농학석사학위논문

The evolutionary flow of the soybean mosaic virus population based on the analysis of newly sequenced isolates in Korea

국내 콩 모자이크 바이러스 개체군의 진화 양상 분석

2023년 2월

서울대학교 대학원

농생명공학부 식물미생물학전공

오지선

A THESIS FOR DEGREE OF MASTER OF SCIENCE

The evolutionary flow of the soybean mosaic virus population based on the analysis of newly sequenced isolates in Korea

BY

JISEON OH

Department of Agricultural Biotechnology

The Graduate School of Seoul National University

February 2023

농학석사학위논문

국내 콩 모자이크 바이러스 개체군의 진화 양상 분석

지도교수 김 국 형

이 논문을 농학석사학위논문으로 제출함

2023년 2월

서울대학교 대학원

농생명공학부 식물미생물학전공

오지선

오지선의 석사학위논문을 인준함

2023년 2월

위 원 장 _____ (인)

부위원장 _____ (인)

위 원 _____ (인)

A THESIS FOR THE DEGREE OF MASTER OF SCIENCE

The evolutionary flow of the soybean mosaic virus population based on the analysis of newly sequenced isolates in Korea

UNDER THE DIRECTION OF

DR. KOOK-HYUNG KIM

SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL OF SEOUL NATIONAL UNIVERSITY

BY JISEON OH

MAJOR IN PLANT MICROBIOLOGY DEPARTMENT OF AGRICULTURAL BIOTECHNOLOGY

FEBRUARY 2023

APPROVED AS A QUALIFIED THESIS OF JISEON OH FOR THE DEGREE OF MASTER OF SCIENCE BY THE COMMITTEE MEMBERS

CHAIRMAN _____

VICE CHAIRMAN _____

MEMBER _____

The evolutionary flow of the soybean mosaic virus population based on the analysis of newly sequenced isolates in Korea

Jiseon Oh
Major in Plant Microbiology
Department of Agricultural Biotechnology
The Graduate School of Seoul National University

ABSTRACT

Soybean mosaic virus (SMV) is the most chronic and devastating soybean pathogen worldwide. In this study, 12 new isolates were detected from the 41 soybean-growing regions in Korea. Phylogenetic analysis revealed 4 main phylogroups (A to D) in the global SMV population and showed the geographic features of each phylogroup. To identify the population dynamics in Korea, we conducted an additional analysis using Korean isolates. Recombination analysis prevailed 7 of 12 new isolates were recombinants. In addition, 3 showed unique signals suggesting recombination constantly affected the evolution of SMV. In total 27 non-recombinant Korea SMV isolates, only 2 phylogroups, A and B, remained, and the statistically significant genetic difference was checked. Even though isolates of B showed higher genetic diversity than A, both phylogroups are under negative selection pressure in all 11 domains. In addition, phylogroup B, which contains G6 and G7 strains, seems to be preferred with the recently emerged subgroup, B1. In the sequence similarity plot, the novel subgroup (B1) showed the highest distinction in the P1 region, suggesting that host range or

pathogenicity changes occur in the SMV population. These results provide information about the evolutionary flow of the SMV population in Korea.

Keywords: soybean mosaic virus, *Glycine max*, phylogenetic analysis, population genetics, recombination analysis

Student number: 2021-24653

CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

Soybean mosaic virus (SMV) is a chronic soybean pathogen worldwide. SMV induces various symptoms like a mosaic, mottling, necrosis, and stunting, in the economically valuable host, *Glycine max*, causing severe damage to crop quality and yield (Hajimorad et al., 2018). Since the first report in 1921, decades of effort have been made to control this devastating pathogen, and to date, four independent resistant genes, Rsv1, Rsv2, Rsv3, and Rsv5, have been found and used for breeding SMV-resistant cultivars (Gardner and Kendrick, 1921; Usovsky et al., 2022). Nevertheless, resistance-breaking isolates against each R-gene are reported constantly, making it challenging to control SMV (Hajimorad et al., 2018).

SMV, genus *Potyvirus*, family *Potyviridae*, has a flexible filamentous virion consisting of an RNA genome, 5' end joined genome-linked viral protein (VPg), and capsid protein (CP). The positive-sense single-stranded RNA genome encodes large polyprotein, which is self-digested into 10 functional proteins: P1, HC-Pro, P3, 6K1, CI, 6K2, VPg, NIa-Pro, NIb, and CP. During the replication, frameshift on the P3 cistron generates an additional protein, PIPO. These 11 proteins are involved in various functions like host adaptation (P1), genome replication (NIb), cell-to-cell movement (MP), vector transmission (HC-Pro and CP), and so on (Hajimorad et al., 2018; Riechmann et al., 1992; Widyasari et al., 2020).

In the SMV quasispecies, researchers found various pathotypes and classified isolates into several strains according to their pathogenicity (Cho and Goodman, 1979; Li et al., 2010; Takahashi et al., 1980). In Korea, American strains (G1-G7) were adopted with additional Korean strains (G5H, G6H, and G7H) rather than China strains

(SMV-SC strains) (Cho and Chung, 1986; Kim et al., 1991; Kim et al., 2003). Many phylogenetic studies have been conducted based on these strains and found that the dominant strains dynamically switch in the SMV populations over time in Korea, from G5 to G5H and from G5H to G7H (Cho et al., 1983; Kim et al., 2003; Kim, 2000).

Understanding the dynamic of the virus population is essential for effective disease control, especially in breeding resistant cultivars (García-Arenal et al., 2003). However, the resistance can be easily detoured by various methods making it difficult to be maintained. In this study, we evaluated the phylogenetic structure of SMV with newly emerged isolates and estimated the explanation for clade formation. In addition, we checked the evolution of the Korean SMV population using population genetic parameters. I hope my study can provide additional information for a better understanding of the evolution of SMV and eventually contribute to designing management strategies.

MATERIAL AND METHODS

1. Sample collection and total RNA extraction

Soybean leaf samples showing virus-like symptoms were collected and stored in a region-tagged zip-loc plastic bag. All leaf stocks were stored at –80 ºC to minimize deformation. Total RNA was extracted from pooled samples of each region using RNAiso Plus reagent (Takara, Japan) according to the manufacturer's protocol. After RNA extraction, RNA integrity was checked by agarose gel electrophoresis in 1.2% agarose gel, and concentration was quantified using NanoPhotometer® (IMPLEN, USA).

2. Virus detection

Ribosomal RNA was removed from the total RNA samples of 2020 using TruSeq Stranded Total RNA with a Ribo-Zero Plant Kit (Illumina, San Diego, CA, USA). Then, 27 libraries for RNA sequencing were generated using the TruSeq Stranded Total RNA LT Sample Prep Kit (Illumina). Twenty-seven libraries representing each collection region were paired-end (2 × 100 bp) sequenced using a NovaSeq 6000 system (Macrogen, Seoul, Korea). The raw sequence reads were *de novo* assembled by the Trinity program with default parameters (Haas et al., 2013). The obtained contigs were subjected to BLASTX search with a cutoff value of 1×10^{-10} against the plant viral database in the NCBI (https://www.ncbi.nlm.nih.gov/genome/viruses/). To eliminate non-viral sequences, the obtained virus-associated contigs were again subjected to BLASTX search against the NCBI non-redundant (NR) protein database.

Based on the virome information detected in the Korean soybean field in 2020, reverse transcription PCR was performed to detect viruses from soybean samples collected in 2021. First-strand cDNA was synthesized from total RNA (2ug) using GoScript™ Reverse Transcriptase (Promega, USA) according to the manufacturer's protocol. PCR was conducted using r Taq® polymerase (TaKaRa, Japan), with conditions of 3 min at 95 °C followed by 35 cycles of 30 sec at 95°C, 45 sec at primer-specific annealing temperature, 1 min at 72 °C, and finish reaction with 10 min at 72 °C.

3. SMV genome sequencing and assembly

SMV sequencing primer sets were designed after aligning NCBI-registered and newly sequenced SMV isolates (Table 1). Each of the terminal regions was amplified by the RACE system. pGEM®-T Easy Vector (Promega, USA) was used for cloning, and isolated plasmid was sequenced from both directions by sanger sequencing (Macrogen, Seoul, Korea). Sequencing data were assembled by SeqMan II version 5.01 (Burland, 2000).

Table 1. Primer sets used for SMV total genome sequencing.

