ORIGINAL ARTICLE

Genome resequencing reveals genetic loci and genes conferring resistance to SMV-SC8 in soybean

Jiahao Chu[1] · Wenlong Li[1] · Zhanwu Yang[1] · Zhenqi Shao[1] · Hua Zhang[1] · Shaoda Rong[1] · Youbin Kong[1] · Hui Du[1] · Xihuan Li[1] · Caiying Zhang[1]

Abstract

Key message **A soybean natural population genotyped by resequencing and another RIL population genotyped by SoySNP6K were used to explore consistent genetic loci and genes under greenhouse- and field-conditions for SMV-SC8 resistance.**

Abstract *Soybean mosaic virus* (SMV) is a member of the genus *Potyvirus* that occurs in all soybean-growing areas of the world and causes serious losses of yield and seed quality. In this study, a natural population composed of 209 accessions resequenced at an average depth of $18.44 \times$ and another RIL population containing 193 lines were used to explore genetic loci and genes conferring resistance to SMV-SC8. There were 3030 SNPs significantly associated with resistance to SC8 on chromosome 13 in the natural population, among which 327 SNPs were located within an ~0.14 Mb region (from 28.46 to 28.60 Mb) of the major QTL *qRsc8F* in the RIL population. Two genes among 21 candidate genes, *GmMACPF1* and *GmRad60*, were identified in the region of consistent linkage and association. Compared to the mock control, the changes in the expression of these two genes after inoculation with SC8 differed between resistant and susceptible accessions. More importantly, *GmMACPF1* was shown to confer resistance to SC8 by significantly decreasing virus content in soybean hairy roots overexpressing this gene. A functional marker, FMSC8, was developed based on the allelic variation of *GmMACPF1*, and a high coincidence rate of 80.19% between the disease index and marker genotype was identified in the 419 soybean accessions. The results provide valuable resources for studies on the molecular mechanism of SMV resistance and genetic improvement in soybean.

Introduction

Soybean [*Glycine max* (L.) Merr.] is a main source of edible vegetable oil and protein around the world. Many diseases pose a serious threat to soybean production (Hartman et al.

Communicated by Henry T. Nguyen.

Jiahao Chu and Wenlong Li authors contributed equally to this work.

✉ Xihuan Li
    lixihuan@hebau.edu.cn

✉ Caiying Zhang
    zhangcaiying@hebau.edu.cn

[1] State Key Laboratory of North China Crop Improvement and Regulation, North China Key Laboratory for Crop Germplasm Resources of Education Ministry, Key Laboratory for Crop Germplasm Resources of Hebei, Hebei Agricultural University, Baoding, China

2011), among which *Soybean mosaic virus* (SMV) is an important viral pathogen that causes significant yield loss and quality deterioration (Malapi-Nelson et al. 2009). The management of SMV is limited to the use of good agricultural practices and the development of disease-resistant varieties through conventional and molecular breeding (Galvez et al. 2014). To achieve the goal of controlling SMV, it is necessary to understand the genetic basis of soybean resistance to SMV.

Due to the long-term coevolution between SMV and its host, numerous SMV strains with diverse virulence characteristics have arisen worldwide. According to the pathogenicity differences among diverse hosts, SMV has been divided into different strains in different countries, including seven strains in the USA (Cho et al. 1979), five strains in Japan (Takahashi et al. 1980), 22 strains in China and six strains in the Republic of Korea (Cho et al. 1983; Kim et al. 1991; Kim et al. 2003). Some genetic loci for resistance to these SMV strains have been found, including the four

dominant disease resistance loci *Rsv1*, *Rsv3*, *Rsv4* and *Rsv5* identified in the United States (Shi et al. 2008, 2013). *Rsv1* on chromosome 13 in accession PI96983 confers resistance to SMV strains G1~G6 but does not prevent susceptibility to the G7 strain (Chen et al. 1991). *Rsv3* on chromosome 14 in accession L29 is associated with susceptibility to strains G1~G4 but resistance to strains G5~G7 (Ma et al. 2002). *Rsv4* on chromosome 2 in accession V94-5152 confers perfect resistance to strains G1~G7. A new *Rsv5* locus on chromosome 13 from soybean 'York' was assigned as the resistance gene replacing the original *Rsv1-y* allele (Klepadlo et al. 2017).

