Analysis of Soybean mosaic virus genetic diversity in Iran allows the characterization of a new mutation resulting in overcoming Rsv4-resistance

Akbar Ahangaran,1 Mina Koohi Habibi,1 Gholam-Hossein Mosahedi Mohammadi,1 Stephan Winter2 and Fernando García-Arenal³

1Department of Plant Protection, University College of Agriculture and Natural Resources, University of Tehran, Karaj, Iran
2German Collection of Microorganisms and Cell Cultures, DSMZ, Braunschweig, Germany
3Centro de Biotecnología y Genómica de Plantas (UPM-INIA) and E.T.S.I. Agrónomos, Campus Montegancedo, Universidad Politécnica de Madrid, Pozuelo de Alarcón, Madrid, Spain

The genetic variation and population structure of Soybean mosaic virus (SMV) in Iran was analysed through the characterization of a set of isolates collected in the soybean-growing provinces of Iran. The partial nucleotide sequence of these isolates showed a single, undifferentiated population with low genetic diversity, highly differentiated from other SMV world populations. These traits are compatible with a population bottleneck associated with the recent introduction of SMV in Iran. Phylogenetic analyses suggest that SMV was introduced into Iran from East Asia, with at least three introduction events. The limited genetic diversification of SMV in Iran may be explained by strong negative selection in most viral genes eliminating the majority of mutations, together with recombination purging deleterious mutations. The pathogenicity of Iranian SMV isolates was typified on a set of soybean differential lines either susceptible or carrying different resistance genes or alleles to SMV. Two pathotypes were distinguished according to the ability to overcome Rsv4 resistance in line V94-5152. Amino acid sequence comparisons of virulent and avirulent isolates on V94-5152 (Rsv4), plus site-directed mutagenesis in a biologically active cDNA clone, identified mutation S1053N in the P3 protein as the determinant for virulence on V94-5152. Codon 1053 was shown to be under positive selection, and S1053N-determined Rsv4-virulence occurred in isolates with different genealogies. The V94-5152 (Rsv4)-virulence determinant in Iranian isolates maps into a different amino acid position in the P3 protein than those previously reported, indicating different evolutionary pathways towards resistance breaking that might be conditioned by sequence context.

INTRODUCTION

Soybean [Glycine max (L.) Merr.], an annual plant species belonging to the Fabaceae, is the most important legume crop worldwide (Hartman et al., 2011). Soybean mosaic virus (SMV), a member of genus Potyvirus, is a devastating soybean pathogen found in all soybean-growing regions of the world. Like other potyviruses, SMV has a monopartite, single-stranded, positive-sense RNA genome of about 9.6 kb nucleotides, with a 5’ viral protein covalently linked (VPg) and a 3’ poly-A tail. The SMV genome encodes a large polyprotein of about 350 kDa, which is cleaved to yield at least nine mature proteins (Adams et al., 2005; Jayaram et al., 1992; Urcuqui-Inchima et al., 2001). An additional small ORF overlapping with the P3 cistron of potyviruses, termed pipo, has also been reported in SMV (Chung et al., 2008; Wen & Hajimorad, 2010). An additional small ORF overlapping with the P3 cistron of potyviruses, termed pipo, has also been reported in SMV (Chung et al., 2008; Wen & Hajimorad, 2010). SMV has a narrow natural host range, and is transmitted by aphids in a non-persistent manner, and through the seed (Domier et al., 2007).

Great effort has been made to breed soybean varieties resistant to SMV. Alleles at three independent loci, Rsv1,
Rsv3 and Rsv4, have been identified in soybean germplasm conferring dominant resistance to SMV (Gunduz et al., 2002; Hayes et al., 2000; Zheng et al., 2005). Rsv1 and Rsv3 encode coiled-coil, nucleotide-binding, leucine-rich repeat (CC-NB-LRR) resistance proteins (Hayes et al., 2004; Suh et al., 2011), and recognition of SMV results in extreme resistance (ER) (Hajimorad & Hill, 2001; Zhang et al., 2009). Rsv4 appears to belong to a new class of resistance genes that interfere with viral infection and cell-to-cell movement, and delay vascular movement (Gunduz et al., 2004; Saghai Maroof et al., 2010). It has been shown that the Rsv4 allele in PI88788 (Rsv4) expresses late susceptibility to the standard SMV pathotype G1 (Gunduz et al., 2004), whereas V94-5152 (Rsv4) displays broad resistance to all the standard SMV pathotypes (Gunduz et al., 2002). In contrast with PI88788, resistance to systemic movement of SMV in V94-5152 is robust and this could be a result of the genetic differences between the Rsv4 alleles in these two soybean genotypes. However, it is equally possible that the difference in the strength of resistance is a reflection of difference(s) in the genetic background of PI88788 and V94-5152 (Khatabi et al., 2012).

