Complete Nucleotide Sequence of Ryegrass Mottle Virus: A New Species of the Genus Sobemovirus

Feiyun ZHANG1,2, Shigemitsu TORIYAMA1,3* and Mami TAKAHASHI1

ABSTRACT

The genome of Ryegrass mottle virus (RGMoV) comprises 4210 nucleotides. The genomic RNA contains four open reading frames (ORFs). The largest ORF 2 encodes a polyprotein of 947 amino acids (103.6 kDa), which codes for a serine protease and an RNA-dependent RNA polymerase. The viral coat protein is encoded on ORF 4 present at the 3'-proximal region. Other ORFs 1 and 3 encode the predicted 14.6 kDa and 19.8 kDa proteins of unknown function. The consensus signal for frameshifting, heptanucleotide UUUAAAC and a stem-loop structure just downstream is in front of the AUG codon of ORF 3. Analysis of the in vitro translation products of RGMoV RNA suggests that the 68 kDa protein may represent a fusion protein of ORF 2-ORF 3 produced by frameshifting. The protease region of the polyprotein and coat protein have a low similarity with that of the sobemoviruses (approximately 25% amino acid identity), while the RNA-dependent RNA polymerase region has particularly strong similarity (54 to 60% of more than 350 amino acid residues). The sequence similarities of RGMoV to the sobemoviruses, together with the characteristic genome organization indicate that RGMoV is a new species of the genus Sobemovirus.

(Received June 28, 2000; Accepted November 14, 2000)

Key words: Ryegrass mottle virus, nucleotide sequence, genome organization, sobemovirus.

INTRODUCTION

Ryegrass mottle virus (RGMoV) was first isolated from stunted Italian ryegrass (Lolium multiflorum) and cocksfoot (Dactylis glomerata) having mottling and necrotic symptoms on leaves. The isometric particle, 28 nm in diameter contains a species of single-stranded RNA with a molecular weight of 1.5 × 10^6. The physical properties and some biological ones of the virus are similar to Cocksfoot mottle virus (CfMV), which is prevalent in cocksfoot pastures in Japan. However, RGMoV is serologically distinct from CfMV, Cocksfoot mild mosaic virus, Cynosurus mottle virus and Phleum mottle virus, which occur in European countries. Last year, an isometric virus, isolated from Italian ryegrass in Germany was found to be serologically related to RGMoV; in agar gel double diffusion tests, a spur formed between RGMoV and the Germany isolate (Frank Rabenstein, Germany; personal communication).

In spite of serological differences between RGMoV and sobemoviruses, general properties of RGMoV are similar to those of grass viruses that belong to sobemoviruses. The genome sequence of sobemoviruses has been determined in Southern bean mosaic virus (SBMV)2,3,4, Cocksfoot mottle virus (CfMV)5,6,7, Rice yellow mosaic virus (RYMV)8,9, and Lucerne transient streak virus (LTSV, accession number U31286). The genomic RNA of sobemoviruses is a single-stranded molecule, approximately 4100 to 4500 nucleotides (nt) in size. The 5' terminus has a genome-linked viral protein (VPg) and the 3' end does not have a poly (A) tail. The genome encodes four ORFs: the largest ORF encodes the polyprotein of approximately 100 kDa, which contains protease and RNA polymerase motifs. Only the polyprotein of CfMV is encoded by two smaller overlapping ORFs, by –1 frameshifting. Recently, we determined the complete nucleotide sequence of the Japanese isolate of CfMV (CfMV/JP) (Zhang and Toriyama, unpublished data; accession number AB040447). The nucleotide sequence is 96.8% identical to the Norwegian isolate of CfMV(CfMV/NOR)10 and 95.8% identical to the Russian isolate11. Its
genome organization is identical to that of CfMV. So far, the genome sequence of RGMoV and the Germany isolate has not been determined, so the genus is still unknown. In this paper, we report the complete nucleotide sequence of RGMoV and compare it to that of the sobemoviruses.