In China, 22 strains (SC1~SC22) were identified according to the pathogenicity differences in ten hosts from soybean planting regions worldwide, and SC8 was identified as one of the predominant SMV strains in the Yangtze River valley and the Huang-Huai-Hai soybean-growing eco-region (Guo et al. 2005; Li et al. 2010; Yang et al. 2014). However, there have been few reports on SMV-SC8 resistance loci and genes until now. Wang et al. (2011a) constructed F_2, $F_{2:3}$ and $F_{7:11}$ recombinant inbred line (RIL) populations to study the inheritance of SC8 resistance and mapped one resistance locus. Two simple sequence repeat (SSR) markers were identified as flanking markers of QTLs for resistance to SC8. Zhao et al. (2016) generated an F_2 population to pinpoint the SC8 resistance gene, and it was narrowed to a 30.8-kb region containing two annotated genes. Recently, the MADS-box transcription factor GmCAL was identified as a positive regulatory factor conferring soybean resistance to diverse SMV strains (Ren et al. 2021).

To date, only a few studies have identified QTLs associated with SC8 resistance in soybean, and all of these QTLs were derived from family-based segregating populations based on SSR and SNP-chip markers. The QTLs and candidate genes identified via this strategy are usually restricted due to the narrow range of genetic variation and large mapping region. Genome-wide association study (GWAS) combined with a resequencing strategy could overcome this limitation resulting from multiple allelic variations in the natural population (Yan et al. 2017; Zhang et al. 2017, 2018; Li et al. 2018).

In our present work, a genome-wide association study was performed in a natural population containing 209 accessions showing SC8 different resistance phenotypes in different environments that were resequenced with an ~20×depth. QTL mapping for SC8 resistance was also performed in an RIL population including 193 lines based on SoySNP6K array genotypes. The purpose of this study was to identify stable genetic loci, candidate genes and functional markers significantly associated with resistance to SC8 and to provide elite germplasms with more favourable alleles and higher resistance to SC8 for the genetic improvement of resistance in soybean varieties.

Materials and methods

Plant materials and SMV strain

A natural population composed of 209 soybean accessions (50 landraces and 159 cultivars) (Table S1) and an RIL population ($F_{6:8}$, 193 lines) derived from a cross of the susceptible accession C813 and the resistant accession KN7 (Chu et al. 2021) were used for experiments conducted under greenhouse and field conditions. SC8, a moderately virulent SMV strain, was provided by the National Center for Soybean Improvement, Nanjing Agricultural University, China.

SC8 inoculation and resistance evaluation

The natural soybean population was planted under three different environments in 2021 and 2022 and designated as E1–E3. E1 and E2 represented plants grown under greenhouse conditions in 2021 and 2022, while E3 represented plants grown under field condition in 2022. The RIL population was planted in two environments in 2022, and were designated as E4–E5, where E4 represented the plants grown under greenhouse condition in 2022, while E5 represented the plants grown under field condition in 2022. In these five experiments conducted in these two populations, a complete randomized block design with three replications was used.

In the greenhouse experiment, seeds of the soybean population were planted in round plastic pots (diameter × depth: 30 cm × 25 cm) filled with vermiculite and nutrient soil (3:1) and grown in an aphid-free greenhouse with 14 h of supplemental light (~10,000 lx) per day at room temperature (26 °C ± 1 °C) at Hebei Agricultural University. Twelve seeds of each accession were planted in one plastic pot, and three pots were included for each accession. Seven days after planting, ten plants were left in each pot.

In the field experiment, six seeds were planted in one porcelain pot (diameter × depth: 50 cm × 45 cm), with two pots per accession, and the plants were grown at Hebei Agricultural University (Baoding City, N 38°10′, E 113°40′) from May to July in 2022. Seven days after planting, five plants were left in each pot.

Fresh leaves of the NN1138-2 soybean accession infected with the SC8 virus were ground in 0.01 mol/L sodium phosphate buffer (approximately 10 mL per g leaf tissue, pH 7.4) with a small amount of carborundum (600 mesh). Seedlings were inoculated with the obtained inoculum by rubbing unfolded primary leaves at the soybean V1 stage and were inoculated again on the first trifoliate

leaf at the V2 stage. The inoculated leaves were rinsed with tap water. In greenhouse and field experiments, resistance to SC8 was evaluated by calculating the disease index (DI) score one month after SC8 inoculation as previously described (Zhi et al. 2005):

$$DI = \left[\sum f_i S_i \middle/ (n \times S_{max}) \right] \times 100$$

where S_i is the disease severity; f_i is the number of plants with S_i; S_{max} is the highest S_i; and n is the total number of surveyed plants.

