Plant Pathol. J. 19(3) : 171-176 (2003)

Complete Genome Sequences of the Genomic RNA of *Soybean mosaic virus* Strains G7H and G5

Won-Seok Lim, Yul-Ho Kim[1] and Kook-Hyung Kim*

School of Agricultural Biotechnology and Center for Plant Molecular Genetics and Breeding Research, Seoul National University, Suwon 441-744, Korea

[1]National Crop Experimental Station, Rural Development Administration, Suwon 441-100, Korea

(Received on May 10, 2003; Accepted on June 18, 2003)

The complete nucleotide sequences of the genomic RNAs of *Soybean mosaic virus* strains G5 (SMV-G5) and G7H (SMV-G7H) were determined and compared with sequences of other SMV strains. Each viral RNA was determined to be 9588 nucleotides in length excluding the poly (A) tail and contained an open reading frame to encode a polyprotein subsequently processed into up to ten proteins by proteolytic cleavage. Comparison of the amino acid sequences with those of other SMV strains showed high percentage of amino acid sequence homology with the same genome organization. The nucleotide and the deduced amino acid sequences between SMV-G5 and SMV-G7H were greater than 99% identity. When compared with those of other SMV strains in a phylogenetic analysis of the nucleotide and deduced amino acid sequences, they formed a distinct virus clade showing over 97% amino acid identity, but were more distantly related to the other potyvirus (44.1-69.6% identity). Interestingly, SMV G7H strain caused a severe mosaic or necrosis symptom in soybean cultivars including Jinpum-1, Jinpum-2, and Sodam, whereas, no symptom was observed in SMV-G5 inoculation. Complete nucleotide sequences of these strains will give clues for determining symptom determinant(s) in future research.

Keywords : SMV, strains G5 and G7H, genome sequence, phylogeny

Soybean mosaic virus (SMV), which causes soybean mosaic disease, is a member of the genus *Potyvirus* in the family *Potyviridae*. The potyviral genome is a single-stranded, positive-sense RNA of approximately 10,000 nucleotides (nt) in length. The viral genomic RNA has a 5'-terminal genome-linked viral protein (VPg) and contains a 3'-poly(A) tail (Dougherty and Carrington, 1988). The genome contains one large open reading frame (ORF) translated into a large polyprotein, which is subsequently cleaved into mature polypeptides including P1 (protease), the helper component/protease (HC-Pro); P3 protein, a 6 kDa peptide (6K1), the cylindrical inclusion (CI) protein with RNA helicase activity, a second 6 kDa peptide (6K2), the VPg, nuclear inclusion 'a' protein (NIa), the nuclear inclusion 'b' protein (NIb) with RNA dependant RNA polymerase (RdRp) activity, and the coat protein (CP) (Dougherty and Semler, 1993; Reichmann et al., 1992). SMV can be transmitted mechanically by aphids and by seeds and, therefore, is considered as the most economically important plant virus in soybean [*Glycine max* (L.) Merr] causing mosaic and severe necrosis in many soybean cultivars. SMV has been first described in 1921 and reported in many countries throughout the world (Gardner and Kendrick, 1921; Jayaram et al., 1992; Liu et al., 1993).

Since the first description of seven strain groups (G1-G7) of SMV based on symptom development of each isolate on five cultivars including 'Buffalo', 'Ogden', 'Marshall', 'Davis', and 'Kwanggyo' by Cho and Goodman (1979, 1982), diversities and strain variations and their resistance on different soybean cultivars have been studied extensively (Gunduz et al., 2001, 2002; Jayaram et al., 1992; Lim, 1985). In Korea, diseases caused by SMV and the viral nature of the cause have been reported in the early 1970s (Chung et al., 1973, 1974). Cho and Chung (1976) reported the difference in symptom development depending on soybean cultivars and virus isolates. SMV strain G5 was the most prevalent SMV strain (about 80%) infecting soybean cultivars in the early 1980s, but the disease incidence caused by SMV-G5 was decreased as resistant soybean cultivars were grown in the fields (Kim, 2000). Recently, a new SMV strain, G7H, causing mosaic and necrosis on recommended soybean cultivars that are resistant to SMV strains G5 and G5H was found (Kim, 2000). Presence of many SMV strains has been reported and the continual variation is generally believed to happen in soybean fields in Korea (Cho et al., 1983). This study reported the complete

*Corresponding author.
Phone) +82-31-290-2441, FAX) +82-31-294-5881
Email) kookkim@snu.ac.kr

nucleotide sequence of the two SMV strains G5 and G7H, and the phylogenetic relationships with other reported SMV strains.

