Research Note

The Nucleotide Sequence of the Coat Protein Gene and 3’ Untranslated Region of Azuki Bean Mosaic Potyvirus, a Member of the Bean Common Mosaic Virus Subgroup

Candace W. Collmer,1 Marcia F. Marston,2 Susanne M. Albert,1 Sonia Bajaj,1 Heather A. Maville,1 Sarah E. Ruuska,1 Erika J. Vesely,1 and Molly M. Kyle2

1Department of Biological and Chemical Sciences, Wells College, Aurora, NY 13026 U.S.A.; 2Department of Plant Breeding and Biometry, Cornell University, Ithaca, NY 14853 U.S.A.

Received 16 January 1996. Accepted 28 June 1996.

The relationship of azuki bean mosaic potyvirus (AzMV) to members of the bean common mosaic virus ( BCMV ) subgroup has been unclear. Degenerate oligonucleotide primers and the polymerase chain reaction were used to amplify and clone the coat protein (CP) gene and 3’ untranslated region (UTR) of AzMV. The deduced amino acid sequence of the CP is 94% identical to that of dendrobium mosaic virus, establishing the two as strains of the same virus. While the CP amino acid identities between AzMV and potyviruses of the BCMV species are at or below 90%, the 91 to 94% identity between their UTRs suggests that AzMV could be considered a strain of BCMV. Interestingly, the grouping of potyviruses within the greater BCMV subgroup on a coat protein amino acid tree correlates with a grouping based on the response elicited on bean containing the I gene for resistance to BCMV.

Additional keywords: coat protein phylogeny, I gene of Phaseolus vulgaris, phyllogenetic tree, plant virus, taxonomy, temperature-sensitive resistance.

It has been well recognized that the taxonomy of the potyviruses, the largest group of plant viruses, is in a “very unsatisfactory state”; the size of the group and the extensive variation therein have made it difficult to reliably identify individual members and particularly to find characters that distinguish distinct potyviruses from strains (Ward et al. 1992 and references therein). While some have predicted that a “continuum” of strains may make such a distinction impossible (Bos 1970; Matthews 1979; Hollings and Brunt 1981; Harrison 1985), others have presented analyses of sequence homologies that argue for a demarcation between distinct viruses. Shukla and Ward (1988) found that pairwise coat protein (CP) amino acid sequence comparisons among 17 strains of eight distinct potyviruses showed a bimodal distribution—sequence identities ranged from 38 to 71% for distinct viruses (species) as opposed to 90 to 99% for strains of the same virus. Further work has revealed an additional cluster in the 75 to 89% range, which Ward and Shukla have grouped with the former (distinct species) rather than the latter (strains of the same species) (see Fig. 3 in Mink et al. 1994). Frenkel et al. (1989) had earlier suggested that the 3’ untranslated region (UTR) may be useful for determining relationships among viruses whose CP amino acid sequence homology falls midway between the original bimodal distribution. In that work they found that nucleotide sequence homology ranged from 39 to 53% for distinct viruses, in contrast to 83 to 99% for strains.

One group of potyviruses for which sequence comparisons could prove particularly useful includes the many biologically distinct isolates of several potyviruses that naturally infect legumes. While it is clear that a number of these, including bean common mosaic virus ( BCMV ), blackeye cowpea mosaic virus ( BICMV ), azuki bean mosaic virus ( AzMV ), cowpea aphid-borne mosaic virus ( CABMV ), peanut stripe virus ( PSTV ), and soybean mosaic virus ( SMV ) are closely enough related to be included in a BCMV subgroup of potyviruses (Dijkstra and Khan 1992), there has been disagreement about which sets of isolates constitute distinct viruses. Recently, BCMV serotypes A and B were formally separated into two distinct species, BCMNV (bean common mosaic necrosis virus) and BCMV, respectively (Mink et al. 1994). Numerous studies of the biology, serology, and CP peptides of the other potyviruses named above have failed to find “characteristics that can be used to distinguish unequivocally” among them (Mink and Silbernagel 1992, and references therein), and recent data have led to the proposal that AzMV, BICMV, and PSTV be considered isolates of the newly defined BCMV species (McKern et al. 1992; Huguenot et al. 1994; Mink et al. 1994).