Statistical analysis of disease index of SC8 infection

The genetic variation analysis of the two populations was performed with SPSS V19.0 software. The broad-sense heritability was estimated based on the arcsine-transformed data according to the method described by Nyquist and Baker (Nyquist et al. 1991). The best linear unbiased prediction (BLUP) value was estimated in R (lme4 package) (Che et al. 2017).

Resequencing of the natural population

Sequence alignment and SNP calling

Genomic DNAs of the accessions were extracted via the modified CTAB method, and approximately 350-bp libraries were constructed for resequencing. The low-quality paired reads were filtered according to criteria similar to those applied in a previous study, and the high-quality reads were aligned to the Williams82 reference genome (Wm82a4v1) with BWA software. SNP calling was performed with the Genome Analysis Toolkit (GATK, version v3.1), and the filtration parameters were set as in a previous study (Shao et al. 2022). SNPs were annotated according to Williams82 (Wm82a4v1) with the package ANNOVAR and were classified into four different types as described in a previous study (Shao et al. 2022).

Population structure and linkage disequilibrium analysis

To analyse the phylogenetic relationships of the natural population, an individual-based NJ tree was constructed based on the P distance with TreeBestv1.9.2 software. The population structure was analysed by the expectation maximization algorithm, as implemented in the Admixture program. The number of assumed genetic clusters (K) was set at 2–8, with 10,000 iterations for each run. PCA was carried out with GCTA software. The linkage disequilibrium (LD) coefficient (r^2) between pairwise high-quality SNPs was determined with Haploview with the parameters (maxdistance 500,

$MAF > 0.05$, miss ≤ 0.2). LD was evaluated according to half of the maximum value at 75 kb for soybean accessions.

GWAS analysis of SC8 resistance

The GWAS analysis of SC8 resistance was conducted by genome-wide efficient mixed-model association (GEMMA) software via a similar method to that described in a previous study (Shao et al. 2022). In the present study, the top three PCs were used to build the S matrix for population-structure correction, and the matrix of simple matching coefficients was used to build the K matrix. According to the physical distance of significant SNPs and LD, the SNPs within the LD range were attributed to the same locus, as in a previous study (Huo et al. 2019).

Regression analysis of associated SNPs

To further verify the associated SNPs obtained in the present study, the soybean accessions were divided into different groups based on the numbers of elite alleles for SNPs with the most significant P value at each locus. Then, regression analysis between the elite allele numbers and disease index was conducted with SPSS V25.0 software.

Linkage mapping of the disease index of SC8 infection in the RIL population

The genotypes of the RIL population were analysed by using the SoySNP6K array based on 5403 SNPs (core subset from the SoySNP50K Illumina Bead Chip array), and the genetic linkage map was constructed by using Joinmap v4.0 software in our previous study (Chu et al. 2021). In the present study, linkage mapping of the disease index of SC8 infection was conducted via the inclusive composite interval mapping method in ICIMapping v4.2 software, similar to our previous description (Chu et al. 2021).

Screening of candidate genes for SC8 resistance in the consistent region

To further reveal candidate genes controlling SC8 resistance in the consistent region of GWAS and linkage mapping, the genes in this region were searched, and the variant types of associated SNPs in these genes were analysed. Meanwhile, the full sequence alignments of these genes between the RIL parents KN7 and C813 were analysed via the resequencing data of these two varieties in our previous study (Sun et al. 2022). Thereafter, the expression of these genes in soybean (transcriptome data of Williams 82 from https://phytozome.jgi.doe.gov) and the positions of variants in gene promoter regions (https://www.dna.affrc.go.jp) as well as the corresponding protein-encoding regions (https://www.

ncbi.nlm.nih.gov) were analysed. Finally, by combining the gene annotations (Wm82.a4.v1, https://www.soybase.org) and SMV resistance-related reports, candidate genes were screened.

qRT−PCR analysis of candidate genes for SC8 resistance

For the expression analysis of candidate genes, the leaves of the resistant accession KN7 and the susceptible accession C813 were sampled at 0, 4, 8, 12, 24, 48, and 72 h post-inoculation (hpi) with SC8. The leaves were also sampled at the same time points after inoculation with sodium phosphate buffer, which was used as a control. Subsequently, real-time qPCR (qRT−PCR) was conducted via a method similar to that applied in our previous study (Kong et al. 2018), with three technical replicates for each sample. The qRT−PCR primers of the candidate genes and the internal control gene are listed in Table S2.

