Virology 393 (2009) 91–103

Contents lists available at ScienceDirect

Virology

journal homepage: www.elsevier.com/locate/yviro

Molecular variability and genetic structure of the population of *Soybean mosaic virus* based on the analysis of complete genome sequences

Jang-Kyun Seo [a,b], Kazusato Ohshima [c], Hyeok-Geun Lee [a], Moonil Son [a], Hong-Soo Choi [d], Su-Heon Lee [d], Seong-Han Sohn [d], Kook-Hyung Kim [a,b,*]

[a] *Department of Agricultural Biotechnology and Plant Genomics and Breeding Institute, College of Agriculture and Life Sciences, Seoul National University, Seoul 151-921, Republic of Korea*
[b] *Institute of Molecular Biology and Genetics, Seoul National University, Seoul 151-747, Korea*
[c] *Faculty of Agriculture, Saga University, Saga 840-8502, Japan*
[d] *National Academy of Agricultural Science, Rural Development Administration, Suwon 441-707, Republic of Korea*

ARTICLE INFO

Article history:
Received 6 May 2009
Returned to author for revision
1 July 2009
Accepted 6 July 2009
Available online 27 August 2009

Keywords:
SMV
Recombination
Genetic structure
Phylogenetic analysis
Genetic variation

ABSTRACT

The complete genomes of 30 *Soybean mosaic virus* (SMV) isolates and strains were sequenced in this study. Together with fourteen previously reported sequences, we analyzed the genetic structure of the SMV population. Analyses of genetic diversity showed that different genomic regions of SMV are under different evolutionary constraints and that there was no significant genetic differentiation between East Asian and North American populations of SMV. Phylogenetic analyses revealed a significant correlation between phylogeny of the cylindrical inclusion (CI) gene of SMV and SMV resistance gene 3 (*Rsv3*)-relating pathogenicity of SMV, suggesting CI might be a pathogenic determinant in *Rsv3*-mediated disease response. Interestingly, recombination analyses identified 19 'clear' recombination events in the SMV population. Furthermore, as several resistance-breaking strains were identified as recombinants, it appears that recombination might contribute to overcome host resistance in SMV–soybean pathosystem. Our finding suggests that recombination as well as mutation is an important evolutionary process in the genetic diversification of SMV population.

Introduction

Studying the molecular variability and genetic structure of viruses helps to provide understanding of the molecular evolutionary history of viruses in relation to their virulence, dispersion, and emergence of new epidemics. During the last two decades, many studies have been undertaken to understand the mechanisms which drive the evolution and geographical dispersion of plant viruses (reviewed by Garcia-Arenal et al., 2001). Such studies have been concerned to disentangle phylogenetic relationships among virus isolates because most viruses continue to evolve through genetic exchanges (recombination and reassortment of genome segments) as well as accumulation of mutations. RNA viruses, which are the largest group of plant viruses, are known to exist as quasispecies because of the low fidelity of viral RNA-dependent RNA polymerase (Domingo et al., 1985). Recombination is also a major evolutionary force for plant RNA viruses and has been shown to have important roles in changing viral virulence as well as emergence of new viral pathotypes (Moreno et al., 2004; Ogawa et al., 2008; Ohshima et al., 2002). The analysis of complete genomic sequence data of viral populations may provide information to understand the genetic stability of virus population and the evolutionary mechanism of genome dynamics. However, there have been only a few reports studying the population genetics and evolutionary history of potyviruses with analysis of the complete genomic sequence data at geographical scale (Desbiez and Lecoq, 2008; Ogawa et al., 2008; Ohshima et al., 2007; Tomimura et al., 2003).

Potyviruses constitute the largest genus of plant viruses and include many economically important species (Shukla et al., 1994). *Soybean mosaic virus* (SMV) is a member of the genus *Potyvirus* (Mayo and Pringle, 1998). Like all potyviruses, SMV has a positive-stranded RNA genome of approximately 10 kb in length, a genome-linked viral protein (VPg) covalently bound to the 5′ end, and a poly (A) tail at the 3′ end (Riechmann et al., 1992). SMV genome encodes one large polyprotein, which is cleaved to yield at least 10 mature proteins by virus-encoded proteases (Jayaram et al., 1992). SMV causes severe symptoms such as mosaic or necrosis in many soybean cultivars, and is easily transmitted by aphids in fields, thus resulting in significant reductions in soybean yield and quality. SMV isolates can produce very different disease reactions on various soybean cultivars. In the United States, at least ninety-eight SMV isolates were classified into seven strains (G1–G7) based on the symptoms developed on various resistant soybean cultivars (Cho

* Corresponding author. Department of Agricultural Biotechnology and Plant Genomics and Breeding Institute, College of Agriculture and Life Sciences, Seoul National University, Seoul 151-921, Republic of Korea. Fax: +82 2 873 2317.
E-mail address: kookkim@snu.ac.kr (K.-H. Kim).

0042-6822/$ – see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.virol.2009.07.007

and Goodman, 1979). Concomitantly, three distinct resistance genes, *Rsv1* (several alleles at the *Rsv1* locus), *Rsv3*, and *Rsv4* conferring resistance to different strains of SMV have been reported (Buss et al., 1999; Chen et al., 1991; Gunduz et al., 2002; Hayes et al., 2000). Accumulating evidence suggests that a variety of disease reactions caused by SMV is the result of interactions between viral pathogenic determinants and host resistance genes (Eggenberger et al., 2008; Hajimorad et al., 2005; Hajimorad et al., 2006; Seo et al., 2009c).

In Korea, diseases caused by SMV were first reported in the early 1970s (Chung et al., 1974). Since then, the emergence of viral strain variations and their virulence on soybean cultivars have been reported continuously (Cho, 1995; Cho et al., 1977; Cho et al., 1983; Choi et al., 2005; Kim and Lee, 1991; Kim, 2000; Kim et al., 2003). At least five strains, G1, G2, G5, G6, and G7 existed in soybean fields in Korea during 1980s and about 80% of the SMV damage was caused by SMV-G5 in the early 1980s (Cho et al., 1983; Kim, 2000). However, in the late 1980s, more than 65% of the SMV damage in the middle area of Korea was due to by SMV-G5H, which was first identified in 1983 (Cho et al., 1983; Kim, 2000). Furthermore, a new severe SMV isolate causing systemic mosaic symptom or necrosis in G5H-resistant soybean cultivars emerged in the late 1990s (Kim, 2000; Kim et al., 2003). This new strain, G7H, was the most prevalent strain (about 50%) in soybean fields in the early 2000s while the relative incidence caused by G5H decreased (about 4%) (Kim, 2000; Kim et al., 2003). It has been considered that resistant soybean cultivars to SMV were developed and cultivated continuously in Korea and this appears to have imposed selection pressure for the emergence of resistant-breaking SMV strains (Cho, 1995; Kim, 2000; Kim et al., 2003). However, it is still unclear which evolutionary mechanism was implicated in changing virulence of SMV; accumulation of mutations or recombination, or combination of these processes.

In the present study, we have determined complete genome sequences of 24 SMV isolates collected from wild soybean (*Glycine soja*) around the Korean Peninsula and 6 SMV strains including G1, G3, G4, G6, G6H, and G7a. We have analyzed variability and genetic structure of this SMV population, together with 14 previously reported sequences (Table 1), using a range of methods to understand their evolutionary relationships. Our results suggest that the emergence of resistance-breaking SMV strains in Korea might be due to recombination rather than accumulation of mutations, and that recombination is an important evolutionary process for genetic diversity of SMV population.

Table 1
SMV strains and isolates analyzed in this study.

Strain or isolate	Country of origin	Collecting host	Year of sampling	GenBank accession no.	Reference
Previous studies					
G2	USA	*Glycine max*	–	S42280	Jayaram et al. (1992)
G5	Korea	*Glycine max*	2001	AY294044	Lim et al. (2003)
G5H	Korea	*Glycine max*	2005	FJ807701	Seo et al. (2009a)
G7	USA	*Glycine max*	–	AY216010	Hajimorad et al. (2003)
G7d	USA	*Glycine max*	–	AY216987	Hajimorad et al. (2003)
G7H	Korea	*Glycine max*	2001	FJ807700	Seo et al. (2009b)
N	USA	*Glycine max*	–	D00507	Eggenberger et al. (1989)
Aa	Japan		–	AB100442	Unpublished
Aa15-M2	Japan		–	AB100443	Unpublished
L	Canada	*Glycine max*	2005	EU871724	Gagarinova et al. (2008a)
L-RB	Canada	*Glycine max*	2005	EU871725	Gagarinova et al. (2008a)
HH5	China	*Glycine max*	–	AJ310200	Chen et al. (2004)
HZ	China	*Glycine max*	–	AJ312439	Chen et al. (2004)
CN18	Korea	*Glycine max*	–	AJ619757	Choi et al. (2005)
This study					
G1	USA	*Glycine max*	2003	FJ640977	
G3	USA	*Glycine max*	2003	FJ640978	
G4	USA	*Glycine max*	2003	FJ640979	
G6	USA	*Glycine max*	2003	FJ640980	
G6H	Korea	*Glycine max*	2003	FJ640981	
G7a	USA	*Glycine max*	2003	FJ640982	
WS32	Korea	*Glycine soja*	2006	FJ640954	
WS37	Korea	*Glycine soja*	2006	FJ640955	
WS84	Korea	*Glycine soja*	2006	FJ640956	
WS101	Korea	*Glycine soja*	2006	FJ640957	
WS105	Korea	*Glycine soja*	2006	FJ640958	
WS109	Korea	*Glycine soja*	2006	FJ640959	
WS110	Korea	*Glycine soja*	2006	FJ640960	
WS116	Korea	*Glycine soja*	2006	FJ640961	
WS117	Korea	*Glycine soja*	2006	FJ640962	
WS128	Korea	*Glycine soja*	2006	FJ640963	
WS132	Korea	*Glycine soja*	2006	FJ640964	
WS135	Korea	*Glycine soja*	2006	FJ640965	
WS144	Korea	*Glycine soja*	2006	FJ640966	
WS145	Korea	*Glycine soja*	2006	FJ640967	
WS149	Korea	*Glycine soja*	2006	FJ640968	
WS151	Korea	*Glycine soja*	2006	FJ640969	
WS155	Korea	*Glycine soja*	2006	FJ640970	
WS156	Korea	*Glycine soja*	2006	FJ640971	
WS160	Korea	*Glycine soja*	2006	FJ640972	
WS162	Korea	*Glycine soja*	2006	FJ640973	
WS200	Korea	*Glycine soja*	2006	FJ548849	
WS202	Korea	*Glycine soja*	2006	FJ640974	
WS205	Korea	*Glycine soja*	2006	FJ640975	
WS209	Korea	*Glycine soja*	2006	FJ640976	

