



Article

Genome-Wide Association Study and Genomic Prediction of Soybean Mosaic Virus Resistance

Di He ^{1,2,†}, Xintong Wu ^{1,†}, Zhi Liu ¹ , Qing Yang ¹, Xiaolei Shi ¹, Qijian Song ³, Ainong Shi ⁴ , Dexiao Li ^{5,*} and Long Yan ^{1,2,*}

¹ Institute of Cereal and Oil Crops, Hebei Academy of Agricultural and Forestry Sciences, Shijiazhuang 050035, China; ternuraeyes@163.com (D.H.); xintongwu66@163.com (X.W.); zhiliulin@sina.com (Z.L.); qyang0807008@163.com (Q.Y.); shixiaolei59@163.com (X.S.)

² College of Life Sciences, Hebei Agricultural University, Baoding 071001, China

³ Soybean Genomics and Improvement Laboratory, Agricultural Research Service, Beltsville, MD 20705, USA; qijian.song@usda.gov

⁴ Department of Horticulture, University of Arkansas, Fayetteville, AR 72701, USA; ashi@uark.edu

⁵ College of Agronomy, Northwest A&F University, Yangling 712100, China

* Correspondence: lidexiao@nwsuaf.edu.cn (D.L.); dragonyan1979@163.com (L.Y.)

† These authors contributed equally to this work.

Abstract: Soybean mosaic virus (SMV), a pathogen responsible for inducing leaf mosaic or necrosis symptoms, significantly compromises soybean seed yield and quality. According to the classification system in the United States, SMV is categorized into seven distinct strains (G1 to G7). In this study, we performed a genome-wide association study (GWAS) in GAPIT3 using four analytical models (MLM, MLMM, FarmCPU, and BLINK) on 218 soybean accessions. We identified 22 SNPs significantly associated with G1 resistance across chromosomes 1, 2, 3, 12, 13, 17, and 18. Notably, a major quantitative trait locus (QTL) spanning 873 kb (29.85–30.73 Mb) on chromosome 13 exhibited strong association with SMV G1 resistance, including the four key SNP markers: Gm13_29459954_ss715614803, Gm13_29751552_ss715614847, Gm13_30293949_ss715614951, and Gm13_30724301_ss715615024. Within this QTL, four candidate genes were identified: *Glyma.13G194100*, *Glyma.13G184800*, *Glyma.13G184900*, and *Glyma.13G190800* (3Gg2). The genomic prediction (GP) accuracies ranged from 0.60 to 0.83 across three GWAS-derived SNP sets using five models, demonstrating the feasibility of GP for SMV-G1 resistance. These findings could provide a useful reference in soybean breeding targeting SMV-G1 resistance.

Keywords: soybean mosaic virus (SMV); genome-wide association study (GWAS); genomic prediction (GP); quantitative trait locus (QTL); marker-assisted selection (MAS)



Academic Editors: Diego Rubiales, Nicolas Rispail and Eleonora Barilli

Received: 6 December 2024

Revised: 8 February 2025

Accepted: 20 February 2025

Published: 27 February 2025

Citation: He, D.; Wu, X.; Liu, Z.; Yang, Q.; Shi, X.; Song, Q.; Shi, A.; Li, D.; Yan, L. Genome-Wide Association Study and Genomic Prediction of Soybean Mosaic Virus Resistance. *Int. J. Mol. Sci.* **2025**, *26*, 2106. <https://doi.org/10.3390/ijms26052106>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Soybean (*Glycine max* (L.) Merr.), a cultivated oilseed crop, serves as a vital source of plant-based protein and edible oil, contributing substantially to agricultural economies worldwide [1]. However, soybean productivity is severely compromised by soybean mosaic virus (SMV), a member of the Potyvirus genus that induces characteristic symptoms such as leaf mosaic, localized necrosis at infection sites, systemic necrosis vein browning, and apical necrosis [2].

In the United States, SMV isolates are classified into seven distinct strains (G1–G7) based on differential interactions with soybean genotypes [3]. Four major SMV resistance loci (*Rsv1*, *Rsv3*, *Rsv4*, and *Rsv5*) have been identified across genetic studies [2,4,5]. *Rsv1*,

located on chromosome 13, which encodes a cluster of nucleotide-binding leucine-rich repeat (NLR) proteins conferring resistance to G1–G6 strains, with multiple allelic variants (*Rsv1*, *Rsv1-t*, *Rsv1-m*, *Rsv1-k*, *Rsv1-s*, *Rsv1-h*, *Rsv1-r*, and *Rsv1-n*) has been reported across the germplasm [2,6–8]. *Rsv3*, located on chromosome 14, mediates resistance to G5–G7 through a coiled-coil NLR encoded by *Glyma.14G204700* [9], while *Rsv4*, located on chromosome 2, employs an RNase H-mediated antiviral mechanism targeting double-stranded RNA to confer broad-spectrum resistance against all seven strains [10]. Notably, SMV-G1 exhibits the narrowest virulence spectrum, selectively infecting susceptible accessions while failing to establish infection in resistant cultivars, making it an ideal model for dissecting resistance mechanisms [1].

In China, SMV isolates are categorized into 22 strains (SC1–SC22), with resistance governed by the *RSC* (Resistance of SMV in China) locus, encompassing multiple sub-loci such as *RSC3Q*, *RSC3*, *RSC4*, *RSC5*, *RSC7*, *RSC8*, *RSC12*, *RSC14Q*, *RSC18*, and *RSC20*, primarily chromosomes 2, 13, and 14 [4,5]. Several novel resistance loci have been identified on various chromosomes, including 1, 2, 6, 8, 9, 11, 12, 13, 14, 15, and 19 [4,11–14]. Most SMV resistance loci have been identified through QTL mapping [4,6,9,10,15].

Genome-wide association study (GWAS) and genomic prediction (GP) have accelerated the identification of SMV resistance determinants. For instance, Che et al. [16] identified 24 relevant SNPs associated with strain SC3 resistance, including five novel loci on chromosomes 8 and 12, while Pu et al. [4] mapped *Glyma.15G015700* and six novel loci as key players in SC7 resistance. These identified SNPs and candidate genes have the potential for marker-assisted selection (MAS) in breeding programs aimed at enhancing resistance to SMV. In this investigation, ridge regression best linear unbiased prediction (rr-BLUP) in the rrBLUP package and Bayesian models, including Bayes A (BA), Bayes B (BB), Bayes LASSO (BL), and Bayes ridge regression (BRR), were employed for predicting genomic estimated breeding value (GEBV) in genomic prediction [17].

The *Rsv1* locus in soybeans contains a cluster of nucleotide-binding site/leucine-rich repeat (NBS-LRR) genes that collectively elicit diverse responses to SMV strains [6]. Functional studies implicate defense regulators *GmEDR1*, *GmEDS1*, and *GmHSP90* in *Rsv1*-mediated extreme resistance, while *elF5A* has been identified as a crucial element in the *Rsv1*-triggered lethal systemic hypersensitive response to SMV-G7 [18,19]. PCR-based markers have been developed for detecting *Rsv1* alleles [20]. The *Rsv1* gene provides resistance to a wide range of SMV strains through an extreme resistance mechanism that does not rely on hypersensitive response (HR) at the inoculation site [21]. The integration of *Rsv1* with *Rsv3* and *Rsv4* in elite lines such as Chinese cultivar 8101 exemplifies the successful deployment of broad-spectrum resistance [22–24].