**MATERIALS AND METHODS**

**Viruses**  
Ryegrass mottle virus (RGMoV) was propagated in barley plants (cv. Shunsei) and purified as described previously. A purified preparation of CfMV/JP, was stored at -80°C and used for the in vitro translation experiment.

**Terminal sequence of viral RNA**  
In a preliminary experiment, we found that RGMoV RNA does not have a poly (A) tail at the 3'-terminus. Thus, we determined the 3'-terminal sequence by two-dimensional mobility shift analysis as described previously. In this experiment, the homomix (alkaline digested yeast RNA mixture) was prepared by using the RNA from Torula utilis, a product of Fluka (Riedel-de Haén ; Seelze, Germany). The 5'-terminal sequence of RGMoV RNA was identified by sequencing the PCR clones amplified by using the 5' RACE abridged anchor primer system (Gibco BRL, Gaithersburg, USA).

**Cloning and DNA sequence**  
CDNA synthesis was done as described previously using M-MLV reverse transcriptase (Gibco BRL), random hexanucleotide primer and synthetic oligonucleotide primer (P1), 5'-ACTAGTCGACACGAAAACCCC-3': the sequence at the 3' end underlined was analyzed by two-dimensional sequence analysis. The synthesized second strand cDNA was blunt-ended with T4 DNA polymerase and ligated into SmaI-digested pUC18. Recombinant plasmids were transformed into competent *Escherichia coli* DH5α (TOYOBO, Osaka, Japan). The cDNA clones shown in Fig. 1 were made by primer extension and PCR amplification and used for sequencing of RGMoV RNA. The ambiguous nucleotide sequence was confirmed by using PCR clones prepared independently (not shown in Fig. 1). Nucleotide sequences were determined using the Pharmacia DNA sequencing kit and an ALFred DNA sequencer (Pharmacia, Uppsala, Sweden). The sequence data were assembled and analyzed using the DNASIS (Macintosh) program (Hitachi Software Engineering Co., Yokohama, Japan). GeneBank/EMBL, NBRF and PIR databases were searched for nucleic acid and amino acid sequence identity.

**Cell-free translation**  
In vitro translation using wheat germ extract (Promega, Madison, USA) was performed as described by the manufacturer's manual in a final volume of 50 µl in the presence of redive L-[35S]methionine (Amersham Pharmacia Biotech, Buckinghamshire, UK) for 1 hr at 25°C. Translation products were separated by SDS-PAGE (10% polyacrylamide) and detected using a Molecular Imager System (BioRad, Richmond, USA). A set of prestained SDS-PAGE standards (BioRad) was used as protein size markers.

**N-terminal amino acid sequence**  
Purified RGMoV was electrophoresed on 10% polyacrylamide-SDS gels and electro-blotted onto the PVDF membrane (Immobilon-P®; Millipore, Middlesex, UK). The portion corresponding to the coat protein on the PVDF membrane was excised, and the N-terminal sequence of the coat protein was analyzed using a gas-phase protein sequencer (model 477A/120A, Applied Biosystems, Foster City, USA).

**RESULTS AND DISCUSSION**

**Nucleotide sequence and genome organization**  
The complete nucleotide sequence of RGMoV com-

---

**Fig. 1.** cDNA clones used to determine the nucleotide sequence of RGMoV RNA. Solid lines with each number, cDNA clones synthesized using synthetic primer P1 or by primer extension; Slant lines, PCR clones.
The sequence contains four major ORFs flanked by 5'- and 3'-untranslated sequences of 99 and 198 nt, respectively. Database searches indicated that the genome sequence of RGMoV is significantly similar to that of sobemoviruses, for which the genome organization is summarized in Fig. 2. As shown in Fig. 2, ORFs 1 and 2 of RGMoV are separated by the intergenic region of 143 nt, but other ORFs overlap each other. RGMoV ORF 1 encodes a predicted 14.6 kDa protein of 133 amino acids. No distinct sequence similarity was found between RGMoV and the corresponding ORF 1 of sobemoviruses (Table 1). The ORF 1 of RYMV/ Ivory Coast isolate (P1 protein) is required for viral replication and virus spread. So far the function of the 14.6 kDa protein of RGMoV is unknown.