Resistance conferred by all these genes and alleles might not be durable due to variation among SMV isolates. Indeed, pathogenic variability among SMV isolates has been widely observed, and classification of SMV pathotypes based on their ability to overcome the resistance genes varies depending on the set of differential soybean genotypes used and the geographical region studied. Thus, Cho & Goodman (1979) classified 98 isolates of SMV from the USA soybean germplasm collection into seven strain groups (G1–G7). In Japan, 102 SMV isolates were classified into five strains, A, B, C, D and E, based on differential symptoms induced in four soybean cultivars (Takahashi et al., 1980). In China, up to 21 pathotypes have been described (Li et al., 2010). Consequently, the pathotypic relationships among groups of SMV strains from different regions of the world are not clear (Gagarinova et al., 2008b).

The molecular determinants in SMV for Rsv-resistant breaking have been characterized to various extents. Different alleles have been reported at the Rsv1 locus, with different specificities, as shown by being overcome by different SMV strains (Chen et al., 2001; Zheng et al., 2005), and the virulence determinants on Rsv1 have been mapped to the HC-Pro and P3 proteins (Eggenberger et al., 2008; Hajimorad et al., 2008, 2011). Determinants for virulence on Rsv3 have been mapped to the cylindrical infiltration (CI) protein (Seo et al., 2009a; Zhang et al., 2009). Virulence determinants on Rsv4 were defined based on interaction with V94-5152 and they have been mapped to the P3 protein (Chowda-Reddy et al., 2011a; Khatabi et al., 2012). In addition, a recent study has shown the essential role of CI for breaking down Rsv1 resistance, the involvement of P3 for virulence on Rsv3 and the requirement of other viral proteins or domain(s) than P3, including CI, for virulence of SMV on Rsv4-conferrred resistance (Chowda-Reddy et al., 2011b).

In 1938, for the first time, food grade soybean seeds were imported to Gilan province in northwest Iran, but their cultivation was unsuccessful. Later, in 1962, a second, successful attempt to introduce the soybean crop in Iran was made in Golestan province, using seed from Japan (Amirshahi, 1976; Shurtleff & Aoyagi, 2008). Presently, soybean cultivation in Iran amounts to about 76 000 ha (Anonymous, 2011) and the area is increasing, with the crop concentrated into four provinces, Golestan, Mazandaran, Ardabil and Lorestan (Fig. 1), with very different crop areas (57 732, 12 262, 5937 and 135 ha, respectively). SMV was first reported in the Caspian Sea area (Mazandaran and Golestan provinces) in Iran in 1978 (Eskandari, 1978), and later it was found in all soybean production regions (Fazlizzad et al., 2002; Karimi & Noaparast, 1989). There is no information about the genetic variation of SMV in Iran, the possible origin of the Iranian population(s) or the occurring pathotypes, which is a hindrance for planning a long-term sustainable strategy for its control.

The goal of the present study was to characterize the genetic variation and population structure of SMV in Iran. Results showed that SMV isolates from all soybean growing areas in Iran had reduced genetic variation. Two pathotypes were identified that differed in their ability to overcome Rsv4 resistance in line V94-5152. Molecular analyses of these pathotypes allowed identification of new determinants for Rsv4-resistance breaking.

**RESULTS**

**Genetic diversity of the SMV population from Iran**

During the summer of 2011, leaf samples were collected from a total of 269 soybean plants showing possible symptoms of virus infection (see Methods) from the four main soybean-growing provinces of Iran, representing 51 different fields (Fig. 1, Table S1 available in JGV Online). Fifty-eight samples were SMV positive both by double antibody sandwich-ELISA (DAS-ELISA) and by reverse transcriptase (RT)-PCR amplification of the CP gene using primers CPf and CPR (see Table S5 and Methods). The fraction of symptomatic plants infected by SMV ranged between 10 and 36.9%, and was significantly higher in Mazandaran ($\chi^2 > 4.82; P < 0.04$) (Table S1). Out of 58 SMV-infected plants, 25 were selected to represent the different geographical areas and field symptoms for SMV isolation for further studies (Table S1).

These 25 isolates were molecularly characterized on the basis of the nucleotide sequence of the genomic regions encoding proteins P1, P3, 6K1, CI and CP. These genes were chosen as representing different genomic regions regardless of their biological function, for an unbiased estimate of genetic diversity over the SMV genome. The nucleotide sequence of the complete genome was determined for four isolates: Lo3 and Ar33, which were unable...
to overcome Rsv4 resistance (see below); Ar13, which differed much in the partial sequences from other isolates; and Go11, which caused severe symptoms. The genome of isolates Lo3, Go11 and Ar33 consisted of 9584 nt, being 1 nt smaller than most reported SMV isolates (Gagarinova et al., 2008a; Seo et al., 2009b) due to a single-nucleotide deletion in the 3' UTR, while the genome of isolate Ar13 had 9587 nt, due to a triplet insertion at the P1 gene. The nucleotide sequences determined in this study are available at GenBank with the accession numbers listed in Table S2.