In this study, we performed GWAS and GP analyses on 218 USDA soybean accessions to elucidate the genetic determinants of SMV-G1 resistance. These findings advance marker-assisted breeding strategies and provide insights into the genomic architecture underlying resistance to SMV-G1 in soybean.

2. Results

2.1. SMV-G1 Resistance Phenotyping

Evaluation of 218 soybean accessions revealed 72 resistant (R) (33%) and 146 susceptible (S) (67%) genotypes (Supplementary Table S1, Figure 1). Resistance (R) and susceptibility (S) classifications were maintained consistently across the Supplementary Materials.

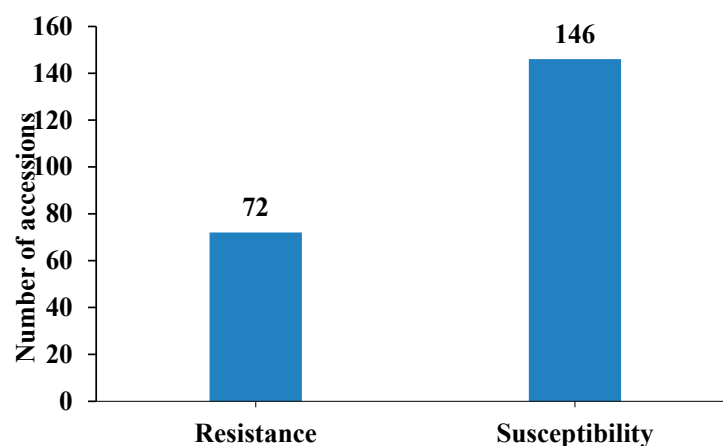


Figure 1. The distribution of soybean mosaic virus reactions among 218 soybean accessions.

2.2. Population Structure and GWAS

Employing GAPIT3, the 218 soybean accessions were categorized into four distinct clusters (subpopulations), designated Q1 to Q4 (Figure 2). This division was based on the following analyses: (1) a 3D graphical plot of the principal component analysis (PCA) (Figure 2A), (2) a PCA eigenvalue plot (Figure 2B), and (3) phylogenetic trees (Figure 2C,D), which were constructed using the neighbor-joining (NJ) method. Additionally, the kinship plot confirmed the existence of these four groups among the 218 accessions (Supplementary Figure S1). Each of the 218 accessions was assigned to one of the four clusters (Q1 to Q4), and the Q-matrix with four matrices was subsequently applied to the GWAS analysis.

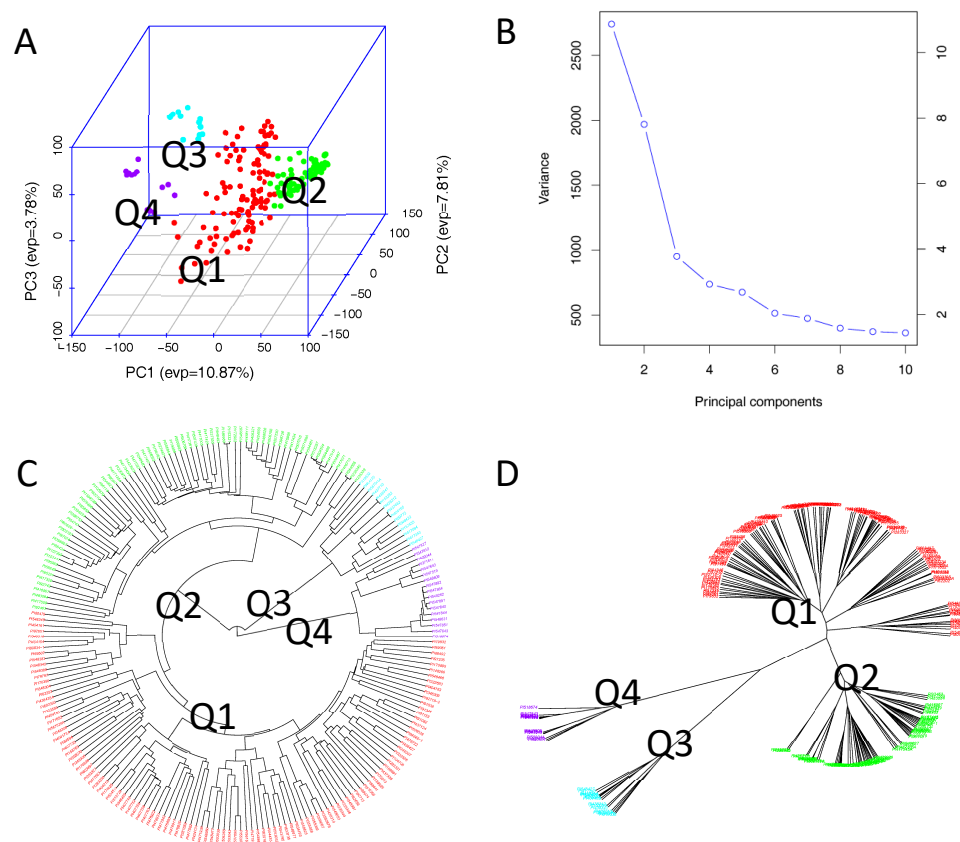


Figure 2. Analysis of genetic diversity in the association panel consisting of 218 USDA soybean accessions: (A) Three-dimensional graphical plot of the principal component analysis (PCA). (B) PCA eigenvalue plot. (C,D) Phylogenetic trees ((C)—fan, (D)—unrooted) drawn using the neighbor-joining (NJ) method in four sub-populations (Q1–Q4) by GAPIT3.

Based on the analysis in GAPIT3, the multiple QQ plot distribution showed a significant deviation from the expected distribution (Figure 3, left half; Supplementary Figure S2, right), indicating the presence of SNPs associated with resistance to SMV G1. The multiple Manhattan plot, covering all 34,631 tested SNPs, revealed several SNPs with LOD values greater than 5.84, primarily located on chromosome 13, suggesting the presence of SNPs associated with SMV G1 resistance in the panel (Figure 3, right half; Supplementary Figure S2, left).