Primer		Sequences (5′ → 3′)	Product size
Fragment 1	Fw	GAT TGG AAG CAT GGC GAT TT	1209bp
	Rv	TTC ACA TAC YTC ATG CCG TCA A	
Fragment 2	Fw	AAG CCA ATC AAT CTT TCC AG	2064bp
	Rv	CCY TGC AGY ACA CAC TAG TCA TTT G	
Fragment 3	Fw	CAG GTG CTA CAG TGA TAT AG	2535bp
	Rv	TAT CTG CCT CTT CTT TCC TTG A	
Fragment 4	Fw	CTG GAC CAT TTG TTG AGT GA	2233bp
	Rv	TCA GGG ATC CAT TCC AGA CT	
Fragment 5	Fw	GAA TAT GAA AGC TGC AGT TGG TG	2236bp
	Rv	GAC TAG TTT TTT TTT TTT TTT TTT TTT GGA CAA CAA ACA TTG	
5’ RACE	Fw1	CAG CCA ACG GAA TTC CTC ACT AAA CCC CCC CCC CCC CC	580bp
	Fw2	CAG CCA ACG GAA TTC CTC ACT AAA	
	Rv	AAC CCA ACA CTC CCT CCT TCA GAC	
3’ RACE	Fw	GCT ATG CTT TTG ATT TCT ATG AG	471bp
	Rv1	GAC TAG CTG GAA TTC GCG GTT AAA TTT TTT TTT TTT TTT T	
	Rv2	GAC TAG CTG GAA TTC CGC GTT AAA	

4. Sequence alignment and phylogenetic analysis

One hundred and thirty-two SMV isolates (present study=12, NCBI database=120) and 2 outgroup viruses (BCMV and WMV) were used to analyze the phylogenetic structure of the SMV population (Supplementary Table S1). Since most of the SMV sequences in the NCBI database were incomplete at the terminal site, only the polyprotein coding nucleotide sequence was used in this study. Phylogenetic analysis, sequence alignment, best-fit model selection for the tree, and tree construction were accomplished using MEGA11 (Tamura et al., 2021). After sequence alignment using ClustalW (codon) method, the maximum likelihood (ML) tree was constructed with the GTR + Γ_5 + I model and 1000 bootstrap replicates. Finally, the phylogenetic tree was visualized using FigTree v1.4.4 (http://tree.bio.ed.ac.uk/software/figtree/).

5. Haplotype network analysis

A minimum spanning network was generated using the alignment of 132 SMV isolates by PopART v1.7 (Leigh and Bryant, 2015). The collection region of each isolate was acquired from the NCBI database and previous studies (Hajimorad et al., 2018; Seo et al., 2009).

6. Recombination analysis

The existence of recombination was analyzed from Korea SMV isolates (Korea=47, outgroup=2) by the NeighborNet method using SplitsTree4 (Huson and Bryant, 2006). Then, specific recombination signals were analyzed using seven different algorithms (RDP, GENECONV, Chimaera, MaxChi, BootScan, SiScan, 3Seq) as default settings using RDP4.101 (Martin et al., 2015). In this step, we added the sequence of 7

American SMV pathogenicity strains (N, G1, G2, G3, G4, G6, and G7) to the previous dataset to analyze the phylogeny of each sequence. Only events supported by more than 4 methods, with an associated P-value < 1×10^{-6}, were considered significant to minimize deceitful signals. Lastly, putative recombination events were rechecked using the recombination analysis tool (RAT) (Etherington et al., 2005). Recombinant sequences were eliminated from datasets for further phylogenetic analysis (Fuentes et al. 2022).

7. Calculation of the population genetic parameter

Population structure and selection pressure of Korean SMV isolates were estimated from several population genetic parameters using DnaSP v6 (Rozas et al., 2017). Nucleotide diversity (π) and haplotype diversity (h) were evaluated for each domain to estimate the genetic variability in phylogroups (Nei, 1987). Tajima's D, Fu & Li's D, and Fu's Fs statistics were calculated for the neutrality test (Fu, 1997; Fu and Li, 1993; Tajima, 1989). Selection pressure was estimated from nonsynonymous (dN) to synonymous (dS) substitution ratio (ω=dN/dS) (Kryazhimskiy and Plotkin, 2008). Three permutation statistics (Ks*, Z*, and Snn) were calculated with 1000 replicates to estimate the genetic differentiation between phylogroups (Hudson, 2000). The null hypothesis (no genetic differentiation) will be rejected if statistics are strongly supported by *P* values < 0.05. The degree of gene flow between phylogroups was evaluated from Fst (allele frequency across populations) and Nm (migration rate) values (Sun et al., 2021). Fst and Nm values can be interpreted as if |Fst|> 0.33 or |Nm|< 1, gene flow between phylogroups is infrequent, and if |Fst|< 0.33 or |Nm|> 1, gene flow between phylogroups is frequent (Hudson et al., 1992; Lu et al., 2021). The sequence

similarity plot was constructed using SimPlot++ software (Samson et al., 2022) with default settings to detect significant genomic differences between phylogroups.

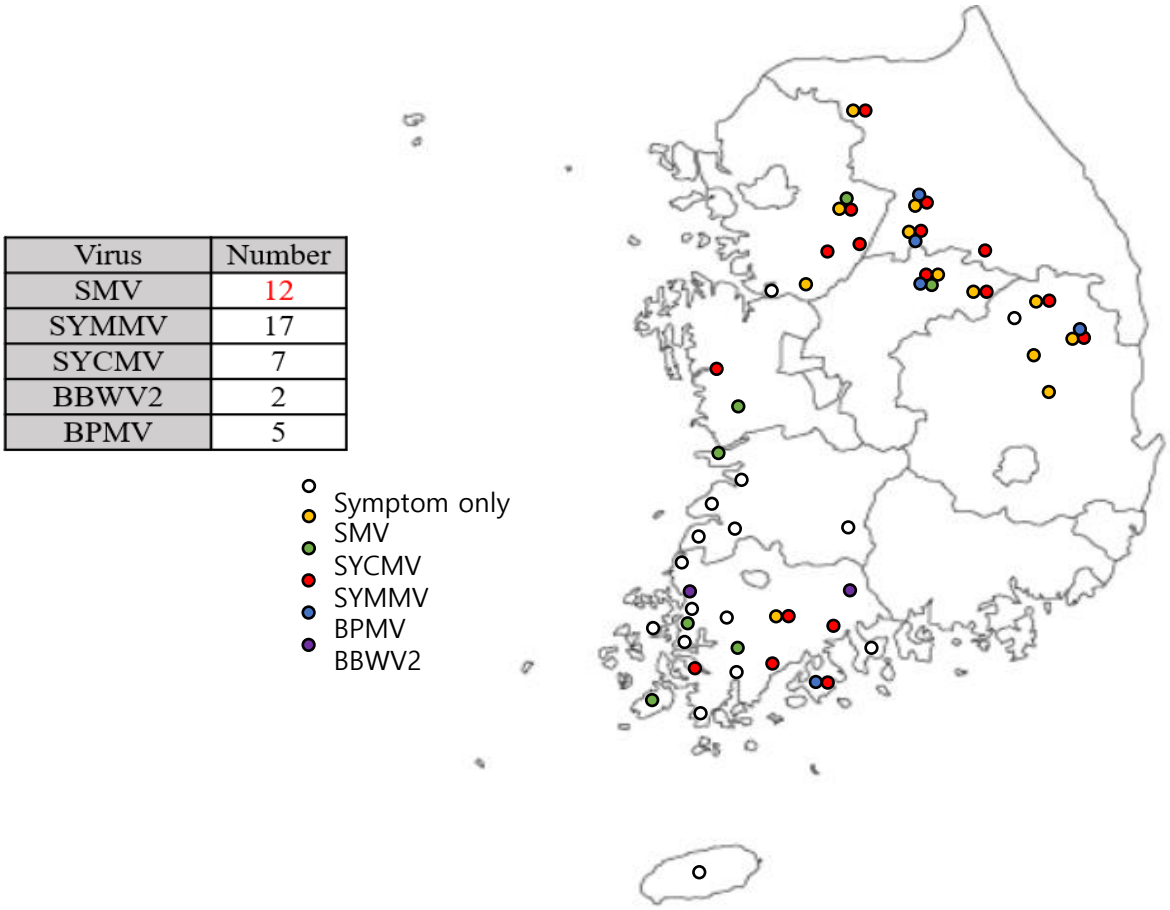
RESULTS

1. Virus detection and SMV genome sequencing

From 2020 to 2021, symptomatic soybean leaves were collected from 41 regions in Korea. From the RNA sequencing and RT-PCR detection, 12 new SMV isolates and other soybean infecting viruses (SYMMV, SYCMV, BBWV2, and BPMV) were detected from the 27 local samples (Figure 1). In 2020, a nearly complete genome sequence of 5 SMV isolates was obtained from RNA sequencing data. In 2021, SMV was detected from 7 of the 11 collection regions, and total genome sequences were obtained from the alignment of 5 overlapping PCR fragments. The sequence of the 5' and 3' terminal regions was confirmed by RACE PCR. The complete nt sequences of 12 new SMV isolates were registered in the GenBank database under the accession number OQ161628-OQ161639.