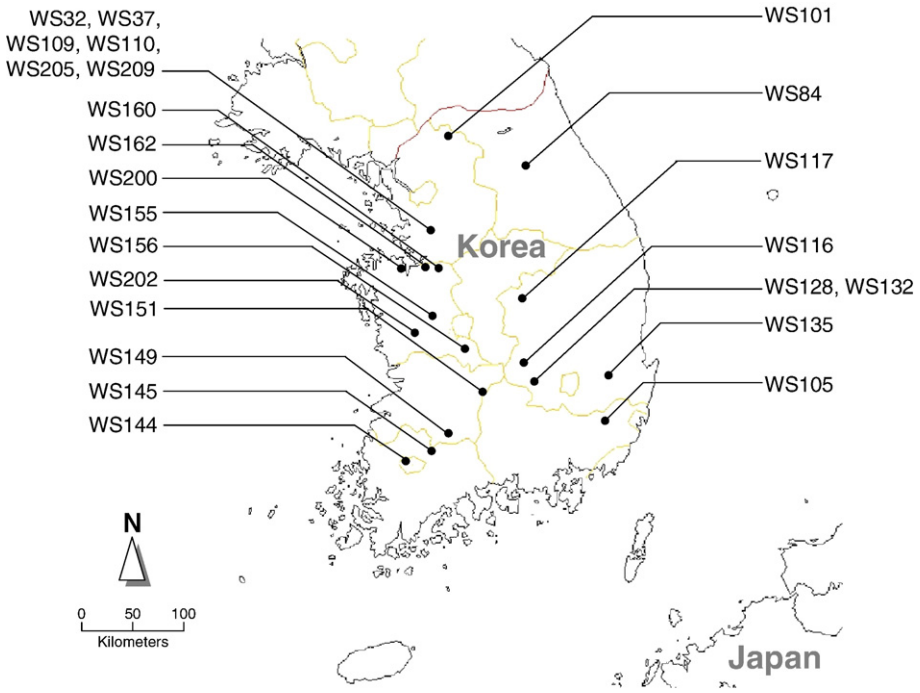


Fig. 1. Geographical locations of the collected SMV isolates in Korea.

Results

Molecular and biological characterization of the SMV population

To examine genetic diversity of the SMV population in Korea, field surveys were performed throughout the country in 2006. Because, in commercial fields, SMV-resistant soybean cultivars have been cultivated continuously and homogeneous strains can be prevalent (Cho, 1995; Kim, 2000), sampling SMV from tillage fields might not represent the latent genetic diversity of the SMV population at geographical scale. Therefore, we have sampled SMV from wild soybean (*G. soja*) growing in natural fields without cultivation. Forty wild soybean samples showing typical mosaic symptoms were collected throughout the country and all collected samples were identified by RT-PCR as infected with SMV (data not shown). Out of 40 SMV isolates, 24 isolates have been further analyzed to determine complete genomic sequences (Fig. 1 and Table 1). In addition, we obtained 6 SMV strains from National Institute of Crop Science (NICS) of Korea (Initially, G1, G3, G4, G6 and G7a purchased from the American Type Culture Collection and G6H collected from soybean fields in Korea) and their complete genome sequences were also determined in this study. Complete genomes of most isolates were either 9588 or 9585 nt long, whereas those of isolate WS32 and WS155 were 9584 and 9582 nt long, respectively. The sequences are available in the GenBank database with Accession codes listed in Table 1.

We assessed disease reactions on several soybean cultivars for 10 isolates sampled from wild soybean, together with G5H and G7H as reference strains, to determine their pathotypes (Table 2). Lee68 (*rsv*) was susceptible to all 10 collected isolates while PI96983 (*Rsv1*), V94-5152 (*Rsv4*), and Jinpumkong-2 (*Rsv*-genotype has not yet been characterized) were resistant to all of them. Suwon97 (*Rsv1-h*) reacted with necrosis to an isolate, WS200, but with resistance to other 9 isolates. Two types of disease responses were observed in L29 (*Rsv3*); WS37, WS149, WS160, WS200 and WS209 infected L29 systemically while WS37, WS110, WS132, WS135 and WS155 could not infect this cultivar. To check whether the initial sequenced SMV populations were maintained during inoculation experiments, we isolated progeny viruses from the infected soybean cultivars and resequenced their P1 cistronic regions. We found that there were fewer than 4 nucleotide differences between the parental and newly determined progeny viral sequences, suggesting that initial viral populations were maintained during serial infections on tested plants and that the disease responses on tested cultivars were induced by the initially sequenced isolates. Disease reactions of G1 through G7a,

Table 2
Differential responses of soybean cultivars to SMV isolates and strains.

Isolate or strain	Cultivar response to SMV inoculation[a]					
	Lee68 (*rsv*)	PI96983 (*Rsv1*)	Suwon97 (*Rsv1-h*)	L29 (*Rsv3*)	V94-5152 (*Rsv4*)	Jinpumkong-2 (Unknown)
WS32	S	R	R	R	R	R
WS37	S	R	R	S	R	R
WS110	S	R	R	R	R	R
WS132	S	R	R	R	R	R
WS135	S	R	R	R	R	R
WS149	S	R	R	S	R	R
WS155	S	R	R	R	R	R
WS160	S	R	R	S	R	R
WS200	S	R	N	S	R	R
WS209	S	R	R	S	R	R
G1[b]	S	R	R	S	R	R
G2[b]	S	R	R	S	R	R
G3[b]	S	R	R	S	R	R
G4[b]	S	N	R	S	R	R
G5[b]	S	R	R	R	R	R
G5H	S	R	N	R	R	R
G6[b]	S	R	R	R	R	R
G6H[b]	S	R	N	R	R	R
G7[b]	S	N	R	R	R	R
G7a[b]	S	S	R	S	R	R
G7H	S	R	N	S	R	N
CN18[c]	S	S	S	S	S	–
L[d]	–	R	R	S	R	–
L–RB[d]	–	R	R	S	S	–

[a] S, susceptible (systemic mosaic); R, resistant; N, systemic necrosis; '–', not determined.
[b] Disease responses were analyzed by Kim (2000) and Kim et al. (2003).
[c] Disease responses were analyzed by Choi et al. (2005).
[d] Disease responses were analyzed by Gagarinova et al. (2008).

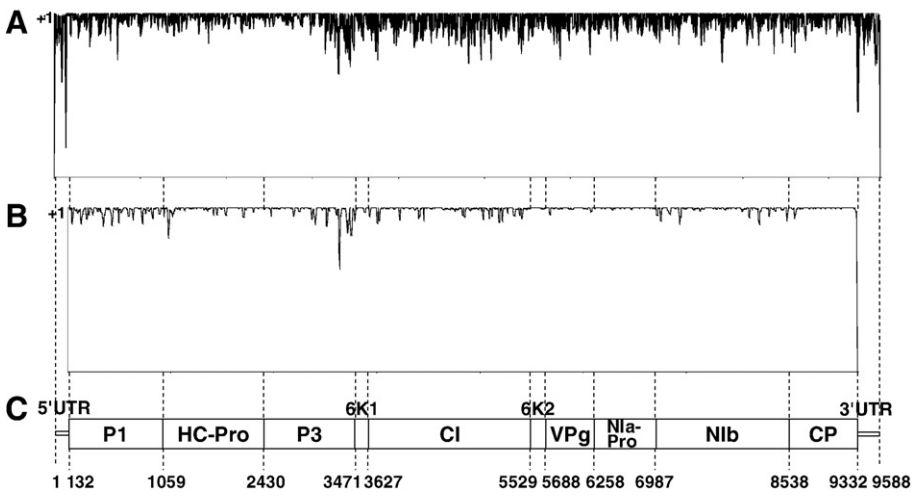


Fig. 2. Nucleotide and encoded amino acid sequence similarities in the SMV population. Forty-four SMV full-length sequences aligned by ClustalX2 were analyzed using AlignX by setting window site to 5 to estimate sequence similarities. '+ 1' on *y*-axis means that sequences are perfectly conserved. (A) Similarity profile of SMV genomic nucleotide sequences. (B) Similarity profile of SMV polyprotein amino acid sequences. (C) The genome map of SMV is shown to scale under the graphs.

CN18, L, and L-RB were analyzed in previous reports (Choi et al., 2005; Gagarinova et al., 2008a; Kim, 2000; Kim et al., 2003) and summarized in Table 2 to compare those of 10 collected isolates. Based on disease responses of 6 soybean cultivars, 10 isolates sampled from wild soybean can be divided into three pathotypes; similar to the less virulent strains (G1 to G3) or the more virulent strain (G6), or new and unique pathotype (WS200).

Genetic diversity and structure of the SMV population

We assessed the molecular variability of the SMV population by comparing either complete nucleotide sequences or deduced amino acid sequences of 44 SMV strains and isolates. Widespread nucleotide variations were detected throughout the genome and more significant variations were observed in 5′ untranslated region (UTR) and 3′ UTR of SMV genome (Fig. 2). However, comparison of the deduced amino acid sequences revealed that the P1 and P3 cistronic regions exhibited high variability while other cistronic regions of SMV were relatively conserved, considering the nucleotide variations throughout the genome (Fig. 2).

We estimated genetic distances for each cistronic region according to the Kimura's two-parameter method. The overall mean value of nucleotide diversity for the polyprotein-coding region was 0.0202 ± 0.0015 (Table 3). Low nucleotide diversity values were found in 6K2, NIa-Pro, and CP cistronic regions while other cistronic regions exhibited relatively high diversity values (Table 3). We also checked pairwise genetic differences at nonsynonymous (d_N) and synonymous (d_S) nucleotide positions using the Pamilo–Bianchi–Li (PBL) method. The d_N/d_S ratio can be used as an index representing the pattern of selective constraint in evolutionary relationships (Nei and Gojobori, 1986). The d_N/d_S ratios of P1 and P3 cistronic regions were significantly higher than those of the other cistronic regions (Table 3), consistent with the above result of nucleotide variation analysis (Fig. 2). These results suggest that SMV P1 and P3 are under looser evolutionary constraints than the other cistrons. We divided the SMV population into two geographic subpopulations (Far East Asia and North America) to estimate genetic distances within and between these subpopulations for all cistronic regions. Nucleotide diversities were calculated according to the Kimura's two-parameter method. For P1 cistronic region, within-population diversities (0.0177 ± 0.0035 in Far East Asia and 0.0240 ± 0.0059 in North America) were slightly different to the between-population diversity (0.0256 ± 0.0054). However, we could not find any meaningful evidence for population subdivision in the other cistronic regions including the polyprotein-coding region (data not shown), suggesting that there was no significant genetic differentiation between the two subpopulations.

