Virology **264,** 344–349 (1999)
Article ID viro.1999.0018, available online at http://www.idealibrary.com on IDEAL®

The Nucleotide Sequence and Genome Organization of *Sclerophthora macrospora* Virus B

Toshiro Yokoi, Yutaka Takemoto, Masashi Suzuki, Shuichi Yamashita, and Tadaaki Hibi[1]

Laboratory of Plant Pathology, Department of Agricultural and Environmental Biology, The University of Tokyo, Yayoi 1-1-1, Bunkyo-ku, Tokyo 113-8657, Japan

Received August 23, 1999; returned to author for revision August 25, 1999; accepted September 21, 1999

Sclerophthora macrospora Virus B (SmV B) found in *S. macrospora,* the pathogenic fungus responsible for downy mildew in gramineous plants, is a small icosahedral, monopartite virus containing a positive-strand ssRNA genome. In the present study, the complete nucleotide sequence of the SmV B genome was determined. The viral genome consists of 5533 nucleotides and has two large open reading frames (ORFs). ORF1 encodes a putative polyprotein containing the motifs of chymotrypsin-related serine protease and RNA-directed RNA polymerase. ORF2 encodes a capsid protein. The deduced amino acid sequence shows some similarity to those of certain positive-strand RNA viruses, but the genome organization is characteristic and distinct from those of other known fungal RNA viruses. These results suggest that SmV B should be classified into a new group of mycoviruses.

INTRODUCTION

Sclerophthora macrospora Virus B (SmV B), first described by Hibi *et al.* (1976), is a 32-nm icosahedral monopartite virus composed of a plus-sense single-stranded (ss) RNA genome of approximately 5.5 kb and a 41-kDa capsid protein (Honkura *et al.,* 1983; Shirako and Ehara, 1985). The host of SmV B is *S. macrospora,* which causes downy mildew in gramineous plants, including rice, wheat, and so forth. The single-stranded nature of the SmV B genome is exceptional, as most mycoviruses have double-stranded RNA genomes (Buck, 1986; Hollings, 1978). Mushroom bacilliform virus (MBV) is the only known ssRNA mycovirus (Tavantzis *et al.,* 1980; Hollings, 1982), but SmV B differs from MBV in particle shape because MBV is a bullet-shaped particle (Hollings, 1962).

Another species of virus designated SmV A has been found also in *S. macrospora.* SmV A is a 30-nm icosahedral virion with spikes of 4 nm on the periphery and is serologically distinct from SmV B (Honkura *et al.,* 1983). Either or both viruses are detected frequently in downy mildew-diseased gramineous plants, but not in all cases (Shirako and Ehara, 1985). These viruses have been found abundantly in the cytoplasm of fungal oospores and mycelia, and the amount of the virus in the cells was found to decrease with increased vacuolation in the cytoplasm as the cells aged (Ehara, 1989). The role of these viruses in downy mildew disease and the phenotypic changes evoked in *S. macrospora* are unclear. In this study, the complete nucleotide sequence and genome organization of SmV B RNA were determined and compared with those of some other positive-strand ssRNA viruses.

RESULTS AND DISCUSSION

Nucleotide sequence

An RT–PCR product corresponding to the 5′ end of SmV B RNA was not obtained when the RNA was treated beforehand with bacterial alkaline phophatase (BAP) and tobacco acid pyrophophatase (TAP) followed by oligoribonucleotide ligation, although the product was obtained when BAP and TAP treatments were omitted, indicating that the 5′ end of the viral RNA was most likely uncapped (data not shown). The nucleotide sequence of the 3′ end of the SmV B RNA was determined by sequencing several independent 3′-terminal RT–PCR clones, and the results showed that the 3′ end of the RNA genome was not polyadenylated.

The sequence of the noncoding junction region between the two open reading frames (ORFs) was found to be an A-rich (43%) region. As some differences were observed in this region, we obtained several clones containing this region by RT–PCR and sequenced a number of them (Fig. 1). More than half of the clones showed an identical sequence, but the other clones showed a difference in three bases with A instead of U at nucleotide positions 3929, 3936, and 3940, and the amino acid sequence deduced from the nucleotide sequence of these different clones was four residues smaller than that of the majority. Furthermore, the A-rich region (nts 3929–3945) of some of these different clones contained

The nucleotide sequence data reported in this paper have been deposited in the DDBJ, GenBank, and EMBL nucleotide sequence databases under Accession No. AB012756.