The largest ORF 2 extends from nucleotides 643 to 3486. The predicted 103.6 kDa protein consists of 947 amino acids. Database searches revealed a significant similarity to the polyproteins of sobemoviruses; SBMV (accession number, M23021), RYMV (accession number, L20893), CFMV (accession number, Z48630) and LTSV (accession number, U31286). The polyprotein of RGMoV contains serine and P3C proteases and an RNA-dependent RNA polymerase (Fig. 3). A conserved sequence, GxPxFDPxYG, is found in the N-terminal

---

**Table 1. Amino acid sequence similarity between coding regions of RGMoV and sobemoviruses**

| Viruses   | ORF 1 | ORF 2: polyprotein | ORF 3 | ORF 4: coat protein |
|-----------|-------|---------------------|-------|---------------------|
|           |       | Protease region     | Polymerase region |                     |
| SBMV      | 31.6% (38 aa) | 26.0% (346 aa) | 54.2% (369 aa) | 42.1% (122 aa) | 26.8% (157 aa) |
| LTSV      | 29.0% (62)    | 29.7% (353)     | 56.0% (357)     | 40.3% (120)     | 25.3% (154)   |
| RYMV      | 26.7% (15)    | 20.1% (259)     | 54.2% (331)     | 34.3% (48)      | 23.8% (67)    |
| CFMV      | 26.3% (19)    | 23.4% (397)     | 59.0% (332)     | 48.3% (89)      | 14.9% (210)   |

a) References of sequence data: SBMV (M23021), LTSV (U31286), RYMV (L20893), and CFMV (Z48630).
b) The percentage values indicate the identity over the stretch of amino acid residues indicated in parentheses.
c) This similarity was found between the N-terminal region of the 56.3 K ORF of CFMV (refer to Fig. 2).
Fig. 3. Schematic representation of RGMoV genome organization. The ORFs are shown as boxes on the genomic RNA (a line with vertical bars, 1000 nt scale). PRO, region coding for serine protease and P3C protease; VPg, a putative region for genome-linked viral protein. Putative E/S cleavage sites around the VPg are shown with vertical lines; POL, region for RNA-dependent RNA polymerase; CP, ORF coding for RGMoV coat protein; F/S with an arrow indicates the characteristic frameshifting site. Shaded box under F/S represents the putative ORF 2-ORF 3 transframe fusion product.

region (amino acids 70 to 90 residues) of the 103.6 kDa polyprotein. The protease motif appears immediately downstream of the conserved sequence: serine protease, in amino acids 148 to 220 from N-terminus and P3C protease, in amino acids 272 to 300 (Fig. 3). The serine protease motif is well conserved between RGMoV, sobemoviruses and poliovirus\(^2\). In addition, the P3C protease motif is completely within the ORF 3. The serine protease motif ... xGxS*/C*GxxxxxxxGxxxxGxH* ... (the catalytic amino acid residue is marked with asterisks), is present just downstream. However, instead of serine (S*) or cysteine (C*), alanine is found in RGMoV. Thus, it is uncertain whether the P3C protease domain is catalytic in RGMoV or not.

The RNA-dependent RNA polymerase is encoded near the C-terminal region of the polyprotein. This region showed very strong similarity, 54 to 60% identity over a 350 amino acid stretch (Table 1). The RNA polymerase motifs\(^3\) are distributed between amino acids 680 to 810. The sequence of this domain is conserved in particular, with approximately 75% identity between RGMoV and sobemoviruses. Database searches also showed that the sequence of RGMoV polymerase is highly conserved between the RNA polymerases of Beet mild yellowing virus (S65829), Cucurbit aphid-borne yellowing virus (X76931), Potato leaf roll virus (X74789) and Barley yellow dwarf virus (L25299) of the family Luteoviridae. The similarity is approximately 50% identity over a 240 amino acid stretch, suggesting an evolutionary close relationships between RGMoV, sobemovirus and luteovirus (subgroup II)\(^5\).