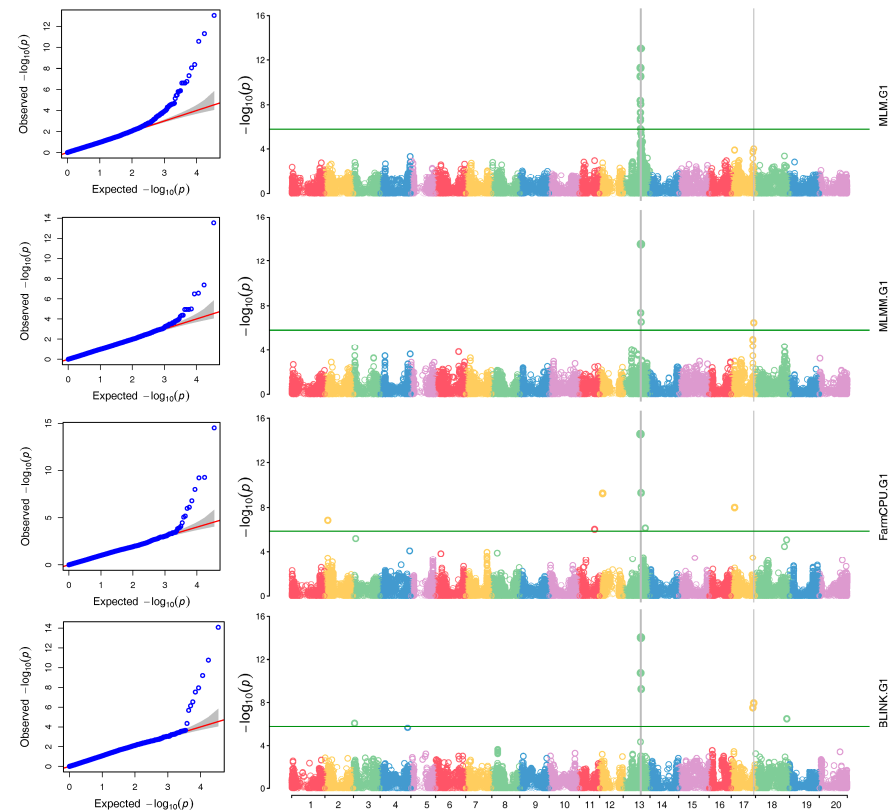


Figure 3. QQ-plots (**left**) and Manhattan plots (**right**) of the results of GWASs on resistance to soybean mosaic virus (SMV) in strain G1 in 218 accessions based on MLM, MLMM, FarmCPU, and BLINK in GAPIT3. In the QQ-plot (**left**), the x-axis is the LOD [$-\log(p\text{-value})$] value and the y-axis is the expected LOD [$-\log(p\text{-value})$] value. In the Manhattan plot (**right**), the x-axis is the soybean chromosome and the y-axis is the LOD [$-\log(p\text{-value})$] value.

Twenty-two SNPs were significantly associated ($\text{LOD} > 5.84$) with G1 resistance across chromosomes 1, 2, 3, 12, 13, 17, and 18 (Figure 3; Supplementary Table S2). Four high-confidence SNPs ($\text{LOD} > 6.5$; detected by ≥ 2 models), Gm13_29459954_ss715614803, Gm13_29751552_ss715614847, Gm13_30293949_ss715614951, and Gm13_30724301_ss715615024, anchored a 1.27 Mb QTL on chromosome 13 (29.46–30.73 Mb), overlapping the Rsv1 locus (Table 1).

Table 1. List of four SNPs with LOD [$-\log(p\text{-value})$] greater than 5.84 detected by two or more models (BLINK, FarmCPU, MLMM, or MLM) in GAPIT3 and their position on chromosomes based on the soybean genome assembly Wm82.a2.v1.

SNP	Chr	Position	LOD [$-\log(p\text{-Value})$]	MAF%	Model
Gm13_29459954_ss715614803	13	29459954	10.58	19.5	MLM
			14.53	19.5	FarmCPU
Gm13_29751552_ss715614847	13	29751552	8.06	44	MLM
			7.37	44	MLMM
			10.76	44	BLINK

Table 1. *Cont.*

SNP	Chr	Position	LOD [$-\log(p\text{-Value})$]	MAF%	Model
Gm13_30293949_ss715614951	13	30293949	13.04	13.5	MLM
			13.54	13.5	MLMM
			14.09	13.5	BLINK
Gm13_30724301_ss715615024	13	30724301	6.56	22.48	MLMM
			9.21	22.48	BLINK

2.3. Candidate Gene Identification

Thirty-eight genes resided within 5 kb flanking regions of 20 significant SNPs (Supplementary Table S3).

Notably, four disease resistance gene analogs were detected in a region spanning 873 kb, from 29.85 Mb to 30.73 Mb, with strongly associated SNP markers on chromosome 13 (Supplementary Table S3). *Glyma.13G194100*, located from 30,726,074 bp to 30,730,604 bp on chromosome 13, is closely linked to SNP marker Gm13_30724301_ss715615024 (chr 13: 30,724,301 bp) within a distance of less than 2 kb. *Glyma.13G184800* and *Glyma.13G184900* are situated between the two SNP markers, Gm13_29751552_ss715614847 and Gm13_30293949_ss715614951, with distances ranging from 107 kb to 417 kb. Another gene, called 3Gg2 [6,20], is also located in this region between Gm13_30293949_ss715614951 and Gm13_30462695_ss715614981, with distances of approximately 34 kb and 130 kb (Supplementary Table S3).

2.4. Genomic Prediction Accuracy

GP in the reference: The genomic prediction achieved very high accuracy (maBLUP: $r = 0.95$; gBLUP: $r = 0.96$; sBLUP: $r = 0.83$) (Figure 4). These estimates were obtained using a training set to predict SMV G1 resistance in a panel of 218 soybean accessions with all 34,631 SNPs in GAPIT3. This indicates that genomic selection is efficient for selecting SMV G1 resistance.

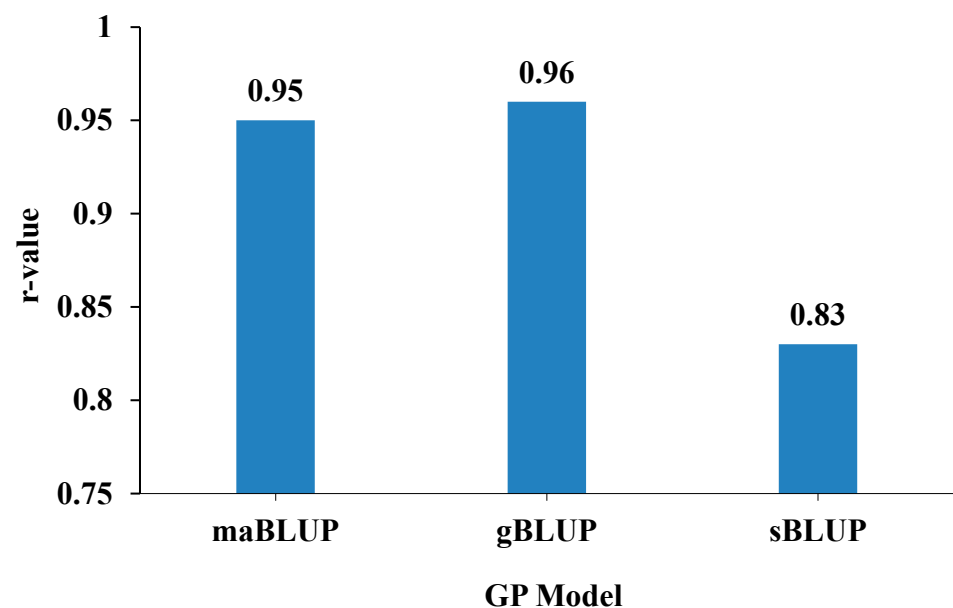


Figure 4. GP model r-value of SMV strain G1 resistance in training sets within a panel of 218 soybean accessions assayed with 34,631 SNPs. The predictions were conducted using three models: maGLUP, gBLUP, and sBLUP in GAPIT3.

GP in cross-prediction using randomly selected SNP markers: The genomic prediction via cross-prediction using randomly selected SNP markers showed the following average *r*-values: 0.19 in *r*4, ranging from 0.10 to 0.22; 0.35 in *r*100, ranging from 0.31 to 0.39; 0.41 in *r*500, ranging from 0.39 to 0.43; 0.48 in *r*1000, ranging from 0.45 to 0.50; 0.47 in *r*2000, ranging from 0.46 to 0.48; 0.52 in *r*5000, ranging from 0.51 to 0.54; and 0.53 in *r*10000, ranging from 0.49 to 0.58 (Table 2, Figure 5). These results showed that the *r*-value increased as more randomly selected SNPs were used, from an average of 0.19 in the 4-SNP set to 0.53 in the 10,000-SNP set. This suggests that random SNP sets required ≥ 1000 markers to reach $r \geq 0.48$ in genomic selection (GS) of SMV-G1 resistance. GWAS-derived SNP marker sets: The average *r*-value was 0.60 in *m*3, ranging from 0.56 to 0.62; 0.63 in *m*4, ranging from 0.58 to 0.64; and 0.80 in *m*22, ranging from 0.72 to 0.83. These findings revealed that GWAS-derived SNP markers can be effectively used to estimate GP and to select for SMV-G1 resistance in soybean breeding through MAS and GS (Table 2, Figure 5).