Consistent with this proposal, sequence data provided the basis for a claim to finally “settle the taxonomic position” of certain BCMV and BICMV isolates as strains of the same virus (Khan et al. 1993) and for inclusion of a new member, dendrobium mosaic virus ( DeMV; Hu et al. 1995), in the
BCMV subgroup. The present study was undertaken to provide similar sequence data for AzMV, a potyvirus related in our studies to certain BCMV serotype B isolates, CABMV, and BICMV in terms of the response elicited on bean containing the I gene for resistance to BCMV (Fisher and Kyle 1994).

AzMV was obtained from M. J. Silbernagel (USDA WSU-IAREC, Prosser, WA) and maintained on Phaseolus vulgaris ‘BT-2’. A crude nucleic acid extract from systemically infected leaf tissue served as the starting material for a two-step procedure (modified from Pappu et al. 1993) that included reverse transcription and PCR-based amplification of approximately 1,370 nucleotides of the 3’ end of AzMV RNA (containing the complete CP gene and 3’ UTR). Ten microliters of the Sephadex G-50 column flowthrough was mixed with 2 μl of a 50 μM, 25-base poly-T oligonucleotide primer (CWC-111) containing a 3’ G (previously determined to prime reverse transcription on AzMV RNA), incubated at 70°C for 5 min, cooled, and then incubated at 42°C for 50 min following addition of the following components to a total reaction volume of 30 μl: 6 μl 5x reverse transcription buffer (Promega, Madison, WI), 3 μl dNTPs (10 mΜ each), 3 μl 100 mM DTT, 3 μl 40 mM sodium pyrophosphate (prewarmed to 42°C), 30 units RNAsin (Promega), and 18 units AMV reverse transcriptase (Promega).

Following reverse transcription, 3 μl of the reaction mix was used directly in a polymerase chain reaction containing the following additional components in a total volume of 50 μl: 10 mM Tris-HCl (pH 8.8), 50 mM KCl, 2.0 mM MgCl2, 120 μM each dNTP, 75 pmole primer CWC-111, 75 pmole of a degenerate oligonucleotide primer (CWC-107) for the Nib gene (5’-CAYKGYNTNGARCGNATYTGYGC-3’; Cassidy et al. 1993), and 2.5 units of Taq polymerase (Boehringer-Mannheim). (Degenerate positions in oligonucleotide primers are represented by the following single letter codes: Y = C or T; R = A or G; K = A or C; H = A, C, or T; N = A, C, G, T.)

