The Complete Nucleotide Sequence of Bean Yellow Mosaic Virus Genomic RNA*

Shigeo NAKAMURA**, Ryoso HONKURA**, Takayoshi IWAI**, Masashi UGAKI*** and Yuko OHASHI***

Abstract

The complete nucleotide sequence of the RNA genome of bean yellow mosaic potyvirus (BYMV) has been determined. The RNA genome of BYMV is 9532 nucleotides long excluding the poly(A) tract. The RNA contains one open reading frame (ORF) of 9171 nucleotides encoding a large polypeptide of 3056 amino acids with a calculated Mr of 347,571. Nine putative proteolytic cleavage sites, one by P1 protease, one by HC protease and seven by NIa protease, are identified by comparison with those identified for other sequenced potyviruses. The genetic organization of BYMV genome is proposed to be 5′UTR (190 nt)-P1 (33 K)/HC-Pro (52 K)/P3 (41 K)/6K1/CI (71 K)/6K2/NIa-VPg (22 K)/NIa-Pro (27 K)/NIb (59 K)/CP (31 K)-3′UTR (171 nt)-poly(A). Comparing the amino acid sequences of individual BYMV proteins with the corresponding proteins of other potyviruses, the BYMV proteins, excepting P1, have a high or moderate homology.

(Received May 7, 1996; Accepted June 20, 1996)

Key words: bean yellow mosaic potyvirus, nucleotide sequence.

INTRODUCTION

Bean yellow mosaic virus (BYMV) is a member of the genus Potyvirus. The potyviral genome consists of a positive-sense RNA of approximately 10 kb with a genome-linked protein (VPg) at the 5′ terminus and a poly(A) tract at the 3′ terminus. The complete nucleotide sequences of several potyvirus genomes have also been determined: tobacco etch virus (TEV), tobacco vein mottling virus (TVMV), plum pox virus (PPV), potato virus Y (PVY), pea seed-borne mosaic virus (PSbMV), pepper mottle virus (PepMoV), papaya ringspot virus (PRSV), turnip mosaic virus (TuMV), soybean mosaic virus (SMV), Johnson grass mosaic virus (JGMV), potato virus A (PVA) and peanut stripe virus (PStV) (see Table 1). These reports indicate that the genome organization of all potyviruses is similar and the polypeptide is proteolytically processed into at least eight mature proteins: the N-terminal protein (P1), the helper component-protease (HC-Pro), the P3 protein, the cytoplasmic inclusion protein (CI), the nuclear inclusion protein a (NIa), which consists of VPg domain (NIa-VPg) and protease domain (NIa-Pro), the nuclear inclusion protein b (NIb) and the coat protein (CP), by three virus-encoded proteases. The polypeptides also contain conserved protein motifs, such as ‘G-D-D’ involved in polymerase activity, described for plant RNA viruses.

BYMV is a member of a subgroup consisting of BYMV and clover yellow vein virus (CYVV), pea mosaic virus and sweet pea mosaic virus. They are serologically related and have similar host ranges. Sequence comparisons of the 3′-part of the genomes of several isolates of BYMV and CYVV have shown that the two viruses are distinct but closely-related genetically. The complete nucleotide sequence of these viruses, however, has not been determined. In this study, the complete nucleotide sequence of the BYMV RNA is presented and its genetic organization is compared with those of other potyviruses.

MATERIALS AND METHODS

cDNA synthesis and cloning. The MB4 isolate of BYMV, originally isolated from broad bean in Miyagi Prefecture, was propagated in broad bean (Vicia faba). Viral RNA was isolated from freshly prepared virus as described by Robaglia et al. Two types of cDNA were synthesized. The 3′ terminal part of the RNA was

---

* The nucleotide sequence data reported in this paper will appear in the GSDB, DDBJ, EMBL and NCBI nucleotide sequence database with the following accession number D83749.
** Department of Plant Biotechnology, Miyagi Prefecture Agricultural Research Center, Takadate-kawakami, Natori 981-12, Japan
*** Department of Molecular Biology, National Institute of Agrobiological Resources, Kannondai 2-1-2, Tsukuba 305, Japan
cloned with oligo(dT) primers and clone pMB625 containing a 4.3 kbp insert was sequenced as described previously16. cDNAs corresponding to the remaining part of the RNA were synthesized using a TimeSaver cDNA Synthesis Kit (Pharmacia Biotech) with random primers and cloned into pBluescript II SK+.
Fig. 1. Nucleotide sequence of bean yellow mosaic virus RNA and the deduced amino acid sequence of the putative polyprotein. The putative proteolytic cleavage sites for P1 protease, HC protease and NIa protease are underlined.