Table 2. GP model *r*-value and standardized errors of the *r*-values (SE) of SMV strain G1 resistance using ten SNP sets: seven random SNP sets with SNP number from 4 to 10,000 and three GWAS-derived SNP marker sets (3 markers—*m*3; 4 markers—*m*4; and 22 markers—*m*22). Predictions were performed using five GP models: BA, BB, BL, BRR, and rrBLUP.

SNP Set	GP Model <i>r</i> -Value						SE					
	BA	BB	BL	BRR	rrBLUP	SNP Set Mean	BA	BB	BL	BRR	rrBLUP	SNP Set Mean
<i>r</i> 4	0.22	0.10	0.20	0.21	0.20	0.19	0.011	0.013	0.012	0.012	0.011	0.012
<i>r</i> 100	0.39	0.33	0.35	0.37	0.31	0.35	0.012	0.014	0.014	0.011	0.013	0.013
<i>r</i> 500	0.43	0.42	0.41	0.39	0.40	0.41	0.010	0.011	0.012	0.013	0.012	0.011
<i>r</i> 1000	0.50	0.47	0.48	0.50	0.45	0.48	0.011	0.011	0.012	0.011	0.012	0.012
<i>r</i> 2000	0.48	0.46	0.48	0.48	0.46	0.47	0.012	0.013	0.010	0.012	0.011	0.012
<i>r</i> 5000	0.52	0.51	0.51	0.52	0.54	0.52	0.012	0.010	0.012	0.009	0.011	0.011
<i>r</i> 10000	0.54	0.58	0.50	0.49	0.53	0.53	0.011	0.010	0.011	0.013	0.010	0.011
<i>m</i> 3	0.61	0.60	0.61	0.62	0.56	0.60	0.011	0.015	0.009	0.009	0.011	0.011
<i>m</i> 4	0.64	0.64	0.63	0.64	0.58	0.63	0.011	0.009	0.008	0.009	0.012	0.010
<i>m</i> 22	0.83	0.81	0.82	0.82	0.72	0.80	0.004	0.005	0.005	0.005	0.007	0.005
Model Mean	0.52	0.49	0.50	0.51	0.48	0.50	0.010	0.011	0.010	0.010	0.011	0.011

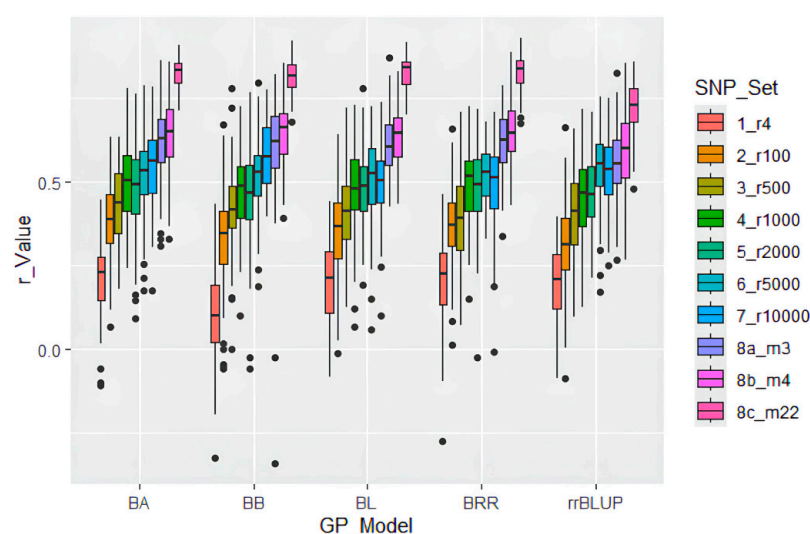


Figure 5. GP model *r*-value of SMV strain G1 resistance using ten SNP sets: seven random SNP sets, with 4 to 10,000 SNPs, plus three GWAS-derived significant SNP marker sets (3 markers—*m*3; 4 markers—*m*4; and 22 markers—*m*22). Predictions were estimated using five genomic prediction (GP) models: BA, BB, BL, BRR, and rrBLUP.

GP Model: Bayesian models (BA, BB, BL, BRR) exhibited comparable efficacy ($r \approx 0.53\text{--}0.80$), surpassing rrBLUP ($r \approx 0.49\text{--}0.58$). This suggests that BA, BB, BL, and BRR can be effectively utilized for selecting resistance to SMV-G1 in genomic selection (GS).

2.5. Genetic Diversity and Utilization of the SMV-Resistant Accessions

The 72 resistant accessions originated from 15 countries, including China (21 accessions), Japan (13), the United States (12), South Korea (7), North Korea (4), Uruguay (3), Georgia (2), South Africa (2), and 1 each from 7 other countries, plus 1 of unknown origin. The majority (57 accessions; 79.2%) of SMV G1-resistant accessions were from China, Japan, the United States, South Korea, and North Korea (Supplementary Table S4).

Maximum-likelihood phylogeny resolved three genetic clusters (Figure 6, Supplementary Table S4), as follows.