We performed phylogenetic analysis to determine the relationships of all 44 SMV strains and isolates including an isolate of *Watermelon mosaic virus* (WMV-CHI87-620; GenBank accession code EU660580) as the outgroup. The phylogenetic trees were reconstructed by the neighbor joining (NJ) and maximum parsimony (MP) methods based on either the full-length nucleotide or polyprotein amino acid sequence alignments, and only the NJ trees are shown (Fig. 3, panels A and B). However, the resulting trees were inconsistent and did not support geographic clustering of the SMV isolates. Moreover, it was difficult to find clear relationships between the phylogeny of the isolates and their pathogenicity on various soybean cultivars although various phylogenetic trees for either nucleotide or deduced amino acid sequence of each viral cistronic region were analyzed based NJ and MP methods (data not shown). Interestingly, however, we could find significant relationship between phylogeny for the CI cistron and pathogenicity of SMV on soybean L29 (*Rsv3*). When the phylogenetic trees were reconstructed for either nucleotide or deduced amino acid sequences of the CI cistron, SMV isolates (excepting G7a and CN18) inducing mosaic symptom on L29 fell into the Group A while other SMV isolates, which cannot infect L29, belonged to the Group B (Figs. 3C and D). This strongly suggests that CI may play a role as one of the pathogenic determinants in SMV–soybean pathosystem via direct or indirect interaction with the *Rsv3* gene.

Table 3
Nucleotide diversity for different genomic regions of the SMV population.

Genomic region	Nucleotide diversity[a]			
	d	d_N	d_S	d_N/d_S
P1	0.0211 ± 0.0037	0.0234 ± 0.0030	0.0841 ± 0.0081	0.2788
HC-Pro	0.0171 ± 0.0032	0.0092 ± 0.0017	0.1140 ± 0.0080	0.0811
P3	0.0301 ± 0.0057	0.0227 ± 0.0043	0.1459 ± 0.0132	0.1559
6K1	0.0246 ± 0.0139	0.0085 ± 0.0057	0.2023 ± 0.0508	0.0421
CI	0.0306 ± 0.0043	0.0118 ± 0.0016	0.1959 ± 0.0120	0.0603
6K2	0.0026 ± 0.0019	0.0016 ± 0.0008	0.1895 ± 0.0489	0.0083
VPg	0.0143 ± 0.0053	0.0049 ± 0.0015	0.1840 ± 0.0252	0.0264
NIa-Pro	0.0063 ± 0.0030	0.0016 ± 0.0004	0.1662 ± 0.0194	0.0099
Nib	0.0201 ± 0.0034	0.0089 ± 0.0015	0.1485 ± 0.0111	0.0597
CP	0.0085 ± 0.0027	0.0049 ± 0.0014	0.1478 ± 0.0143	0.0329
Polyprotein	0.0202 ± 0.0015	0.0111 ± 0.0008	0.1471 ± 0.0044	0.0758

[a] *d*, nucleotide diversity estimated by the Kimura's two-parameter method; d_N and d_S, nucleotide diversity at nonsynonymous and at synonymous position, respectively estimated by the Pamilo–Bianchi–Li method; Values are means ± SEM.

To obtain a network structure for this phylogenetic incompatibility of the SMV population, we next performed a split decomposition analysis using the SplitsTree4 program. Such network analyses are particularly useful in studying the reticulate evolution of organisms caused by recombination or gene conversion (Holmes et al., 1999; Huson and Bryant, 2006). The split decomposition analysis revealed that some of the SMV population formed a reticulate network structure (Fig. 4A). Especially, SMV G5, G5H, G7H, and WS200 exhibited strong conflicting phylogenetic signals, highly suggestive of recombination. Moreover, phylogenetic correlation profiles of the SMV population analyzed using the PHYLPRO program showed more than 10 clear peaks indicating potential recombination breakpoints appeared in various parts of the SMV genome (Fig. 4B). Thus, the multiple independent recombination events in the SMV population can be considered as one of the possible reasons for the phylogenetic inconsistency of the SMV population. However, it is still insufficient to explain geographical and pathogenical relationships among the SMV population by phylogenetic analyses. Other possible reasons for this phylogenetic inconsistency are further considered in the Discussion section.

Recombination analysis

To further investigate the putative recombination signals detected by the SplitsTree4 and PHYLPRO programs, 44 full-length sequences of SMV were aligned using ClustalX2 and were analyzed using the recombination detection program (RDP), GENECONV, Chimaera, MaxChi, BOOTSCAN, and SISCAN methods implemented in the RDP3 program with default settings and a Bonferroni corrected *P*-value cut-off of 0.01. This analysis identified many unique recombination signals. To omit unreliable signals, we selected recombination events supported by 3 different methods with an associated *P*-value of $<1.0\times10^{-6}$, and analyzed each of the recombination events by phylogenetic approach to verify the parent/daughter assignment made by RDP3. Moreover, the recombination events identified by the RDP3 program were confirmed by the PHYLPRO and another recombination analysis programs, original SiScan v. 2 and the Recombination Analysis Tool (RAT). In accordance with the parsimony principle, we finally identified 19 'clear' recombination events and 28 out of 44 sequences (~64%) were found to be clear recombinants (Fig. 5 and Table 4). As summarized in Table 4, most recombination events were detected by more than 5 methods of the RDP3 program with a high degree of confidence and supported by original SiScan v. 2 with *Z*-values greater than 3. For example, WS132 seemed to have resulted from recombination between WS116 as the major parent and WS135 as the minor parent. In this recombination event, the genomic region (nt 1–800) of WS116 was replaced with the homologous region of WS135. This recombination event was detected by *P*-value (6.8×10^{-25}) in SISCAN methods of the RDP3 program and by *Z*-value (4.64) calculated by the original SiScan v. 2 program. In addition, this event was supported by the PHYLPRO program. The recombination events identified by RDP3 and SiScan v. 2 were schematically represented in Fig. 5A and compared with recombination events analyzed using the RAT program (Fig. 5B). Comparison of recombination events analyzed by two different programs clearly showed phylogenetically inconsistent regions of recombinants and patterns of recombination in the SMV population.

It seems likely that some recombinants were driven by multiple crossover recombinations. Especially, WS117 and WS162 were found to have more than two phylogenetically inconsistent regions in their genomes (Fig. 5). In the case of WS117, three parts of the genome (nt 1–2200, 4550–5500 and 6200–9585) showed high similarity to WS145, WS155 and WS144, respectively. In WS162, three genomic regions (nt 1–2250, 4150–4850 and 5650–6250) showed high similarity to WS155, WS37 and N, respectively. These indicate that WS 117 and WS162 are multiple crossover recombinants.

Some recombination events showed identical patterns. For instance, G6H, HZ, and HH5 identified to be derived from same parental isolates and had two recombination sites at nt 3000–3050 and at nt 4650 and 4700 and (Table 4). This may indicate that recombination sites of G6H, HZ, and HH5 originated from a common ancestral virus. Furthermore, another recombination site at 3000–3050 identified as a recombination between G1 and WS101 was found in 18 SMV isolates including WS37, WS109, WS128, and others (Table 4). Interestingly, these isolates were found within the same clade in the tree calculated from on full-length nucleotide sequences (Fig. 3A). This suggests that an ancestor of these isolates might have emerged by recombination in the early stage of SMV differentiation.

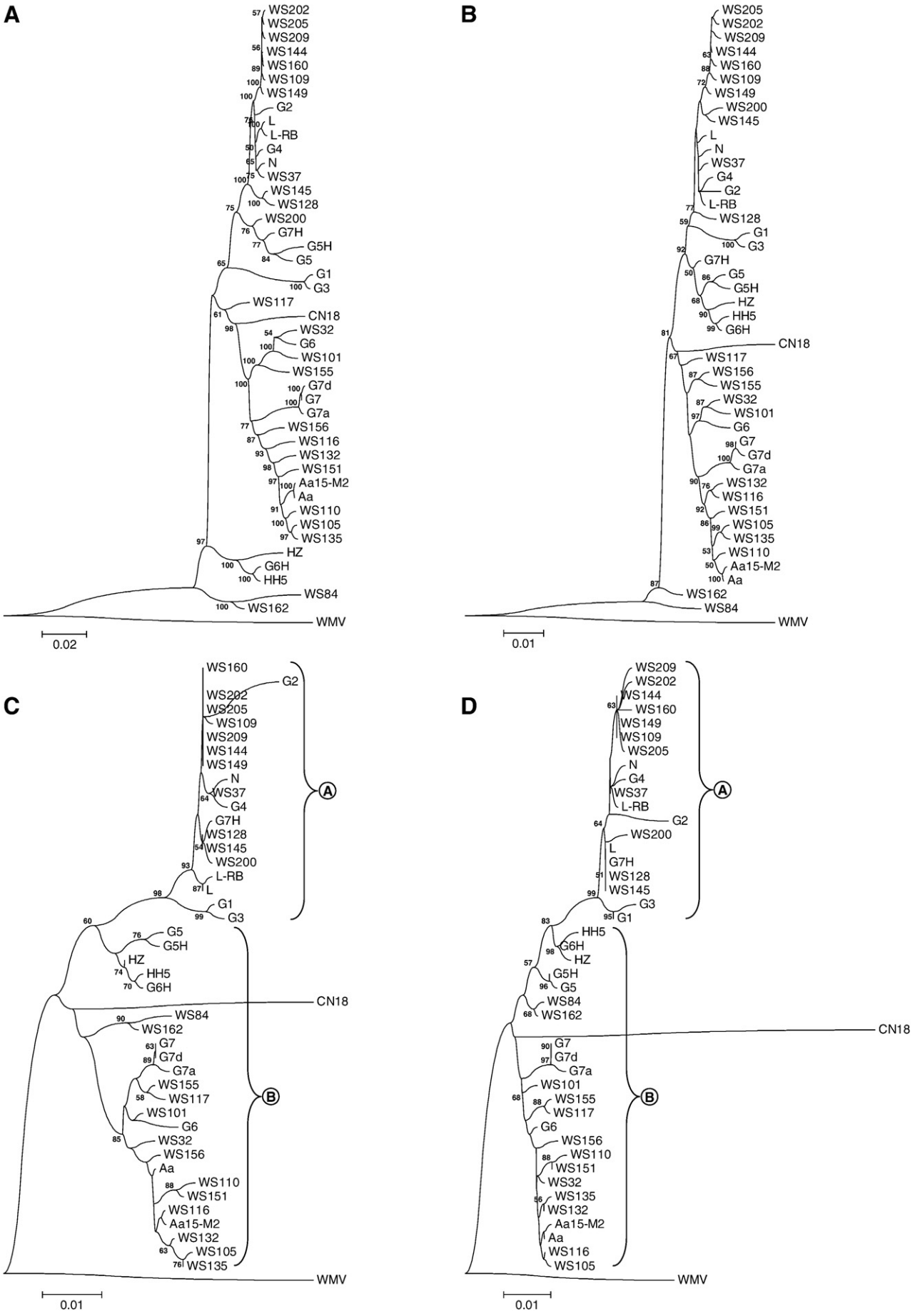
Our recombination analyses also detected several 'tentative' recombination sites (Table 4). These were discarded as 'tentative' because of unclear recombination sites, weak supports by recombination-detecting programs, or because one of the parental isolates was unknown. Further analyses using much more diverse isolates will be required to clarify these 'tentative' sites.