Materials and Methods

Virus strains and purification. Soybean (Dawonkong) was inoculated with SMV G7H and G5 strains and with carborundom and was grown in the glasshouse. After 15-day post-inoculation (dpi), infected leaves were harvested. Virus purification was performed following the procedures outlined by Ross (1966). For viral RNA extraction, 0.1 vol of 10% SDS was added to virus preparation and samples were extracted twice with phenol: chloroform: isoamyl alcohol (25:24:1) and twice with chloroform. Viral RNA was ethanol precipitated and re-suspended in DEPC-treated water.

Genomic RNA sequencing. Total RNAs were extracted from infected tissues using TRIZOL (Gibco BRL, USA) according to the manufacturer's protocol. RT-PCR and cloning were performed as described previously (Jung et al., 2002; Yun et al., 2002). PCR primers for amplification were designed based on the sequences of the SMV strains G2, G7 and N (Table 1). cDNA clone containing the 5' end of genome was obtained by following the 5'/3' RACE kit protocol (Boehringer Mannheim, Germany) according to the manufacturers instructions. cDNA was synthesized using SMV sequence specific oligonucleotide primer (5'-GTCTTCCTCTTCTTGCCTTCGCTTTG-3') complementary to nt 469 to 494.

Recombinant plasmids containing cDNA inserts were sequenced by the dideoxynucleotide chain termination method using the ABI prismTM Terminator Cycle Sequencing Ready Reaction Kit and an ABI Prism 3700 Genetic Analyzer (Perkin Elmer, USA) located at the NICEM (SNU) according to the manufacturer's instructions. Every base was determined by sequencing at least two independent clones or by sequencing twice from a single clone. The complete genome sequences of the SMV strain G5 and G7H have been deposited in the GenBank under accession numbers AY294044 and AY294045, respectively.

Phylogenetic analysis. The sequence data were analyzed using the LaserGene program (DNASTAR, USA). Comparative analysis including multiple alignments of nucleotide sequences was done by the ClustalW method with weighted residue weight table and the phylogeny was analyzed using PAUP software (Sinauer

Table 1. SMV specific primers used for cloning and sequencing

Primer	Sequence	Position	Purpose
prS1(+)	5'-GCAGTTTCACWTTTCCTCAC-3'	96-115	Cloning & Seq.
prS1(−)	5'-CCCATGACAGCCCATATGAG-3'	1296-1315	
prS2(+)	5'-AAGCCAATCAATCTTTCCAG-3'	1188-1207	
prS2(−)	5'-CCAAAAGAGTCAATCACGTG-3'	2322-2341	
prS3(+)	5'-TAGCAACAGCTGCATACATG-3'	2207-2226	
prS3(−)	5'-GCAGCACACTAGTCATTTGG-3'	3310-3329	
prS4(+)	5'-CAGGTGCTACAGTGATATAG-3'	3168-3187	
prS4(−)	5'-GAGCTTGCACRAAGTTCTGA-3'	4311-4330	
prS5(+)	5'-YCAACSATTGCATTCAACTG-3'	4168-4187	
prS5(−)	5'-AGYTGCAAAAATTTACTCAC-3'	5538-5557	
prS6(+)	5'-TTGATTGCAACAAAGTTGAC-3'	5450-5469	
prS6(−)	5'-TTGTGTATCACAAACTCACC-3'	6439-6458	
prS7(+)	5'-ATGTTTGGGGTYGGCTATGG-3'	6349-6368	
prS7(−)	5'-TCAGGGATCCATTCCACACT-3'	7479-7498	
prS8(+)	5'-CGCTTGCTTTATCTCAGTTG-3'	7370-7389	
prS8(−)	5'-TATCTTTGTAGCTCACTTGC-3'	8457-8476	
prS9(+)	5'-CACCATATATTGCAGAGA-3'	8402-8421	
prS9(−)	5'-GAAATGATAACTGTGRCC-3'	9348-9367	
prS5end	5'-AAATTAAAACTMSYYATAAAGACAACAAAC-3'	1-30	
prS3end	5'-GCTCTAGAT$_{14}$AGGACAACAAACATTGC-3'	9572-9588	
oligodT$_{18}$	5'-TTTTTTTTTTTTTTTTTT		
prS1(−)	5'-TCAACCACTTCAAGAGCG-3'	1014-1031	Seq.
prS2(−)	5'-AGATCAATGTATTTCGGG-3'	2019-2036	
prS3(−)	5'-TTTCCTGAAGATGCGCTG-3'	3018-3035	
prS4(−)	5'-CAATCCTCTCATTCTCAG-3'	4021-4038	
prS5(−)	5'-CAACTTGTCTGGTATTCC-3'	5113-5130	
prS6(−)	5'-GGCAGAGAAGTGTGGGTC-3'	6119-6136	
prS7(−)	5'-GGCTTGGTTGATACGCAC-3'	7154-7171	
prS8(−)	5'-GCTCCTGTCCCACTCTAG-3'	8239-8256	