[1] To whom correspondence and reprint requests should be addressed. Fax: +81 3 5841 5090. E-mail: akihibi@ims.u-tokyo.ac.jp.

0042-6822/99 $30.00

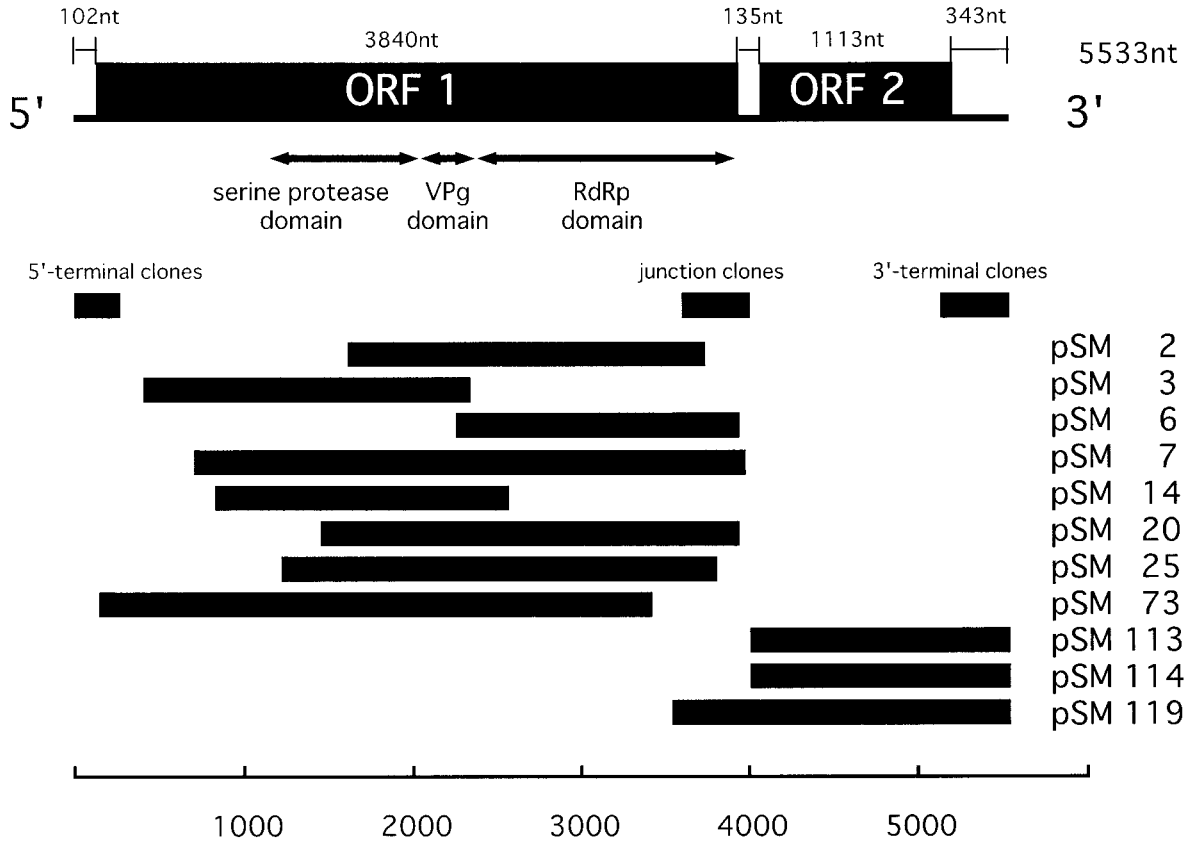

FIG. 1. The genome organization of SmV B and alignment of the cDNA clones derived from SmV B genomic RNA. The number of nucleotides is indicated by the scale at the bottom of the figure.

a large number of additional A residues comprising internal poly(A) tracts.

The complete SmV B genome consists of 5533 nt, with a base composition of A (25.7%), C (21.5%), G (25.1%), and U (27.8%).