Discussion

Understanding the evolutionary history of viruses and the evolutionary mechanisms driving their selection and diversification is an important aspect of evolutionary biology and would help us to manage viral diseases and the risk of emerging new viruses. There have been several studies attempting to understand evolutionary history of potyviruses by analyzing genetic structures of their populations at geographical scales, for instance of those of *Turnip mosaic virus* (Ohshima et al., 2002; Tomimura et al., 2003; Tomimura et al., 2004), *Papaya ringspot virus* (Bateson et al., 2002), WMV (Moreno et al., 2004), *Plum pox virus* (Glasa et al., 2004), *Yam mosaic virus* (Bousalem et al., 2000), and *Potato virus Y* (Ogawa et al., 2008). SMV, a member of the potyvirus group, causes major diseases of soybean worldwide, Korea being no exception. As noted above, in Korea, outbreaks of SMV disease caused by variant strains have been occurred periodically, but few studies have performed to understand the causal evolutionary mechanisms. Here, we present a study attempted to investigate the mechanism of the emergence of resistance-breaking SMV strains based on analysis of diversity and genetic structure of the SMV population in Korea. Such regional-scale analyses for the viral populations can provide other detailed aspects of evolution comparing with those from the worldwide-scale analyses.

In this study, we have determined complete genome sequences of 30 SMV isolates and strains in total and analyzed genetic diversity and population structure of SMV together with 14 previously reported sequences (Table 1). As noted above, in order to increase the chance of collecting genetically diverse isolates, we have collected 24 SMV isolates from wild soybean (*G. soja*) growing in natural fields in Korea. Wild soybean is the progenitor of cultivated soybean (*G. max*) and is also a major host of SMV. In addition, wild soybean has much higher genetic diversity than cultivated soybean and East Asia including the Korean Peninsula is the original home of wild soybean as well as cultivated soybean (Xu, 1989; Yu and Kiang, 1993). Therefore, it can be seen that our SMV collection including field isolates and strains is representative of the genetic population of SMV in Korea.

In characterizing virus isolates, biological cloning is necessary because the possibility of mixed infections in fields with different viral strains cannot be ruled out. In many cases, biological cloning can be accomplished through single-lesion passage. However, we could not isolate the cloned viruses using this method because most of the collected SMV did not induce lesions in various tested soybean cultivars (Table 2 and data not shown). Instead, we have characterized the collected SMV as isolates by molecular cloning. Complete sequences of each isolate were determined using three large fragments overlapped each other by at least 200 bp. This convinced us that the assembled complete sequence was from the one major


A
WS202
WS205
WS209
WS144
WS160
WS109
WS149
G2
L
L-RB
G4
N
WS37
WS145
WS128
WS200
G7H
G5H
G5
G1
G3
WS117
CN18
WS32
G6
WS101
WS155
G7d
G7
G7a
WS156
WS116
WS132
WS151
Aa15-M2
Aa
WS110
WS105
WS135
HZ
G6H
HH5
WS84
WS162
WMV
0.02
B
WS205
WS202
WS209
WS144
WS160
WS109
WS149
WS200
WS145
L
N
WS37
G4
G2
L-RB
WS128
G1
G3
G7H
G5
G5H
HZ
HH5
G6H
CN18
WS117
WS156
WS155
WS32
WS101
G6
G7
G7d
G7a
WS132
WS116
WS151
WS105
WS135
WS110
Aa15-M2
Aa
WS162
WS84
WMV
0.01
C
WS160
G2
WS202
WS205
WS109
WS209
WS144
WS149
N
WS37
G4
G7H
WS128
WS145
WS200
L-RB
L
G1
G3
A
G5
G5H
HZ
HH5
G6H
CN18
WS84
WS162
G7
G7d
G7a
WS155
WS117
WS101
G6
WS32
WS156
Aa
WS110
WS151
WS116
Aa15-M2
WS132
WS105
WS135
B
WMV
0.01
D
WS209
WS202
WS144
WS160
WS149
WS109
WS205
N
G4
WS37
L-RB
G2
WS200
L
G7H
WS128
WS145
G3
G1
A
HH5
G6H
HZ
G5H
G5
WS84
WS162
CN18
G7
G7d
G7a
WS101
WS155
WS117
G6
WS156
WS110
WS151
WS32
WS135
WS132
Aa15-M2
Aa
WS116
WS105
B
WMV
0.01

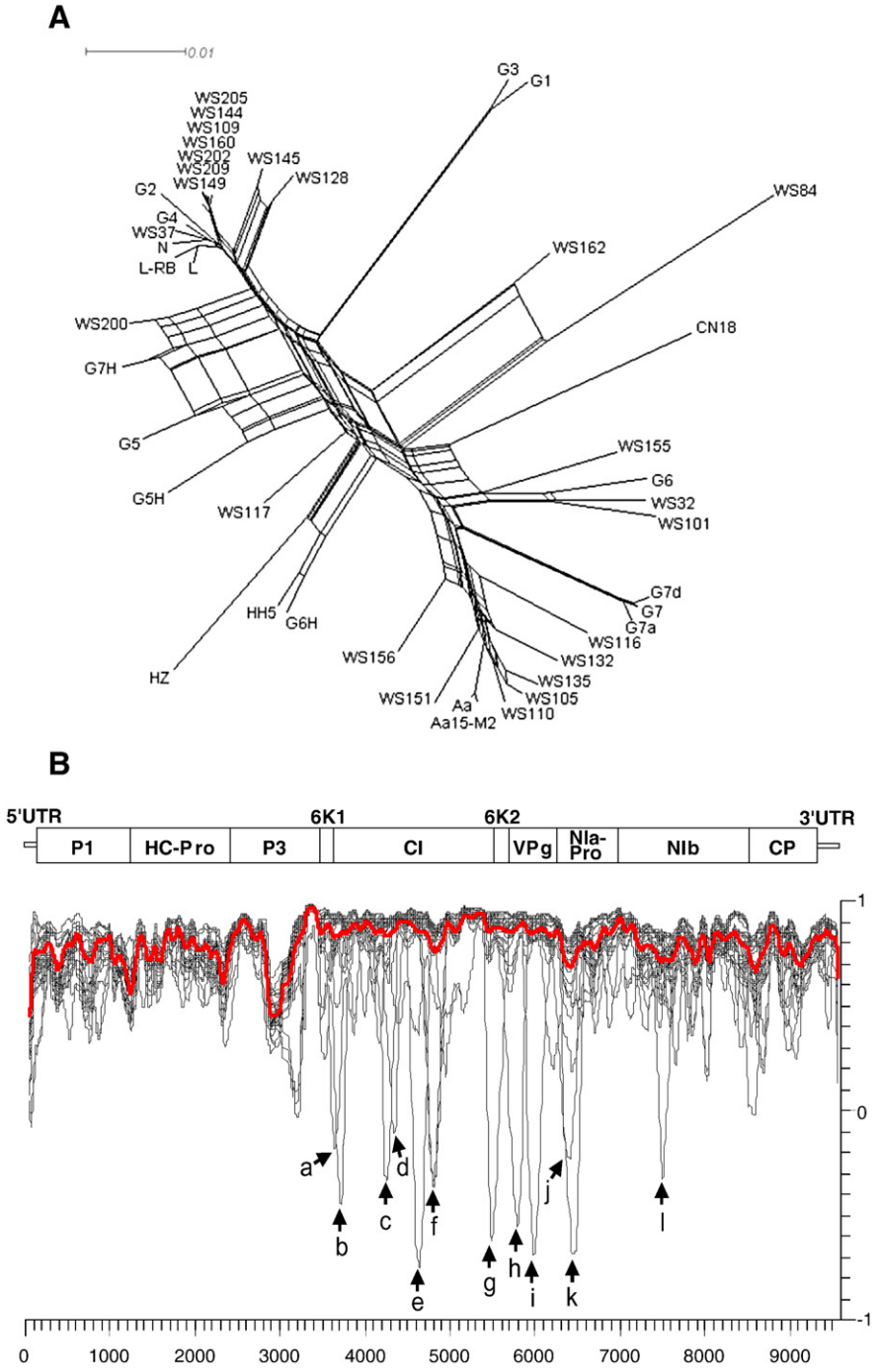


Fig. 4. Phylogenetic evidence for recombination in the SMV population. (A) Split decomposition network of the SMV population. Network tree was generated using SplitsTree4. Forming an reticular network rather than a single bifurcating tree is suggestive of recombination. (B) Graph showing phylogenetic correlations in the SMV population detected by the PHYLPRO method. The phylogenetic correlation (*y*-axis) was obtained from pairwise distance analysis of all aligned sequences. Significant recombination signals (appearing as single sharp downward peaks) are indicated by arrows; a (at nt 3631 for WS162), b (at nt 3709 for WS84), c (at nt 4243 for G5), d (at nt 4336 for WS162), e (at nt 4627 for G5H), f (at nt 4801 for HH5, Hz and G6H), g (at nt 5482 for G5), h (at nt 5784 for WS162), i (at nt 5972 for G5H), j (at nt 6403 for WS117), k (at nt 6442 for WS162), and l (at nt 7492 for G5H). The genome map of SMV is shown to scale above the graph.

genome. Furthermore, we resequenced P1 cistronic regions of progeny viruses, which were extracted from the infected soybean cultivars in pathotype testing experiments, and found that there were no changes of the initial viral populations during serial pathotype testing inoculations. Therefore, we concluded that disease reactions of SMV isolates on the tested cultivars were caused by the sequenced isolates.

Analyses of the pairwise nucleotide and encoded amino acid similarities among the SMV population revealed that nucleotide

Fig. 3. Phylogenetic analyses for the complete genome sequences (A), the encoded polyprotein amino acid sequences (B), and nucleotide and amino acid sequences of CI cistron (C and D, respectively) of the SMV population. An isolate of WMV was included as the outgroup. The phylogenetic trees were reconstructed by the NJ method applying Kimura's two-parameter method for nucleotide sequence analyses and *p*-distance model amino acid sequence analyses. The numbers on the branches indicate bootstrap percentages based on 1000 replications (only values >50% are shown).