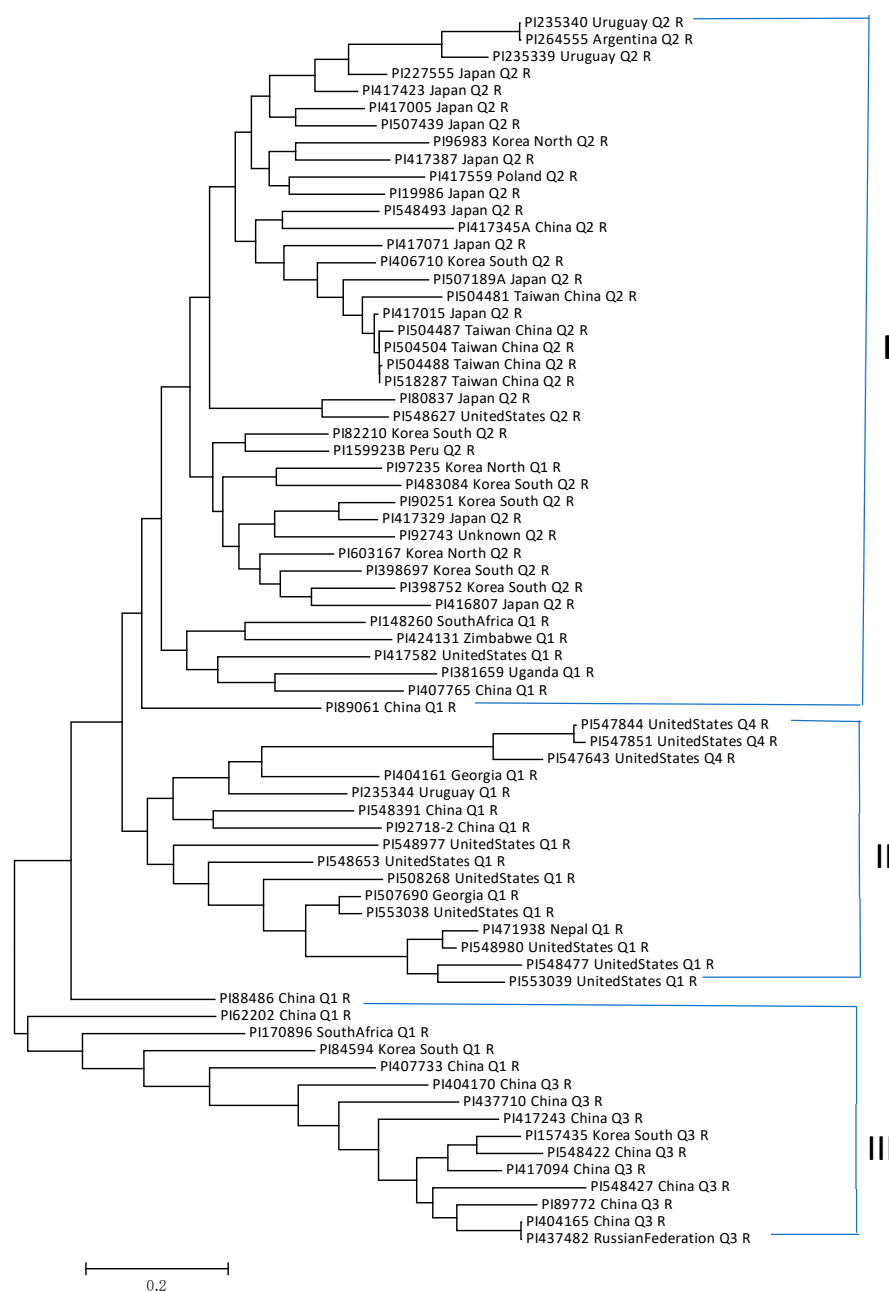


Figure 6. Phylogenetic tree created from MEGA 7 of 72 soybean germplasm accessions resistant to SMV strain G1 using the maximum likelihood (ML) method.

Group I consists of 41 accessions collected from 12 countries, including Japan (13 accessions), China (8), South Korea (6), North Korea (3), the United States (2), Uruguay (2), and 1 each from 6 other countries, plus 1 of unknown origin. The majority (30 accessions; 73.2%) were from China, Japan, South Korea, and North Korea (Supplementary Table S4b).

Group II consists of 16 accessions collected from five countries: the United States (10), Georgia (2), China (2), Uruguay (1), and Nepal (1). The majority (10 accessions; 62.5%) were from the United States (Supplementary Table S4b).

Group III consists of 15 accessions collected from four countries, including China (11), South Korea (2), the Russian Federation (1), and South Africa (1). The majority (11 accessions; 73.3%) were from China (Supplementary Table S4b).

Geographic analysis showed that all 13 SMV-resistant accessions from Japan were clustered into Group I. Among the 12 accessions from the United States, 10 (83.3%) were classified in Group II. Of the 21 accessions from China, 11 (52.4%) were in Group III, while 8 (19.5%) and 2 (12.5%) were in Groups I and II, respectively (Supplementary Table S4). These findings indicate that geographic factors influence the distribution of SMV-G1-resistant accessions.

Phylogenetic tree analysis of all 218 tested materials showed that 72 G1-resistant materials were located on different branches of the phylogenetic tree using 6000 randomly selected SNPs (Supplementary Figure S3). However, they clustered into several groups when using the four associated SNP markers, Gm13_29459954_ss715614803, Gm13_29751552_ss715614847, Gm13_30293949_ss715614951, and Gm13_30724301_ss715615024 (Supplementary Figure S4). This indicates that accessions with the same G1 resistance gene clustered together, but there were different G1 resistance genes among these accessions.

3. Discussion

3.1. Phenotypic Evaluation and Genetic Diversity of SMV-G1 Resistance

The screening of 218 soybeans identified 72 accessions (33%) exhibiting resistance to SMV-G1 (Figure 1, Supplementary Tables S1 and S4), indicating that many SMV-G1-resistant accessions were available in the USDA Soybean Germplasm Collections. The 72 resistant accessions were distributed across all branches of the phylogenetic tree of the 218 tested soybean accessions (Supplementary Figure S3). Phylogenetic analysis revealed broad genetic diversity among SMV-resistant accessions, with distinct clustering patterns reflecting geographic origins. Notably, the accessions from China exhibited more variation in genetic backgrounds.

The 72 accessions showed differential reaction patterns for resistance to SMV strains from G1 to G7 in previous studies [2,7,8,20]. Of these, 58 accessions have been reported to carry one gene at the *Rsv1* locus for SMV-G1 resistance (Supplementary Table S5). Further GWAS is ongoing to explore these additional resistance factors across multiple SMV strains.

3.2. Genome-Wide Association Study

The GWAS identified four high-confidence SNPs (LOD > 6.5) co-localizing with the *Rsv1* locus on chromosome 13 (29.46–30.73 Mb). Three other SNP markers (Gm13_29751552_ss715614847, Gm13_30293949_ss715614951, and Gm13_30724301_ss715615024) flanked SSR markers (BARCSOYSSR_13_1114, BARCSOYSSR_13_1128, and BARCSOYSSR_13_1155) previously linked to *Rsv1*-mediated resistance [25–27]. Yang et al. [26] mapped two *Rsv1*-associated candidate genes between BARCSOYSSR_13_1128/1136 (SC3, SC6, SC17 resistance) and BARCSOYSSR_13_1140/1155 (SC7 resistance), while Ma et al. [27] delineated *Rsv1-h* within BARCSOYSSR_13_1114/1115. These SSR markers have commonly verified *Rsv1* as the primary SMV G1 resistance locus.

A phylogenetic tree showed that SMV-G1-resistant accessions cluster together in distinct groups, separate from susceptible accessions. Phylogenetic clustering of resistant

accessions suggested that the identified SNP markers could be used to verify new resistant accessions (Supplementary Figure S4). However, incomplete separation of resistant and susceptible subgroups indicated that these resistant accessions have likely carried additional resistance genes beyond the *Rsv1* alleles. Further analysis is ongoing to identify additional resistance genes for SMV resistance.

3.3. Functional Characterization of Candidate Genes

Four NLR-class genes were prioritized within the 873 kb QTL interval: *Glyma.13G184800*, *Glyma.13G184900*, *Glyma.13G194100*, and *3Gg2*. *Glyma.13G184800*. *Glyma.13G184900* encoded CC-NBS-LRR proteins, one of the largest resistance gene families in plants [28,29]. *Glyma.13G194100* contained an NB-ARC domain for regulating the activity of resistance protein [30]. *3Gg2* was a candidate gene for *Rsv1*, encoding an NBS-LRR similar resistance protein [6,31]. Additionally, using *Rsv1*-f/r, a pair of DNA primers from the *3Gg2* gene sequence, and conducting PCR in 101 soybean genotypes and an F₂ population segregated at the *Rsv1* locus made it possible to distinguish resistant lines with the *Rsv1* gene from lines with other SMV resistance genes, thereby identifying the *3Gg2* gene at the *Rsv1* locus [20].

In conclusion, the four candidate genes are involved in the translation of plant resistance proteins or the encoding of similar resistance proteins at the *Rsv1* locus, providing a robust tool for marker-assisted pyramiding.

3.4. Genomic Prediction for Breeding Applications

Comparative analysis of GP models revealed Bayesian approaches (BA, BB, BL, BRR) All five models exhibited similar high r-values across all of SNP sets and outperformed rrBLUP ($r \approx 0.53$ – 0.80), with three GWAS-derived SNP sets ($r \approx 0.60$ – 0.83) achieving higher accuracy than seven random SNP sets ($r \approx 0.10$ – 0.58) using the same GP model (Table 2; Figure 5).