Genome organization

Analysis of the determined sequence of SmV B RNA revealed the existence of two ORFs (Fig. 1). The putative polyprotein encoded by ORF1 has a calculated molecular mass of 145 kDa. Analysis of the predicted amino acid sequence of ORF1 revealed the presence of the putative RNA-directed RNA polymerase domain containing the motif (D_{X3}[FYWLCA]$_{X0-1}D_{Xn}$[STM]$G_{X3}T_{X3}$[NE]$_{Xn}$[GS]DD) known to occur in other positive-strand RNA viruses (Koonin and Dolja, 1993) near the C terminus of the ORF1 protein. The central part of the SmV B ORF1 product, upstream of the putative RdRp domain, represents a domain with a motif typical of chymotrypsin-related serine protease found in some positive-strand RNA viruses such as sobemovirus (Gorbalenya *et al.,* 1988).

The genome has a 5′-nontranslated leader sequence (5′-NTR) of 102 nucleotides and a 3′-terminal NTR of 343 nucleotides.

Capsid protein

Direct sequencing of the N terminus of SmV B capsid protein and comparison with the deduced amino acid sequences of the ORFs showed that the capsid protein was coded by ORF2. The amino acid sequence (TKKNVKTSGAIVVYKPQTRL . . .) obtained by Edman degradation matches the deduced amino acid sequence of ORF2 at the N terminus (MTKKNVKTSGAIVVYKPQTRL . . .). The molecular mass of the predicted protein encoded by ORF2 (38.7 kDa) is similar to that obtained by polyacrylamide gel electrophoresis analysis of the protein in a purified virion preparation (41 kDa).

Comparisons of ORFs and genome organization with those of other viruses

The putative RdRp domain containing the GDD motif in ORF1 shows similarity to those of viruses of the "sobemovirus lineage in supergroup 1" among the positive-strand RNA viruses classified by Koonin *et al.* (1993). After pairwise alignment for maximum matching, sequence identities of 17–22% were observed between the putative RdRp domain of SmV B and those of some other viruses, including MBV (*barnavirus*), southern bean mosaic virus (SBMV) (*sobemovirus*), and potato leafroll virus

(PLRV) (*luteovirus* subgroup II). Multiple alignment with those viruses shows that the eight RdRp motifs identified by Koonin and Dolja (1993) are highly conserved (Fig. 2A).

When the putative serine protease domain was compared with those of MBV, SBMV, and PLRV (Fig. 2B) near the third putative catalytic residue, SmV B shows significant identity with MBV (STLRGWSG . . .). The domains of VPg of PLRV (Van der Wilk *et al.,* 1997), SBMV (Van der Wilk *et al.,* 1998), and MBV (Revill *et al.,* 1998) are located between the putative serine protease domain and the polymerase domain. A sequence partially similar to those domains (amino acid sequence identity of approximately 19–21%) was detected in the same region of SmV B ORF1 (Fig. 2C). The amino acid sequence of ORF2, encoding the virus capsid protein, does not show significant similarity to proteins of any other viruses.

The genome arrangement of SmV B is similar to those of *sobemovirus, barnavirus,* and *luteovirus* subgroup II. The putative serine protease domains, the VPg domains, the putative RdRp domains, and the capsid protein domains are located in this order from the 5′ terminus to the 3′ terminus. However, as there are only two ORFs in the genome of SmV B, the genome structure of SmV B is distinctive, whereas that of MBV resembles that of *luteovirus* subgroup II (Revill *et al.,* 1994) (Fig. 3). As for both termini, poly(A) tails are lacking at the 3′ termini of those viruses, while VPgs are covalently linked to the RNA of the viruses. The 5′ terminus of SmV B RNA is uncapped, but it is possible that VPg may be linked to the RNA because a region homologous to the VPg domain was detected in ORF1 of SmV B RNA.

In conclusion, SmV B differs from the other known mycoviruses in terms of particle shape and the nature of its genome. Notably, the genome organization of SmV B does not resemble that of any other mycoviruses. These results suggest that SmV B should be classified into a new group of mycoviruses.

Further investigation is required to understand the expression of the ORFs of SmV B and to identify their products. The replication strategy of SmV B and its role in downy mildew disease also remain to be elucidated.