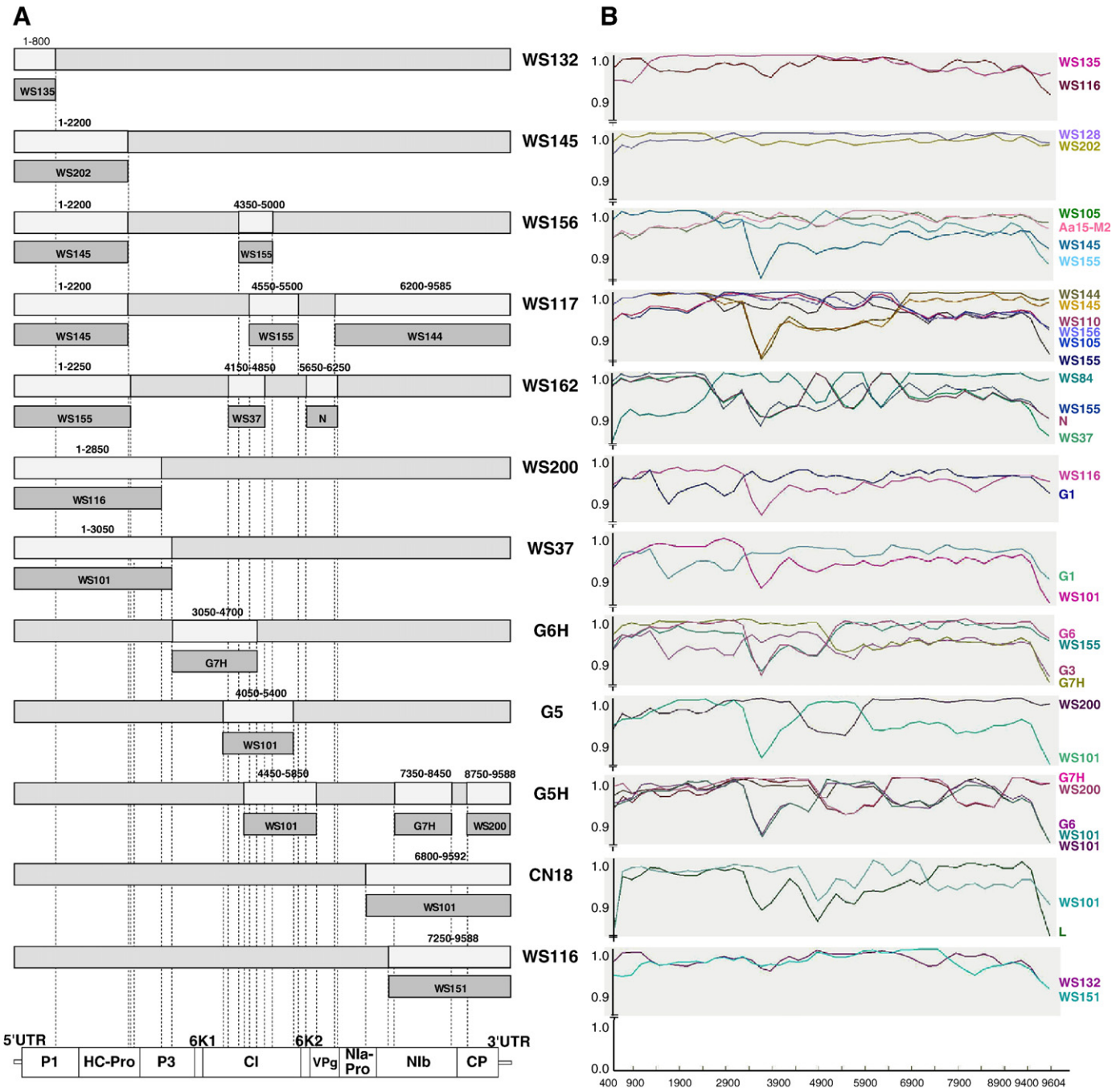


Fig. 5. Recombination events in the SMV population. (A) Locations of recombination events identified by RDP3 and SiScan v. 2 programs. Each full-length genome of a recombinant is represented by a long light gray box and the corresponding isolate name is given to the right of the box. Each recombination event is demarcated by a dark gray box below the recombinant genome. The genome map of SMV is shown to scale on the bottom of the figure. (B) Analyses of recombination events using RAT. The outputs for each recombinant (the isolate name is shown on the left of the graph) are shown, along with contributing parental sequences (the names are shown on the right of the graph). The *x*-axis and *y*-axis represent the location on the sequence and the genetic distance of each sequence, respectively.

variations were distributed throughout the entire genome, but P1 and P3 protein sequences were significantly more variable than other mature protein sequences (Fig. 2). This suggests that different viral components of SMV are under different evolutionary constraints. This was clearly demonstrated by estimating the d_N/d_S ratios of each cistronic region. Second to P1 cistron, P3 had the largest d_N/d_S ratio compared to those of the other cistrons (Table 3). These results are consistent with those from the analyses of the other potyviruses (Aleman-Verdaguer et al., 1997; Moreno et al., 2004; Ogawa et al., 2008; Oruetxebarria et al., 2000; Tomimura et al., 2004). Different evolutionary dynamics for different genomic regions have been considered to be associated with stronger functional or structural constraints on the conserved region (Garcia-Arenal et al., 2001). Many of viral proteins play multifunctional and essential roles in the virus life cycle in association with various host factors. Therefore, different selective constraints corresponding to various functions seem to act on viral proteins and genetic diversity of viral populations may be limited by such negative selection. On the other hand, some viral proteins such as P1 are polymorphic and not strictly required for viral infectivity, and hence these features seem likely to give the possibility of differentiation into genetically distinct groups or host adaptation

Table 4
Recombination in the SMV population detected by recombination detection programs.

Recombination site	Recombinant[a]	Parental isolates[b]	Recombination detecting programs		
			RDP3 (*P*-value)[c]	*Z*-value[d]	PHYLPRO[e]
P1					
nt 750–800	WS132	WS116 × WS135	RGBMC**S** (6.8×10^{-25})	4.64	D
HC-Pro					
nt 1050–1100	G1[f], G3[f]	Unknown × G7	RG**B**MCS (3.7×10^{-38})	NA	ND
nt 2140–2200	WS156, WS117	WS105 × WS145	RG**B**MCS (2.5×10^{-58})	5.58	D
	WS145	WS128 × WS202	RG**B**MCS (8.1×10^{-20})	4.50	D
nt 2200–2250	WS162	WS84 × WS155	**R**GBMCS (7.8×10^{-84})	6.24	D
nt 2350–2950	WS155[g]	WS145 × G7A	**R**GBMCS (1.1×10^{-42})	5.70	D
P3					
nt 2800–2850	WS200	G1 × WS116	**R**GBS (1.6×10^{-14})	5.40	D
nt 3000–3050	G6H[h], HZ[h], HH5[h]	G3 × G6	RGBMC**S** (2.9×10^{-15})	4.77	D
	WS37[i], WS109[i], WS128[i], WS144[i], WS145[i], WS149[i], WS160[i], WS202[i], WS205[i] WS209[i], G2[i], G4[i], G5[i], G5H[i], G7H[i], L[i], L–RB[i], N[i]	G1 × WS101	RG**B**MCS (1.5×10^{-39})	5.89	D
CI					
nt 4050–4100	G5	WS200 × WS101	RG**B**MCS (1.9×10^{-41})	6.78	D
nt 4150–4200	WS162	WS84 × WS37	RG**B**MCS (7.5×10^{-32})	6.00	D
nt 4350–4400	WS156	Aa15-M2 × WS155	RG**B**MCS (1.3×10^{-15})	4.37	ND
nt 4450–4500	G5H	WS200 × WS101	RG**B**MCS (2.9×10^{-23})	6.28	D
nt 4550–4600	WS117	WS110 × WS155	**R**GBMCS (1.1×10^{-15})	4.79	ND
nt 4650–4700	G6H[j], HZ[j], HH5[j]	WS155 × G7H	RG**B**MCS (2.5×10^{-31})	5.69	D
nt 4800–4850	WS162	WS84 × WS37	RG**B**MCS (7.5×10^{-32})	6.00	D
nt 4950–5000	WS156	Aa15–M2 × WS155	RG**B**MCS (1.3×10^{-15})	4.46	ND
nt 5250–5300	WS32k	Aa × WS101	RGBM**S** (5.2×10^{-5})	3.22	ND
nt 5350–5400	G5	WS200 × WS101	RG**B**MCS (1.9×10^{-41})	6.55	D
nt 5450–5500	WS117	WS110 × WS155	**R**GBMCS (1.1×10^{-15})	4.38	ND
6K2					
nt 5650–5700	WS162	WS84 × N	RG**B**MCS (1.9×10^{-23})	6.68	D
VPg					
nt 5800–5850	G5H	WS200 × WS101	RG**B**MCS (2.9×10^{-23})	5.40	D
nt 5850–6000	WS32[k]	Aa × WS101	RGBM**S** (5.2×10^{-5})	3.22	ND
nt 6200–6250	WS117	WS156 × WS144	RG**B**MCS (1.9×10^{-56})	5.73	D
	WS162	WS84 × N	RG**B**MCS (1.9×10^{-23})	5.57	D
NIa-Pro					
nt 6250–6500	WS32[g]	G6H × Aa	RGM**S** (1.8×10^{-8})	4.69	D
nt 6800–6850	CN18	L × WS101	R**G**MCS (3.4×10^{-18})	4.93	D
Nib					
nt 7250–7300	WS116	WS132 × WS151	RGBMC**S** (4.5×10^{-25})	4.46	D
nt 7350–7400	G5H	G6 × G7H	RGBMC**S** (4.2×10^{-21})	5.51	D
nt 8400–8450	G5H	G6H × WS200	**R**GBMCS (3.5×10^{-11})	5.51	D
CP					
nt 8750–8850	G5H	G6H × WS200	**R**GBMCS (3.5×10^{-11})	3.51	D

[a] Some recombination sites seemed to be 'tentative' sites or originated from a common ancestral virus; f, 'tentative' because one of two parental isolates is 'unknown'; g, 'tentative' because the cross-over site is wide and unclear; f, h, i and j, recombination sites might be originated from a common ancestral virus; k, 'tentative' because supported by less than 3 methods with an associated *P*-value of $<1.0\times10^{-6}$ implemented in RDP3.

[b] 'Parental isolate' means the most likely isolate among analyzed isolates; Major parent × minor parent.

[c] RDP3-implemented methods supporting the corresponding recombination sites; R (RDP), G (GENECONV), B (BOOTSCAN), M (MAXCHI), C (CHIMAERA), and S (SISCAN). The reported *P*-value in parentheses is the greatest *P*-value among the calculated *P*-values using RDP3-implemented methods and the corresponding method is shown boldface.

[d] The *Z*-valuses were calculated by original SiScan v. 2 program. Smaller *Z*-values of one of the parents are shown. The *Z*-values greater than 3 represent the strong support of recombination. NA, not available.

[e] PHYLPRO v. 1.1 program was used to detect recombination site. D, detected; ND, not detected.

of viral populations (Ohshima et al., 2002; Oruetxebarria et al., 2000; Roossinck, 1997).