TTTGAATTAT GAGATTGG CGTCGACG TCTGTCGAC GCATACAG GACAGGAC AAGAGTATG CTTATCATCA GAGATGTTT CATCTGCT 63
SGINQP7

GAAGCTAT GGCAGGAC TACGTGCTA CCAAGACA CCAATACTG GACAGAC AAGAGTTT TACGAGTC AAGAGGAC 126

EVRAGIDTGKDEKKSSKKGK

CAGCAGAC AAGAAGTC TACGAGAC AAGAGGAC 189

FGKSEKGSGNNSRGAENSTMRD

GAAGATCTA AATCTGCTG TCAAAAGGG AGATTTGTAC TCTGCGTT CATCAAGG ATCAAGAG 252

KBVNAVSGKKGKVKVPRLOQRITRK

ATGATATG CAGGACATGA AGGCGAAAT AGATTCTAA ATATTCTG TATCTATG TATCTAGA AAGATGAC 315

NMNPTVGRNMNILNLDHHLDYK

CCAGACAA AAGAGATTT TTACTACCA AGAGCAAC AGATGAC AAGAGATT TCTGAGAT CGTCAGAT 378

FEQTDLFTNTRAKMQFEMWYN

GCTGTCGAG GAGGATGCT GATGACAG ATGATGATG ATGATGATG GCTGTCTGCT 441

AVGKEYEDIDEQMSIVMNGF

GTCCTGCT ATGATGAG CATGTACCA CGTGTCACG TACGATGATG TACGATGATG 504

VWCIIDNGTSVDPVNGTWVMMDG

GATGACAG GGTGATAG CCAGCTACAA CCAATGTT AAGATGAC AAGAGGAC AAGAGGAG 567

DEQVEYPLKFMVENDAERKLQ

ATCCTGACCA CTTTCTCACA CAGAAGCTT AAGCTACG AAGAGGAC 630

IMMMHFSDAEEAYIEMENSFRP

TATACGCTA CGTCTCGAG AATGTCAGG GACAGACCA CTAGCTGCT GCTCTCTT 693

YMFRYGLRMRNLRDKLNRLLAYAF

GATTCTCA GACAGTACAC TCAAAACA TCGTGTGAC GACAGGAC 756

DFYEVTSKTSDDRAREAVAQMK

GCAGCAGAC ATCAGCAAC GTTACAACG AAGCTTATT GCAGATGTG GATGATGATG 819

AAALSNVTSKFLGDGDNVATT

AGGAGATTT AAGAAGCC AAGCTACG AATGATGAC 882

SENTESLARVDVQNMHSLLG

ATGCGCCCG CCGAGCTAA AATGTTGTG CAGATCAGCAC 945

MGSPQ#

ATTGTACAT AGTACTTCTT TATCTTCTTTT AATGGCTTT TTATGTTGT 1008

CTATGTCGATATGAAAACTCGCAGC CGAGATGATG 1071

GGAAACAG TAGATGAC AAGGAGTG CGTCAGACG CTACGATGAC GAGTATGAC 1134

GCAGATGCTCTTCGAGCCTTCC

Fig. 1. DNA sequence corresponding to the 3’ terminus of azuki bean mosaic virus (AzMV) RNA, beginning 42 nucleotides 5’ to the assumed start of the CP gene (codon underlined) and continuing to the 3’ poly-A tail. Underlined individual nucleotides are sites where variation was observed in partial clones (see text). The # denotes the stop codon at the end of the CP gene. The deduced CP sequence is presented; the DAG sequence is underlined. GenBank (EMBL) Data Library accession no. U60100.
<table>
<thead>
<tr>
<th>Viruses</th>
<th>AzMV</th>
<th>DeMV</th>
<th>BCMV-US-1</th>
<th>BCMV-NL-4</th>
<th>BCMV-NY-15</th>
<th>BCMV-NL-3</th>
<th>SMV</th>
<th>BCMV-NL-3</th>
<th>PWV-K</th>
<th>WMV</th>
<th>ZYMV</th>
<th>SAPV</th>
<th>BYMV</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>93.9</td>
<td>94.6</td>
<td>90.3</td>
<td>89.7</td>
<td>89.4</td>
<td>88.7</td>
<td>88.0</td>
<td>86.9</td>
<td>78.2</td>
<td>77.2</td>
<td>76.4</td>
<td>74.2</td>
<td>70.9</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>90.6</td>
<td>89.9</td>
<td>89.6</td>
<td>89.2</td>
<td>88.5</td>
<td>87.8</td>
<td>79.1</td>
<td>78.1</td>
<td>77.3</td>
<td>75.1</td>
<td>71.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>93.6</td>
<td>92.0</td>
<td>91.3</td>
<td>91.7</td>
<td>94.1</td>
<td>96.5</td>
<td>96.2</td>
<td>78.5</td>
<td>77.3</td>
<td>76.3</td>
<td>74.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>90.8</td>
<td>91.6</td>
<td>90.6</td>
<td>89.6</td>
<td>88.9</td>
<td>87.8</td>
<td>76.7</td>
<td>77.4</td>
<td>75.3</td>
<td>73.8</td>
<td>72.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>94.0</td>
<td>92.0</td>
<td>98.8</td>
<td>91.0</td>
<td>91.0</td>
<td>89.6</td>
<td>88.9</td>
<td>77.8</td>
<td>76.7</td>
<td>74.9</td>
<td>73.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>93.3</td>
<td>93.0</td>
<td>95.5</td>
<td>90.4</td>
<td>96.7</td>
<td>92.4</td>
<td>91.3</td>
<td>78.1</td>
<td>77.8</td>
<td>75.3</td>
<td>73.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>93.4</td>
<td>90.7</td>
<td>97.9</td>
<td>89.3</td>
<td>97.5</td>
<td>94.5</td>
<td>95.8</td>
<td>77.4</td>
<td>77.4</td>
<td>75.3</td>
<td>74.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>92.3</td>
<td>90.8</td>
<td>96.9</td>
<td>89.8</td>
<td>98.0</td>
<td>95.1</td>
<td>95.9</td>
<td>76.7</td>
<td>76.4</td>
<td>75.3</td>
<td>72.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>59.8</td>
<td>58.4</td>
<td>61.6</td>
<td>58.9</td>
<td>60.8</td>
<td>60.0</td>
<td>60.6</td>
<td>60.9</td>
<td>78.9</td>
<td>79.6</td>
<td>87.6</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>52.7</td>
<td>57.3</td>
<td>54.2</td>
<td>57.0</td>
<td>54.2</td>
<td>55.7</td>
<td>55.1</td>
<td>52.6</td>
<td>50.7</td>
<td>80.0</td>
<td>77.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>51.7</td>
<td>50.3</td>
<td>51.6</td>
<td>51.2</td>
<td>50.7</td>
<td>52.5</td>
<td>50.8</td>
<td>49.7</td>
<td>49.2</td>
<td>74.4</td>
<td>72.9</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>63.3</td>
<td>60.7</td>
<td>63.9</td>
<td>60.4</td>
<td>64.3</td>
<td>63.5</td>
<td>63.3</td>
<td>64.0</td>
<td>76.6</td>
<td>55.9</td>
<td>50.9</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>38.4</td>
<td>37.1</td>
<td>38.3</td>
<td>37.2</td>
<td>37.6</td>
<td>38.2</td>
<td>38.0</td>
<td>37.8</td>
<td>33.7</td>
<td>36.0</td>
<td>32.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>46.7</td>
<td>45.1</td>
<td>46.7</td>
<td>46.8</td>
<td>45.6</td>
<td>45.0</td>
<td>44.9</td>
<td>43.5</td>
<td>40.3</td>
<td>62.3</td>
<td>39.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>34.7</td>
<td>33.6</td>
<td>35.1</td>
<td>35.2</td>
<td>35.1</td>
<td>35.4</td>
<td>34.7</td>
<td>35.4</td>
<td>32.4</td>
<td>28.5</td>
<td>28.8</td>
</tr>
</tbody>
</table>