Van der Wilk et al.\(^7\) found that the VPg of SBMV is encoded by ORF 2, downstream of the protease domain and in front of the RNA polymerase. We compared the amino acid sequence similarity between the VPg region of SBMV ORF 2 and the corresponding region of RGMoV ORF 2. The search revealed no significant similarity. Sequence diversities in the VPg region\(^2\) are also shown between SBMV, CfMV and RYMV. However, the conserved sequence, WAG + E/D rich sequence is detected in the region, and putative E/S cleavage sites are present on both sides of the region: proteolytic cleavage would result in a protein of 9 kDa. Possibly, the VPg of RGMoV is located between the protease and the RNA-dependent RNA polymerase domains in the same order as in the SBMV ORF 2\(^2\) (Fig. 3).

RGMoV ORF 3 is completely within the ORF 2. The predicted 19.8 kDa protein has distinct similarity, 40% identity to the corresponding ORFs of SBMV and LTSV. However, it is unknown whether the 19.8 kDa protein is independently translated \textit{in vivo}, because ORF 3 may be expressed as a fusion protein as will be discussed. ORF 4 comprises 198 amino acids encoding a 25.6 kDa coat protein. The 16 amino acid sequence of the N-terminus of the viral coat protein was identical to that deduced from the ORF 4 nucleotide sequence (data not shown). Sequence similarity searches indicated that the RGMoV coat protein revealed a weak but significant similarity, 24 to 27% identity with that of SBMV, LTSV and RYMV, but only 15% identity with CfMV (Table 1).

\textit{In vitro} translation of RGMoV RNA

In the wheat germ extract system, RGMoV RNA directs the synthesis of two products of 103 kDa and 68 kDa, but no other distinct product was detected. In contrast, the translational products synthesized \textit{in vitro} with CfMV/JP RNA are four major proteins with sizes almost identical to those previously reported for CfMV/NO\(^1\) (Fig. 4). The translational activity of CfMV/JP RNA was low in our present system, as reported for other sobemoviruses\(^4\). RGMoV RNA is a poorer message in our wheat germ extract system.

The largest product of RGMoV RNA was 103 kDa and seems to be derived from the largest ORF 2 for the polyprotein. In the RGMoV RNA sequence, no ORF corresponds to the second largest product of 68 kDa. The putative replicase of CfMV is translated as part of a single polyprotein by -1 ribosomal frameshifting
**Figure 4.** *In vitro* translation products made by RGMoV RNA and CfMV RNA. Lane 1, Heat-denatured, 3 μg RGMoV RNA. Lane 2, Heat-denatured, 3 μg RGMoV RNA and the labeled product was immunoprecipitated with antiserum to RGMoV and loaded. Lane 3, Heat-denatured, 3 μg CfMV RNA. Positions of protein size markers are shown on the left.

between two overlapping ORFs having a coding capacity for 60.9 kDa and 56.3 kDa proteins\(^7,18\). Translational frame shifts are known in coronavirus IBV\(^2\), polymerase genes of retroviruses\(^7,19\) and plant viruses\(^7,20\). As consensus signals for frameshifting, the heptanucleotide sequence (e.g., UUUAAAC sequence) and the stem-loop structure immediately downstream have been proposed by Jacks et al.\(^5\). As found in CfMV, SBMV and RYMV\(^7\), identical signals are found in RGMoV RNA just preceding the initiation codon of the ORF 3 (Fig. 5). Tamm et al.\(^18\) proposed a possible mechanism that the 70 kDa *in vitro* translation product of SBMV and RYMV RNAs may represent the ORF 2-ORF 3 transframe fusion protein. Thus, the 68 kDa translational product of RGMoV RNA is probably derived from −1 ribosomal frameshifting (Fig. 3), not from proteolytic cleavage of the polyprotein\(^14\). In this experiment, we tried to detect the RGMoV coat protein in the *in vitro* translation products by immunoprecipitation. However, we could not detect any signal for the coat protein. The coat protein of SBMV is translated only from a smaller, subgenomic RNA, which is detected in virus-infected tissues as well as virus particles\(^14\). As smaller RNAs were not detectable in our RGMoV RNA preparation, the amount of subgenomic RNA, if any, may have been insufficient for the detection of the *in vitro* translated coat protein.