Based on the RNA sequencing data, detection primer sets for SYMMV, SYCMV, BBWV2, and BPMV were designed and used for virus detection from 2021 collected samples. SYMMV was detected in ten of 11 regions, while BPMV and SYCMV were detected from 4 and 2 regions, respectively. In addition, more than two kinds of viruses were detected from 7 regional samples suggesting the possibility of mixed infection (Supplementary Table S2).

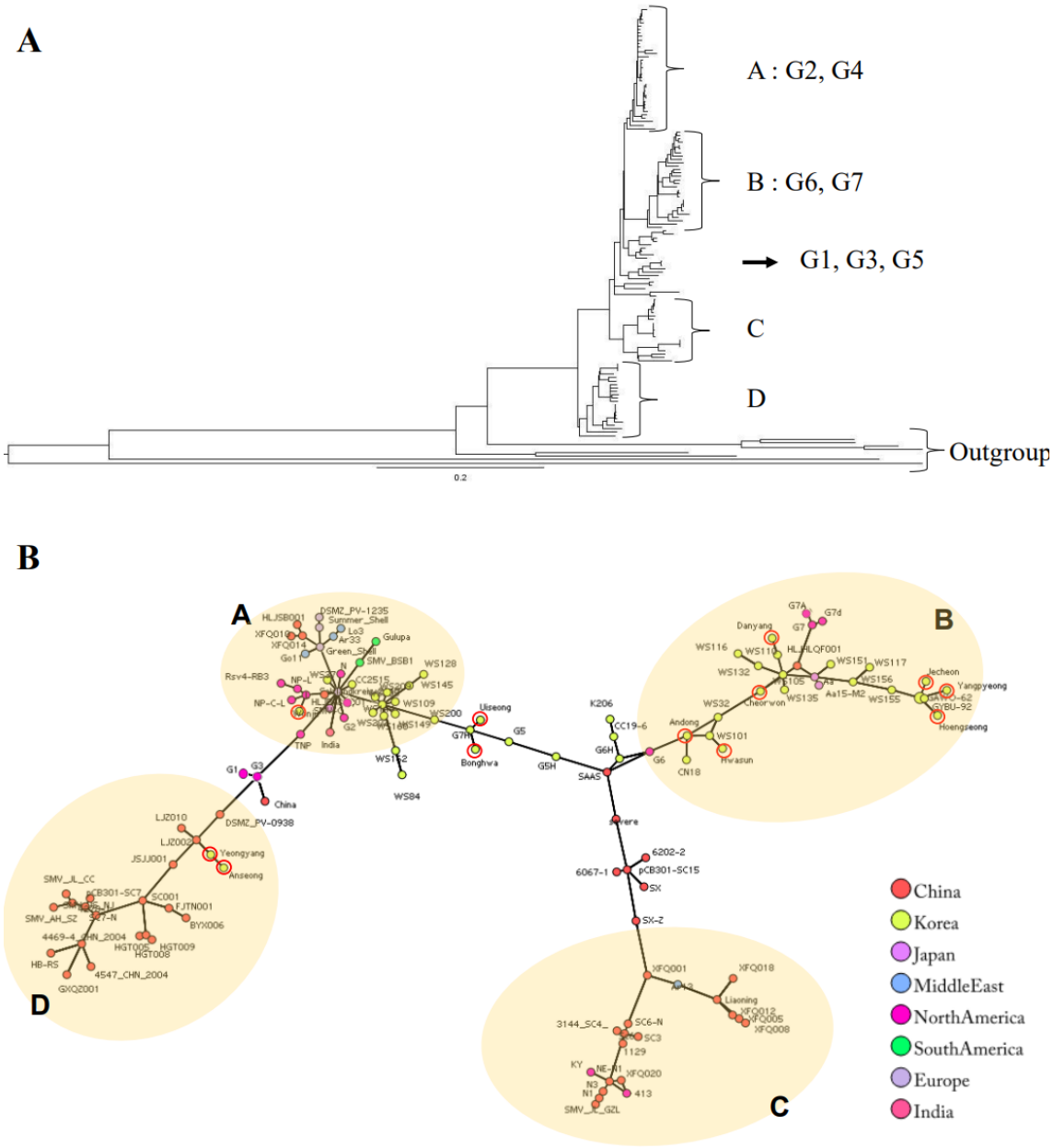
Figure 1. The geographical location of soybean-infecting viruses in Korea.



2. Phylogenetic analysis of the SMV population

The ML tree shows four major clades in the SMV population: A, B, C, and D (Figure 2A). Group A consists of G2, G4, and N strains, and isolates belonging to this group have been found in various countries of the world. Group B consisted of G6 and G7 strains, most of which were Korean isolates (22 out of 29). G1, G3, and G5 strains did not belong to any major clade. Most of the C and D groups were isolated from China, and in particular, the D groups were presumed to have occurred by the recombination events of SMV and BCMV (Yang et al., 2011; Yang et al., 2014; Zhou et al., 2015). Haplotype network analysis reveals the presence of the same clades as in the ML tree, with isolates not belonging to these clades appearing between them (Figure 2B). The new SMV isolates obtained in this study belong to 1 in group A, 7 in B, 2 in D, and 2 in an ungrouped region close to the G7H strain. Additionally, most of the new isolates were located at the end of each branch suggesting continuous sequence variation in the SMV population.

Figure 2. Maximum likelihood tree (A) and Haplotype network (B) of global SMV population.



3. Detection of recombination signal

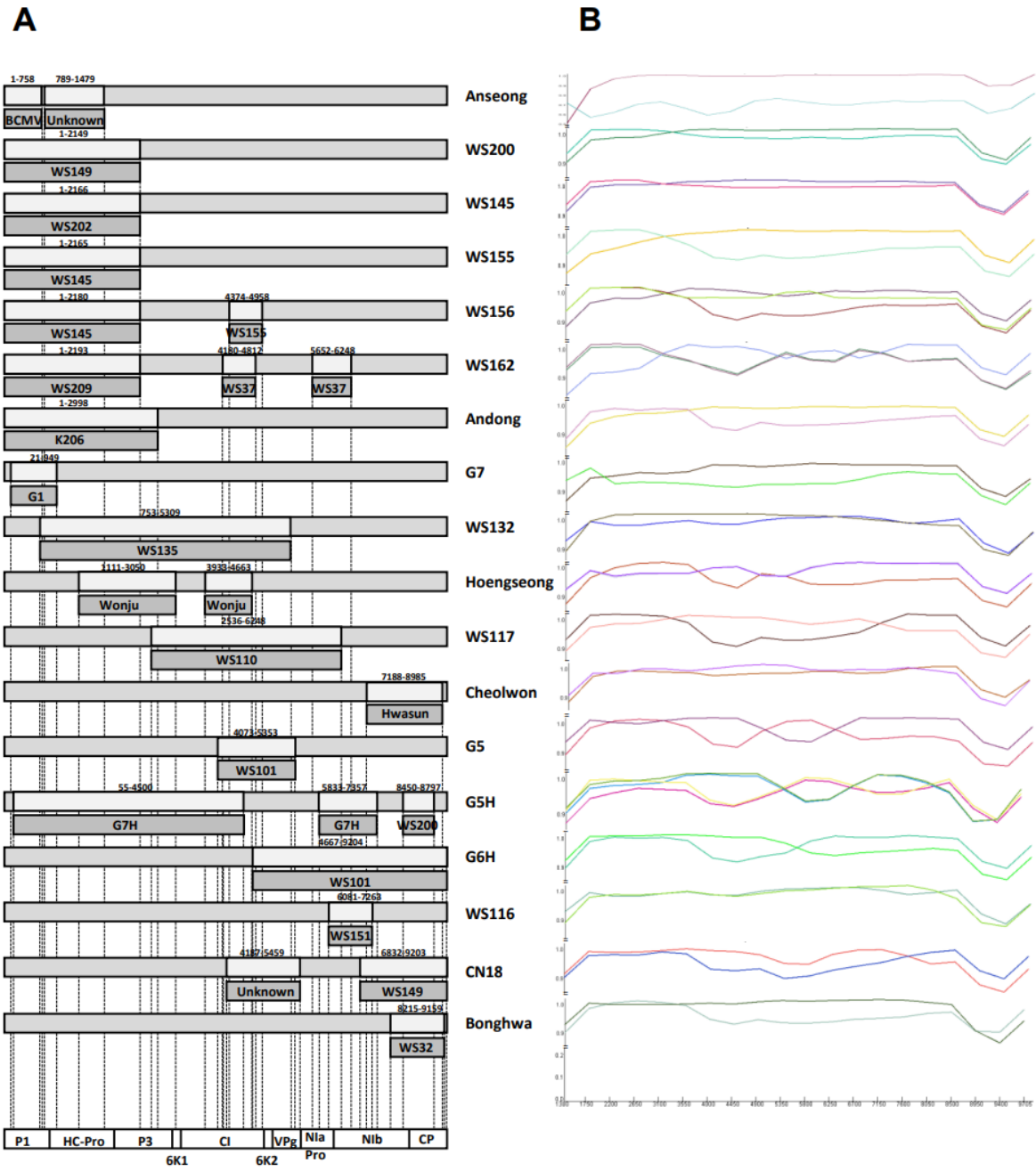
SplitsTree4 network analysis showed a reticulated network in the Korean SMV isolates suggesting the recombination signal exists in the population (Data not shown). The RDP4 program detected 26 unique recombination events from 56 SMV isolates (Table 2). RAT software confirmed each recombination event, major, and minor parent (Figure 3B). In the SMV population, recombination breakpoints were not restricted to the specific genomic region, and the length of the recombination region varied from 348bp to 4557bp. Both inter- and intragroup recombination were detected, and interspecific recombination events with BCMV were also detected from 2 isolates of this study (Figure 3A).