MATERIALS AND METHODS

Virus purification and nucleic acid extraction

Wheat plants that exhibited typical downy mildew symptoms, such as stunting and chlorotic flecks on the foliage, were collected from Saitama Prefecture, Japan, in 1995. SmV B was purified from these diseased leaves by differential centrifugation followed by sucrose density gradient centrifugation as described previously (Honkura *et al.,* 1983). Immunoelectron microscopy was employed for detection and identification of the virus. Nucleic acid was extracted from purified virions by treatment with 1% sodium dodecyl sulfate (SDS) and 0.2 mg/ml proteinase K followed by phenol/chloroform extraction, and the size was determined by electrophoresis through denaturing formaldehyde–agarose gels (Sambrook *et al.,* 1989).

cDNA synthesis and cloning

Synthesis of cDNA from SmV B RNA was carried out using the cDNA synthesis module RPN 1256 (Amersham) according to the manufacturer's instructions. Briefly, first-strand cDNA was synthesized with random hexanucleotide primers and AMV reverse transcriptase, and second-strand DNA was synthesized with *Escherichia coli* DNA polymerase I in the presence of RNase H. After the dsDNA was blunt-ended with T4 DNA polymerase, the resulting DNA fragment was ligated into the *Hin*cII site of the plasmid vector pUC118.

DNA sequencing and sequence analysis

The inserts from eight selected cDNA clones (pSM 2, 3, 6, 7, 14, 20, 28, and 73) were sequenced (both strands) (Fig. 1). Sequence analysis was performed with an Applied Biosystems Model 377 automated sequencer, employing a cycle-sequencing kit with AmpliTaq DNA polymerase (Applied Biosystems). Multiple alignments for sequence comparison with other viruses were done using CLUSTAL W (Thompson *et al.,* 1994) as incorporated into DNASIS version 3.6 for Macintosh (Hitachi Software). GenBank searches were done using the BLAST programs (Altschul *et al.,* 1990).

Determination of the 5′- and 3′-terminal nucleotides

To obtain cDNA clones representing the 5′ end of the viral RNA and to determine whether the 5′ end has a cap structure, SmV B RNA, which was treated beforehand with BAP and subsequently with TAP, was ligated to an oligoribonucleotide (r-oligo) (5′-GAGCACUGUUGGCCUACUGG-3′) using T4 RNA ligase (Maruyama and Sugano, 1994). As a control, untreated SmV B RNA was also ligated to r-oligo. After this procedure, the sequence of the 5′ end was amplified by RT–PCR using OSP primer (5′-GAGCACTGTTGGCCTACTGG-3′; homologous to the r-oligo sequence) and S5-GSP primer (5′-GGGGCCCATAAACCTTTTG-3′, complementary to the viral RNA sequence at nucleotide positions 483–501), and the PCR product was cloned into plasmid pGEM-T (Promega).

The 3′-terminal sequence of SmV B RNA was determined by poly(A) tailing using poly(A) polymerase. cDNA was synthesized with oligo(dT)Sph anchor primer (5′-AACTGGAAGAATTGCATGCAGGAA$(T)_{18}$-3′), the 3′-end sequence was amplified by RT–PCR using TSP primer (5′-TGGAAGAATTGCATGCAGGAA-3′, homologous to the anchor sequence of oligo(dT)Sph anchor primer) and S-GSP primer 1 (5′-GTTCTCGAAGCACCTCGAAA-3′, homologous to the viral RNA sequence at nucleotide positions 3654–3673) or S-GSP primer 2 (5′-GGTGTCTTTGACACCCTGTG-3′; 3951–3970), and the PCR product di-