Our phylogenetic analyses of the 44 SMV isolates and strains did not provide clear clues for classifying the SMV population according to geographical origins. Our SMV collection listed in Table 1 can be largely divided into two geographic subpopulations, Far East Asia and North America. Actually, soybean originated in East Asia and has been cultivated for several thousands of years (Wu et al., 2004; Xu, 1989), and it was recently suggested that SMV, like soybean itself, originated in South East Asia when it diverged from WMV around 1500 years ago, and further diverged into the soybean and Pinellia lineages around a 1000 years ago (Gibbs et al., 2008). However, in North America, soybean was first introduced in 1765 and rose to agricultural prominence in the early 1900s (Hymowitz and Harlan, 1983; Probst and Judd, 1973). As soybean is the natural host of SMV in fields, the history of SMV in North America is believed to be similar to that of soybean. Comparison of nucleotide diversities of within- or between-population also revealed that little genetic differentiation has occurred between two geographical subpopulations (data not shown). Thus, it is likely that few sequence diversity

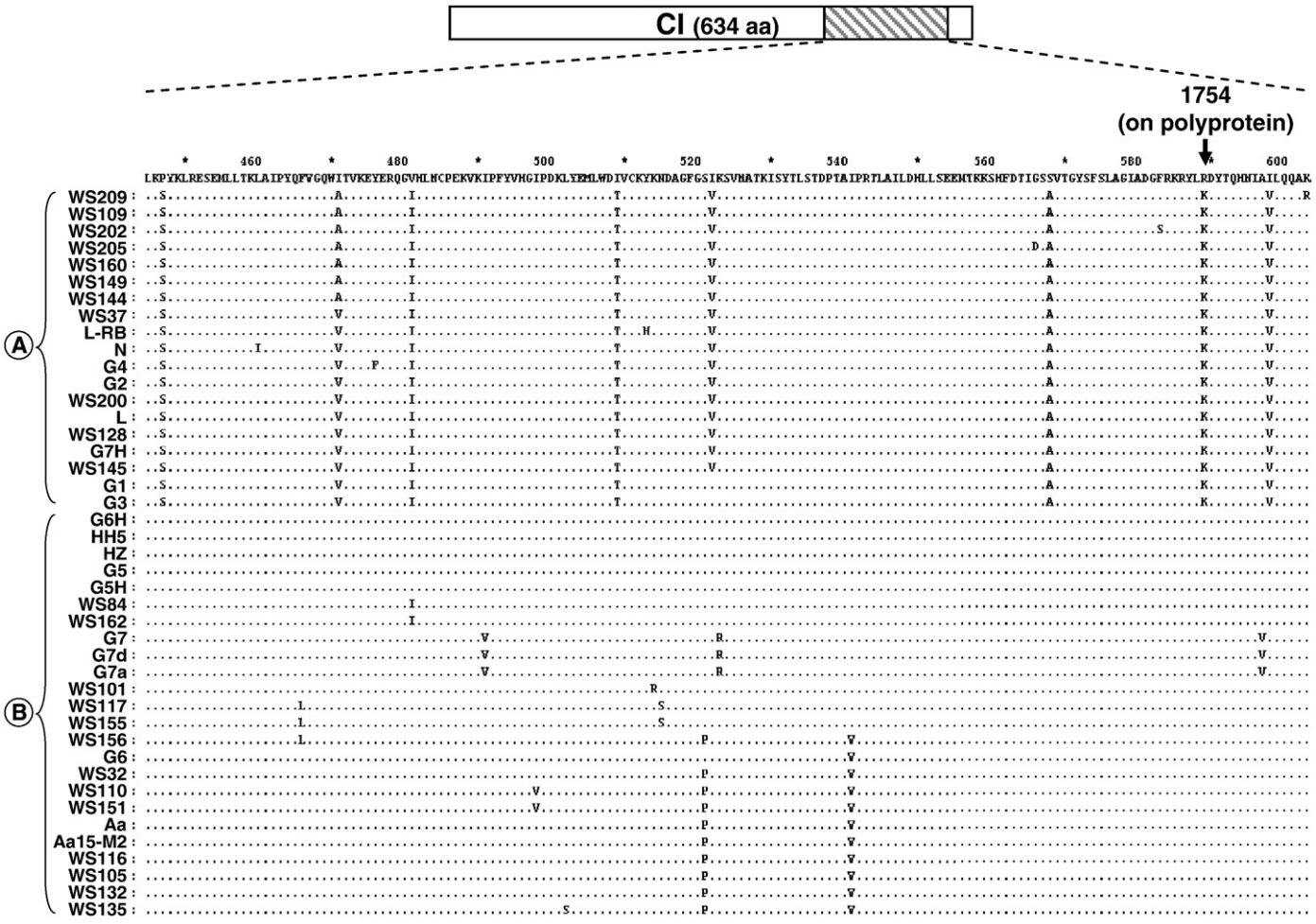


Fig. 6. Alignment of the C-terminal amino acid sequences of CI cistron of the SMV population. Sequences were aligned by ClustalX2. Amino acids that are the same are indicated with a '·'. Arrow indicates the amino acid position 1754 on the polyprotein precursor.

have yet accumulated in SMV population of North America. This, taken together with our phylogenetic results, also suggests that not much genetic drift has occurred in the SMV population. On the other hand, soybean is one of the major food crops and is traded worldwide. Since SMV is transmitted efficiently through seed, the possibility of SMV spread via soybean trade still exists. The involvement of such human activities in SMV distribution can be another possible reason for phylogenetic inconsistency on geographical clustering. In addition, we cannot exclude the possibility that the analyzed population size is too small to detect geographic differentiation between SMV subpopulations.

Our phylogenetic analyses did not show clear relationships between the phylogeny of the isolates and their pathogenicity, excepting the case of CI. It has been reported that soybean contains three independent resistance genes with several alleles at each locus (Buss et al., 1999; Chen et al., 1991; Gunduz et al., 2002; Hayes et al., 2000). Accumulating evidence suggests that different disease responses on soybean cultivars result from interactions among these resistance genes and viral pathogenic determinants (Eggenberger et al., 2008; Hajimorad et al., 2005; Hajimorad et al., 2006). As in many cases of potyvirus–host pathosystem (Eggenberger et al., 2008; Jenner et al., 2002; Jenner et al., 2003; Mestre et al., 2000), it seems likely that the pathogenic determinant of SMV functions at the amino acid level and that a single or several differences in viral amino acid sequences results in loss or gain of function as the pathogenic determinant. Furthermore, as reported previously (Gagarinova et al., 2008a; Hajimorad et al., 2003), resistance-breaking SMV can emerge by introductions of a few mutations changing amino acid sequences in the viral genome. In these cases, the parental and mutant viruses should be found to be most closely related to each other in phylogenetic trees and we cannot classify them according to their pathogenicity. In all, the difficulty in analyzing relationships between phylogeny and pathogenicity of SMV might be due to the high level of complexity in the SMV–soybean pathosystem. On the other hand, the phylogenetic tree of CI showed significant correlation between groupings based on the sequence similarity and groupings based on the pathogenicity (Fig. 3, panels C and D). Recently, we have demonstrated that strain-specific CI of SMV elicits *Rsv3*-mediated extreme resistance (Seo et al., 2009c). Although in that study we found that a single amino acid substitution at polyprotein position 1754 is responsible for pathogenicity of SMV, the amino acid sequence alignment of CI showed that there were two distinct types of sequence conservation in the C-terminal region of CI (Fig. 6). Furthermore, when we estimated amino acid sequence diversities for CI according to *p*-distance method, within-population diversities (0.0068 ± 0.0013 in Group A and 0.0118 ± 0.0021 in Group B) were significantly different to the between-population diversity (0.0296 ± 0.0051) as well as the overall diversity (0.0231 ± 0.0031), indicating the higher conservation of CI amino acid sequences within groups. These features of CI may make it possible to classify SMV isolates and strains according to their pathogenicity.

Accumulating evidence suggests that recombination is a common and important source of genetic variation in potyviruses (Bousalem et al., 2000; Moreno et al., 2004; Ogawa et al., 2008; Ohshima et al., 2007; Tomimura et al., 2003). An earlier study reported *in silico* analysis of recombination between distinct SMV

pathotypes based on 14 full-length SMV sequences retrieved from GenBank (Gagarinova et al., 2008b). In this study, we have analyzed 44 full-length SMV sequences to evaluate recombination in the SMV population using several recombination-detecting programs and finally identified 19 'clear' recombination events (Fig. 5). Interestingly, all identified recombination sites, excepting 'tentative' sites and recombination sites originated from a common early ancestor, were found in SMV isolates from Far East Asia. Thirteen of nineteen recombination events were detected in SMV isolates from wild soybean and the other six recombination events were found in SMV isolates from cultivated soybean. From an evolutionary point of view, it can be interpreted as comparing ecological differences in soybean cultivation between in the Far East Asia and in North America. It was reported that Asian SMV isolates showed more diverse symptoms and severities than those of North American isolates (Xu et al., 1986). Moreover, various soybean aphids, which can transmit SMV, as well as wild soybeans are widely distributed throughout the Far East Asia (Wu et al., 2004). Soybean has usually been cultivated in scattered small holdings in the Far East Asia. In this ecological condition, it seems possible that SMV might be cross-transmitted between wild soybean and cultivated soybean by aphids. Furthermore, mixed infections of soybean with more than two distinct viral genomes can occur occasionally. Consequently, it is likely that a high possibility of recombination in SMV exists in Far East Asia. On the other hand, in North America, the principle host of SMV is cultivated soybean and soybean aphids were discovered in North America in the early 2000s (Wu et al., 2004). In addition, soybean has been produced by large-scale cultivation in North America. Therefore, it can be considered that the incidence of horizontal transmission of SMV in North America was much lower than that in the Far East Asia, and hence recombination occurred rarely, if at all, in North America. Indeed, it is likely that no recombination has recently occurred in North America group since SMV originated from East Asia and the recombination signals at nt 3000–3050 in North America group might be a trace of recombination in early SMV differentiation as noted above. In addition, the existence of only one kind of trace of recombination in several SMV variants in North America group might be due to founder effects that resulted from the introduction of less diverse SMV variants into North America. It is logical to hypothesize that one of those introduced SMV was the ancestral virus harboring recombination at nt 3000–3050 and that this recombinant differentiated to produce several SMV variants still keeping on recombination at nt 3000–3050 in North America. However, as soybean aphids invade North America, the possibility of emergence of resistance-breaking SMV recombinants could increase.

It is interesting that G5, G5H, and G7H, which were prevalent in Korea, were identified as either a recombinant or its parental isolate. This result strongly suggests that recombination plays an important role in resistant-breaking of SMV in certain ecological conditions. Since G5H and G7H can infect several wild soybeans collected from natural fields in Korea while most of SMV isolated from wild soybean cannot infect resistant soybean cultivars (Table 2 and data not shown), we can suppose that, in Korea, recombination of SMV might have occurred in wild soybean infected with more than two SMV populations, and then resistant-breaking recombinants might be transmitted to resistant soybean cultivars. This further suggests that resistant-breaking SMV can emerge continuously in Korea. On the other hand, because G7H more recently emerged in soybean fields, it was unexpected that G7H was not identified as a recombinant, but as a parental isolate of G5H and G6H (Fig. 5 and Table 4). It is possible that G7H might have existed in natural fields before G5H and G6H emerged. However, we cannot exclude the possibility of obtaining false evidence in *in silico* assignments of parent/daughter designations because recombination analysis is affected by the phylogenetic order of analyzed sequences. To clarify this ambiguity, further analyses of recombination events using more diverse SMV sequence data may be needed.

In the present study, we have analyzed the genetic diversity and structure of the SMV population. The results showed that the SMV population is genetically conserved in spite of the high potential for genetic variation through accumulation of mutations and recombination. This report also revealed that recombination occurred naturally in the SMV population, and might play an important role in the emergence of resistance-breaking strains in Korea. Our study, together with previous studies (Gagarinova et al., 2008a,b; Hajimorad et al., 2003), shows that both mutation and recombination were involved in SMV evolution and contributed to overcoming host resistance. This should be considered when designing strategies for controlling SMV disease.