Recombination was analysed among Iranian SMV isolates plus a set of 52 SMV isolates from other parts of the world, for which the whole-length genome sequence was reported in databases. Either full-length genome sequences or the concatenated sequences encoding proteins P1, P3, 6K1, CI and CP were analysed using six methods as implemented in RDP4 (see Methods). The inclusion of the Iranian isolates in this analysis allowed detection of new recombination sites in the SMV genome, in addition to those reported previously by others (Gagarinova et al., 2008b, Seo et al., 2009b). A summary of all recombination sites detected in the SMV genome in this and in previous analyses appears in Fig. 2. Note that the two recombination sites identified in P3 genes were located in the overlapping gene encoding P3N+PIPO (Fig. 2). As for P3, recombination affected the sequence of pipo, but did not result in the introduction of a premature stop codon.

Phylogenetic analyses were carried out using only genomes or parts of genomes in which no recombination was identified. Thus, the concatenated regions encoding P1, P3, 6K1 and CI proteins of the 25 Iranian isolates, plus the corresponding regions of the 24 non-recombinant SMV isolates from the databases were used. Iranian isolates clustered into three clades: one included most Iranian isolates from all geographical regions; the second included isolates Ar33 and Lo3; and the third clade included isolate Ar13, which differed broadly from most Iranian and world SMV isolates (Fig. 3). A similar clustering of Iranian isolates was obtained when the non-recombinant portions of the full-length sequenced Iranian and world isolates were used for the analyses (not shown).
In agreement with phylogenetic analyses, the genetic diversity of the Iranian SMV population was quite low (Table 1), being $0.0088 \pm 0.0004$ for the concatenated genomic fragments. Diversity was highest for the P1 gene and lowest for the CP gene (Table 1), suggesting differential selection pressures along the genome.

**Fig. 2.** Recombination map of SMV. Recombination sites reported in previous studies (Gagarinova et al. 2008b; Seo et al. 2009a) are shown below the genome, and new recombination sites identified in this work are shown above the genome. Nucleotide positions are numbered according to the sequence of isolate G2 (accession no. S42280).

**Fig. 3.** ML phylogenetic tree for the concatenated regions encoding proteins P1, P3, 6K1 and CI of 25 Iranian SMV isolates (in italics) and 24 isolates from other regions of the world. Significance of nodes in bootstrap analysis (100 replicates) is indicated. Bar, the number of substitutions per base.
Selection pressures on the genome of SMV

To analyse selection pressures on SMV proteins, the $d_N/d_S$ ratio was estimated separately for the P1, P3, 6K1, CI and CP genes using the set of Iranian isolates. Results (Table 1) show that the proteins encoded by these genes are under negative selection. However, $d_N/d_S$ ratios differed largely among genes, the genes encoding proteins CI and CP were the most constrained for diversification, and the genes encoding proteins P3 and, particularly, P1 the least constrained. Similar results were obtained from the analysis of all SMV genes using the set of 56 fully sequenced SMV isolates from the world (not shown). Thus, for a more precise analysis of selection pressures at the codon level we used the four fully sequenced Iranian isolates plus the 52 fully sequenced world isolates in Table S2, as a larger set of isolates will increase the resolution of the analysis. Results (Table 2) show that most codons in SMV genome were under neutral evolution or negative selection, and only four or three codons were rated as under positive selection, depending on the analysis. Two codons were rated as under positive selection by both methods, one in the P1 protein gene, encoding amino acid 140, and the other in the P3 protein gene, encoding amino acid 288. In addition, codon 52 in the P1 protein gene, 96 in the HC-Pro gene and 460 in the Nl b gene were found to be under positive selection by just one method (Table 2).

Spatial structure of the Iranian and world population of SMV

Although SMV isolates from Iran or worldwide show low genetic diversity (Tables 1 and 2; Seo et al., 2009b), we analysed whether genetic diversity was structured at two spatial scales, among provinces within Iran and among

---

Table 1. Nucleotide diversity for different genomic regions of the Iranian population of SMV

$d_N$ and $d_S$ indicate nucleotide diversity at non-synonymous and at synonymous positions, respectively, and were estimated by the Pamilo–Bianchi–Li method. Values given in the table are means ± SEM.

<table>
<thead>
<tr>
<th>Genomic region</th>
<th>Nucleotide diversity</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$d^*$</td>
</tr>
<tr>
<td>P1</td>
<td>0.0164 ± 0.0014</td>
</tr>
<tr>
<td>P3</td>
<td>0.0083 ± 0.0010</td>
</tr>
<tr>
<td>6K1</td>
<td>0.0072 ± 0.0020</td>
</tr>
<tr>
<td>CI</td>
<td>0.0078 ± 0.0007</td>
</tr>
<tr>
<td>CP</td>
<td>0.0038 ± 0.0008</td>
</tr>
<tr>
<td>Concatenated</td>
<td>0.0088 ± 0.0004</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Nucleotide diversity was estimated by Kimura’s two-parameter method.

