produce empty virus-like particles. J. Gen. Virol. 72,1801–1809.

Blok, V.C., Wardell, J., Jolly, C.A., Manoukian, A., Robinson, D.J., Edwards, M.L. and Mayo, M.A. (1992) The nucleotide sequence of RNA2 of raspberry ringspot nepovirus. J. Gen. Virol. 73, 2189–2194.

Bouzoubaa, S., Ziegler, V., Beck, D., Guilley, H., Richards, K., and Jonard, G. (1986) Nucleotide sequence of beet necrotic yellow vein virus RNA2. J. Virol. 67, 1689–1700.

Brault, V., Hibrand, L., Candresse, T., Le Gall, O. and Dunez, J. (1989) Nucleotide sequence and genetic organization of Hungarian grapevine chrome mosaic nepovirus RNA 2. Nucl. Acids Res. 17(19), 7809–7819.

Demangeat, G., Greif, C., Hemmer, O. and Fritsch, C. (1990) Analysis of the in vitro cleavage products of tomato black ring virus RNA-1 encoded 250K polyprotein. J. Gen. Virol. 71, 1649–1654.

Demangeat, G., Hemmer, O., Fritsch, C., Le Gall, O. and Candresse T. (1991) In vitro processing of the RNA-2-encoded polyprotein of two nepoviruses: tomato black ring virus and grapevine chrome mosaic virus. J. Gen. Virol. 72, 247–252.

Devereux, J., Haeberli, P. and Smithies, O. (1984) A comprehensive set of sequence analysis programs for the VAX. Nucl. Acids Res. 12(1), 387–395.

Domingo, E. and Holland, J.J. (1988) High error rates, population equilibrium and evolution of RNA replication systems. In: E. Domingo, J.J. Holland and P. Ahlquist (eds.), RNA Genetics, Vol 3, pp. 3–36. CRC Press, Boca Raton, Florida.

Doolittle, R.R. 1981. Similar amino acid sequences: chance or common ancestry? Science 214(9), 149–159.

Fraenkel-Conrat, H., Singer, B. and Tsugita, A. (1961) Purification of viral RNA by means of bentonite. Virology 14, 51–58.

Greif, C., Hemmer, O. and Fritsch, C. (1988) Nucleotide sequence of tomato black ring virus RNA 1. J. Gen. Virol. 69, 1517–1529.

Gubler, U. and Hoffman, B.J. (1983) A simple and very efficient method for generating cDNA libraries. Gene 25, 263–269.

Harrison, B.D. and Murant, A.F. (1977) Nepovirus group. CMI/AAB Descriptions of Plant Viruses, no. 185, unpaged. Kew, Surrey, UK.

Hunkapiller, M.W., Lajun, E., Ostrander, F. and Hood, L.E. (1983) Isolation of microgram quantities of protein from polyacrylamide gels for amino acid sequence analysis. Meth. Enzymol. 91, 227–236.

Kamer, G. and Argos, P. (1984) Primary structural comparison of RNA-dependent polymerases from plant, animal and bacterial viruses. Nucl. Acids Res. 12, 7269–7782.

Lai, M.M. (1990) Coronaviruses: Organization, replication and expression of genome. Ann. Rev. Microbiol. 44, 303–333.

Le Gall, O., Candresse, T., Brault, V. and Dunez, J. (1989) Nucleotide sequence of Hungarian grapevine chrome mosaic nepovirus RNA1. Nucl. Acids Res. 17, 7795–7807.

Martelli, G.P. (1975) Nematode Vectors of Plant Viruses. Edited by F. Lamberti, C.E. Taylor and J.W. Seinhorst, London, 223 pp.

Matthews, R.E.F. (1991) Plant Virology, Third Edition, 835 pp. London, UK and New York, USA.

Meshi, T., Kiyama, R., Ohno, T., AND Okada, Y. (1983) Nucleotide sequence of coat protein cistron and the 3′ non-coding region of cucumber green mottle virus (watermelon strain) RNA. Virology 127, 54–64.

Meyer, M., Hemmer, O., Mayo, M.A. and Fritsch, C. (1986) The nucleotide sequence of tomato black ring virus RNA 2. J. Gen. Virol. 67, 1257–1271.

Ramsdell, D.C. and Stace-Smith, R. (1981) Physical and chemical properties of the particles of ribonucleic acid of blueberry leaf mottle virus. Phytopathol. 71(4), 468–472.

Rott, M.E., Tremaine, J.H. and Rochon, D.M. (1991a) Comparison of the 5′ and 3′ termini of tomato ringspot virus RNA 1 and RNA 2: evidence for RNA recombination. Virology 185, 468–472.

Rott, M.E., Tremaine, J.H. and Rochon, D.M. (1991b) Nucleotide sequence of tomato ringspot virus RNA2. J. Gen. Virol. 72, 1505–1514.

Sambrook, J., Fritsch, E.F. and Maniatis, T. (1991) Molecular Cloning: A Laboratory Manual. Cold Springs Harbor Laboratory, Cold Spring Harbor, N.Y.

Sanger, F. (1981) Determination of nucleotide sequence in DNA. Science 214, 1205–1210.

Scott, N.W., Cooper, J.I., Liu, Y.Y. and Hellen, C.U.T. (1992) A 1.5 kb homology in the 3′-terminal regions of RNA1 and RNA2 of birch isolate of cherry leaf roll nepovirus is also present, in part, in a rhubarb isolate. J. Gen. Virol. 73, 481–485.

Serghini, M.A., Fuchs, M., Pinck, M., Reinbolt, J., Walter, B. and Pinck, L. (1990) RNA 2 of grapevine fanleaf virus: sequence analysis and coat protein cistron location. J. Gen. Virol. 71, 1433–1441.

Steinhauer, D.A. and Holland, J.J. (1987) Rapid evolution of RNA viruses. Ann. Rev. Microbiol. 41, 409–433.

Vieira, J. and Messing, J. (1987) Meth. Enzymol. 153, 3–11.