Table 2. Nucleotide and amino acid sequence identity (%) between SMV-G7H and other potyviruses

Virus	Amino acid										nucleotide				Genbank Accession#
	Poly-Protein	P1	HC-Pro	P3	6K1	CI	6K2	NIa-VPg	NIa-Pro	NIb	CP	Full-Length	5'NTR	ORF	3'NTR	
SMV-G5	99.0	99.7	99.3	99.4	100	97.6	100	97.9	99.2	99.0	100	97.8	96.9	97.8	99.2	
SMV-G2	97.8	92.2	98.5	98.0	100	98.3	100	97.9	99.2	97.9	99.2	96.4	94.7	96.5	94.5	S42280
SMV-G7	96.4	89.6	97.4	94.2	98.1	96.8	100	98.9	99.2	96.7	98.9	94.0	92.4	94.1	91.7	AF241739
SMV-N	98.6	97.4	98.0	97.7	100	99.2	100	98.4	99.6	98.6	98.9	96.6	93.1	96.7	95.3	NC-002634
SMV-Aa	97.2	94.5	98.5	95.1	98.1	96.8	100	98.9	99.6	96.7	98.9	93.9	94.7	94.0	91.3	AB100442
SMV-Aa15-M2	97.1	94.2	98.2	95.1	98.1	96.7	100	98.9	99.6	96.7	98.9	93.9	95.4	93.9	90.9	AB100443
BCMV	69.6	38.5	72.6	51.0	76.9	77.9	62.3	78.4	80.7	77.3	77.7	67.0	31.5	65.3	66.8	NC-003397
ZYMV	62.1	23.9	68.2	39.6	63.5	71.8	62.3	65.8	67.1	74.7	71.3	59.3	42.0	61.4	5.6	NC-003224
PRSV	44.5	19.1	46.4	22.3	32.7	53.8	30.2	49.2	45.4	55.7	55.5	48.4	69.4	50.5	17.5	NC-001785
PVY	44.1	14.4	42.1	21.0	44.2	54.3	23.1	46.3	42.0	57.1	60.4	47.7	55.0	37.3	7.5	NC-001616

Associates, USA). Databases were searched using the BLAST program (Altschul, 1997). The sequences of previously reported SMV strains were retrieved from the GenBank database and used for analyses. All data sets were subjected to bootstrap analysis by performing 1000 replications by neighbor-joining method and used to build a phylogenetic tree using the TreeView Win32 program (Page, 1996).

Results and Discussions

This paper presents the complete nucleotide sequence of the SMV strains G5 and G7H. Comparison of the nucleotide sequence, as well as the deduced amino acid sequence with those of other SMV strains shows a similar genomic organization and high sequence similarities. Genome sequences of the SMV G7H strain causing severe mosaic or necrosis symptom in certain soybean cultivars where no symptom was observed with SMV-G5 inoculation will give clues for determining symptom determinant(s) in future research.