Table 2. Analysis of selection in codons of SMV genome

Analyses based on 56 full-length sequenced isolates using FEL and IFEL.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Sites</th>
<th>FEL</th>
<th>IFEL</th>
<th>Location*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>P1, 640†</td>
</tr>
<tr>
<td>P1</td>
<td>308</td>
<td>54</td>
<td>253</td>
<td>2</td>
</tr>
<tr>
<td>HC-Pro</td>
<td>457</td>
<td>126</td>
<td>331</td>
<td>0</td>
</tr>
<tr>
<td>P3</td>
<td>347</td>
<td>89</td>
<td>257</td>
<td>1</td>
</tr>
<tr>
<td>6K1</td>
<td>52</td>
<td>10</td>
<td>42</td>
<td>0</td>
</tr>
<tr>
<td>CI</td>
<td>634</td>
<td>181</td>
<td>453</td>
<td>0</td>
</tr>
<tr>
<td>6K2</td>
<td>53</td>
<td>13</td>
<td>40</td>
<td>0</td>
</tr>
<tr>
<td>Vpg</td>
<td>190</td>
<td>4</td>
<td>186</td>
<td>0</td>
</tr>
<tr>
<td>Nla-Pro</td>
<td>243</td>
<td>68</td>
<td>175</td>
<td>0</td>
</tr>
<tr>
<td>Nlb</td>
<td>517</td>
<td>140</td>
<td>376</td>
<td>1</td>
</tr>
<tr>
<td>CP</td>
<td>265</td>
<td>63</td>
<td>202</td>
<td>0</td>
</tr>
</tbody>
</table>

*Significant sites, $P$ value < 0.05.
†Sites detected to be under positive selection by both FEL and IFEL methods.
‡Sites detected to be under positive selection by only IFEL.
larger geographical regions in the world, using in both cases the concatenated genes encoding proteins P1, P3, 6K1, CI and CP. Genetic diversity was generally low for the populations sampled at each of the four Iranian provinces, in the range 0.003–0.029 (Table S4), the higher diversity within the Ardabil population (0.029) being due only to the highly divergent isolate Ar13 (not shown). Similarly, the fixation index $F_{ST}$ showed low values between the three Iranian populations for which five or more sequences were available (Table 3).

At the world level, spatial genetic structure was analysed considering as different populations those from countries for which sequences of five or more isolates had been reported. The analysed set consisted of 29 SMV isolates from Korea, 10 from the USA, six from China and five from Canada, plus the 25 isolates from Iran. The diversity of the Iranian and Canadian populations was lowest, 10–15 % of the diversity of the populations from Korea, China or USA (Table S4). The degree of population differentiation, as estimated by the $F_{ST}$ index, varied largely among these regions of the world, but showed quite high values indicating a strong spatial structure (Table 3b). $F_{ST}$ values were unrelated to geographical distance, as they were lowest between Korea and USA, and between China and these two countries. The Iranian population was the most strongly differentiated from the rest (Table 3b).

**Pathogenicity diversity of SMV in Iran**

Fifteen isolates, representing the three analysed regions in Iran, were characterized for their pathogenicity to a set of nine differential soybean cultivars that were either susceptible, or carrying different resistance genes and alleles against SMV. The results of this assay (Table 4) show that, as expected, all 15 isolates were able to infect the two cultivars that did not carry any resistance allele, i.e. Williams 82 and Essex (rsv). Infection in the susceptible cultivars resulted in a systemic mosaic, except for isolate Go11, which caused a clearly identifiable severe mosaic. No isolate was able to infect plants carrying any of the resistance alleles at Rsv1. However, minor differences in reaction on Rsv1 could be detected among isolates. Notably, for isolate Go37, one out of 11 inoculated plants of cv. York showed systemic necrosis at 14 days post-infection (p.i.), and SMV was detected by ELISA in upper non-inoculated leaves. However, retroinoculation of seven York plants with sap from symptomatic leaves of this plant did not result in systemic infection of any of them, discarding the possibility of selection of a resistance-breaking variant. Also, one out of seven Marshall (Rsv1-m) plants inoculated with isolate Go10 showed a delayed systemic necrosis at 24 days p.i., but it was negative in ELISA testing. All isolates were able to overcome Rsv3 resistance causing a systemic mosaic, with isolates Go11 and Ar13 causing more severe symptoms. Lastly, isolates were split into two groups according to their ability to overcome Rsv4 resistance: most isolates infected plants of line V94-5152 (Rsv4) systemically, but inoculation with isolates Lo3 and Ar33 did not result in infection. The biological characterization of these SMV isolates in the same soybean genotypes was repeated in a second independent experiment, with the same results (not shown).

In summary, two major SMV pathotypes were identified in Iran: one pathotype, including isolates Lo3 and Ar33, was able to overcome Rsv3 resistance, and another pathotype, including the other 13 isolates, was able to overcome both Rsv3 and Rsv4 resistance.

**Table 3. Analysis of spatial structure in the SMV population of Iran (a) and of the world (b)**

Data are $F_{ST}$ values.