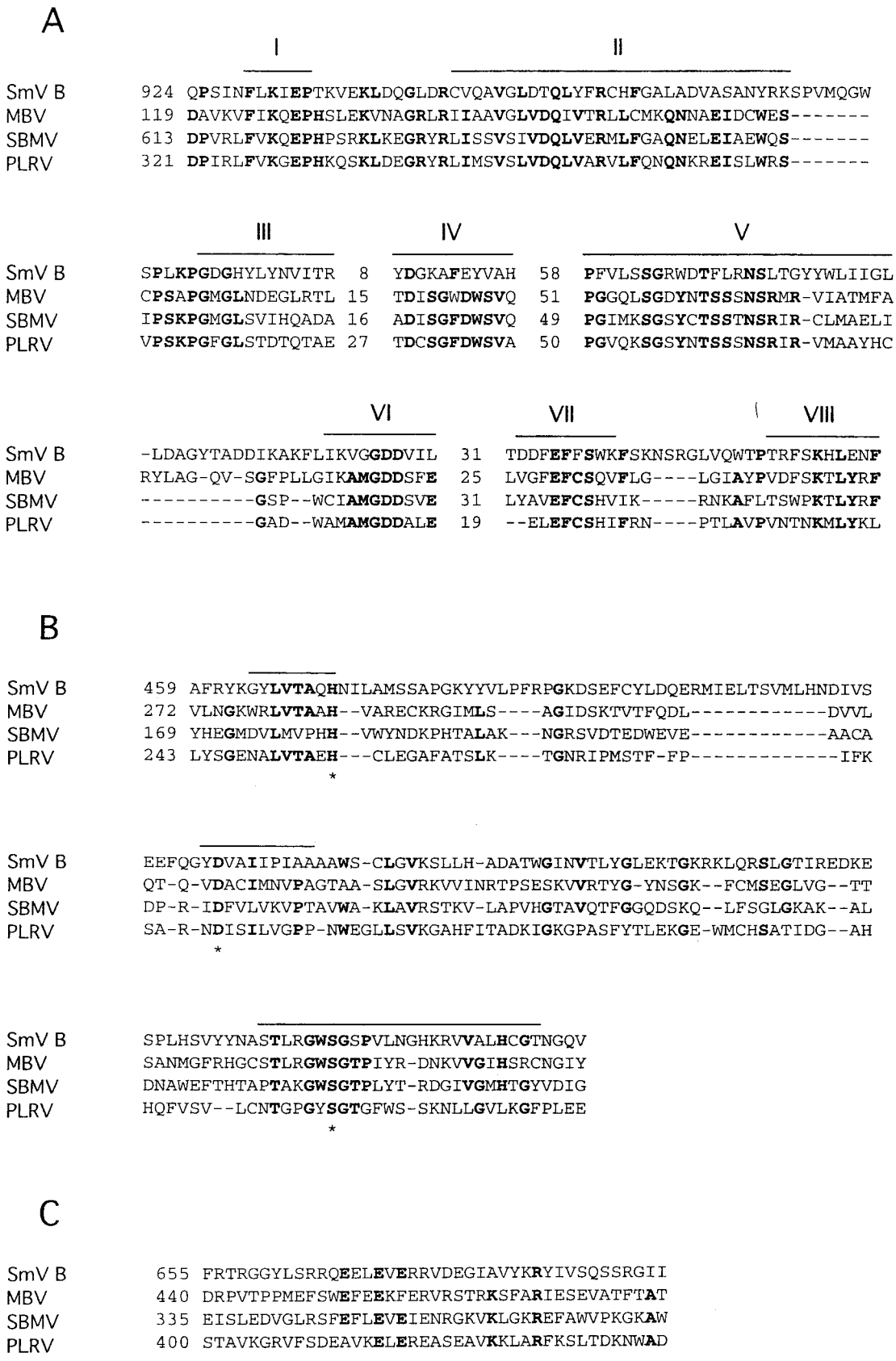

FIG. 2. Amino acid alignment of the putative RdRp domains (A), the putative serine protease domains (B), and the VPg domains (C) of SmV B, mushroom bacilliform virus (MBV) (Revill *et al.,* 1994), southern bean mosaic virus (SBMV) (Wu *et al.,* 1987), and potato leafroll virus (PLRV) (Mayo *et al.,* 1989). Gaps are introduced for better alignment. Numbers on the left refer to the amino acid position in the ORF. The conserved motifs of protease and RdRp domains according to Koonin and Dolja (1993) are overlined. The putative catalytic residues are shown by asterisks. Residues conserved in at least three of the viruses are shown in bold.