The scale of SNP sets could also be an important factor influencing GP [32,33]. Through an analysis of seven randomly selected SNP sets ranging from 4 to 10,000 SNPs using five models, we observed that the average r-value decreased significantly with smaller SNP sets. SNP sets with fewer than 2000 randomly selected SNPs may not be suitable for GS related to resistance to SMV strain G1. This finding aligns with Shi et al. [34], who observed similar trends in their study. They used nine randomly selected SNP sets ranging from 9 to 4836 SNPs to conduct cross-prediction in 346 USDA spinach accessions for white rust resistance and observed that the r-value increased from 0.25 to 0.79 as the number of randomly selected SNPs increased from 9 to 4836.

These results suggest that the SMV-G1 resistance in soybean is primarily controlled by major gene(s) or QTLs, indicating that selecting SNPs through GWAS can be highly effective.

4. Materials and Methods

4.1. Plant Material

A diversity panel of 218 soybean accessions from the USDA Germplasm Collection representing 25 geographic regions, including 1 accession of unknown origin, was used for this study (Supplementary Table S1).

4.2. Phenotyping and Genotyping

SMV-G1 resistance data were curated from published screens [7,35] and the USDA GRIN database (<https://npgsweb.ars-grin.gov/gringlobal/descriptordetail?id=51159>, accessed on 24 February 2024) [7,35]. The accessions were genotyped with Soy50K SNP Infinium Chips [36]. All 36,431 SNPs across 218 soybean accessions were downloaded from

SoyBase (<https://www.soybase.org/tools/snp50k/>, accessed on 24 February 2024) and distributed across all 20 chromosomes of the soybean genome (Supplementary Figure S5) [37].

Virus inoculum was prepared by homogenizing systemically infected leaves from virus-maintaining stock plants in 0.05 M potassium phosphate buffer (pH 7.2) at an approximate dilution of 1:10 (*w/v*) using a mortar and pestle. Inoculation was performed by rubbing the inoculum with a pestle onto both unifoliolate leaves that had been previously dusted with carborundum [7,35,38]. Plants of each accession were monitored for SMV symptoms regularly and their reactions were classified according to six types: ER, N, N/R, R, S, and S/R (described above). We used only 218 accessions, consisting of 72 resistant accessions and 146 susceptible accessions, for GWAS and GP in this study.

4.3. Principal Component Analysis and Genetic Diversity

Population stratification was assessed via PCA, neighbor-joining phylogeny, and kinship analysis in GAPIT3 [39].

PCA and genetic diversity were analyzed with within R software version 4.3.1 (<https://www.r-project.org/>, accessed on 1 January 2024) [39], with PCA components set from 2 to 10 and the NJ tree settings from 2 to 10 [39,40]. Phylogenetic trees were drawn using the neighbor-joining method in R package GAPIT3 (GAPIT v.3.1.0).

Genetic diversity was assessed for (1) the 218 accessions and (2) the SMV G1-resistant accessions using 6000 randomly selected SNPs from total 36,431 and using GWAS-derived SNP markers associated with SMV resistance. The phylogenetic trees were generated using MEGA 7 [41] based on the maximum likelihood method with the parameters described in Shi et al. [32,42].

In the multi-model GWAS (MLM, MLMM, FarmCPU, BLINK), a Bonferroni-corrected threshold ($p < 1.37 \times 10^{-6}$; LOD = 5.84), was employed for significance.

4.4. Candidate Gene Annotation

Following the methodology outlined by Zhang et al., candidate genes within 5 kb on both sides of significant SNPs were annotated using Glycine max Wm82.a2.v1 (https://phytozome-next.jgi.doe.gov/info/Gmax_Wm82_a2_v1, accessed on 2 February 2024) [43].

4.5. Genomic Prediction

Three methods in the GAPIT3 packages (maGLUP, gBLUP, and sBLUP) were employed to predict GEMV, utilizing all 34,631 SNPs within R software [40]. Then, GEMV prediction was executed using a set of 1000 randomly selected SNPs and two GWAS-derived SNP marker sets (3 and 22 markers—M3 and m22) for the resistance to SMV-G1 in the panel, estimated using four Bayesian models GP models (BA, BB, BL, and BRR) in the BGLR package and rrBLUP in the rrBLUP packages [44,45]. Subsequently, the training and validation sets were randomly created 100 times, and r-value was estimated each time. The average r-value across the 100 iterations was calculated for the SMV-G1 resistance [46]. For both GP scenarios, a higher r-value indicates higher prediction accuracy and better selection efficiency in GS, reflecting the reliability of the genomic predictions for SMV-G1 resistance [47,48].

5. Conclusions

This genome-wide association study and genomic prediction analysis using 36,431 SNPs delineates a 1.27 Mb QTL on chromosome 13 harboring four key SNP markers (Gm13_29459954_ss715614803, Gm13_29751552_ss715614847, Gm13_30293949_ss715614951, and Gm13_30724301_ss715615024) and four SMV-G1-resistant candidate genes (*Glyma.13G184800*, *Glyma.13G184900*, *Glyma.13G194100*, and *3Gg2*) underlying SMV-G1 resistance using 218 soybean germplasm accessions from the USDA Germplasm Collection. High genomic prediction accuracy was observed using three GWAS-derived SNP sets. This

study provides valuable gene information for selecting SMV-G1-resistant plants and lines in soybean breeding and enables precision pyramiding of *Rsv1* alleles with complementary resistance loci to accelerate the development of durable SMV-G1-resistant cultivars.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/ijms26052106/s1>.

Author Contributions: Formal analysis, A.S.; data curation, D.H., Q.S. and D.L.; software, A.S.; validation, Q.S. and A.S.; writing—original draft preparation, D.H.; writing—review and editing, D.H., X.W., Z.L., Q.Y. and X.S.; investigation, D.L.; conceptualization, supervision, L.Y.; project administration, X.W.; funding acquisition, L.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This study is supported by the China Agriculture Research System of MOF and MARA (CARS-04-PS06), Hebei Agriculture Research System (HBCT2023040101), HAAFS Agriculture Science and Technology Innovation Project (2022KJCXZX-LYS-7), and HAAFS Agriculture Science and Technology Innovation Project (2023KJCXZX-LYS-23) and Scientific and Technological Innovation Talent Team Construction (C24R0302; C24R0308).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Phenotype data of 218 soybean accessions can be accessed through the USDA GRIN website (<https://npgsweb.ars-grin.gov/gringlobal/descriptordetail?id=51159>, accessed on 24 February 2024). Genotype data of 36,431 SNPs can be accessed through SoyBase (<https://www.soybase.org/tools/snp50k/>, accessed on 24 February 2024). Phenotype data are available in Supplementary Table S1. Imputed genotype data are available in Supplementary Table S2. Supplementary Table S6 lists 36,431 SNP alleles, chromosome, positions, and genotypes for different soybean accessions.

Conflicts of Interest: The authors declare no conflicts of interest.