<table>
<thead>
<tr>
<th></th>
<th>Golestan</th>
<th>Mazandaran</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mazandaran</td>
<td>0.02671</td>
<td>0.04959</td>
</tr>
<tr>
<td>Ardabil</td>
<td>0.14959</td>
<td>0.04959</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Iran</th>
<th>Korea</th>
<th>USA</th>
<th>China</th>
</tr>
</thead>
<tbody>
<tr>
<td>Korea</td>
<td>0.38870</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>USA</td>
<td>0.49240</td>
<td>0.04952</td>
<td></td>
<td></td>
</tr>
<tr>
<td>China</td>
<td>0.56838</td>
<td>0.21267</td>
<td>0.16041</td>
<td></td>
</tr>
<tr>
<td>Canada</td>
<td>0.66698</td>
<td>0.19853</td>
<td>0.21400</td>
<td>0.41612</td>
</tr>
</tbody>
</table>

**Identification of determinants for overcoming Rsv4 resistance in line V94-5152**

It has been reported that mutations in SMV P3 protein are responsible for overcoming Rsv4 resistance (Chowdareddy *et al.*, 2011; Khatabi *et al.*, 2011), notwithstanding that other viral proteins could also be involved (Chowdareddy *et al.*, 2011). However, the reported mutations were not present in the P3 protein of any of the Iranian isolates. A comparison of the P3 protein sequence of Iranian SMV isolates that were or were not able to infect the Rsv4-carrying line V94-5152 (Table 4) showed that they differed at the amino acid position 1053 of the polyprotein (Fig. S2): avirulent isolates Lo3 and Ar33 had a serine (S), while virulent isolates had an asparagine (N), resulting from the nucleotide substitution G3289A (not shown). Amino acid position 1053 of SMV polyprotein corresponds to position 288 in the P3 protein, which was shown to be under positive selection (Table 2). Hence, we determined whether mutation S1053N could have an effect on V94-5152 Rsv4-resistance breaking.

For this, an infectious cDNA clone of isolate Ar33 was obtained (pSMV-Ar33). Infectious transcripts from pSMV-Ar33 had the same pathotype as the original isolate: they infected systemically plants of the susceptible cultivar Williams 82 as well as of the Rsv3-carrying line L29, but did not infect plants of the various cultivars carrying Rsv4 or Rsv3 (not shown and Fig. S1). Then, pSMV-Ar33 was
Table 4. Responses of differential soybean genotypes after inoculation with different SMV isolates, or derived mutants, sampled in Iran

Inoculated plants were evaluated at 20 days p.i. by ELISA.

<table>
<thead>
<tr>
<th>SMV isolate</th>
<th>Soybean cultivar (genotype)*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Williams 82 (rsv)</td>
</tr>
<tr>
<td>Lo2</td>
<td>6/7</td>
</tr>
<tr>
<td>Lo3</td>
<td>7/8</td>
</tr>
<tr>
<td>Lo4</td>
<td>7/7</td>
</tr>
<tr>
<td>Go5</td>
<td>9/10</td>
</tr>
<tr>
<td>Go6</td>
<td>7/8</td>
</tr>
<tr>
<td>Go9</td>
<td>9/9</td>
</tr>
<tr>
<td>Go10</td>
<td>9/9</td>
</tr>
<tr>
<td>Go11</td>
<td>12/12‡</td>
</tr>
<tr>
<td>Go12</td>
<td>10/10</td>
</tr>
<tr>
<td>Go37</td>
<td>8/8</td>
</tr>
<tr>
<td>Ar13</td>
<td>11/11</td>
</tr>
<tr>
<td>Ar15</td>
<td>9/9</td>
</tr>
<tr>
<td>Ar16</td>
<td>7/9</td>
</tr>
<tr>
<td>Ar18</td>
<td>7/10</td>
</tr>
<tr>
<td>Ar33</td>
<td>9/10</td>
</tr>
<tr>
<td>Ar33S1053N</td>
<td>9/9</td>
</tr>
</tbody>
</table>

*Number of plants systemically infected/plants inoculated on primary leaves.
†One of seven plants showed delayed systemic necrosis at 24 days p.i. but was negative in the ELISA test.
‡Severe mosaic symptoms.
§One of eleven inoculated plants showed systemic necrosis and was positive in the ELISA test.
mutagenized to introduce the substitution S→N at position 1053, yielding plasmid pSMV-Ar33S1053N. The nucleotide sequence of the SMV genome in this clone (accession no. KF297335) was identical to that determined for the parental isolate Ar33, except for the introduced mutation. Inoculation of transcripts from this plasmid into plants of different soybean cultivars showed that the S1053N mutant was able to systemically infect the susceptible cultivar Williams 82 and the L29 Rsv3 resistant line, and unable to infect Rsv1 resistance cultivars. Inoculation of pSMV-Ar33S1053N transcripts on plants of the V94-5152 (Rsv4) line resulted in a systemic mild mosaic (not shown) that was apparent 10 days p.i., and systemic infection was confirmed by ELISA and RT-PCR amplification of the P3 gene using primer pair P3F/P3rb (Fig. S1). The stability of the introduced mutation was confirmed by sequencing the P3 cistron from RNA extracts of systemically infected leaves. No new mutation at P3 was found in the progeny virus.