*CP amino acid sequence alignments were conducted in the GCG program PILEUP (Devereux et al. 1984) with a gap weight of 3 and a gap length weight of 0.1. Nucleotide sequence alignments of the 3' UTRs were constructed initially in PILEUP with a gap weight of 4 and a gap length weight of 0.1 and then were edited manually for better alignment. Pairwise identity between each of the viruses was determined using the GCG DISTANCES program with the average sequence length as the denominator. CP amino acid comparisons are above the diagonal; 3'-UTR nucleotide comparisons are below the diagonal. The fourteen potyviral coat protein and 3'-UTR sequences are from the following GenBank accession numbers or publications: dendrobium mosaic virus (DeMV; U23564); bean common mosaic virus (BCMV) strain US-1 (L12740); BCMV-NL-4 (L12766); BCMV-NY-15 (S66252); BCMV-NL-3 (S66251); bean common mosaic necrosis virus (BCMV) strain NL-3, formerly designated BCMV NL-3 (U19227); peanut stripe virus blotch isolate (PSV; U05771); blackeye cowpea mosaic virus (BCMV; S66253); soybean mosaic virus strain G2 (SMV; S42280); passion-fruit woodiness virus strain K (PWV-K; Gough and Shukla, 1992); watermelon mosaic virus (WMV; D00592); zucchini yellow mosaic virus (ZYMV; L29569); South African passiflora virus (SAPV; D10053); bean yellow mosaic virus (BYMV; X63338).

The AzMV CP sequence and the 3' UTR nucleotide sequence were compared with sequences of 14 other potyviruses, selected from available databases as representative of either closely related potyviruses, legume-infecting potyviruses, or other potyviruses that had been tested for the response(s) elicited on bean cultivars possessing the f gene for resistance to BCMV (Fisher and Kyle 1994). In an attempt to apply Shukla and Ward's (1988) criteria for distinguishing strains of the same virus from distinct viruses, the pairwise percent identity of CP amino acids and of the 3' UTR nucleotides between each of the viruses was determined (Table 1) using the GCG Distances program from the University of Wisconsin Genetics Computer Group (Devereux et al. 1984). Phylogenetic (evolutionary) relationships among the 15 potyviruses were examined by parsimony analysis of aligned CP amino acid sequences using the Phylogenetic Analysis Using Parsimony (PAUP) software package (Swoford 1991; see Fig. 2).