Complete nucleotide sequences of SMV G7H and SMV G5 strain. The complete RNA genome of SMV G5 and G7H was determined to be 9588 nucleotides in length, excluding the poly (A) tail. The overall base compositions of the sequence of SMV G7H and G5 were found to be 31.82% A, 24.04% G, 26.29% T and 17.85% C; and 31.76% A, 24.03% G, 26.30% T and 17.91% C, respectively. Analysis of putative ORF shows a single large ORF (nt 132-9381), encoding a 3067 amino acid polypeptide of 349.9 kDa. The 5' and 3' non-translated regions (NTRs) were determined as 131 and 253 nucleotides in length. Computer analysis of the G5 and G7H RNA sequences revealed no other ORF in neither plus-strand nor minus-strand RNAs. The nucleotide sequence and genome organization of both G5 and G7H strains were similar to that of other previously reported SMV strains (Table 2). The overall sequence identity of the seven SMV genomes

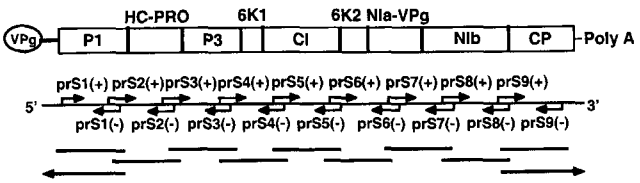

Fig. 1. Schematic representation of the genome organization of SMV and nucleotide sequencing strategy. Open bars represent an open reading frame (ORF) encoding the polyprotein. The vertical line within ORF represents mature proteins produced from proteolytic process. The non-translated regions are indicated as single line. Horizontal lines and arrows represent clones obtained by RT-PCR using each primer set and 5'/3' RACE, respectively.

was over 93.9%. In general, the nucleotide sequence in the coding region was more conserved than the sequences in the 5' and 3' NTR. The nucleotide sequence identity between SMV and the other potyviruses, however, was only 47.7-67.0%.

Amino acid sequence analysis. The polyprotein of a potyvirus is cleaved by the virus-encoded protease P1, HC-Pro, and NIa, resulting to at least nine functional, mature protein (Shukla et al., 1994). The amino acid sequence was deduced from nucleotide sequence and used for amino acid similarity analysis. The amino acid sequence identity of polyprotein among SMV strains was over 96.4%. The deduced amino acid sequences of the 6K1, 6K2, NIa-Pro, and CP showed higher sequence identity (over 98.9%) while those of the other proteins were lower (Table 2). The deduced amino acid sequences of the 6K1, 6K2, and CP were 100% identical while CI and NIa-VPg contained the most different sequences sharing 97.6% and 97.9% sequence identity, respectively, between G7H and G5.

SMV G7H strain caused severe mosaic or necrosis symptom in certain soybean cultivars including Jinpum-1,