**DISCUSSION**

The purpose of this study was to characterize the variability of the SMV population infecting soybean in the major soybean producing areas of Iran. This information should be valuable for setting up sustainable strategies for the control of a major pathogen in an expanding crop in Iran. In addition, comparison of the obtained results with reports from other regions of the world could elucidate SMV long-distance movement and diversification. Iranian isolates were characterized molecularly by total or partial determination of their genome’s nucleotide sequence, and biologically by the ability to infect a panel of differential soybean cultivars or lines.

The Iranian SMV population is a single, spatially undifferentiated population, with effective gene flow between regions, as shown by low FST values (Table 3a). It showed very low genetic diversity, one order of magnitude lower than for regions such as China, Korea or the USA, which have a long history of soybean cultivation and SMV infection. Low genetic diversity strongly suggests a recent population bottleneck that could be associated with the recent introduction and expansion of the virus in Iran, in the late twentieth century (Eskandari 1978; Karimi & Noaparast, 1989). Low diversity after a population bottleneck would also explain the strong genetic differentiation of the Iranian population relative to those from other world regions (Table 3b). Accordingly, most (22 out of 25) characterized SMV isolates formed a single, monophyletic cluster that is sister to a cluster mostly including isolates from Korea, strongly suggesting a major introduction from East Asia. A second cluster, formed by isolates Lo3 and Ar33, again sister to clusters of isolates from East Asia and North America, belongs to the same subgroup of SMV isolates (Fig. 3), and may represent a second introduction from the same region. Last, the broadly divergent isolate Ar13 would indicate a third introduction from an unknown origin. Efficient seed transmission of SMV (Domier et al., 2007) suggests that these introductions could have occurred through imported lots of soybean seed. Seed transmission could also explain the presence of isolates from the second phylogenetic cluster, which represents a minor fraction of the Iranian SMV population, in the distant provinces of Lorestan and Ardabil.

The limited genetic diversification of the Iranian SMV population, as has also been reported for those of many other viruses (e.g. Garcia-Arenal et al., 2001), might be in apparent contradiction with the high nucleotide substitution rates reported for RNA viruses, including poty-viruses (Gibbs et al., 2008, 2010; Jenkins et al., 2002). High mutation rates would be countered by strong selection against amino acid change in the proteins encoded by all SMV genes, as shown by dN/dS ratios (Table 1) and by the large fraction of codons under negative selection. Indeed, most codons in the SMV ORF are under negative selection or neutral variation. Thus, the evolution of SMV seems to be strongly conditioned by purifying selection in combination with neutral variation, a trait in common with most RNA viruses (Chare & Holmes, 2004; Elena & Sanjuán, 2005; García-Arenal et al., 2001; Hughes & Hughes, 2007). Constraints on genetic variation were weaker for P1 and P3 proteins, in agreement with previous studies of other sets of SMV isolates (Seo et al. 2009b) or other potyviruses such as Turnip mosaic virus (TuMV), Potato virus Y and Watermelon mosaic virus (Cuevas et al., 2012; Moreno et al., 2004; Ohshima et al., 2002; Tomimura et al., 2004), indicating different evolutionary dynamics acting on the various potyviral genes. The second major mechanism in the evolution of SMV is recombination, which may also have a role in maintaining the effectiveness of purifying selection by preventing the accumulation of deleterious mutations due to ‘Muller’s ratchet’ (Chao, 1990; Duarte et al., 1992; Fraile et al., 1997). However, recombination is a key mechanism in the generation of genetic variation, and its role in the evolution of SMV (Gagarinova et al., 2008b, Seo et al., 2009b) and in other potyviruses (Chare & Holmes, 2006; Cuevas et al., 2012; Ohshima et al., 2007) has been well documented. The inclusion of the Iranian SMV isolates in an analysis of recombination, using all available SMV genome sequences, allowed identification of new recombination sites, refining the recombination map of SMV. Thus, the genomes of most reported SMV isolates (55 of 77 isolates) show evidence of recombination events. The fact that putative parental genomes are often from distant geographical regions (see examples in Table S3) suggests that some recombinants arose a long time ago, and spread later on, but this is a subject that needs more detailed analyses. The recombination map of SMV (Fig. 2) indicates an uneven distribution of recombination sites over the genome, with a noticeable concentration in the CI cistron, as reported also in TuMV (Ohshima et al., 2007). High recombination in the CI cistron is suggestive of a role of recombination as a factor of purifying selection, as this gene is among those with a lowest dN/dS ratio of the whole genome.
Pathotyping of 15 Iranian SMV isolates on a collection of indicator soybean cultivars, or lines carrying different resistance genes, showed a low phenotypic diversity: most (13 of 15) isolates belonged to the same pathotype, which was able to overcome *Rsv*3 and *Rsv*4 resistances, but not *Rsv1* resistances. Overcoming the resistance provided by *Rsv1* alleles is infrequent in SMV (Hajimorad et al., 2003), and has been reported only from South Korea (Choi et al., 2005). This may be due to the fact that *Rsv1* is a multigene locus, and that the gain of pathogenicity on *Rsv1* is dependent on different cistrons, P3, HC-Pro (Eggenberger et al., 2008; Hajimorad et al., 2008, 2011; Wen et al., 2013) and CI (Chowda-Reddy et al., 2011b), and requires different combinations of mutations in these various genes. However, both pathogenicity on *Rsv3*, which also is dependent on the CI (Seo et al., 2009a; Zhang et al., 2009) and P3 proteins (Chowda-Reddy et al., 2011b), and pathogenicity on *Rsv4*, involving mutations in the P3 protein (Chowda-Reddy et al., 2011a; Khatabi et al., 2012) and possibly other viral proteins or domain(s) including CI (Chowda-Reddy et al., 2011b), is widespread, and has been reported both in East Asia and North America (Choi et al., 2005; Gagarinova et al., 2008a; Khatabi et al., 2012; Seo et al., 2009b). It is interesting to note that the breaking of V94-5152 *Rsv4* resistance in the Iranian SMV population seems to have evolved at least twice, as it occurs in the broadly divergent isolate Ar13 and in the major phylogenetic cluster of Iranian isolates several times (compare Fig. 3 and Table 4). Unfortunately, we have been unable to establish whether *Rsv4* resistance has been deployed in Iran and whether it et al., 2005; Gagarinova et al., 2008) and has been reported in bean (*Phaseolus vulgaris*), in which local necrotic lesions developed 5 days p.i. Individual lesions were excised and used to inoculate soybean Williams 82 for isolate multiplication. Interestingly, none of the isolates belonging to the minor (Lo3, Ar33) cluster were able to overcome V94-5152 *Rsv4* resistance. The capacity of overcoming *Rsv4* resistance in line V94-5152 has been mapped to the P3 protein (Chowda-Reddy et al., 2011a; Khatabi et al., 2012), a protein that is often the avirulence factor for resistance to different potyviruses in various plant species (Hjulsager et al., 2006; Jenner et al., 2002, 2003; Kim et al., 2010). Comparisons of the P3 amino acid sequence of Iranian SMV isolates virulent or avirulent on *Rsv4* showed that the two mutations reported as involved in overcoming *Rsv4* resistance, Q1033K and G1054R (Chowda-Reddy et al., 2011a; Khatabi et al., 2012), were not present in the resistance-breaking Iranian isolates. Instead, they had the mutation S1053N, shown by site-directed mutagenesis to suffice for *Rsv4* resistance breaking. Interestingly, codon 1053 was shown to be under positive selection, strongly suggesting that there is a selection pressure to overcome *Rsv4* resistance. Thus, our data complement previous reports showing different pathways towards *Rsv4* resistance breaking. The *Rsv4* resistance-breaking isolate L-RB does not cluster with SMV isolates from Iran (Fig. 3). This result allows speculation about different evolutionary pathways to resistance breaking being favoured by different genomic contexts, as reported for other viruses (Pinel-Galzi et al., 2007; Pouliard et al., 2012). Thus, our results could provide another instance of limitations to genome variation in RNA viruses due to epistatic effects.