To calculate population genetic parameters in the Korea SMV population, putative recombinants were eliminated from the datasets. Twelve of group A, 10 of B, and 5 of ungrouped isolates remain in non-recombinant Korea SMV isolates (Figure 4A).

Table 2. Recombination data of Korean SMV isolates using the RDP4 program.

Event	Breakpoint	Recombinant Sequence(s)	Minor Parent	Major Parent	RDP	GENECONV	Bootscan	Maxchi	Chimaera	SiSscan	3Seq
1	1-758	Anseong, Yeongyang	BCMV NL1	G3	1.49E-142	1.98E-113	NS	7.21E-24	4.25E-26	1.98E-26	1.99E-61
2	4073-5353	G5	WS101	G7H	3.17E-05	1.84E-47	NS	2.33E-19	5.87E-20	1.78E-19	2.60E-12
3	1-2149	WS200	WS149	G7H	2.69E-11	1.16E-08	2.44E-09	3.68E-13	NS	2.37E-23	1.12E-43
4	1111-3050	Hoengseong	Wonju	Jecheon	2.47E-38	6.85E-36	1.90E-38	2.14E-15	5.10E-16	4.05E-19	2.60E-12
5	4180-4812	WS162	WS37	WS84	9.25E-37	7.57E-36	NS	2.67E-11	3.20E-11	2.67E-14	2.60E-12
6	753-5309	WS132	WS135	WS116	4.77E-26	1.30E-21	2.16E-26	5.39E-18	3.49E-10	9.90E-24	2.51E-24
7	3933-4663	Hoengseong	Wonju	GYBU-92	5.27E-16	7.1215E-08	NS	5.01E-05	1.80E-06	9.46E-08	5.23E-10
8	6832-9203	CN18	WS149	WS101	2.49E-27	NS	NS	2.53E-07	1.78E-15	1.28E-20	1.48E-31
9	1-2165	WS155	WS145	Jecheon	1.38E-25	2.87E-22	2.74E-26	7.38E-20	5.13E-20	2.85E-52	4.63E-93
10	2536-6248	WS117	WS156	WS109	2.44E-24	6.21E-20	1.73E-22	1.33E-21	4.29E-03	1.99E-38	8.04E-64
11	1-2180	WS156	WS145	WS105	1.16E-57	1.25E-55	4.44E-58	2.19E-19	1.79E-10	5.06E-21	1.66E-50
12	4667-9204	G6H, CC19-6, K206	WS101	G7H	6.32E-27	2.27E-28	NS	5.98E-21	3.68E-24	6.30E-36	2.60E-12
13	5652-6248	WS162	WS37	WS84	3.79E-36	9.70E-37	NS	1.15E-10	8.42E-11	4.31E-14	2.60E-12
14	1-2166	WS145	WS202	WS128	1.85E-16	5.58E-14	NS	1.44E-09	2.61E-11	3.22E-15	2.32E-23
15	6081-7263	WS116	WS151	WS132	1.17E-13	7.62E-12	1.17E-13	7.90E-08	3.48E-05	4.57E-10	1.85E-10
16	1-2193	WS162	WS209	WS84	1.54E-93	1.69E-82	1.55E-93	6.63E-34	6.48E-34	8.86E-41	2.60E-12
17	8215-9159	Bonghwa	WS32	G7H	7.98E-23	1.26E-19	3.43E-21	3.98E-09	1.96E-09	1.46E-12	2.60E-12
18	21-949	G7	G1	Unknown	1.78E-36	7.43E-39	1.01E-40	8.41E-09	3.50E-11	8.11E-21	7.80E-12
19	5833-7357	G5H	G7H	G6	1.98E-11	2.19E-21	1.09E-21	3.42E-17	5.38E-17	2.80E-19	NS
20	4374-4958	WS156	WS155	WS105	3.49E-13	1.06E-11	3.25E-13	3.09E-04	2.55E-04	1.66E-03	4.04E-08
21	789-1479	Anseong, Yeongyang	Unknown	G3	6.89E-22	1.29E-05	1.91E-21	1.05E-07	3.15E-09	3.66E-08	5.20E-12
22	7188-8985	Cheorwon	Hwasun	Danyang	2.80E-06	NS	NS	1.00E-08	4.27E-10	4.97E-07	5.46E-11
23	8450-8797	G5H	WS200	G6	8.17E-12	2.97E-10	7.46E-12	1.50E-04	2.61E-04	3.16E-04	4.09E-08
24	55-4500	G5H	G7H	GYBU-92	1.83E-19	5.21E-28	1.00E-26	6.09E-18	8.34E-14	2.80E-33	7.68E-49
25	4187-5459	CN18	Unknown	Hwasun	2.28E-10	1.04E-06	3.35E-07	7.12E-11	1.06E-10	2.41E-09	5.20E-12
26	1-2998	Andong, Hwasun, WS32, WS101, G6,K206 CN18		GYBU-92	1.22E-23	1.20E-16	2.63E-07	9.18E-18	2.10E-19	3.06E-39	2.23E-41

Figure 3. Recombination analysis using RDP4 (A) and RAT (B) program.



4. Genetic variation in the Korean SMV population

In total 27 non-recombinant Korea SMV isolates, haplotype diversity (h) and nucleotide diversity (π) values of all domains ranged from 0.903 to 1.000 and 0.01842 to 0.05664, respectively, showing different evolutionary features in each domain (Table 3). Especially, the nucleotide diversity (π) value of B was higher than A in all domains, suggesting the substantial genetic variation in B. In both groups, the P1 domain showed the highest genetic diversity, while the PIPO domain showed the lowest genetic diversity.

5. Neutrality test and selection pressure analysis

The demographic characteristics of each phylogroup were estimated from 3 neutrality parameters: Tajima's D, Fu & Li's D, and Fu's Fs statistics. All 11 SMV protein domains of each phylogroup (Total, A, and B) were used for analysis (Table 3). The values of 3 statistics were negative in most protein domains of A, while positive for B. These results suggest positive selection or perhaps a recent bottleneck on A and balancing selection or sudden population contraction on B. However, these results were not statistically significant in many domains, especially in B, suggesting the possibility of neutral selection in the population. The non-synonymous (dN) to synonymous (dS) substitution ratio (ω=dN/dS) was less than 1 in all domains of all phylogroups, suggesting that negative or purifying selections are affecting entire genomic regions of SMV.

Table 3. Population genetic parameters using 27 non-recombinant SMV isolates.