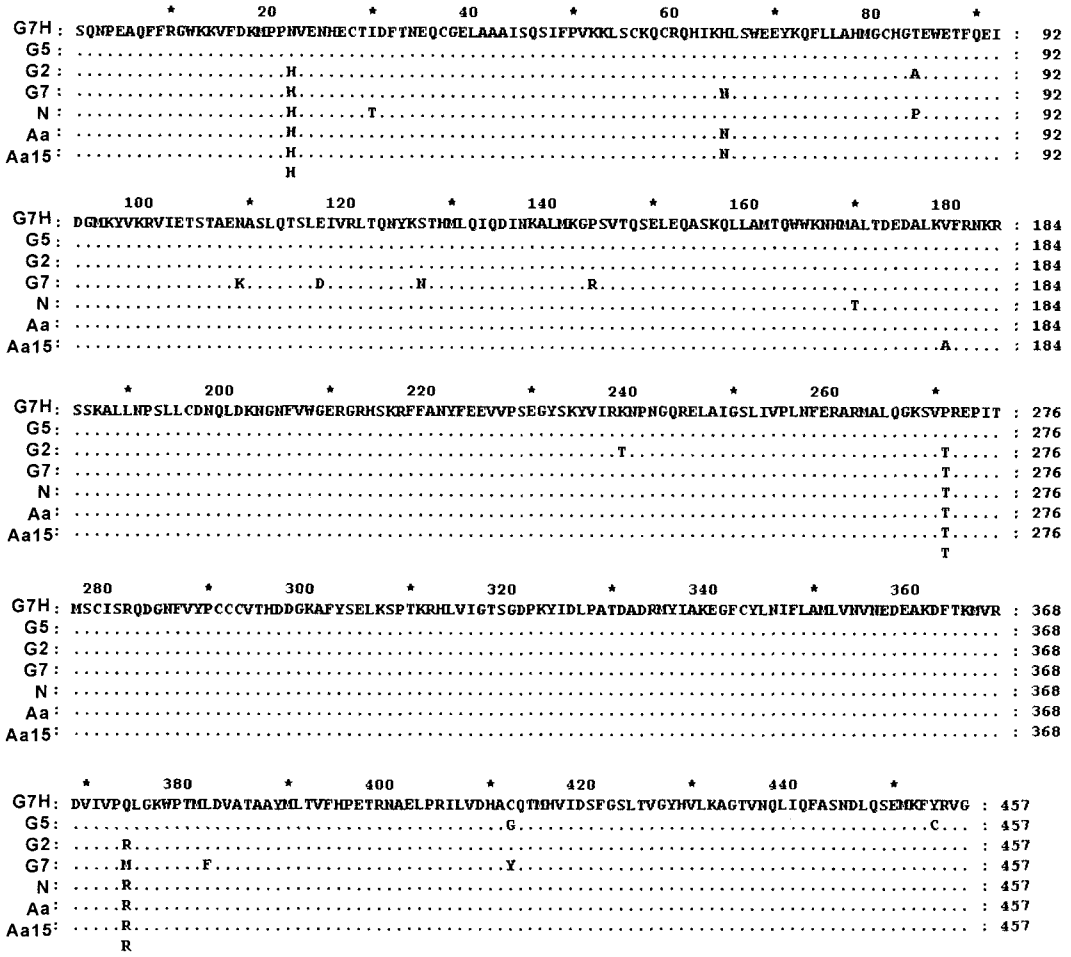

Fig. 2. Multiple alignments of the helper component-protease (HC-Pro) amino acid of SMV strains. Numbers on top represent the deduced HC-Pro amino acid position. Only the differences are shown.

Jinpum-2, and Sodam, while these cultivars developed no distinct symptom when inoculated with SMV G5 strain (data not shown). It has been shown that the HC-Pro of potyvirus may greatly affect symptom development (Atreya et al., 1992). HC-Pro is also reported to be involved in different steps of the viral cycle, aphid transmission, replication, virus cell-to-cell and systemic movement, and is a suppressor of posttranscriptional gene silencing (Anandalakshmi et al., 1998). Interestingly, amino acid sequences of HC-Pro gene of the seven SMV strains were highly conserved and only two amino acids differed between G7H and G5 (Fig. 2). It was not clear whether these differences of the amino acid sequence were significant in showing different symptom development in certain soybean cultivars. It is worth noting that some other regions of SMV genome including 5' and 3' NTR regions could be responsible for the biological differences between SMV-G5 and SMV-G7H. Studies are underway to further define the effect of two amino acid substitution in the HC-Pro gene of SMV RNA on symptom development.

Phylogenetic analyses. To determine the relationship among SMV G7H, G5 strains and other SMV strains, phylogenetic analyses were performed using PAUP4.0 program. Phylogenetic analysis of both nucleotide and predicted amino acid sequences indicated that both SMV-G5 and SMV-G7H were closely related to other SMV strains. Phylogenetic trees derived from the genomic RNA sequence, the deduced amino acid sequence, P1, and HC-Pro comparison are shown in Fig. 3. When the genomic RNA sequence and polyprotein sequence of SMV strains were aligned, similar results were obtained (Fig. 3, panels A & B). The phylogenetic trees were divided into two large subgroups on genomic RNA sequence, as well as amino acid alignment of polyprotein, G5-G7H and the other SMV types. Some distinct amino acid variations in the putative polyprotein protein within each subgroup were also observed indicating the continuous sequence variations among SMV strains although whole genomic sequences contain over 95% identify regardless of the soybean cultivar and location where they were isolated.