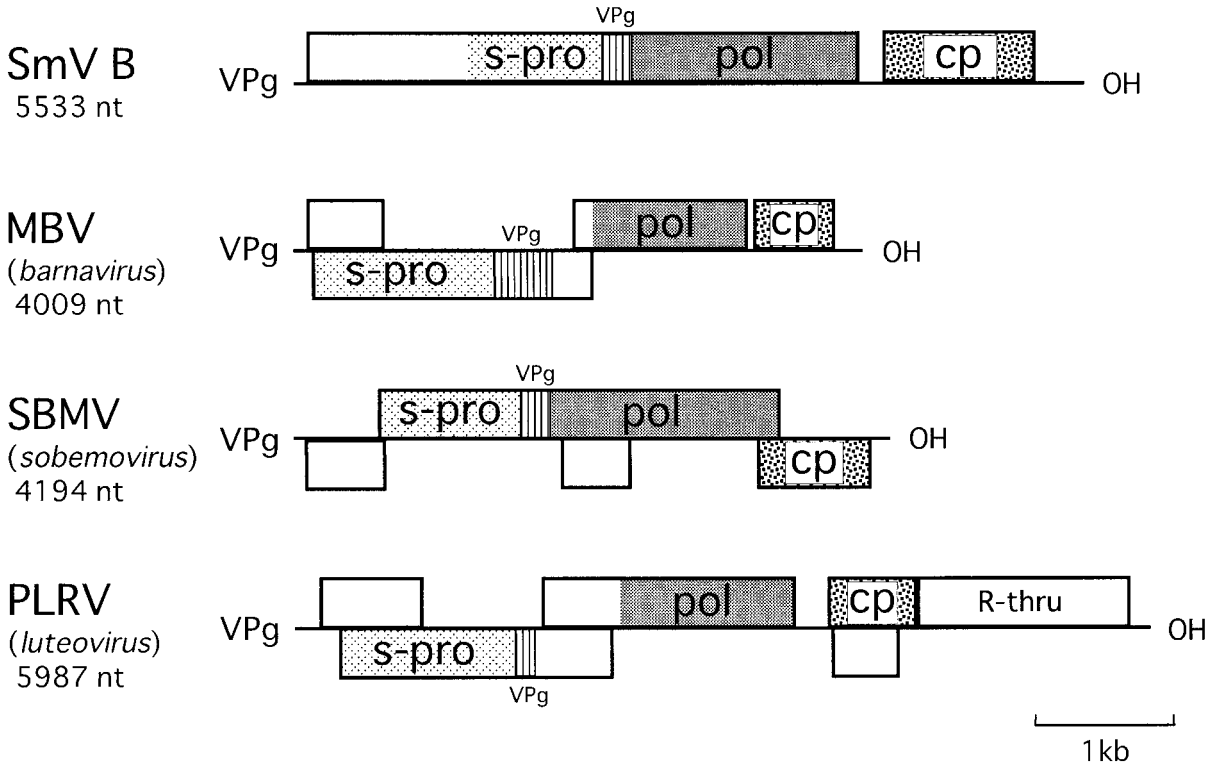

FIG. 3. Comparison of the genome organization of SmV B, SBMV, MBV, and PLRV. S-pro, putative chymotrypsin-related serine protease domain; pol, putative RNA-directed RNA polymerase domain; cp, capsid protein; R-thru, domain expressed by translational readthrough.

gested by the combination of *Sph*I and *Nde*I or *Sph*I and *Eco*RV was cloned into pGEM 5.

For accurate identification of the 3′ end of the viral RNA, the anchor oligonucleotide (5′-AGCCGCATGCA-CAGGCCCATGTACAGGAAGCC-3′), phosphorylated at the 5′ end and blocked by ddATP at the 3′ end, was ligated to the 3′ end of the viral RNA using T4 RNA ligase (Weng and Xiong, 1995). After this procedure, the 3′ end sequence was amplified by RT–PCR using ASP primer (5′-TTGTGAGCTCGGCTTCCTGTACATGGGCCTGT-3′; sequence complementary to the anchor oligonucleotide is underlined) and S-GSP primer 1, and the PCR product digested by *Sph*I and *Xba*I was cloned into pUC 18.

Amino acid sequencing

The purified virus was dissolved in 62.5 mM Tris–HCl (pH 6.8) containing 2.5% SDS, 5% 2-mercaptoethanol, and 10% glycerol and then heated at 100°C for 3 min. The viral protein was electrophoresed through a SDS–polyacrylamide gel (Laemmli, 1970; Hirano, 1989), visualized by staining with Coomassie brilliant blue (Matsudaira, 1987), and blotted onto an Immobilon P membrane (Millipore); the protein band was then excised from the membrane and sequenced directly by sequential Edman degradation using a protein sequencer (Applied Biosystems 477).

ACKNOWLEDGMENTS

We gratefully thank Professor Y. Shirako of The University of Tokyo for supplying antiserum to SmV B and Mr. S. Tomioka for the amino acid sequence.

REFERENCES

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* **215,** 403–410.

Buck, K. W. (1986). Fungal virology—An overview. *In* "Fungal Virology" (K. W. Buck, Ed.), pp. 1–84. CRC Press, Boca Raton, FL.