Materials and methods

Viral origin and biological characteristics

Twenty-four isolates of SMV were collected from wild soybean (*G. soja*) in Korea in the summer of 2006. Details of the isolates are shown in Table 1 and Fig. 1. All isolates collected were identified as SMV by RT-PCR using SMV-specific primers (5′-AAGCCAATCAATCTTTCCAG-3′ and 5′-CCAAAAGAGTCAATCACGTG-3′). Ten of 24 isolates were inoculated into several soybean cultivars including Lee68 (*rsv*), PI96983 (*Rsv1*), Suwon97 (*Rsv1-h*), L29 (*Rsv3*), V94-5152 (*Rsv4*), and Jinpumkong-2 (*Rsv*-genotype was unknown) to determine their pathotypes. Symptom development was observed for 3 weeks. Progeny viruses were recovered from infected plants and their P1 nucleotide sequences were determined by direct sequencing of RT-PCR products to confirm whether the initial inoculated SMV population was maintained. SMV strains G1, G3, G4, G6, G6H, and G7a were provided by Dr. Y.-H. Kim (NICS, Suwon, Korea). Initially, G1 (PV-571), G3 (PV-616), G4 (PV-572), G6 (PV-612), and G7a (PV-614) were purchased from the American Type Culture Collection (Rockville, USA) and G6H was collected from soybean fields in Korea. These strains were maintained in Lee68 (*rsv*) through several plant passages. Their pathotypes were determined as described by Kim (2000) and Kim et al. (2003) and infected seeds were deposited at NICS as virus stocks.

RT-PCR, cloning and sequencing

Total RNAs were extracted from SMV-infected tissues using the TRI Reagent method (MRC, USA) according to the manufacturer's instructions. cDNAs were synthesized using the M-MLV reverse transcriptase (Promega, USA), two SMV-specific primers (5′-CCYTGCAGYACACTAGTCATTTG-3′, which is complementary to nt 3231–3253; 5′-CCCAACCATACAAACCCGTTC-3′, which is complementary to nt 6576–6596), and a tagged oligo(dT) primer (5′-*GACTAGCTGGAATTCGCGGTTAAA*TTTTTTTTTTTTTTT-3′, the tag sequence is italicized) as described previously (Kim et al., 2008). Three large fragments covering SMV full-length genome were amplified using Ex *Taq* polymerase (TaKaRa, Japan) and appropriate primer pairs (5′-AAATTAAAACTMSTYATAAAGA-3′ and 5′-CCYTGCAGYACACTAGTCATTTG-3′, for 5′ fragment PCR; 5′-CTCCACATACGGARAAATG-3′ and 5′-CCCAACCATACAAACCCGTTC-3′, for middle fragment PCR; 5′-ATGTTTGGGGTYGGCTATGG-3′ and 5′-*GACTAGCTGGAATTCGCGGTTAAA*-3′, for 3′ fragment PCR). Adjacent regions of these PCR fragments were overlapped by at least 200 bp to ensure that they were from the same genome. Each PCR fragment was cloned into pGEM-T Easy vector (Promega, USA). Subsequently, more than 2 clones for each fragment were completely sequenced and sequence discords among clones were verified by direct sequencing of RT-PCR product of corresponding regions. 5′ terminal sequences of SMV were determined by the classic protocol (Sambrook and Russell, 2001) and

direct sequencing of PCR products. All sequences were assembled using DNA Star v. 5.02 (Lasergene, USA). DNA sequencing was performed using the dideoxynucleotide termination method and an ABI Prism 3730 XL DNA Analyzer (Applied Biosystems, USA) located at the National Instrumentation Center for Environmental Management of Seoul National University according to manufacturer's instructions. The nucleotide sequence data have been submitted to GenBank database with Accession numbers listed in Table 1.

Variation analyses

Multiple sequence alignments used in this study were obtained using the ClustalX2 program (Larkin et al., 2007). Nucleotide and encoded amino acid sequence similarities were analyzed using AlignX implemented in Vector NTI Suite (Invitrogen, USA). Pairwise genetic distances and pairwise synonymous and nonsynonymous substitutions were analyzed by the Kimura's two-parameter method (Kimura, 1980) and the Pamilo–Bianchi–Li method (Li, 1993; Pamilo and Bianchi, 1993), respectively, using MEGA v. 4.1 program (Tamura et al., 2007).

Phylogenetic analyses

The phylogenetic relationships of the SMV sequences were analyzed by the NJ and MP algorithms implemented in MEGA v. 4.1 program using WMV (GenBank Acc. code EU660580). In NJ analyses, the Kimura's two-parameter method and *p*-distance model were applied for nucleotide and amino acid sequence analyses, respectively. For MP analyses, the Close-neighbor-interchange model was used with default settings. Bootstrap values were calculated using 1000 random replications. The calculated trees were displayed using TreeExplorer implemented in MEGA v. 4.1 program. Phylogenetic network analysis was performed using SplitsTree v. 4.1 program (Huson and Bryant, 2006).

Recombination analyses

The phylogenetic correlation in the SMV population was analyzed using PHYLPRO v. 1.1 program with default settings (except for smooth degree 20) (Weiller, 1998) on the whole set of 44 full-length sequences, which were aligned by ClustalX2 and refined manually to maintain the degapped alignment. Recombination events, likely parental isolates of recombinants, and recombination break points were analyzed using the RDP, GENECONV, Chimaera, MaxChi, BOOTSCAN, and SISCAN methods implemented in the RDP3 program (Martin et al., 2005) with default settings and a Bonferroni corrected *P*-value cut-off of 0.01. To reduce the possibility of obtaining false recombination signals, only recombination events supported by 3 different methods with an associated *P*-value of $<1.0\times10^{-6}$ were further analyzed. Next, all likely recombinants identified by the RDP3 program were checked again using PHYLPRO v. 1.1, original SiScan v. 2, and RAT v. 1.0 (by setting window size to 400) programs (Etherington et al., 2005; Gibbs et al., 2000).

Acknowledgments

This research was supported in part by grants from the Korea Science and Engineering Foundation grant (No. R01-2008-000-10087-0) funded by the Ministry of Education, Science and Technology (MEST) and the Rural Development Administration (20090101-030-144-001-05-00 and 20090101-060-029-001-03-00). JKS, HGL and MS were supported by the graduate research fellowships from the MEST through Brain Korea 21 Project.

References

Aleman-Verdaguer, M.E., Goudou-Urbino, C., Dubern, J., Beachy, R.N., Fauquet, C., 1997. Analysis of the sequence diversity of the P1, HC, P3, Nlb and CP genomic regions of several yam mosaic potyvirus isolates: implications for the intraspecies molecular diversity of potyviruses. J. Gen. Virol. 78, 1253–1264.

Bateson, M.F., Lines, R.E., Revill, P., Chaleeprom, W., Ha, C.V., Gibbs, A.J., Dale, J.L., 2002. On the evolution and molecular epidemiology of the potyvirus *Papaya ringspot virus*. J. Gen. Virol. 83, 2575–2585.

Bousalem, M., Douzery, E.J., Fargette, D., 2000. High genetic diversity, distant phylogenetic relationships and intraspecies recombination events among natural populations of *Yam mosaic virus*: a contribution to understanding potyvirus evolution. J. Gen. Virol. 81, 243–255.

Buss, G.R., Ma, G., Chen, P., Tolin, S.A., 1999. A new allele at the *Rsv3* locus for resistance to soybean mosaic virus. In: Kauffman, H.E. (Ed.), Proceedings of the World Soybean Research Conference VI. Superior Printing, Chicago, IL, p. 490.

Chen, P., Buss, G.R., Roane, C.W., Tolin, S.A., 1991. Allelism among genes for resistance to soybean mosaic virus in strain-differential soybean cultivars. Crop Sci. 31, 305–309.

Chen, J., Zheng, H.Y., Lin, L., Adams, M.J., Antoniw, J.F., Zhao, M.F., Shang, Y.F., Chen, J.P., 2004. A virus related to *Soybean mosaic virus* from *Pinellia ternata* in China and its comparison with local soybean SMV isolates. Arch. Virol. 149, 349–363.

Cho, E.K., 1995. Identification of soybean mosaic virus strains and a consideration on genetics of soybean for resistance to SMV strains. Plant Dis. Agr. 1, 22–25.

Cho, E.K., Goodman, R.M., 1979. Strains of soybean mosaic virus: classification based on virulence in resistant soybran cultivars. Phytopathology 69, 467–470.

Cho, E.K., Chung, B.J., Lee, S.H., 1977. Studies on identification and classification of soybean disease in Korea. II. Etiology of a necrotic disease of *Glycine max*. Plant Dis. Rep. 61, 313–317.

Cho, E.K., Choi, S.H., Cho, W.T., 1983. Newly recognized soybean mosaic virus mutants and sources of resistance in soybeans. Res. Rep. ORD (SPMU) 25, 18–22.

Choi, B.K., Koo, J.M., Ahn, H.J., Yum, H.J., Choi, C.W., Ryu, K.H., Chen, P., Tolin, S.A., 2005. Emergence of *Rsv*-resistance breaking *Soybean mosaic virus* isolates from Korean soybean cultivars. Virus Res. 112, 42–51.

Chung, B.J., Lee, S.H., Cho, E.K., Park, H.C., 1974. Identification of soybean viruses and soybean varietal reactions. Annu. Rep. CIRC 137–145.

Desbiez, C., Lecoq, H., 2008. Evidence for multiple intraspecific recombinants in natural populations of *Watermelon mosaic virus* (WMV, *Potyvirus*). Arch. Virol. 153, 1749–1754.

Domingo, E., Martinez-Salas, E., Sobrino, F., de la Torre, J.C., Portela, A., Ortin, J., Lopez-Galindez, C., Perez-Brena, P., Villanueva, N., Najera, R., et al., 1985. The quasispecies (extremely heterogeneous) nature of viral RNA genome populations: biological relevance—a review. Gene 40, 1–8.

Eggenberger, A.L., Stark, D.M., Beachy, R.N., 1989. The nucleotide sequence of a soybean mosaic virus coat protein-coding region and its expression in *Escherichia coli*, *Agrobacterium tumefaciens* and tobacco callus. J. Gen. Virol. 70, 1853–1860.

Eggenberger, A.L., Hajimorad, M.R., Hill, J.H., 2008. Gain of virulence on *Rsv1*-genotype soybean by an avirulent *Soybean mosaic virus* requires concurrent mutations in both P3 and HC-Pro. Mol. Plant-Microb. Interact. 21, 931–936.

Etherington, G.J., Dicks, J., Roberts, I.N., 2005. Recombination Analysis Tool (RAT): a program for the high-throughput detection of recombination. Bioinformatics 21, 278–281.