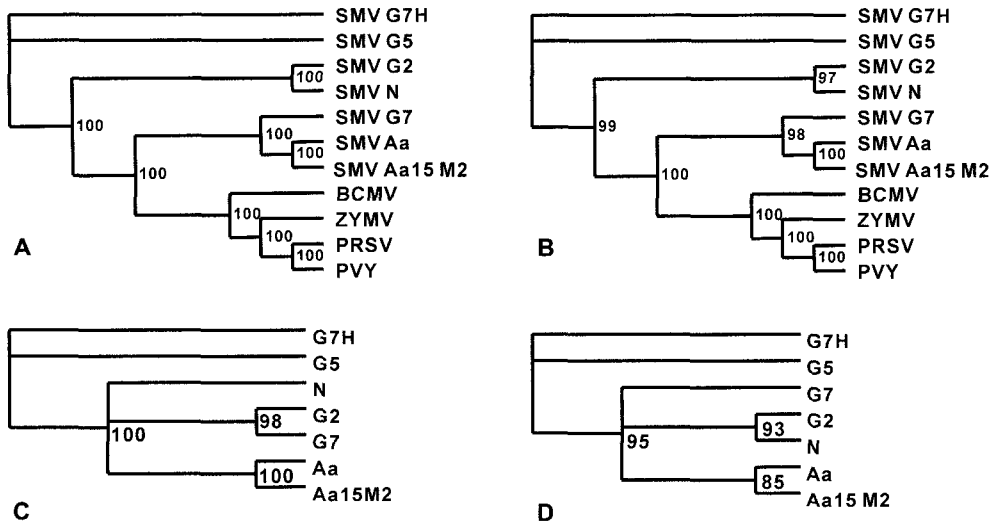

Fig. 3. Phylogenetic tree constructed from alignment of the genomic RNA sequences (A), the deduced polyprotein amino acid sequences (B), and predicted P1 and HC-Pro amino acid sequences (C and D, respectively) of SMV strains. Phylogenic analysis was conducted by parsimony using PAUP. Statistical reliability of the nodes was obtained by bootstrap analysis (1000 replications). Numbers on branches are percentages of bootstrap analyses supporting the grouping of each branch.

Lin et al. (2001) proposed that the diversity of P1 protein of Potyvirus is useful as a standard to distinguish different strains of viruses and investigate the evolutionary relationships. For SMV, P1 protein has relatively low identity compared with other protein-like CP, NIb and comparison with other potyvirus show the same result (data not shown). Phylogenetic tree shows that P1 protein of G7H has 99.7% homology to amino acid sequence of G5, while it has relatively low identity of 92.2% and 89.6% to G2 and G7 (Fig. 3C and Table 2). Interestingly, amino acids from 92 to 107 contained one nucleotide deletion and an addition at 405 to 456 nucleotide position. These mutations may be useful in classifying SMV strains based on P1 gene. It has been reported that the P1 protein may also affect symptom development, host range, and geographical distribution (Lee and Wong, 1998; Tordo et al., 1995). Complete nucleotide sequences of SMV strains G7H and G5 reported in this paper will help in finding symptom determinants and in better understanding virus replication and virus-host interaction for future research.

Acknowledgments

This research was supported in part by grant M101-KG010001-01K070104910 from the Crop Functional Genomics Center of the 21st Century Frontier Research Program and by a grant from the Center for Plant Molecular Genetics and Breeding Research funded by the Ministry of Science and Technology of Republic of Korea. LWS was supported by the graduate fellowship from the Ministry of Education through Brain Korea 21 Project.

References

Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D. J. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25:3389-3444.

Anandalakshmi, R., Pruss, G. J., Ge, X., Marathe, R., Mallory, A. C., Smith, T. H. and Vance, V. B. 1998. A viral suppressor of gene silencing in plant. *Proc. Natl. Acad. Sci. USA* 95:13079-13084.

Atreya, C. D., Atreya, P. L. and Thornbury, D. W. 1992. Site-directed mutations in the potyvirus HC-PRO gene affect helper component activity, virus accumulation, and symptom expression in infected tobacco plants. *Virology* 191:106-111.

Carrington, J. C. and Dougherty, W. G. 1988. A viral cleavage site cassette: identification of amino acid sequences required for tobacco etch virus polyprotein processing. *Proc. Natl. Acad. Sci. USA* 85:3391-3395.

Cho, E. K., Choi, S. H. and Cho, W. T. 1983. Newly recognized soybean mosaic virus mutants and sources of resistance in soybean. *Res. Rept. ORD (S.P.M.U.)* 25:18-22.

Cho, E. K. and Goodman, R. M. 1979. Strains of soybean mosaic virus: Classification based on virulence in resistant soybean cultivars. *Phytopathology* 69:467-470.