Functional analysis of the candidate genes for SC8 resistance

The open reading frames (ORFs) of the candidate genes (*GmMACPF1* and *GmRad60*, screened later in the present study) were PCR amplified from KN7 leaf cDNA using the primers listed in Table S2. Subsequently, the ORF was inserted into the plant expression vector pCam::GFP (Wu et al. 2014) to construct the recombinant overexpression vectors pCam-*GmMACPF1* and pCam-*GmRad60*, while the empty vector was used as a negative control.

The functions of the *GmMACPF1* and *GmRad60* candidate genes in the SC8 strain were analysed by the "*Agrobacterium rhizogenes*-induced soybean hairy roots versus Soybean mosaic virus (ARISHR-SMV)" method as described in a previous study (Jiang et al. 2019), and the procedure is presented in Fig. S1.

First, the constructed pCam-*GmMACPF1* and pCam-*GmRad60* vectors were transformed into the *Agrobacterium rhizogenes* K599 strain via the electroporation method, and the mature seeds of C813 were surface-sterilized and soaked in sterilized water overnight. Thereafter, the seeds of C813 were placed on agar medium (0.7%) for germination after the seed coats were removed. Subsequently, the cotyledons were harvested and used for *Agrobacterium rhizogenes*-mediated transformation in White's medium. After ~ 10 days, the calli grown from the incubated cotyledon pieces were used for SC8 virus inoculation. After 7 ~ 10 days, the hairy roots from the calli were observed for green fluorescence via fluorescence microscopy. Nearly twenty days later, the transgenic soybean hairy roots were screened and used for the evaluation of SC8 virus accumulation and target gene expression,

with three technical replicates for each sample with five hairy roots per replicate. Moreover, a reverse back-inoculation experiment in which the SC8 virus was applied to NN1138-2 leaves was performed to verify the presence of the virus in the transgenic soybean hairy roots.

Development of the functional gene marker FMSC8

First, primer pairs were designed for the functional marker analysis of SC8 (FMSC8) based on a nonsynonymous mutation (T/C) in the functional gene *GmMACPF1* related to SC8 infection in KN7 and C813 (Table S2). Subsequently, the primers were used to amplify the DNA of these two varieties and 417 other soybean accessions. The KN7 accession genotype was recorded as "T", while the C813 accession was recorded as "C". Next, by combining the phenotypes of these accessions (DI < 15, resistant accession; DI ≥ 15, susceptible accession), the detection rate of FMSC8 was calculated using the following formula: (resistant accessions with T + susceptible accessions with C)/number of detected accessions.

Results

Genetic variation of the disease index of SC8 infection in two soybean populations

The genetic variation and broad heritability of the disease index were analysed in the natural population under three environments (Table 1). The average disease indices of the accessions ranged from 14.54 to 24.04. The natural population showed relatively broad genetic variation to SC8, with a disease indices ranging from 0 to 100 and coefficients of variation (CV) ranging from 129.2 ~ 181.77%. Histogram analysis showed that the segregation of the disease index presented a skewed distribution in the population. In addition, the broad-sense heritability for the disease index was as high as 88.45% in this population, which showed that the disease index could be stably inherited in soybean and that some resistant accessions with a lower disease index might be used as parents in breeding programs.

The comparison of the RIL parents showed that C813 presented a higher disease index (65.91) than KN7 (9.69) under two environments (Table 1). The leaves of C813 were severely curled after inoculation with SC8, while KN7 showed slight necrotic spotting (Fig. S2). The descriptive statistics of the RIL population showed that the disease index ranged from 0 ~ 100, and the CV ranged from 99.09 ~ 125.71%, which indicated wide genetic variation in the disease index in the RIL population (Table 1).