The pairwise comparisons of the CP sequence data of the related potyviruses (Table 1) show that AzMV is most closely related to DeMV, with 94% coat protein identity consistent with the view that these isolates are strains of the same virus. Further, the phylogenetic tree generated from the CP amino acid sequence data (Fig. 2) clearly shows a close relationship between AzMV and DeMV and their more distant relationship to the other members of the newly defined BCMV species (including BCMV NY-15, BCMV NL-1, BCMV US-1, BCMV NL-4, and PSV). When the first four of these latter isolates are considered in pairwise comparisons of sequence identity, the values are greater than 90%. When BCMV NL-4 and PSV are added to the group, values drop to 89% or 88%, respectively. When AzMV is included, percent identities range from as low as 87% (with BCMV NL-1) to 90% (with BCMV US-1), somewhat below Ward and Shukla's cutoff point for strains of the same virus (Fig. 3 of and T.) The reaction was overlaid with mineral oil and cycled 30 times (2 min 94°C, 2 min 42°C, 2 min 72°C) followed by elongation at 72°C for 10 min. Alternatively, partial, overlapping clones containing 5' nucleotides 1-764 or 3' nucleotides 445-1153 (numbering as in Fig. 1) were made using degenerate oligonucleotide primers CWC-107 (5'; above) and CWC-106 (3'; GCNGCNNGYTTCTCATYTG); or CWC-104 (5'; TGGTGYAHTGHANAAYGG) and the poly-T primer (3'; above), respectively. Primers CWC-104 and CWC-106 were modified slightly from Langeveld et al. (1991) and Pappu et al. (1993). For the 5' partial clone, the annealing temperature in the PCR reaction was changed from 42°C to 46°C. Each reaction product was purified using GeneClean II (BIO 101 Inc., Vista, CA), ligated into a pCR II vector and transformed into competent E. coli using a TA Cloning Kit (Invitrogen, San Diego, CA). Clones were screened for inserts of the correct size on agarose gels following EcoRI digestion of plasmid DNA purified by an alkaline lysis procedure. The cloned AzMV cDNA was sequenced using a Sequenase 2.0 DNA Sequencing Kit (U.S. Biochemical, Cleveland, OH) and 5% Long Ranger sequencing gels (AT Biochem) according to the manufacturer's directions.

The sequence of the coding region of the CP (855 nucleotides) and the 3' UTR (256 nucleotides excluding the poly-A tail), as determined from a full-length clone of the region, is presented in Figure 1, along with the deduced 285 amino acids of the AzMV CP. Two noteworthy features of the N-terminal region of the CP are the presence of the DAG sequence associated with aphid transmissibility and the fact that the AzMV CP is two amino acids shorter than BCMV CPs (e.g., BCMV NY-15) in this region. Partial clones made a year earlier of the 5' or 3' region of AzMV as described above contained the following nucleotide substitutions, respectively: C to T at position #675; A to G at #575, C to T at #675, G to A at #708, C to T at #731, C to T at #1030.
Mink et al. (1994). When only the amino acids of the variable N-terminus of the CPs are compared (e.g., the first 50 amino acids of AzMV CP), the percent identity drops substantially (e.g., 65% identity for AzMV and BCMV US-1; data not shown). The 91 to 94% identities between the 3’ UTRs of these seven isolates (Table 1), however, support the idea that they all could be considered strains of BCMV.

We have noted previously a correlation between relationships among leghume potyviruses based on CP sequence and similarity of host responses on various genotypes of P. vulgaris (Fisher and Kyle 1994, 1996). All viruses that group by CP sequence with the newly defined BCMV species (including AzMV) and that have been tested carefully on bean cultivars with and without the I gene (e.g., ‘BT-1’ and ‘BT-2’) show a temperature-sensitive resistance/systemic necrosis response (i.e., AzMV, BICMV, and BCMV strains NY-15, NL-1, US-1, NL-4). In contrast, potyviruses that group with the newly defined BCMV species (e.g., BCMV NL-3, formerly designated BCMV NL-3) cause temperature-insensitive necrosis on BT-1 bean. Even when only the variable N-terminus of the CP of the nine BCMV subgroup viruses included on Figure 2 is used, the resulting tree is almost identical (data not shown). The availability of appropriate infectious clones will allow a direct examination of the basis for this correlation between CP sequence and host response.

ACKNOWLEDGMENTS

We gratefully acknowledge E. Rybicki, S. Pappu, and B. Cassidy for technical advice and information prior to publication, and P. Berger and G. Mink for helpful discussions. This research has been supported by U. S. Dept. of Agriculture NRI Competitive Research Grant #9237311-8062 to C.W. Collmer, and by grants to Wells College from the Bristol-Myers Squibb Foundation and the Hewlett-Mellon Program. M.F.M. and C.W.C. received partial support from the NSF/DOE/USDA Cornell University Plant Science Center.

LITERATURE CITED