**Figure 5.** Consensus signals for frameshifting in RGMoV and sobemoviruses: heptanucleotide, UUUAAAC and a stem-looped structure. The number pointing to the first nucleotide in a slippery site indicates the actual site in the genome. The initiation codon of the RGMoV ORF 3 is boxed.

We conclude that RGMoV is a member of the genus *Sobemovirus* based on sequence similarities. The similarity level of nucleic acid (approximately 50% identity) and protein (Table 1) is low enough for virus species demarcation between any species of sobemoviruses, whereas the genome organization of RGMoV is closely related among sobemoviruses. Biological and serological properties of RGMoV are distinct from those of other characterized grass viruses\(^20\). Thus, RGMoV is a unique species of the genus *Sobemovirus*\(^4,23\). The polyprotein gene organization of RGMoV is the same as that of SBMV, RYMV and LTSV, but different from that of CfMV, for which a polyprotein is produced as a single fusion-protein by the frameshifting of two ORFs\(^7\).
ACKNOWLEDGMENTS

We wish to thank the late Professor Dr. D. Hosokawa, Tokyo University of Agriculture and Technology, for his encouragement and Dr. T. Teraoka for his help with the amino acid sequence analysis.

LITERATURE CITED

1. Bonneau, C., Brugidou, C., Chen, L., Beachy, R.N. and Fauquet, C. (1998). Expression of rice yellow mottle virus P1 protein in vitro and in vivo and its involvement in virus spread. Virology 244: 79-86.
2. Brierley, I., Bourne, M., Buis, M.M., Biliomotis, B., Blok, V.C., Brown, T.D.K. and Inglis, S.C. (1987). An efficient ribosomal frame-shifting signal in the polymerase-encoding region of the coronavirus IBV. EMBO J. 16: 3779-3785.
3. Gorbatenya, A.E., Koestin, E.V., Buiinov, V.M. and Donchenko, A.P. (1988). Sobemovirus genome appears to encode a serine protease related to cytosine proteases of picornaviruses. FEBS Lett. 236: 287-290.
4. Hull, R. (1995). Genus sobemovirus. In Virus Taxonomy (Murphy, F.A., Fauquet, C.M., Bishop, D.H.L., Ghabrial, S.A., Jarvis, A.W., Martelli, G.P., Mayo, M.A. and Summers, M.D., eds.). pp. 376-378, Springer-Verlag, Wien, New York.
5. Jacks, T., Madhani, H.D., Masiarz, F.R. and Varmus, H. E. (1988). Signals for ribosomal frameshifting in the rous sarcoma virus gag-pol region. Cell 55: 447-458.
6. Jacks, T., Power, M.D., Masiarz, F.R., Luciw, P.A., Barr, P.J. and Varmus, H.E. (1988). Characterization of ribosomal frameshift in HIV-1 gag-pol expression. Nature 331: 280-283.
7. Mikkinen, K., Næs, V., Tamm, T., Truve, E. Aaspõllu, A. and Saarma, M. (1995). The putative replicase of the cocksfoot mottle sobemovirus is translated as a part of the polyprotein by -1 ribosomal frameshift. Virology 207: 566-571.
8. Mikkinen, K., Tamm, T., Næs, V., Truve, E., Puurand, Ü., Mante, T. and Saarma, M. (1995). Characterization of cocksfoot mottle sobemovirus genomic RNA and sequence comparison with related viruses. J. Gen. Virol. 76: 2817-2825.