**METHODS**

**Field surveys and isolate collection.** During the summer of 2011, a survey for SMV was carried out in 51 randomly selected soybean fields in the four soybean-growing provinces of Iran. Plants showing symptoms of mosaic, mottling, yellowing, stunting, chlorosis, and vein or top necrosis, were collected, and the presence of SMV was detected by DAS-ELISA using specific polyclonal antibodies (DSMZ), and RT-PCR using SMV-specific primer pair CPf1/CPr in PCR (Table S5). SMV was transferred from field-infected samples to soybean cv. Williams 82 by mechanical inoculation of fully expanded primary leaves with leaf homogenates in 0.01 M sodium phosphate buffer (pH 7.0) including 0.5 % Na2SO3, 2 % Polyvinylpyrrolidone (PVP) and 0.2 % activated charcoal. Plants were maintained in an insect-proof greenhouse at a night–day temperature of 20–25 °C. Biological cloning of isolates successfully transferred to Williams 82 was carried out in bean (*Phaseolus vulgaris*) cv. Topcrop, in which local necrotic lesions developed 5 days p.i. Individual lesions were excised and used to inoculate soybean Williams 82 for isolate multiplication.

**Characterization of SMV isolates.** SMV isolates that were successfully biologically cloned and multiplied were molecularly and biologically characterized. For molecular characterization, the nucleotide sequence of the genes encoding P1, P3, 6K1, CI and CP was determined. For this, total RNA was extracted from leaves of infected Williams 82 plants using the RNeasy plant mini kit (Qiagen) and RT-PCR using SMV-specific primer pair CPf1/CPr in PCR (Table S5). To determine the viral isolate, total RNA was extracted from leaves of infected Williams 82 using the RNeasy plant mini kit (Qiagen) and RT-PCR using SMV-specific primer pair CPf1/CPr in PCR (Table S5). SMV was detected using primers specific for the 6K1, CI and CP genes (http://bioinfo.ut.ee/primer3-4.0.0). Full-length genome sequences were determined for some isolates. A total of 17 pairs of primers were used, listed in Table S5. Since sequences were determined from different RT-PCR amplicons, primers were designed so that fragments overlapped by at least 150 bp. PCR products were sequenced at Macrogen (Netherlands). The 5’ and 3’ terminal genomic regions of isolates that were fully sequenced were determined after amplification with a FirstChoice RLM-RACE kit (Invitrogen).