Ehara, Y. (1989). Cellular localization of virus particles associated with downy mildew-disease rice plants. *J. Phytopathol.* **124,** 107–111.

Gorbalenya, A. E., Koonin, E. V., Blinov, V. M., and Donchenko, A. P. (1988). Sobemovirus genome appears to encode a serine protease related to cysteine proteases of picornaviruses. *FEBS Lett.* **236,** 287–290.

Hibi, T., Komuro, Y., Doi, Y., and Yora, K. (1976). *Ann. Phytopathol. Soc. Jpn.* **42,** 106 (Abstr. in Japanese).

Hirano, H. (1989). Microsequence analysis of the winged bean seed proteins electroblotted from two-dimensional gel. *J. Prot. Chem.* **8,** 115–130.

Hollings, M. (1962). Viruses associated with dieback disease of cultivated mushrooms. *Nature* **196,** 962–965.

Hollings, M. (1978). Mycoviruses: Viruses that infect fungi. *Adv. Virus Res.* **22,** 1–53.

Hollings, M. (1982). Mycoviruses and plant pathology. *Plant Dis.* **66,** 1106–1112.

Honkura, R., Shirako, Y., Ehara, Y., and Yamanaka, S. (1983). Two types of virus-like particles isolated from downy mildew diseased rice plants. *Ann. Phytopathol. Soc. Jpn.* **49,** 653–658.

Koonin, V. K., and Dolja, V. V. (1993). Evolution and taxonomy of positive-

strand RNA viruses: Implications of comparative analysis of amino acid sequences. *Crit. Rev. Biochem. Mol. Biol.* **28,** 375–430.

Laemmli, U. K. (1970). Cleavage of structural proteins during assembly of the head of bacteriophage T4. *Nature* **227,** 680–685.

Maruyama, K., and Sugano, S. (1994). Oligo-capping: A simple method to replace the cap structure of eukaryotic mRNAs with oligoribonucleotides. *Gene* **138,** 171–174.

Matsudaira, P. (1987). Sequence from picomole quantities of proteins electroblotted onto polyvinylidene difluoride membranes. *J. Biol. Chem.* **262,** 10035–10038.

Mayo, M. A., Robinson, D. J., Jolly, C. A., and Hyman, L. (1989). Nucleotide sequence of potato leafroll luteovirus RNA. *J. Gen. Virol.* **70,** 1037–1052.

Revill, P. A., Davidson, A. D., and Wright, P. J. (1994). The nucleotide sequence and genome organization of mushroom bacilliform virus: A single-stranded RNA virus of *Agaricus bisporus* (Lange) Imbach. *Virology* **202,** 904–911.

Revill, P. A., Davidson, A. D., and Wright, P. J. (1998). Mushroom bacilliform virus RNA: The initiation of translation at the 5′ end of the genome and identification of the VPg. *Virology* **249,** 231–237.

Sambrook, J., Fritsch, R. F., and Maniatis, R. (1989). "Molecular Cloning: A Laboratory Manual." Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Shirako, Y., and Ehara, Y. (1985). Composition of viruses isolated from *Sclerophthora macrospora* infected rice plants. *Ann. Phytopathol. Soc. Jpn.* **51,** 459–464.

Tavantzis, S. M., Romaine, C. P., and Smith, S. H. (1980). Purification and partial characterization of a bacilliform virus from *Agaricus bisporus* Mushroom: A single-stranded RNA mycovirus. *Virology* **105,** 94–102.

Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994). CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22,** 4673–4680.

Van der Wilk, F., Verbeek, M., Dullemans, A. M., and Van den Heuvel, J. F. J. M. (1997). The genome-linked protein of potato leafroll virus is located downstream of the putative protease domain of the ORF1 product. *Virology* **234,** 300–303.

Van der Wilk, F., Verbeek, M., Dullemans, A., and Van den Heuvel, J. (1998). The genome-linked protein (VPg) of southern bean mosaic virus is encoded by the ORF2. *Virus Genes* **17,** 21–24.

Weng, Z., and Xiong, Z. (1995). A method for accurate determination of terminal sequences of viral genomic RNA. *Genome Res.* **5,** 202–207.

Wu, S., Rinehart, C. A., and Kaesberg, P. (1987). Sequence and organization of southern bean mosaic virus genomic RNA. *Virology* **161,** 73–80.