Table 1 Genetic variations of disease index to SMV-SC8 in soybean natural and RIL populations

Type	Environment	RIL parents		Mean	SD	CV(%)	Max	Min	Skew	Kurt	h_B^2(%)
		C813	KN7								
Natural population	E1	–	–	14.54	26.43	181.77	95.31	0.00	1.39	1.67	88.45
	E2	–	–	24.04	31.06	129.20	100.00	0.00	1.07	− 0.37	
	E3	–	–	18.57	28.32	152.50	100.00	0.00	1.31	0.12	
RIL population	E4	75.00	8.67	38.35	38.00	99.09	100.00	0.00	1.00	0.18	79.39
	E5	56.82	10.71	25.98	32.66	125.71	100.00	0.00	− 0.42	− 1.70	

E1 ~ E3: Different environments for natural population. E1: Greenhouse condition in 2021; E2: Greenhouse condition in 2022; E3: Field condition in 2022

E4 ~ E5: Different environments for RIL population. E4: Greenhouse condition in 2021; E5: Field condition in 2022

SD Standard deviation; *CV* Coefficient of variation; *Max* Maximum; *Min* Minimum; Skew: Skewness;

Kurt: Kurtosis; h_B^2: Broad heritability

Genome-wide association analysis of SNPs for SC8 resistance

Approximately 4.45 T bases were obtained from the resequenced natural population, with an average depth of 18.44 ×, as described in our previous study (Shao et al. 2022). Based on the 5,057,494 filtered SNPs, the 209 accessions were clustered into two clades, with 50 landraces in clade I and 159 cultivars in clade II, and the LD decreased to half of its maximum value at 75 kb for the population (Fig. 1).

In total, 5930 SNPs on 18 chromosomes were obtained, among which 3035 SNPs on five chromosomes (Gm01, Gm07, Gm09, Gm12 and Gm13) were associated with resistance to SC8 across three environments and BLUP values simultaneously (Fig. 2, Table S3). The 3035 SNPs were attributed to 36 loci and included one stopgain, two stoploss, 120 nonsynonymous, 89 synonymous, 522 intronic, 71 UTR3, 53 UTR5, 278 upstream, 249 downstream and 1650 intergenic SNPs (Table S4). There were 3030 SNPs at 32 loci located on chromosome 13, accounting for 99.84% of the total SNPs. Importantly, 31 loci were first discovered to confer resistance to SMV-SC8 in our study (Table S4).

Verification of SNPs associated with SC8 resistance

To verify the 3035 significantly associated SNPs identified at 36 loci for SC8 resistance, regression analysis between the numbers of elite alleles and the disease indices of soybean accessions was conducted. The regression analysis showed that the proportion of elite alleles was linearly related to the disease index, with determination coefficients (R^2) of 0.43, 0.44 and 0.36 in the three environments, respectively (Fig. S3), which indicated that SC8 resistance was significantly correlated with the proportion of elite alleles.

Moreover, to provide special germplasms for the genetic improvement of SC8 resistance, the numbers of elite alleles of the most significantly associated SNPs at the 36 loci were analysed in 10 accessions with different disease indices (the top five and bottom five indices in 209 accessions). The results showed that the five resistant accessions possessed 33 ~ 36 elite alleles, while the five susceptible accessions possessed 0 ~ 9 elite alleles (Table S5). Simultaneously, the analysis of differences in the proportion of elite alleles between the landrace and cultivar showed that the cultivar presented a higher proportion of elite alleles (84.71%) and a lower disease index (6.45) than the landrace (proportion = 34.72%, DI = 46.55) (Fig. S3). This result indicated that the cultivar had accumulated more elite genes in the long-term breeding selection process.