**Biological characterization.** Biological characterization was based on pathotype determination after inoculation on a set of differential soybean cultivars carrying different resistance alleles at the *Rsv* loci, as in Ma et al. (2003) (Table 4), generously provided by Professor Sue Tolin (Virginia Technical University, USA). Plants were monitored for symptom development for 3–4 weeks post-inoculation, and at 20 days p.i. were tested for SMV systemic infection by DAS-ELISA.

**Nucleotide sequence analyses.** Either full-length genome sequences or the concatenated fragments of partial sequences were analysed. Sequences were aligned using CLC Main Workbench 6.7.1 software (http://www.clcbio.com/products/latest-improvements-main-workbench). Recombination events, most likely parental isolates of recombinants, and recombination break points were identified by rup,
GENECONV, Chimera, MaxChi, BOOTSCAN and SISCAN methods as implemented in RDP4. Beta 4.16 (Martin et al., 2010) with default settings, and a Bonferroni corrected P value cut-off of 0.01. Recombination events identified by at least three different methods with an associated $P<10^{-8}$ were considered as significant.

Phylogenetic analyses were carried out on non-recombinant genomes or fragments using maximum-likelihood (ML) procedures. Substitution models were selected by JModelTest 0.1.1, implemented in Phylogenetic 2.0 (http://phylogenies.bioinf.cipf.es/index.html). GTR + I + G and GTR + G models were selected for full-length and concatenated sequence alignment analyses, respectively. Phylogenetic trees were generated using PhyML 3.0 (Guindon & Gascuel, 2003). Significance of tree topology was analysed by 100 bootstrap pseudoreplicates. The tree was midpoint rooted using FigTree version 1.3.1 software (http://tree.bio.ed.ac.uk/software/).

Selective pressures on proteins were analysed by determining pairwise substitutions at synonymous (dS) and non-synonymous (dN) sites by the Pamilo–Bianchi–Li method (Li, 1993; Pamilo & Bianchi, 1993), using MEGA 5. Selection on individual codons was estimated as the difference between non-synonymous and synonymous substitutions rates per codon using two different codon-based ML methods: the fixed-effects likelihood (FEL) and internal branches fixed-effects likelihood (IFEL), as implemented in the HYphy server (http://www.datamonkey.org).

**Population genetics analyses.** Genetic distances between pairs of isolates were analysed by Kimura’s two-parameter method (Kimura, 1980), using MEGA 5. Within population diversities were estimated on the basis of pairwise distances. Isolates from different provinces of Iran, or from different regions in the world, were considered to constitute different populations for these analyses. Haplotypes were identified using GenAlEx 6.5 (Peakall & Smouse, 2012), and population differentiation was tested by pairwise FST test between the groups of haplotypes from each population using Arlequin v3.5 (Excoffier et al., 2005).

**Construction of an infectious full-length cDNA clone and site-directed mutagenesis.** A full-length cDNA copy of isolate SMV-Ar33 genome was synthesized with SuperScript III reverse transcriptase (Invitrogen) using primer 3Addr (with a NotI restriction site) and, subsequently, primer SMV full (including the 5–26 nt of SMV-Ar33 RNA plus a T7 promoter sequence) and 3’RACE were used for full-length PCR amplification by Phusson high-fidelity DNA polymerase (Thermo Scientific). PCR products were gel-purified and cloned into pCR-XL-TOPO vector using a TOPO XL PCR cloning kit, and constructs were transformed in One Shot TOP10 chemically competent cells, according to the manufacturer’s protocol (Invitrogen). Clones were checked by restriction analysis and nucleotide sequence determination of the inserts. Infectivity of clones was analysed by inoculation of susceptible plants with in vitro transscripts of NotI-linearized constructs. The resulting infectious clone of SMV-Ar33 was named pSMV-Ar33.

For site-directed mutagenesis at the P3 gene, primer mutF was designed to introduce the mutation G3289A, resulting in the S1053N amino acid substitution. PCR was performed with primers mutF and mutR for amplicon synthesis from pSMV-Ar33 template. The amplicons were digested with SpeI and SalI restriction enzymes, and ligated into pSMV-Ar33 to generate pSMV-Ar33S1053N. The resulting nucleotide substitution was confirmed by nucleotide sequence determination of the P3 cistron in pSMV-Ar33S1053N.

**ACKNOWLEDGEMENTS**

We would like to thank Professor Sue Tolin for providing seeds of various soybean genotypes. We also thank Drs Aurora Fraile and Israel Pagán, C. B. G. P., U. P. M., for critical reading of a previous version of this manuscript. This project was funded in part by a grant from Plan Nacional de 1+D+I, Spain, no. AGL2008-02458, to F. G. A.

**REFERENCES**