Cho, E. K. and Goodman, R. M. 1982. Evaluation of resistance on soybeans to Soybean mosaic virus strains. *Crop Sci.* 22:1133-1136.

Cho, E. K. and Chung, B. J. 1976. Studies on identification and classification of soybean virus diseases in Korea. *Kor. J. Plant Prot.* 15:61-68.

Chung, B. J., Lee, S. H., Cho, E. K. and Park, H. C. 1974. Identification of soybean viruses and soybean varietal reactions. *Ann. Rept. Crop Impro. Res. Center* pp. 137-145.

Chung, B. J. Lee, S. H. and Sung, J. M. 1973. An investigation on soybean mosaic virus: Varietal resistance and transmission. *Symp. Pl. Environ. Res. Summ.* pp. 73-77.

Gunduz, I., Buss, G. R., Ma, G., Chen, P. and Tolin, S. A. 2001. Genetic analysis of resistance to *Soybean mosaic virus* in OX670 and Harosoy soybean, *Crop Sci.* 41:1785-1791.

Gunduz, I., Buss, G. R., Chen, P. and Tolin, S. A. 2002. Characterization of SMV resistance genes in Tousan 140 and Hourei soybean. *Crop Sci.* 42:90-95.

Jayaram, C., Hill, J. H. and Miller, W. A. 1992. Complete nucleotide sequences of two soybean mosaic virus strains differentiated by response of soybean containing the Rsv resistance gene. *J. Gen. Virol.* 73:2067-2077.

Jung, H. W., Yun, W. S., Hahm, Y. I. and Kim, K. H. 2002. Characterization of *Tobacco mosaic virus* isolated from potato showing yellow leaf mosaic and stunting symptom in Korea. *Plant Dis.* 86:112-117.

Kendrick, J. B. and Gardner, M. W. 1924. Soybean mosaic; Seed transmission and effect on yield. *J. Agric.* 27:91-98.

Kim, Y. H. 2000. G7H, a new strain of *Soybean mosaic virus*: Its virulence and nucleotide sequence of cylindrical inclusion gene. Ph. D. thesis. Kyungpook National University. 120 p.

Lee, K. C. and Wong, S. M. 1998. Variability of P1 protein of zucchini yellow mosaic virus for strain differentiation and phylogenetic analysis with other potyviruses. *DNA sequence* 9:275-293.

Lin, S. S., Hou, R. F. and Yeh, S. D. 2001. Complete genome sequence and genetic organization of a Taiwan isolate of *Zucchini Yellow mosaic virus*. *Bot. Bull. Acad. Sin.* 42:243-250.

Lim, S. M. 1985. Resistance to *Soybean mosaic virus* in soybean. *Phytopathology* 75:199-201.

Liu, J., Peng, X., Li, L. and Mang, K. 1993. Cloning of coat protein gene of soybean mosaic virus and its expression in *Escherichia coli*. *Chin. J. Biotechnol.* 9:143-149.

Page, R. D. 1996. TreeView: an application to display phylogenetic trees on personal computers. *Comput. Appl. Biosci.* 12:357-358.

Riechmann, J. L., Lasin, S. and Garcia, J. A. 1992. Highlights and prospects of potyvirus molecular biology. *J. Gen. Virol.* 73:1-16.

Ross, J. P. 1967. Purification of Soybean mosaic virus for antiserum production. *Phytopathology* 57:465-467.

Shukla, D. D., Ward, C. W. and Brunt A. A. 1994. The Potyviridae. CAB international. Wallingford, Oxon, UK.

Thompson, J. D., Higgins, D. J. 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.

Tordo, V. M. J., Chachulska, A. M., Fakhfakh, H., LeRomanced, M., Robaglia, C. and Asteir-Manifacier, S. 1995. Sequence polymorphism in the 5'UTR and in the P1 coding region of potato virus Y genomic RNA. *J. Gen. Virol.* 76:939-949.

Yun, W. S., Jung, H. W., Oh, M. H., Hahm, Y. I. and Kim, K. H. 2002. Variation of Potato virus Y isolated from potato, tobacco, pea and weeds in Korea on the C-terminal region of coat protein gene and 3' non-translated region. *Plant Pathol. J.* 18:130-137.