9. Miller, W.A., Waterhouse, P.M. and Gerlach, W.L. (1988). Sequence and organization of barley yellow dwarf virus genomic RNA. Nucleic Acids Res. 16: 6097-6111.
10. Miller, W.A., Dinesh-Kumar, S.P. and Paul, C.P. (1995). Luteovirus gene expression. Crit. Rev. Plant Sci. 14: 179-211.
11. Ngon, A., Yassi, M., Rizenthaler, C., Brugidou, C., Fauquet, C. and Beachy, R.N. (1994). Nucleotide sequence and genome characterization of rice yellow mottle virus RNA. J. Gen. Virol. 75: 249-257.
12. Othman, Y. and Hull, R. (1995). Nucleotide sequence of the bean strain of southern bean mosaic virus. Virology 206: 287-297.
13. Poch, O., Sauvageot, I., Delarue, M. and Tordo, N. (1989). Identification of four conserved motifs among the RNA-dependent polymerases encoding elements. EMBO J. 8: 3867-3874.
14. Rutgers, T., Salerno-Rife, T. and Kaesberg, P. (1980). Messenger RNA for the coat protein of southern bean mosaic virus. Virology 104: 506-509.
15. Ryabov, E.V., Krujtw, A.A., Novikov, V.K., Zhelezni- kova, O.V., Morozov, S. Yu. and Zavriev, S.K. (1998). Nucleotide sequence of RNA from the sobemovirus found in infected cocksfoot shows a luteovirus-like arrangement of the putative replicate and protease genes. Phytopathology 86: 391-397.
16. Salerno-Rife, T., Rutgers, T. and Kaesberg, P. (1980). Translation of southern bean mosaic virus RNA in wheat embryo and rabbit reticulocyte extracts. J. Virol. 34: 51-58.
17. Takahashi, M., Toriyama, S., Kikuchi, Y., Hayakawa, T. and Inahama, A. (1990). Complementarity between the 5'- and 3'-terminal sequences of rice stripe virus RNAs. J. Gen. Virol. 71: 2817-2821.
18. Tamm, T., Mikkinen, K. and Truve, E. (1999). Identification of genes encoding for the cocksfoot mottle virus proteins. Arch. Virol. 144: 1557-1567.
19. Toriyama, S. (1982). Cocksfoot mottle virus in Japan. Ann. Phytopathol. Soc. Jpn. 48: 514-520.
20. Toriyama, S., Mikoshii, Y. and Doi, Y. (1983). Rye- grass mottle virus, a new virus from Lolium multiflorum in Japan. Ann. Phytopathol. Soc. Jpn. 49: 610-618.
21. Toriyama, S., Takahashi, M., Sanou, Y., Shimizu, T. and Ishihama, A. (1994). Nucleotide sequence of RNA 1, the largest genomic segment of rice stripe virus, the prototype of the tenuivirus. J. Gen. Virol. 75: 3569-3579.
22. Van der Wilk, F., VerbEEK, M., Dullenans, A. and van der Houvel, J. (1998). The genome-linked protein (VPg) of southern bean mosaic virus is encoded by the ORF2. Virus Genes 17: 21-24.
23. Van Regenmortel, M.H.V., Bishop, D.H.L., Fauquet, C. M., Mayo, M.A., MonlofF, J. and Calisher, C.H. (1997). Guidelines to the demarcation of virus species. Arch. Virol. 142: 1505-1518.
24. Wa, S., Rinchart, C.A. and Kaesberg, P. (1987). Sequence and organization of southern bean mosaic virus genomic RNA. Virology 161: 73-80.
25. Zimmern, D. (1988). Evolution of RNA viruses. In RNA Genetics. Vol. 2. Retroviruses, Viroids, and RNA Recombination (Domingo, E., Holland, J.J. and Ahlquist, P., eds.). pp. 211-240, CRC Press, Boca Raton, Florida.