Domain	Population	Haplotype diversity	Nucleotide diversity	Tajima's D	Fu & Li's D*	Fu's Fs	dN/dS
Total	All(n=27)	0.997±0.00012	0.047	-0.3649(ns)	-0.5498(ns)	1.5	0.0558
	A(n=12)	1.000±0.00116	0.00844	-1.6324(ns)	-0.5005*	-0.195	0.1090
	B(n=10)	0.978±0.00292	0.02928	0.5987(ns)	-0.4251(ns)	4.538	0.0571
P1	All(n=27)	0.969±0.00064	0.04919	-1.0766(ns)	-1.2344(ns)	-0.028	0.1798
	A(n=12)	0.848±0.01080	0.01058	-1.1942(ns)	-0.6103(ns)	0.679	0.3533
	B(n=10)	0.978±0.00292	0.04850	0.7747(ns)	-0.502(ns)	1.289	0.1998
HC-Pro	All(n=27)	0.974±0.00043	0.04227	-0.9037(ns)	-0.4566(ns)	0.755	0.0496
	A(n=12)	0.909±0.00631	0.00863	-1.3827(ns)	-0.1508(ns)	0.095	0.0934
	B(n=10)	0.933±0.00597	0.02809	1.3046(ns)	-0.0676(ns)	2.629	0.0373
P3	All(n=27)	0.991±0.0002	0.05254	-0.4211(ns)	-0.4662(ns)	-2.174	0.1192
	A(n=12)	1.000±0.00116	0.00899	-1.6332(ns)	-0.4397*	-5.199	0.2137
	B(n=10)	0.933±0.00597	0.02553	0.365(ns)	-0.4228(ns)	1.794	0.0930
6K1	All(n=27)	0.903±0.00153	0.05293	-0.0204(ns)	-0.4753(ns)	-0.78	0.0367
	A(n=12)	0.667±0.01985	0.00952	-1.4065(ns)	-0.6004(ns)	-0.927	0.0955
	B(n=10)	0.733±0.01437	0.02664	2.0256*	-0.7352*	2.429	0.0000
CI	All(n=27)	0.991±0.0002	0.05664	0.2453 (ns)	-0.2444(ns)	-0.185	0.0359
	A(n=12)	1.000±0.00116	0.00915	-1.6298(ns)	-0.6704*	-3.112	0.0536
	B(n=10)	0.933±0.00597	0.02676	0.4506(ns)	-0.5731(ns)	3.299	0.0285
6K2	All(n=27)	0.903±0.00205	0.03753	0.2341(ns)	-0.5736(ns)	-2.917	0.0062
	A(n=12)	0.576±0.02669	0.00514	-1.8309*	-0.871*	-2.373	0.1681
	B(n=10)	0.889±0.00569	0.02812	0.7684(ns)	-0.6872(ns)	0.169	0.0000
VPg	All(n=27)	0.98±0.00029	0.05104	-0.0157(ns)	-0.4437(ns)	-1.63	0.0288
	A(n=12)	0.939±0.00333	0.00986	-1.3179(ns)	-0.7073(ns)	-1.871	0.0601
	B(n=10)	0.933±0.00597	0.03143	0.1414(ns)	-0.5861(ns)	0.93	0.0512
NIa-Pro	All(n=27)	0.974±0.00043	0.03994	0.0949(ns)	-0.2822(ns)	-1.59	0.0047
	A(n=12)	0.909±0.00631	0.00617	-1.805*	-0.4962**	-2.538	0.0218
	B(n=10)	0.933±0.00597	0.02152	0.1779(ns)	-0.3516(ns)	0.668	0.0072
NIb	All(n=27)	0.991±0.0002	0.04087	-0.3008(ns)	-0.4561(ns)	-1.68	0.0550
	A(n=12)	1.000±0.00116	0.00757	-1.8468*	-0.3995*	-4.357	0.1269
	B(n=10)	0.933±0.00597	0.02733	0.5372(ns)	-0.354(ns)	2.856	0.0552
CP	All(n=27)	0.983±0.00028	0.03863	-0.5312(ns)	-0.8724(ns)	-2.272	0.0237
	A(n=12)	0.955±0.00324	0.00844	-1.6348(ns)	-0.6645(ns)	-2.591	0.1071
	B(n=10)	0.933±0.00597	0.02999	0.3041(ns)	-0.6593(ns)	1.572	0.0190
PIPO	All(n=27)	0.920±0.00155	0.01842	-0.6784(ns)	-1.6831(ns)	-7.786	0.1931
	A(n=12)	0.682±0.02196	0.00365	-1.8309*	-1.5538*	-3.945	0.3757
	B(n=10)	0.844±0.01060	0.00595	-0.9344(ns)	-1.3185(ns)	-2.923	0.1931

6. Differentiation of Korean SMV populations

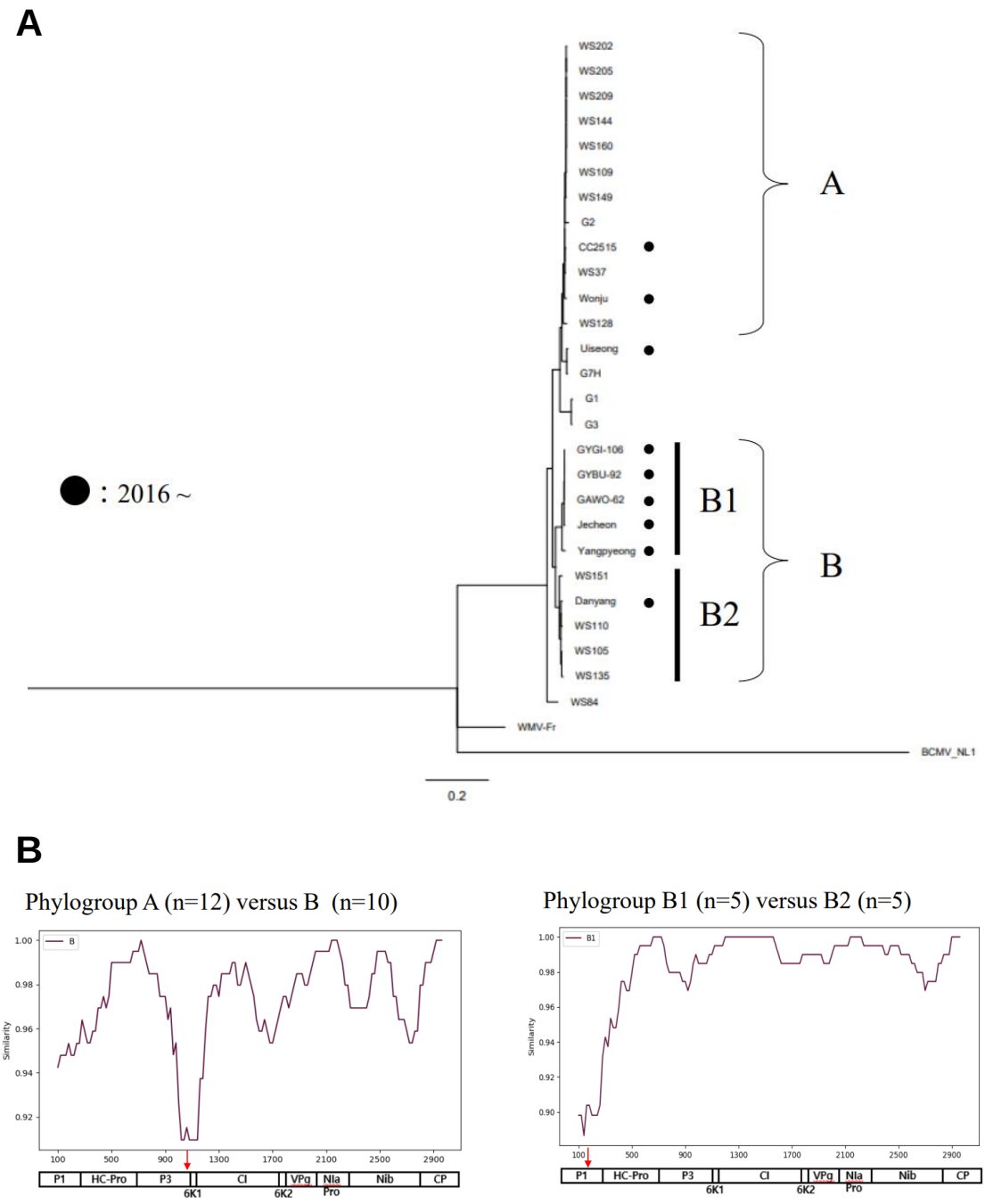
In the ML tree using 27 non-recombinant Korea SMV isolates, I found that most new SMV isolates (6 of 9) belong to group B, while only 2 isolates belong to A. I also found a new subgroup that emerged from B after 2016 and named it B1 and the rest of them B2 (Figure 4. A). Genetic differentiation between A and B, B1 and B2, and new and old isolates of A was estimated by 3 statistical parameters: Ks*, Z*, and Snn (Table 4). The null hypothesis (no genetic differentiation) between A and B, B1 and B2 was rejected by three statistics (*P* values < 0.05). At the same time, the new isolates from A showed no significant genetic variation (*P* value > 0.05). In addition, according to Fst (fixation index) and Nm (number of migrants per generation) values, gene flow was estimated to be infrequent between A and B, B1 and B2 (|Fst|> 0.33 or |Nm|< 1), while frequent between new and old isolates of A (|Fst|< 0.33 or |Nm|> 1). Based on these results, it is estimated that the presence of genetic differentiation between phylogroups (A and B, B1 and B2) in the Korea SMV population. Furthermore, genetic features of B seem to be preferred resulting in the newly emerged subgroup, B1.

The sequence similarity plot was constructed between each phylogroup (A and B, B1 and B2) using the amino acids sequence of the polyprotein (Figure 4B). The most distinct region Between A and B was P3, whereas P1 was the most distinct region between B1 and B2.