RESULTS AND DISCUSSION

Sequencing

The complete nucleotide sequence of the BYMV RNA and the predicted translation product are shown in Fig. 1. The genome consists of 9532 nucleotides (nt) followed by a poly (A) tract. Of five independent clones generated by 5' RACE, three clones had six A residues adjacent to the poly (C) tract added after the first strand synthesis, whereas two clones had five A residues. We cannot declare whether this results from a heterogeneous RNA population of viral RNA or removal of one or more nucleotides with the VPg during the extraction process. The nucleotide sequence reported in this work begins with six adenines because it was the major and the
longer species. The base composition of BYMV RNA is 33% A, 18% C, 23% G and 26% U, which is similar to that of other potyviruses.

**Genome organization**

Computer analysis revealed a single long open reading frame (ORF) of 9171 nt, starting at position 191 and ending with a UAG termination codon at positions 9359-9361. Two in-phase AUG codons were found within the first 300 nt, the first AUG at positions 191-193 and the second at positions 251-253. The base composition of the 190 nt leader sequence is 41% A, 15% C, 11% G and 34% U, whereas that of the 250 nt leader sequence is 40% A, 16% C, 12% G and 32% U. Both potential leader sequences show low G and high A/U contents, which are typical of potyviruses and other plant viruses. The first AUG is likely to be the initiation codon for the polyprotein of BYMV, because it is in the context UUAUAGAC, which is similar to the consensus sequence AACAAUGGC proposed for translation initiation in plants, and the length of the 5' leader sequence is comparable to those of other sequenced potyviruses. Therefore, the ORF potentially encodes a polyprotein of 3056 amino acids with a calculated Mr of 347,571. The ORF is followed by a non-coding region of 171 nt, excluding the poly(A) tract. The length of 5' untranslated region is the longest, apart from TVMV (205 nt) and the 3' non-coding region is the shortest, apart from PSbMV (160 nt). The 5' untranslated region has three blocks of nucleotides similar to box a-like or box b-like, highly conserved regions in several potyviruses. Mature potyviral proteins are expressed by proteolytic processing of the polyprotein by three virus-coded proteases, the P1 protease, the HC protease and the N1a protease. The predicted proteolytic cleavage sites and the genetic map of BYMV polyprotein are shown in Fig. 2. Putative protease cleavage sites were identified by comparison with published potyviral cleavage sites. The putative P1 protease cleavage site between P1 and HC-Pro of the BYMV polyprotein is located at position F284/S285. The surrounding sequence, I-X-X-F-S, is similar to that of TuMV and PSbMV. The HC protease cleavage sequence is highly conserved in all potyviruses, and the consensus sequence, K-X-Y-X-V-G/G, was proposed. A similar sequence exists in the polyprotein of BYMV at positions 736-742, and the sequence K-Y-Y-R-V-G/G is the same as that identified in PsTV.

We have proposed the consensus sequence, F-Q(E)/S(G), recognized by BYMV N1a protease, and identified five putative cleavage sites in the BYMV polyprotein, C1/6K2, 6K2/N1a, N1a/N1b, N1b/CP and N1a internal site. In this work, two additional cleavage sequences by the N1a protease were found in the BYMV polyprotein. One is D-K-Y-T-F-Q1092/A1099, the other is D-A-Y-R-F-Q1142/S1143. Alignment of the surrounding sequences with those of other potyviruses revealed that the former is the boundary between P3 and 6K1, and the latter is the boundary between 6K1 and C1. Seven putative cleavage sequences recognized by the BYMV N1a protease lead to the consensus sequence F-Q(E)/S(A or G), similar to those of TVMV and PVA. From these observations, the BYMV polyprotein is probably processed into ten mature proteins, as shown in Fig. 2. The size of each protein is comparable to those of other potyviruses.