References

- Bandara, A.Y.; Weerasooriya, D.K.; Bradley, C.A.; Allen, T.W.; Esker, P.D. Dissecting the economic impact of soybean diseases in the United States over two decades. *PLoS ONE* **2020**, *15*, e0231141. [CrossRef] [PubMed]
- Usovsky, M.; Chen, P.; Li, D.; Wang, A.; Shi, A.; Zheng, C.; Shakiba, E.; Lee, D.; Canella Vieira, C.; Lee, Y.C.; et al. Decades of genetic research on soybean mosaic virus resistance in soybean. *Viruses* **2022**, *14*, 1122. [CrossRef]
- Cho, E.K.; Goodman, R.M. Strains of soybean mosaic virus: Classification based on virulence in resistant soybean cultivars. *Phytopathology* **1979**, *69*, 467–470. [CrossRef]
- Pu, Y.; Yan, R.; Jia, D.; Che, Z.; Yang, R.; Yang, C.; Wang, H.; Cheng, H.; Yu, D. Identification of soybean mosaic virus strain SC7 resistance loci and candidate genes in soybean [*Glycine max* (L.) Merr.]. *Mol. Genet. Genom.* **2024**, *299*, 54. [CrossRef]
- Hajimorad, M.R.; Eggenberger, A.L.; Hill, J.H. Loss and Gain of Elicitor Function of Soybean Mosaic Virus G7 Provoking *Rsv1*-Mediated Lethal Systemic Hypersensitive Response Maps to P3. *J. Virol.* **2005**, *79*, 1215–1222. [CrossRef]
- Hayes, A.J.; Jeong, S.C.; Gore, M.A.; Yu, Y.G.; Buss, G.R.; Tolin, S.A.; Maroof, M.A. Recombination within a nucleotide-binding-site/leucine-rich-repeat gene cluster produces new variants conditioning resistance to soybean mosaic virus in soybeans. *Genetics* **2004**, *166*, 493–503. [CrossRef] [PubMed]
- Zhang, Y.; Du, H.; Zhao, T.; Liao, C.; Feng, T.; Qin, J.; Liu, B.; Kong, F.; Che, Z.; Chen, L. *GmTOC1b* negatively regulates resistance to Soybean mosaic virus. *Crop J.* **2023**, *11*, 1762–1773. [CrossRef]
- Wu, M.; Wu, W.P.; Liu, C.C.; Liu, Y.N.; Wu, X.Y.; Ma, F.F.; Zhu, A.Q.; Yang, J.Y.; Wang, B.; Chen, J.Q. A bean common mosaic virus (BCMV)—Resistance gene is fine-mapped to the same region as *Rsv1-h* in the soybean cultivar Suweon 97. *Theor. Appl. Genet.* **2018**, *131*, 1851–1860. [CrossRef] [PubMed]
- Tran, P.-T.; Widyasari, K.; Seo, J.-K.; Kim, K.-H. Isolation and validation of a candidate *Rsv3* gene from a soybean genotype that confers strain-specific resistance to soybean mosaic virus. *Virology* **2018**, *513*, 153–159. [CrossRef]
- Ishibashi, K.; Saruta, M.; Shimizu, T.; Shu, M.; Anai, T.; Komatsu, K.; Yamada, N.; Katayose, Y.; Ishikawa, M.; Ishimoto, M.; et al. Soybean antiviral immunity conferred by dsRNase targets the viral replication complex. *Nat. Commun.* **2019**, *10*, 4033. [CrossRef] [PubMed]

11. Che, Y.; Gribenko, A.V.; Song, X.; Handke, L.D.; Efferen, K.S.; Tompkins, K.; Kodali, S.; Nunez, L.; Prasad, A.K.; Phelan, L.M.; et al. Rational design of a highly immunogenic prefusion-stabilized F glycoprotein antigen for a respiratory syncytial virus vaccine. *Sci. Transl. Med.* **2023**, *15*, eade6422. [[CrossRef](#)] [[PubMed](#)]
12. Kim, D.; Lee, J.Y.; Yang, J.S.; Kim, J.W.; Kim, V.N.; Chang, H. The architecture of SARS-CoV-2 transcriptome. *Cell* **2020**, *181*, 914–921.e910. [[CrossRef](#)] [[PubMed](#)]
13. Takahashi, K.; Tanabe, K.; Ohnuki, M.; Narita, M.; Ichisaka, T.; Tomoda, K.; Yamanaka, S. Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* **2007**, *131*, 861–872. [[CrossRef](#)]
14. Yan, H.; Wang, H.; Cheng, H.; Hu, Z.; Chu, S.; Zhang, G.; Yu, D. Detection and fine-mapping of SC7 resistance genes via linkage and association analysis in soybean. *J. Integr. Plant Biol.* **2015**, *57*, 722–729. [[CrossRef](#)]
15. Seo, J.K.; Lee, S.H.; Kim, K.H. 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.* **2009**, *22*, 1151–1159. [[CrossRef](#)] [[PubMed](#)]
16. Che, Z.; Yan, H.; Liu, H.; Yang, H.; Du, H.; Yang, Y.; Liu, B.; Yu, D. Genome-wide association study for soybean mosaic virus SC3 resistance in soybean. *Mol. Breed.* **2020**, *40*, 69. [[CrossRef](#)]
17. Tibbs Cortes, L.; Zhang, Z.; Yu, J. Status and prospects of genome-wide association studies in plants. *Plant Genome* **2021**, *14*, e20077. [[CrossRef](#)] [[PubMed](#)]
18. Zhang, C.; Grosic, S.; Whitham, S.A.; Hill, J.H. The requirement of multiple defense genes in soybean *Rsv1*-mediated extreme resistance to soybean mosaic virus. *Mol. Plant Microbe Interact.* **2012**, *25*, 1307–1313. [[CrossRef](#)] [[PubMed](#)]
19. Chen, H.; Adam Arsovski, A.; Yu, K.; Wang, A. Deep sequencing leads to the identification of eukaryotic translation initiation factor 5A as a key element in *Rsv1*-mediated lethal systemic hypersensitive response to Soybean mosaic virus infection in soybean. *Mol. Plant Pathol.* **2017**, *18*, 391–404. [[CrossRef](#)]
20. Gore, M.A.; Hayes, A.J.; Jeong, S.C.; Yue, Y.G.; Buss, G.R.; Maroof, S. Mapping tightly linked genes controlling potyvirus infection at the *Rsv1* and *Rpv1* region in soybean. *Genome* **2002**, *45*, 592–599. [[CrossRef](#)]
21. Hajimorad, M.R.; Hill, J.H. *Rsv1*-mediated resistance against soybean mosaic virus-N is hypersensitive response-independent at inoculation site, but has the potential to initiate a hypersensitive response-like mechanism. *Mol. Plant Microbe Interact.* **2001**, *14*, 587–598. [[CrossRef](#)]
22. Maroof, M.A.S.; Tucker, D.M.; Skoneczka, J.A.; Bowman, B.C.; Tripathy, S.; Tolin, S.A. Fine mapping and candidate gene discovery of the soybean mosaic virus resistance gene, *Rsv4*. *Plant Genome* **2010**, *3*, 14–22. [[CrossRef](#)]
23. Liao, L.; Chen, P.; Rajcan, I.; Buss, G.R.; Tolin, S.A. Genetic analysis of “8101” soybean containing three genes for resistance to soybean mosaic virus. *Crop Sci.* **2011**, *51*, 503–511. [[CrossRef](#)]
24. Chowda-Reddy, R.V.; Sun, H.; Hill, J.H.; Poysa, V.; Wang, A. Simultaneous mutations in multi-viral proteins are required for soybean mosaic virus to gain virulence on soybean genotypes carrying different R genes. *PLoS ONE* **2011**, *6*, e28342. [[CrossRef](#)]
25. Chang, H.-X.; Lipka, A.E.; Domier, L.L.; Hartman, G.L. Characterization of disease resistance loci in the USDA soybean germplasm collection using genome-wide association studies. *Phytopathology* **2016**, *106*, 1139–1151. [[CrossRef](#)]
26. Yang, Y.; Zheng, G.; Han, L.; Dagang, W.; Yang, X.; Yuan, Y.; Huang, S.; Zhi, H. Genetic analysis and mapping of genes for resistance to multiple strains of Soybean mosaic virus in a single resistant soybean accession PI 96983. *Theor. Appl. Genet.* **2013**, *126*, 1783–1791. [[CrossRef](#)] [[PubMed](#)]
27. Ma, F.; Wu, X.; Chen, Y.-X.; Liu, Y.-N.; Shao, Z.-Q.; Wu, P.; Wu, M.; Liu, C.-C.; Wu, W.-P.; Yang, J.-Y.; et al. Fine mapping of the *Rsv1-h* gene in the soybean cultivar Suweon 97 that confers resistance to two Chinese strains of the soybean mosaic virus. *Theor. Appl. Genet.* **2016**, *129*, 2227–2236. [[CrossRef](#)]
28. Liu, Y.; Li, D.; Yang, N.; Zhu, X.; Han, K.; Gu, R.; Bai, J.; Wang, A.; Zhang, Y. Genome-wide identification and analysis of CC-NBS-LRR family in response to downy mildew and black rot in chinese cabbage. *Int. J. Mol. Sci.* **2021**, *22*, 4266. [[CrossRef](#)]
29. McHale, L.; Tan, X.; Koehl, P.; Micheltore, R.W. Plant NBS-LRR proteins: Adaptable guards. *Genome Biol.* **2006**, *7*, 212. [[CrossRef](#)] [[PubMed](#)]
30. van Ooijen, G.; Mayr, G.; Kasiem, M.M.; Albrecht, M.; Cornelissen, B.J.; Takken, F.L. Structure-function analysis of the NB-ARC domain of plant disease resistance proteins. *J. Exp. Bot.* **2008**, *59*, 1383–1397. [[CrossRef](#)] [[PubMed](#)]
31. Hong, X.; Li, S.; Cheng, X.; Zhi, H.; Yin, J.; Xu, K. Searching for plant NLR immune receptors conferring resistance to potyviruses. *Crop J.* **2024**, *12*, 28–44. [[CrossRef](#)]
32. Shi, A.; Gepts, P.; Song, Q.; Xiong, H.; Michaels, T.E.; Chen, S. Genome-wide association study and genomic prediction for soybean cyst nematode resistance in USDA common bean (*Phaseolus vulgaris*) core collection. *Front. Plant Sci.* **2021**, *12*, 624156. [[CrossRef](#)]
33. Wen, L.; Chang, H.X.; Brown, P.J.; Domier, L.L.; Hartman, G.L. Genome-wide association and genomic prediction identifies soybean cyst nematode resistance in common bean including a syntenic region to soybean *Rhg1* locus. *Hortic. Res.* **2019**, *6*, 9. [[CrossRef](#)]