Table 4. Genetic differentiation statistics

Comparison	Ks* (P value)	Z* (P value)	Snn (P value)	Fst	Nm
A (n=12) versus B (N=10)	4.56717(0.0000 ***)	3.92290(0.0000 ***)	1.00000(0.0000 ***)	0.52937	0.21
A_old (n=10) versus A_new (n=2)	3.93565(0.0680 ns)	2.97977(0.0530 ns)	0.83333(0.4030 ns)	0.16285	2.57
B1 (n=5) versus B2 (n=5)	3.98529(0.0070 **)	2.11966(0.0100 *)	1.00000(0.0120 *)	0.77064	0.15

Figure 4. Maximum likelihood tree using 27 non-recombinant Korean SMV isolates (A) and amino acid sequence similarity plot between each phylogroup (B).



DISCUSSION

Various kinds of viruses are prevalent in the soybean field in Korea, while the most dominant viruses are SMV, SYMMV, and SYCMV (Jo et al., 2020; Lee et al., 2012; Li et al., 2015; Lim, 2013). SYMMV and SYCMV were first reported in Korea in 2009 and 2011, respectively, and their incidence increased rapidly, showing mixed infection with SMV (Lee et al., 2012; Nam et al., 2012; Nam et al., 2009; Seo et al., 2014). In this study, I detected five viruses (SMV: 12, SYMMV: 17, SYCMV: 7, BBWV2: 2, and BPMV: 5) and isolated different types of viruses from the same regional samples, suggesting the possibility of mixed infection (Supplementary Table S2). Therefore, I speculated that the SMV population in Korea might be affected by interaction with other viruses, not just soybean cultivars. In addition, I observed recent geographical traits that SMV is rare in the southwest part of the Korean Peninsula while prevailing in the northeast part. These results, which correspond to the previous study (Jo et al., 2020), suggest that the competition with other viruses impacted on SMV population in Korea.

The origin of SMV, like its host plant *Glycine max*, is presumed to be from East Asia (Gibbs et al. 2008). In the previous study, Cho and Goodman (1979) found that SMV strains (G1 to G7) isolated in the US were mainly detected from soybean accessions introduced from Korea (Cho and Goodman, 1979). In addition, Zhou et al. (2015) found that the genetic diversity of SMV was higher in China compared with other regions suggesting the origin of SMV is China (Zhou et al., 2015). In this study, phylogenetic analysis of SMV using an ML tree and haplotype network reveals four major clades (A, B, C, and D) in the SMV population. Each clade showed geographical features, and 3 of them (B, C, and D) were composed of

isolates from East Asia with few exceptions. Most of the isolates in groups B, C, and D were detected in Korea (22 of 29), China (16 of 19), and China (20 of 22), respectively. In contrast, group A consists of isolates from various countries (Korea: 14, China: 4, Japan: 1, Middle East: 3, North America: 7, South America: 2, Europe: 4, India: 1) (Figure 2B). The distribution pattern of the SMV population supports the previous hypothesis that the origin of SMV is East Asia. Under this hypothesis, group A seems to be constructed by a founder effect with a geographic bottleneck. Besides, in the neutrality test of the Korea SMV population, group A showed negative values in general, suggesting the presence of a bottleneck in the clade construction (Table 3).

Recombination is one of the crucial sources of genetic variation and is frequent in RNA viruses (García-Arenal et al., 2003). In the SMV population, recombination events are commonplace, and many researchers found that recombination serves important functions in the evolution process, especially in resistance breaking (Seo et al., 2009; Zhou et al., 2015). In addition, inter-specific recombinants with Bean common mosaic virus (BCMV) had been detected and prevailed in China, forming an independent clade (D in this study) (Yang et al., 2011; Yang et al., 2014). In the recombination analysis using Korean SMV isolates, I identified that seven of twelve new isolates are recombinant, suggesting that recombination constantly affects the evolution of the SMV population in Korea. Andong and Hwasun isolates seem to have originated from previously existing isolates sharing the identical recombination signal, while Hoengseong, Bongwha, and Cheorwon isolates emerged by the new recombination event. I also found that two novel SMV isolates (Anseong and Yeongyang) showed a recombination signal with BCMV in the P1 coding region. This recombination event was followed by an additional recombinant region with an unknown minor parent (Figure 3A). Therefore, I speculated that BCMV recombinants were not newly emerged in

Korea with the same breakpoint but were from China by seed or insect vectors. This result can be direct evidence for SMV transmission across the border, suggesting that the fittest one can easily spread globally in the SMV population.

In the population genetic analysis, I found that groups A and B showed statistically significant genetic differences in Korea. To find out the motivation of differentiation, I investigated several known motifs about aphid vector transmission and avirulent genes. As a result, I found that groups A and B were expected to show opposite pathogenetic responses in Rsv1 and Rsv3 hosts. Isolates of group A shared 'R1754K', known as Rsv3 resistance breaking mutation, while group B shared 'R1754'. Furthermore, isolates of group B shared 'A947T', which is known as Rsv1 resistance-breaking mutation, while group A shared 'A947' (Hajimorad et al., 2018). From these results, I could estimate that the Avirulent genes played a crucial role in clade construction in Korea, and additional selective sweep and recombination contributed to the ingroup sequence variations. Despite genetic differences among the Korean SMV population, negative selection dominates the whole genome of SMV among all 27 isolates (Table 3). The genome of RNA viruses is believed to show hyper-variation by its error-prone RNA-dependent RNA polymerase (Domingo et al., 2021). However, the virus genome is minimal, while it should satisfy many qualifications for a successful operation (Gopal et al., 2014). Therefore, it is not the surprising result that the genome of the SMV is affected by negative selection.

In Korea, SMV isolates detected after 2016 showed a tendency toward group B (Figure 4. A), suggesting Rsv1 resistance-breaking strains were recently preferred. In addition, an emerging subgroup from B, B1, showed statistically significant genetic differences with B2, which are existing isolates in B. The amino acid sequence similarity plot determined the main difference between B1 and B2 as the P1 region (Figure 4B). The exact

role of potyvirus P1 is not clearly confirmed but speculated to be associated with the host range and pathogenicity of SMV (Mao et al., 2022; Shi et al., 2007). I, therefore, estimated that the emergence of the new subgroup B1 signifies the existence of selection pressures related to the host range of SMV to counter various environmental factors such as interspecies competition.

Overall, I analyzed the phylogenetic structure and estimated the movement of the population over time with 12 new Korean SMV isolates. Mixed infection with other viruses seems to influence the SMV population. In addition, more than half of the new isolates showed recombination signals suggesting that recombination plays an important role in the evolution of SMV. In the Korean SMV population, the previously dominant SMV-G7H phylogeny is detected with reduced prevalence. At the same time, the new isolates mainly emerged in group B, especially near the G7 phylogeny suggesting *Rsv1*-resistant isolates were preferred recently in Korea.

LITERATURE CITED

Burland, T. G. 2000. DNASTAR's Lasergene sequence analysis software. Methods Mol. Biol. 132:71-91.

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., and Chung, K. 1986. Strains of soybean mosaic virus causing soybean necrotic disease in Korea. Korean J. Breed. 18:150-153.

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

Domingo, E., García-Crespo, C., Lobo-Vega, R., and Perales, C. 2021. Mutation rates, mutation frequencies, and proofreading-repair activities in RNA virus genetics. Viruses 13:1882.

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

Fu, Y.-X. 1997. Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. Genetics 147:915-925.

Fu, Y.-X., and Li, W.-H. 1993. Statistical tests of neutrality of mutations. Genetics 133:693-709.

Fuentes, S., Gibbs, A. J., Adams, I. P., Hajizadeh, M., Kreuze, J., Fox, A., Blouin, A. G., and Jones, R. A. 2022. Phylogenetics and Evolution of Potato Virus V: Another Potyvirus that Originated in the Andes. Plant Dis. 106:691-700.

García-Arenal, F., Fraile, A., and Malpica, J. M. 2003. Variation and evolution of plant virus populations. Int. Microbiol. 6:225-232.

Gardner, N., and Kendrick, J. B. 1921. Soybean mosaic. J. Agric. Res. 22:111-114.

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

Gopal, A., Egecioglu, D. E., Yoffe, A. M., Ben-Shaul, A., Rao, A. L., Knobler, C. M., and Gelbart, W. M. 2014. Viral RNAs are unusually compact. PLoS One 9:e105875.

Haas, B.J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P.D., Bowden, J., Couger, M.B., Eccles, D., Li, B., Lieber, M. 2013. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nat. Protoc. 8:1494-1512.

Hajimorad, M., Domier, L., Tolin, S., Whitham, S., and Saghai Maroof, M. 2018. Soybean mosaic virus: a successful potyvirus with a wide distribution but restricted natural host range. Mol. Plant Pathol. 19:1563-1579.

Hudson, R. R. 2000. A new statistic for detecting genetic differentiation. Genetics 155:2011-2014.