**Comparisons of the BYMV proteins with those of other potyviruses**

We have identified common amino acids sequence motifs between the BYMV proteins and other completely sequenced potyviruses. Some of the conserved motifs identified in BYMV proteins and amino acid similarities between the individual proteins of BYMV and those of other potyviruses are shown in Fig. 2 and Table 1, respectively.

The N-terminal protein, P1, was the least conserved...
protein among the potyviruses. The amino acid identity between BYMV P1 and those of other potyviruses is 9 to 22%. The length of P1 proteins varied from 237 aa (JGMV) to 547 aa (PRSV). The positions of the catalytic residues of BYMV P1 protease are H192, D201, S232 and D268. The P1 protein has been proposed to have a function in cell-to-cell movement of potyviruses because of the sequence similarity between the N-terminal region of TVMV and the 30kDa protein of tobacco mosaic virus (TMV)6). However, the amino acid similarity found in the TVMV P1 and the TMV 30kDa protein were not found in the other potyviral P1s11), nor in the BYMV P1. Two conserved motifs have been found in BYMV HC-Pro. The first motif is a ‘zinc-finger’-like metal-binding motif (C310-8aa-C319-13aa-C333-4aa-C338-2aa-C341), which was first noted in PVY HC-Pro20). The second motif (C627-72aa-H700), required for protease activity of HC-Pro18), is located at the C-terminal part of HC-Pro. The putative CI protein of BYMV showed between 47% and 61% sequence identity with that of other potyviruses (Table 1). Gorbalenya et al.9) identified a conserved motif representing the nucleotide-binding site in the CIs of TEV and TVMV. The motif, G(PStV-V)-A(JGMV-N, ZYMV-R) V-G-S-G-K-S-T, which is present in the CIs of all sequenced potyviruses, is also located at positions 1227–1235 of the BYMV polyprotein. In the region of nuclear inclusion proteins, the N1a protease active site (H1202, D1207, C1212) and two conserved motifs involved in the putative RNA-dependent RNA polymerase activity, Y1209-C-D-A-D-G-S1215 and S1215-G-3aa-T-3aa-
N-T-29aa-G-D-3617, are also present.

**Sequence similarity between BYMV and CYVV**
The CP of BYMV MB 4 showed 52 to 65% amino acid identity with those of other potyviruses (Table 1), whereas it showed 89 to 96% and 74 to 76% identity with five other BYMV isolates and three isolates of CYVV, respectively. The amino acid sequence similarity between BYMV and CYVV is consistent with the taxonomic relationship among potyviruses based on four major traits: genome organization, vector, inclusion bodies and host range14). The 3' untranslated sequence of BYMV MB4 also showed greater identity (93 to 100%) with five other BYMV strains than did the CP. However, the identity with three isolates of CYVV was less (64 to 68%) than those of the CP (data not shown). The 3' untranslated regions of potyviruses, therefore, seem to reflect their genetic relatedness better than their coat proteins.

**Literature cited**


和文摘要
中村茂雄・本倉良三・岩井孝志・宇垣正志・大橋祐子：インゲンマ黄斑モザイクウイルス RNA の全塩基配列
本県内のえそモザイク症状ソラマメから分離したインゲンマ黄斑モザイクウイルス（BYMV）MB4 分離株の RNA から、オオギトウとラムダムプライマーを用いたGubler & Hoffman 法、および5′ RACE 法によって全長をカバーするcDNA クローンを得、その塩基配列を決定した。BYMV サブノウRNA は5′poly(A)を含めて9532塩基からなり、第191塩基から第9361塩基までひとつの長いORFが見いだされた。このORFは3656アミノ酸から成るタンパク質（分子量348 K）をコードしており、すでに全塩基配列が報告されているポトivirusとの比較から、ウイルスプロテアーゼによりP1, HC-Pro, P3, 6K1, CI, 6K2, Nia-Vpg, Nia-Pro, Nib, CP の10個のポリペプチドに切断されると予想された。これらのアミノ酸配列を他のpotyvirusと比較したところ、Nibが最も高い相対性を示し（57-65%）、次いでCP(52-65%), CI(47-61%), Nia(43-50%), HC-Pro(39-50%), 6K1(36-33%), 6K2(30-49%), P3(18-26%)の順であり、P1は9-22%と低く、そのサイズもウイルス間で大きな違いがあった。