34. Shi, A.; Bhattarai, G.; Xiong, H.; Avila, C.A.; Feng, C.; Liu, B.; Joshi, V.; Stein, L.; Mou, B.; du Toit, L.J.; et al. Genome-wide association study and genomic prediction of white rust resistance in USDA GRIN spinach germplasm. *Hortic. Res.* **2022**, *9*, uhac069. [[CrossRef](#)] [[PubMed](#)]
35. Cho, E.K.; Goodman, R.M. Evaluation of resistance in soybeans to soybean mosaic virus strains. *Crop Sci.* **1982**, *22*, 1133–1136. [[CrossRef](#)]
36. Song, Q.; Hyten, D.L.; Jia, G.; Quigley, C.V.; Fickus, E.W.; Nelson, R.L.; Cregan, P.B. Development and evaluation of SoySNP50K, a high-density genotyping array for soybean. *PLoS ONE* **2013**, *8*, e54985. [[CrossRef](#)] [[PubMed](#)]
37. Song, Q.; Hyten, D.L.; Jia, G.; Quigley, C.V.; Fickus, E.W.; Nelson, R.L.; Cregan, P.B. Fingerprinting soybean germplasm and its utility in genomic research. *G3 Genes Genomes Genet.* **2015**, *5*, 1999–2006. [[CrossRef](#)]
38. Chen, P.; Buss, G.R.; Roane, C.W.; Tolin, S.A. Allelism among genes for resistance to soybean mosaic virus in strain-differential soybean cultivars. *Crop Sci.* **1991**, *31*, 305–309. [[CrossRef](#)]
39. Wang, J.; Zhang, Z. GAPIT version 3: Boosting power and accuracy for genomic association and prediction. *Genom. Proteom. Bioinform.* **2021**, *19*, 629–640. [[CrossRef](#)] [[PubMed](#)]
40. Meher, P.; Sachin, R.; Anuj, K. Performance of Bayesian and BLUP alphabets for genomic prediction: Analysis, comparison and results. *Heredity* **2022**, *128*, 519–530. [[CrossRef](#)]
41. Kumar, S.; Stecher, G.; Tamura, K. MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* **2016**, *33*, 1870–1874. [[CrossRef](#)] [[PubMed](#)]
42. Shi, A.; Chen, P.; Vierling, R.; Li, D.; Zheng, C. Identification of soybean mosaic virus resistance alleles in Jindou 1 soybean. *Euphytica* **2013**, *192*, 181–187. [[CrossRef](#)]
43. Zhang, H.; Li, C.; Davis, E.L.; Wang, J.; Griffin, J.D.; Kofsky, J.; Song, B.H. Genome-wide association study of resistance to soybean cyst nematode (*Heterodera glycines*) HG type 2.5.7 in wild soybean (*Glycine soja*). *Front. Plant Sci.* **2016**, *7*, 1214. [[CrossRef](#)] [[PubMed](#)]
44. Pérez, P.; de los Campos, G. Genome-wide regression and prediction with the BGLR statistical package. *Genetics* **2014**, *198*, 483–495. [[CrossRef](#)]
45. Jarquín, D.; Kocak, K.; Posadas, L.; Hyma, K.; Jedlicka, J.; Graef, G.; Lorenz, A. Genotyping by sequencing for genomic prediction in a soybean breeding population. *BMC Genom.* **2014**, *15*, 740. [[CrossRef](#)] [[PubMed](#)]
46. Suh, S.J.; Bowman, B.C.; Jeong, N.; Yang, K.; Kastl, C.; Tolin, S.A.; Maroof, M.A.S.; Jeong, S.-C. The *Rsv3* locus conferring resistance to soybean mosaic virus is associated with a cluster of coiled-coil nucleotide-binding leucine-rich repeat genes. *Plant Genome* **2011**, *4*, 55–64. [[CrossRef](#)]
47. Krishnappa, G.; Savadi, S.; Tyagi, B.S.; Singh, S.K.; Mamrutha, H.M.; Kumar, S.; Mishra, C.N.; Khan, H.; Gangadhara, K.; Uday, G.; et al. Integrated genomic selection for rapid improvement of crops. *Genomics* **2021**, *113*, 1070–1086. [[CrossRef](#)]
48. Qu, S.; Lu, S.; Liu, Y.; Li, M.; Chen, S. Accurate genomic selection using low-density SNP panels preselected by maximum likelihood estimation. *Aquaculture* **2024**, *579*, 740154. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.