Hudson, R. R., Slatkin, M., and Maddison, W. P. 1992. Estimation of levels of gene flow from DNA sequence data. Genetics 132:583-589.

Huson, D. H., and Bryant, D. 2006. Application of phylogenetic networks in evolutionary studies. Mol. Biol. Evol. 23:254-267.

Jo, Y., Yoon, Y. N., Jang, Y.-W., Choi, H., Lee, Y.-H., Kim, S.-M., Choi, S. Y., Lee, B. C., and Cho, W. K. 2020. Soybean Viromes in the Republic of Korea Revealed by RT-PCR and Next-Generation Sequencing. Microorganisms 8:1777.

Kim, J. S., Lee, E. J., and Cho, E. K. 1991. A new virulent strain of soybean mosaic virus

infecting SMV resistant soybean cultivar Deogyou. Plant Pathol J. 7:37-41.

Kim, Y.-H., Kim, O.-S., Lee, B.-C., Moon, J.-K., Lee, S.-C., and Lee, J.-Y. 2003. G7H, a new Soybean mosaic virus strain: its virulence and nucleotide sequence of CI gene. Plant Dis. 87:1372-1375.

Kim, Y. 2000. Molecular classification of soybean mosaic virus strains and their interaction with soybeans. Ph. D. Thesis, Kyungpook Natl. Univ., Daegu, Korea.

Kryazhimskiy, S., and Plotkin, J. B. 2008. The population genetics of dN/dS. PLoS Genet. 4:e1000304.

Lee, Y., Lim, S., Yoon, Y., Jeon, M., Yun, H., Ko, J., Lee, S., Lee, K., and Baek, I. 2012. Incidence of soybean viral diseases in Korea. Korea Soybean Dig. 29:7–15

Leigh, J. W., and Bryant, D. 2015. POPART: full-feature software for haplotype network construction. Methods Ecol. Evol. 6:1110-1116.

Li, K., Yang, Q., Zhi, H., and Gai, J. 2010. Identification and distribution of soybean mosaic virus strains in southern China. Plant Dis. 94:351-357.

Li, M., Seo, E.-Y., Cho, S., Kim, J., Chung, J., Lim, H.-J., Gotoh, T., Hammond, J., and Lim, H. S. 2015. A 2014 Nationwide Survey of the Distribution of Soybean Mosaic Virus (SMV), Soybean Yellow Mottle Mosaic Virus (SYMMV) and Soybean Yellow Common Mosaic Virus (SYCMV) Major Viruses in South Korean Soybean Fields, and Changes from 2012 Isolate Prevalence. J. Fac. Agric. Kyushu Univ. 60:339–347.

Lim, S. 2013. Incidence of virus diseases on soybean breeding fields and first report of Bean common mosaic virus on soybean in Korea. Gyeongbuk National University:630.27-13-182.

Lu, J.-J., He, E.-Q., Bao, W.-Q., Chen, J.-S., Sun, S.-R., and Gao, S.-J. 2021. Comparative genomics reveals insights into genetic variability and molecular evolution among

sugarcane yellow leaf virus populations. Sci. Rep. 11:1-11.

Mao, C., Shan, S., Huang, Y., Jiang, C., Zhang, H., Li, Y., Chen, J., Wei, Z., and Sun, Z. 2022. The hypervariable N-terminal of soybean mosaic virus P1 protein influences its pathogenicity and host defense responses. Phytopathol. Res. 4:1-12.

Martin, D. P., Murrell, B., Golden, M., Khoosal, A., and Muhire, B. 2015. RDP4: Detection and analysis of recombination patterns in virus genomes. Virus Evol. 1:1–5.

Nam, M., Kim, S. M., Domier, L. L., Koh, S., Moon, J. K., Choi, H. S., Kim, H. G., Moon, J. S., and Lee, S.-H. 2009. Nucleotide sequence and genomic organization of a newly identified member of the genus Carmovirus, soybean yellow mottle mosaic virus, from soybean. Arch. Virol. 154:1679-1684.

Nam, M., Kim, J.-S., Park, S.-J., Park, C. Y., Lee, J.-S., Choi, H.-S., Kim, J.-S., Kim, H. G., Lim, S., and Moon, J. S. 2012. Biological and molecular characterization of Soybean yellow common mosaic virus, a new species in the genus Sobemovirus. Virus Res. 163:363-367.

Nei, M. 1987. Molecular evolutionary genetics. New york:Columbia university press.

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

Rozas, J., Ferrer-Mata, A., Sánchez-DelBarrio, J. C., Guirao-Rico, S., Librado, P., Ramos-Onsins, S. E., and Sánchez-Gracia, A. 2017. DnaSP 6: DNA sequence polymorphism analysis of large data sets. Mol. Biol. Evol. 34:3299-3302.

Samson, S., Lord, É., and Makarenkov, V. 2022. SimPlot++: a Python application for representing sequence similarity and detecting recombination. Bioinformatics 38:3118-3120.

Seo, J.-K., Ohshima, K., Lee, H.-G., Son, M., Choi, H.-S., Lee, S.-H., Sohn, S.-H., and Kim,

K.-H. 2009. Molecular variability and genetic structure of the population of Soybean mosaic virus based on the analysis of complete genome sequences. Virology 393:91-103.

Seo, J.-K., Park, J.-Y., Kwak, H.-R., Kim, M.-K., Nam, M., Lee, S.-H., Kim, J.-S., and Choi, H.-S. 2014. Severe outbreak of the emerging soybean viruses in Korea: an increasing threat. Plant Health Prog. 15:122-123.

Shi, Y., Chen, J., Hong, X., Chen, J., and Adams, M. J. 2007. A potyvirus P1 protein interacts with the Rieske Fe/S protein of its host. Mol. Plant Pathol. 8:785-790.

Sun, S. R., Chen, J. S., He, E. Q., Huang, M. T., Fu, H. Y., Lu, J. J., & Gao, S. J. 2021. Genetic variability and molecular evolution of maize yellow mosaic virus populations from different geographic origins. Plant Dis. 105.4:896-903.

Tajima, F. 1989. The effect of change in population size on DNA polymorphism. Genetics 123:597-601.

Takahashi, K., Tanaka, T., Iida, W., and Tsuda, Y. 1980. Studies on virus diseases and causal viruses of soybean in Japan. pan. Bull Tohoku Natl. Agric. Exp. Stn. 62:1-130.

Tamura, K., Stecher, G., and Kumar, S. 2021. MEGA11: molecular evolutionary genetics analysis version 11. Mol. Biol. Evol. 38:3022-3027.

Usovsky, M., Chen, P., Li, D., Wang, A., Shi, A., Zheng, C., Shakiba, E., Lee, D., Canella Vieira, C., and Lee, Y. C. 2022. Decades of Genetic Research on Soybean mosaic virus Resistance in Soybean. Viruses 14:1122.

Widyasari, K., Alazem, M., and Kim, K.-H. 2020. Soybean resistance to soybean mosaic virus. Plants 9:219.

Yang, Y., Lin, J., Zheng, G., Zhang, M., and Zhi, H. 2014. Recombinant soybean mosaic virus is prevalent in Chinese soybean fields. Arch. Virol. 159:1793-1796.

Yang, Y., Gong, J., Li, H., Li, C., Wang, D., Li, K., and Zhi, H. 2011. Identification of a novel Soybean mosaic virus isolate in China that contains a unique 5′ terminus sharing high sequence homology with Bean common mosaic virus. Virus Res. 157:13-18.

Zhou, G.-C., Shao, Z.-Q., Ma, F.-F., Wu, P., Wu, X.-Y., Xie, Z.-Y., Yu, D.-Y., Cheng, H., Liu, Z.-H., and Jiang, Z.-F. 2015. The evolution of soybean mosaic virus: an updated analysis by obtaining 18 new genomic sequences of Chinese strains/isolates. Virus Res. 208:189-198.

Supplementary Table S1. 132 SMV isolates (present study=12, NCBI database=120) and 2 outgroup viruses (BCMV and WMV) used for phylogenetic analysis.