Gagarinova, A.G., Babu, M., Poysa, V., Hill, J.H., Wang, A., 2008a. Identification and molecular characterization of two naturally occurring *Soybean mosaic virus* isolates that are closely related but differ in their ability to overcome Rsv4 resistance. Virus Res. 138, 50–56.

Gagarinova, A.G., Babu, M., Stromvik, M.V., Wang, A., 2008b. Recombination analysis of *Soybean mosaic virus* sequences reveals evidence of RNA recombination between distinct pathotypes. Virol. J. 5, 143.

Garcia-Arenal, F., Fraile, A., Malpica, J.M., 2001. Variability and genetic structure of plant virus populations. Annu. Rev. Phytopathol. 39, 157–186.

Gibbs, M.J., Armstrong, J.S., Gibbs, A.J., 2000. Sister-scanning: a Monte Carlo procedure for assessing signals in recombinant sequences. Bioinformatics 16, 573–582.

Gibbs, A.J., Trueman, J.W., Gibbs, M.J., 2008. The *Bean common mosaic virus* lineage of potyviruses: where did it arise and when? Arch. Virol. 153, 2177–2187.

Glasa, M., Palkovics, L., Kominek, P., Labonne, G., Pittnerova, S., Kudela, O., Candresse, T., Subr, Z., 2004. Geographically and temporally distant natural recombinant isolates of *Plum pox virus* (PPV) are genetically very similar and form a unique PPV subgroup. J. Gen. Virol. 85, 2671–2681.

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

Hajimorad, M.R., Eggenberger, A.L., Hill, J.H., 2003. Evolution of *Soybean mosaic virus*-G7 molecularly cloned genome in *Rsv1*-genotype soybean results in emergence of a mutant capable of evading *Rsv1*-mediated recognition. Virology 314, 497–509.

Hajimorad, M.R., Eggenberger, A.L., Hill, J.H., 2005. Loss and gain of elicitor function of *Soybean mosaic virus* G7 provoking *Rsv1*-mediated lethal systemic hypersensitive response maps to P3. J. Virol. 79, 1215–1222.

Hajimorad, M.R., Eggenberger, A.L., Hill, J.H., 2006. Strain-specific P3 of *Soybean mosaic virus* elicits *Rsv1*-mediated extreme resistance, but absence of P3 elicitor function alone is insufficient for virulence on *Rsv1*-genotype soybean. Virology 345, 156–166.

Hayes, A.J., Ma, G., Buss, G.R., Saghai Maroof, M.A., 2000. Molecular marker mapping of *Rsv4*, a gene conferring resistance to all known strains of *Soybean mosaic virus*. Crop Sci. 40, 1434–1437.

Holmes, E.C., Worobey, M., Rambaut, A., 1999. Phylogenetic evidence for recombination in dengue virus. Mol. Biol. Evol. 16, 405–409.

Huson, D.H., Bryant, D., 2006. Application of phylogenetic networks in evolutionary studies. Mol. Biol. Evol. 23, 254–267.
Hymowitz, T., Harlan, J.R., 1983. Introduction of soybean to North America by Samuel Bowen in 1765. Econ. Bot. 37, 371–379.
Jayaram, C., Hill, J.H., 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.
Jenner, C.E., Tomimura, K., Ohshima, K., Hughes, S.L., Walsh, J.A., 2002. Mutations in *Turnip mosaic virus* P3 and cylindrical inclusion proteins are separately required to overcome two *Brassica napus* resistance genes. Virology 300, 50–59.
Jenner, C.E., Wang, X., Tomimura, K., Ohshima, K., Ponz, F., Walsh, J.A., 2003. The dual role of the potyvirus P3 protein of *Turnip mosaic virus* as a symptom and avirulence determinant in brassicas. Mol. Plant-Microb. Interact. 16, 777–784.
Kim, J.S., Lee, E.J., 1991. A new virulent strain of soybean mosaic virus infecting SMV resistant soybean cultivar. Deogyou Korean J. Plant Pathol. 7, 37–41.
Kim, S.S., Ko, Y.J., Jang, J.Y., Lee, T., Lim, M.H., Park, S.Y., Bae, S.C., Yun, C.H., Park, B.S., Hwang, D.J., 2008. Isolation and expression analysis of *Brassica rapa WRKY 7*. Plant Pathol. J. 24, 478–481.
Kim, Y.H., 2000. Molecular Classification of *Soybean Mosaic Virus* Strains and their Interaction with Soybeans. Ph.D. Thesis, Kyungpook Natl. Univ., Daegu, Korea.
Kim, Y.H., Kim, O.S., Lee, B.C., Moon, J.K., Lee, S.C., Lee, J.Y., 2003. G7H, a new *Soybean mosaic virus* strain: its virulence and nucleotide sequence of CI gene. Plant Dis. 87, 1372–1375.
Kimura, M., 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J. Mol. Evol. 16, 111–120.
Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., Thompson, J.D., Gibson, T.J., Higgins, D.G., 2007. Clustal W and Clustal X version 2.0. Bioinformatics 23, 2947–2948.
Li, W.H., 1993. Unbiased estimation of the rates of synonymous and nonsynonymous substitution. J. Mol. Evol. 36, 96–99.
Lim, W.-S., Kim, Y.-H., Kim, K.-H., 2003. Complete genome sequences of the genomic RNA of *Soybean mosaic virus* strains G7H and G5. Plant Pathol. J. 19, 171–176.
Martin, D.P., Williamson, C., Posada, D., 2005. RDP2: recombination detection and analysis from sequence alignments. Bioinformatics 21, 260–262.
Mayo, M.A., Pringle, C.R., 1998. Virus taxonomy—1997. J. Gen. Virol. 79, 649–657.
Mestre, P., Brigneti, G., Baulcombe, D.C., 2000. An *Ry*-mediated resistance response in potato requires the intact active site of the NIa proteinase from *Potato virus Y*. Plant J. 23, 653–661.
Moreno, I.M., Malpica, J.M., Diaz-Pendon, J.A., Moriones, E., Fraile, A., Garcia-Arenal, F., 2004. Variability and genetic structure of the population of *Watermelon mosaic virus* infecting melon in Spain. Virology 318, 451–460.
Nei, M., Gojobori, T., 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol. Biol. Evol. 3, 418–426.
Ogawa, T., Tomitaka, Y., Nakagawa, A., Ohshima, K., 2008. Genetic structure of a population of *Potato virus Y* inducing potato tuber necrotic ringspot disease in Japan; comparison with North American and European populations. Virus Res. 131, 199–212.
Ohshima, K., Yamaguchi, Y., Hirota, R., Hamamoto, T., Tomimura, K., Tan, Z., Sano, T., Azuhata, F., Walsh, J.A., Fletcher, J., Chen, J., Gera, A., Gibbs, A., 2002. Molecular evolution of *Turnip mosaic virus*: evidence of host adaptation, genetic recombination and geographical spread. J. Gen. Virol. 83, 1511–1521.
Ohshima, K., Tomitaka, Y., Wood, J.T., Minematsu, Y., Kajiyama, H., Tomimura, K., Gibbs, A.J., 2007. Patterns of recombination in *Turnip mosaic virus* genomic sequences indicate hotspots of recombination. J. Gen. Virol. 88, 298–315.
Oruetxebarria, I., Kekarainen, T., Spetz, C., Valkonen, J.P., 2000. Molecular characterization of *Potato virus V* genomes from Europe indicates limited spatiotemporal strain differentiation. Phytopathology 90, 437–444.
Pamilo, P., Bianchi, N.O., 1993. Evolution of the *Zfx* and *Zfy* genes: rates and interdependence between the genes. Mol. Biol. Evol. 10, 271–281.
Probst, A.H., Judd, R.W., 1973. Origin, U.S. history, development and world distribution. In: Caldwell, B.E. (Ed.), Soybeans: Improvement, Production, and Uses. Amer. Soc. Agronomy, Madison, WI, pp. 1–15.
Riechmann, J.L., Lain, S., Garcia, J.A., 1992. Highlights and prospects of potyvirus molecular biology. J. Gen. Virol. 73, 1–16.
Roossinck, M.J., 1997. Mechanisms of plant virus evolution. Annu. Rev. Phytopathol. 35, 191–209.
Sambrook, J., Russell, D.W., 2001. Rapid amplification of 5′ cDNA ends. In: Sambrook, J., Russell, D.W. (Eds.), Molecular Cloning: A Laboratory Manual, 2. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 8.54–58.60.
Seo, J.-K., Lee, H.-G., Choi, H.-S., Lee, S.-H., Kim, K.-H., 2009a. Infectious in vivo transcripts from a full-length clone of *Soybean mosaic virus* strain G5H. Plant Pathol. J. 25, 54–61.
Seo, J.-K., Lee, H.-G., Kim, K.-H., 2009b. Systemic gene delivery into soybean by simple rub-inoculation with plasmid DNA of a *Soybean mosaic virus*-based vector. Arch. Virol. 154, 87–99.
Seo, J.-K., Lee, S.-H., Kim, K.-H., 2009c. Strain-specific cylindrical inclusion protein of Soybean mosaic virus elicits extreme resistance and a lethal systemic hypersensitive response in two resistant soybean cultivars. Mol. Plant-Microbe Interact. 22 (9), 1151–1159.
Shukla, D.D., Ward, C.W., Brunt, A.A., 1994. "*The Potyviridae*". CAB International, Wallingford, UK.
Tamura, K., Dudley, J., Nei, M., Kumar, S., 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol. Biol. Evol. 24, 1596–1599.
Tomimura, K., Gibbs, A.J., Jenner, C.E., Walsh, J.A., Ohshima, K., 2003. The phylogeny of *Turnip mosaic virus*; comparisons of 38 genomic sequences reveal a Eurasian origin and a recent 'emergence' in east Asia. Mol. Ecol. 12, 2099–2111.
Tomimura, K., Spak, J., Katis, N., Jenner, C.E., Walsh, J.A., Gibbs, A.J., Ohshima, K., 2004. Comparisons of the genetic structure of populations of *Turnip mosaic virus* in West and East Eurasia. Virology 330, 408–423.
Weiller, G.F., 1998. Phylogenetic profiles: a graphical method for detecting genetic recombinations in homologous sequences. Mol. Biol. Evol. 15, 326–335.
Wu, Z., Schenk-Hamlin, D., Zhan, W., Ragsdale, D.W., Heimpel, G.E., 2004. The soybean aphid in China: a historical review. Ann. Entomol. Soc. Am. 97, 209–218.
Xu, B., 1989. Analysis of ecotypes and their geographical distribution of wild soybean (*G. soja*) in the world. Ecological Acta 13, 114–120.
Xu, Z., Polston, J.E., Goodman, R.M., 1986. Identification of soybean mosaic, southern bean mosaic and tobacco ringspot viruses from soybean in the People's Republic of China. Ann. Appl. Biol. 108, 51–57.
Yu, H., Kiang, Y.T., 1993. Genetic variation in South Korean natural populations of wild soybean (*Glycine soja*). Euphytica 68, 213–221.