Isolate	Collection region	Collection date	GenBank accession no.
G5	South Korea	2003*	AY294044
G5H	South Korea	2005	FJ807701
G6H	South Korea	2003	FJ640981
G7H	South Korea	2003*	AY294045
CN18	South Korea	2003*	AJ619757
WS32	South Korea	2006	FJ640954
WS37	South Korea	2006	FJ640955
WS84	South Korea	2006	FJ640956
WS101	South Korea	2006	FJ640957
WS105	South Korea	2006	FJ640958
WS109	South Korea	2006	FJ640959
WS110	South Korea	2006	FJ640960
WS116	South Korea	2006	FJ640961
WS117	South Korea	2006	FJ640962
WS128	South Korea	2006	FJ640963
WS132	South Korea	2006	FJ640964
WS135	South Korea	2006	FJ640965
WS144	South Korea	2006	FJ640966
WS145	South Korea	2006	FJ640967
WS149	South Korea	2006	FJ640968
WS151	South Korea	2006	FJ640969
WS155	South Korea	2006	FJ640970
WS156	South Korea	2006	FJ640971
WS160	South Korea	2006	FJ640972
WS162	South Korea	2006	FJ640973
WS200	South Korea	2006	FJ548849
WS202	South Korea	2006	FJ640974
WS205	South Korea	2006	FJ640975
WS209	South Korea	2006	FJ640976
GAWO-62	South Korea	2016	MT603829
GYBU-92	South Korea	2016	MT603831
GYGI-106	South Korea	2016	MT603833
CC2515	South Korea	2016	KY986929
K206	South Korea	2019	LC591946
CC19-6	South Korea	2019	LC655955
Andong	South Korea	2020	This study
Anseong	South Korea	2020	This study
Cheorwon	South Korea	2020	This study
Hwasun	South Korea	2020	This study
Uiseong	South Korea	2020	This study
Bonghwa	South Korea	2021	This study
Danyang	South Korea	2021	This study

Hoengseong	South Korea	2021	This study
Jecheon	South Korea	2021	This study
Wonju	South Korea	2021	This study
Yangpyeong	South Korea	2021	This study
Yeongyang	South Korea	2021	This study
G1	USA	2003	FJ640977
G2	USA	1999*	S42280
G3	USA	2003	FJ640978
G4	USA	2003	FJ640979
G6	USA	2003	FJ640980
G7	USA	2000*	AF241739
G7A	USA	2003	FJ640982
G7d	USA	2006*	AY216987
N	USA	2000*	D00507.2
413	USA	2002	GU015011
TNP	USA	2013*	HQ845735
KY	USA	2013*	HQ845736
Aa	Japan	2003*	AB100442
Aa15-M2	Japan	2003*	AB100443
SMV-C	Japan	2017	LC323107
L	Canada	2005	EU871724
NP-L	Canada	2009	HQ166266
NP-C-L	Canada	2009	HQ166265
Rsv4-RB3	Canada	2010	JN416770
SMV/BSB1	Brazil	2003	MN124783
Gulupa	Colombia	2015	KY249378
Salzlandkreis-2 17	Germany	2017	MN399738
DSMZ PV-1235	Germany	2022	ON013906
Summer Shell	Netherlands	2020	MW822167
Green Shell	Netherlands	2020	MW822168
Ar13	Iran	2011	KF135488
Ar33	Iran	2011	KF297335
Go11	Iran	2011	KF135491
Lo3	Iran	2011	KF135490
India	India	2014	KM979229
SAAS	China	2001*	AJ310200
3144(SC4)	China	2002	MN539670
4278-1	China	2004	KT285170
4469-4/CHN/2004	China	2004	HM590055
4547/CHN/2004	China	2004	HQ396725
Sc6	China	2004	HM590054
6067-1	China	2005	JF833015
6202-2	China	2005	JF833014
severe	China	2005*	AJ312439
SC3	China	2005	JF833013
SC6-N	China	2010	KP710867
SC7-N	China	2010	KP710868
SX	China	2012	KC845321

HB-RS	China	2012	KR065437
BYX006	China	2013	KP710861
HGT005	China	2013	KP710862
HGT008	China	2013	KP710863
HGT009	China	2013	KP710864
LJZ002	China	2013	KP710865
LJZ010	China	2013	KP710866
SX-Z	China	2013	KP710870
XFQ001	China	2013	KP710871
XFQ005	China	2013	KP710872
XFQ008	China	2013	KP710873
XFQ010	China	2013	KP710874
XFQ012	China	2013	KP710875
XFQ014	China	2013	KP710876
XFQ018	China	2013	KP710877
XFQ020	China	2013	KP710878
NE-N1	China	2013	KP710869
China	China	2015	KX096578
FJTN001	China	2016	KX834319
GXQZ001	China	2016	KX834322
HLJSB001	China	2016	KX834323
HLJHLQF001	China	2016	KX834325
HLJBADS001	China	2016	KX834324
JSJJ001	China	2016	KX834321
SC001	China	2016	KX834320
Liaoning	China	2017	MK350280
pCB301-SC7	China	2017	MH919385
pCB301-SC15	China	2017	MH919386
N1	China	2018	MN623289
N3	China	2018	MN623290
1129	China	2019	OK105105
SMV AH SZ	China	2019	MW354946
SMV JS NJ	China	2019	MW354948
SMV JL CC	China	2019	MW354951
SMV JL GZL	China	2019	MW354947
DSMZ PV-0938	China	2022*	OK058515
Am***	China	2012	KC845322
SMV-SXBX***	China	2016	MT712111
P***	China	2004*	AJ507388
HZ1***	China	2005*	AJ628750
NN***	China	2011	KF982784
Uraria***	Taiwan	2014	LC037232
WMV-Fr**	China	2004*	AY437609
BCMV NL1**	USA	2011	KM023744

*: Genebank registered date, **: Outgroup viruses, ***: Suspected non-SMV isolates

Supplementary Table S2. Twelve new SMV isolates and other soybean infecting viruses (SYMMV, SYCMV, BBWV2, and BPMV) detected from the 27 local samples of Korea.

Location	Detected Virus in (2020)	Location	Detected Virus (2021)
Andong	SMV	Bonghwa	SMV, SYMMV
Anseong	SMV	Danyang	SMV, SYMMV
Boryeong	SYMMV	Hoengseong	SMV, SYMMV, BPMV
Buyeo	SYCMV	Icheon	SYMMV
Cheolwon	SMV, SYMMV	Jecheon	SMV, SYMMV, SYCMV, BPMV
Goheung	SYMMV, BPMV	Wonju	SMV, SYMMV, BPMV
Gurye	BBWV2	Yangpyeong	SMV, SYMMV, SYCMV
Gunsan	SYCMV	Yeoju	SYMMV
Haenam	SYMMV	Yeongwol	SYMMV
Hampyeong	BBWV2	Yeongyang	SMV, SYMMV, BPMV
Hwasun	SMV, SYMMV		
Jindo	SYCMV		
Jangheung	SYMMV		
Muan	SYCMV		
Suncheon	SYMMV		
Uiseong	SMV		
Yeongam	SYCMV		

국내 콩 모자이크 바이러스 개체군의 진화 양상 분석

오지선

초록

콩 모자이크 바이러스 (SMV)는 전 세계적으로 발견되며, 콩 재배에 있어 만성적이고, 심각한 피해를 미치는 병원균으로 알려져 있다. 이번 연구에서는 한국의 41개 지역의 콩 재배 포장으로부터 12개의 새로운 SMV 분리주를 확보하였고, 기존에 보고된 분리주들과 비교함으로써 개체군 진화 양상을 분석하였다. 계통분석을 통해 전체 SMV 개체군 내에서 4 개의 주요 분기군을 발견하여 A에서 D 로 명명하였으며, 각 분기군이 지리적인 특징을 보이는 것을 확인하였다. 한국에서의 SMV 진화양상 파악을 위해 한국 분리주를 사용하여 개체군 분석이 진행되었다. 재조합 분석을 통해 12개의 새로운 분리주들 중 7개 분리주에서 재조합 신호를 발견하여, SMV 개체군에서 재조합이 매우 빈번하게 일어났음을 확인할 수 있었다. 특히, 3개의 분리주는 기존 서열에서는 발견되지 않았던 새로운 재조합 중단점을 가지고 있어 재조합이 최근까지도 지속적으로 SMV 진화에 영향을 미치고 있었음을 시사하였다. 개체군의 유전적 구조를 분석하기 위해 진행된 27개의 비 재조합 분리주를 사용한 분석에서 대부분의 분리주들은 주로 분기군 A와 B에 속했으며 각 분기군은 통계적으로 유의미한 유전적 차이를 보이는 것을 확인하였다. 유전적 다양성은 A보다 B에서 크게 나타났지만, 두 분기

군에서 모두 11개 단백질 도메인 암호영역에서 공통적으로 음성 선택이 일어나는 것이 확인되었다. 2016년 이후 보고된 분리주는 대부분 B에서 발생한 것을 확인하였고 새로 발생한 분기군을 B1 이라 명명하였다. 아미노산 서열 유사도 분석을 통해 기존 B 분기군인 B2와 새로운 발생한 분기군인 B1의 주된 차이는 P1 단백질 부분에서 발생하였음을 확인할 수 있었는데, 이 결과는 최근 SMV 개체군에서 기주범위나 병원성에 변화가 나타났음을 시사한다.

주요어: soybean mosaic virus, *Glycine max*, phylogenetic analysis, population genetics, recombination analysis

학번: 2021-24653