QTL mapping of the disease index of SC8 infection in the RIL population

Based on the genetic linkage map of the RIL population with 2234 SNP markers and a total genetic distance of 4229.01 cM in 23 linkage groups constructed in our previous study (Chu et al. 2021), additive QTL mapping for the disease index was conducted in the present study. One major QTL, *qRsc8F* (ss715614742 ~ ss715614764), was found to control the disease index with PVEs of 30.08 ~ 51.17% under all environments; this QTL was located on chromosome 13 (physical position 28.46 Mb ~ 28.60 Mb) (Fig. 2, Table S6). Moreover, based on the two flanking markers, the 193 RILs could be divided into resistant and susceptible types with disease indices of 6.84 vs. 50.13 ($P = 3.34 \times 10^{-4}$) and 9.05 vs. 44.38 ($P = 1.09 \times 10^{-6}$), respectively (Fig. 3). The additive effect of *qRsc8F* was derived from the resistant RIL parent KN7 (Table S6). Importantly, 327 associated SNPs were located within the ~ 141.6-kb interval of the *qRsc8F*,
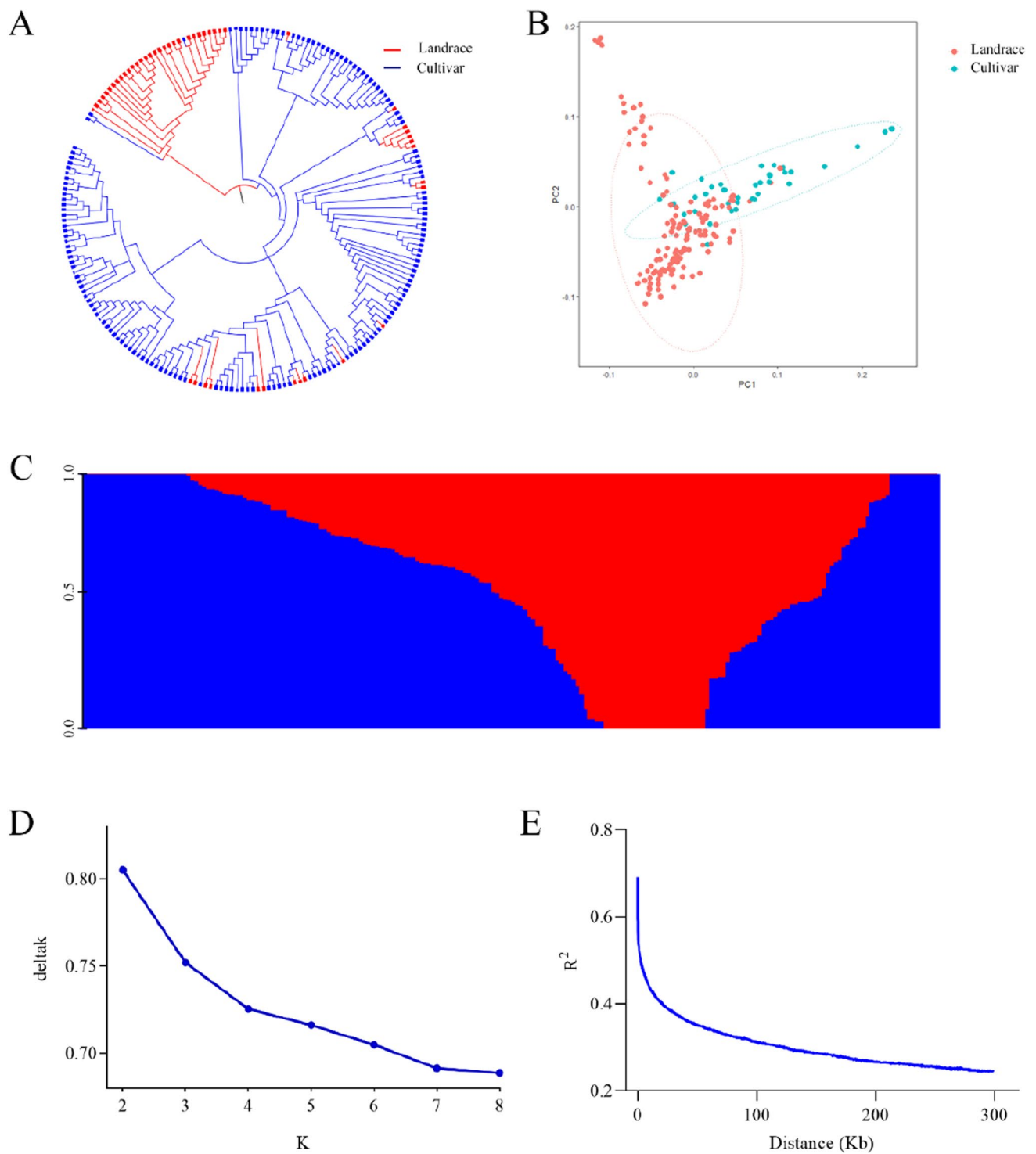
Fig. 1 Phylogenetic tree, PCA, genetic structure and LD decay analysis of soybean natural population. **A** Phylogenetic tree of natural population with 5,053,585 high-quality SNPs. **B** PCA plot of the first two components (PC1 and PC2). **C** Structure analysis with K = 2. The y axis is cluster membership, and x axis is the different accessions. **D** Delta K based on the rate of change of LnP (K) between successive K. **E** Genome-wide average LD decay of 209 accessions
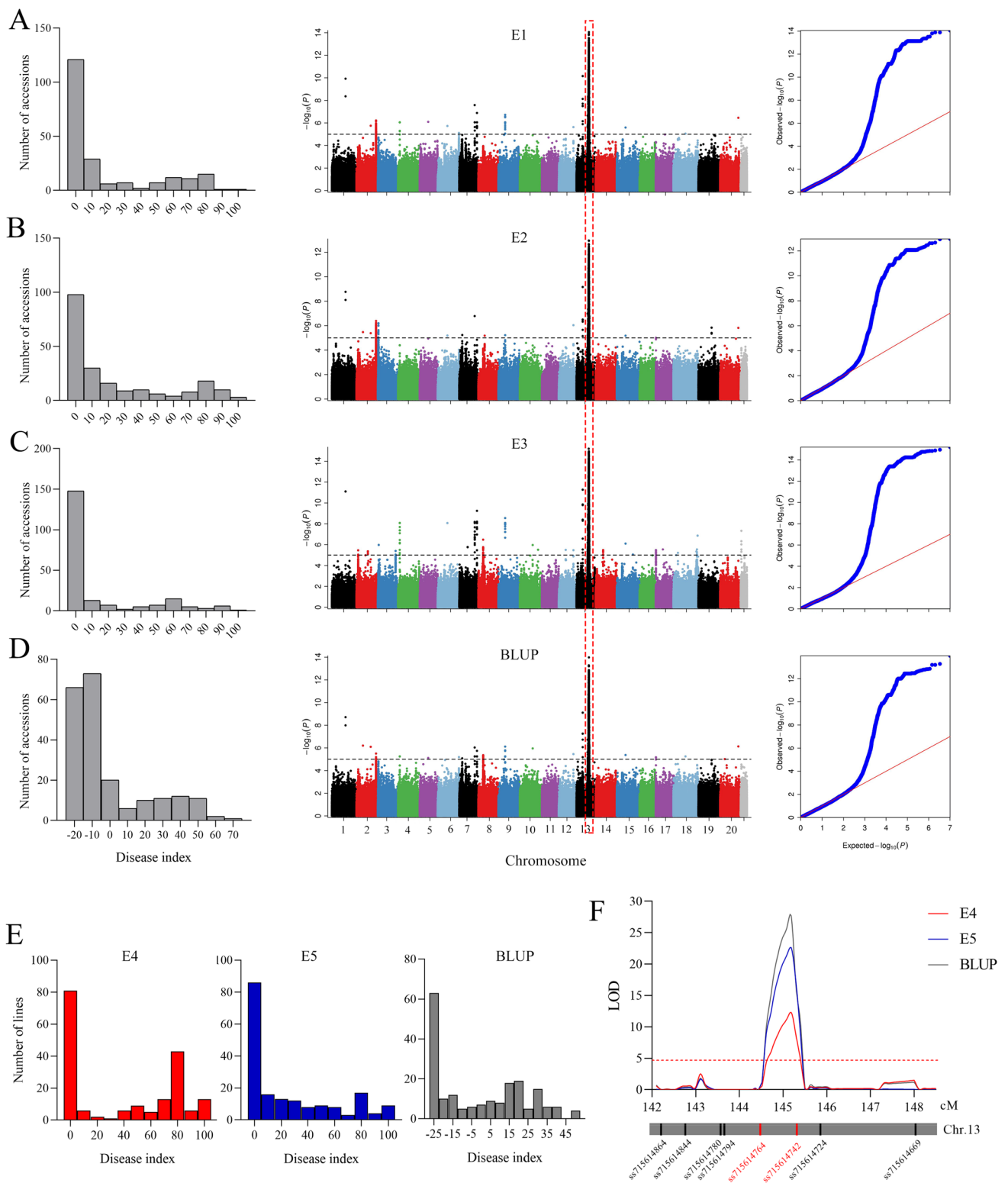
Fig. 2 Association and linkage analysis of disease index to SMV-SC8. **A–D** Distribution, Manhattan and QQ plots for disease index of soybean natural population in three environments (E1 ~ E3) and BLUP-values. **E** Distribution for disease index of soybean RIL pop-ulation in two environments (E4 and E5) and BLUP-values. **F** QTL *qRsc8F* on chromosome 13 for disease index to SC8 in RIL popula-tion
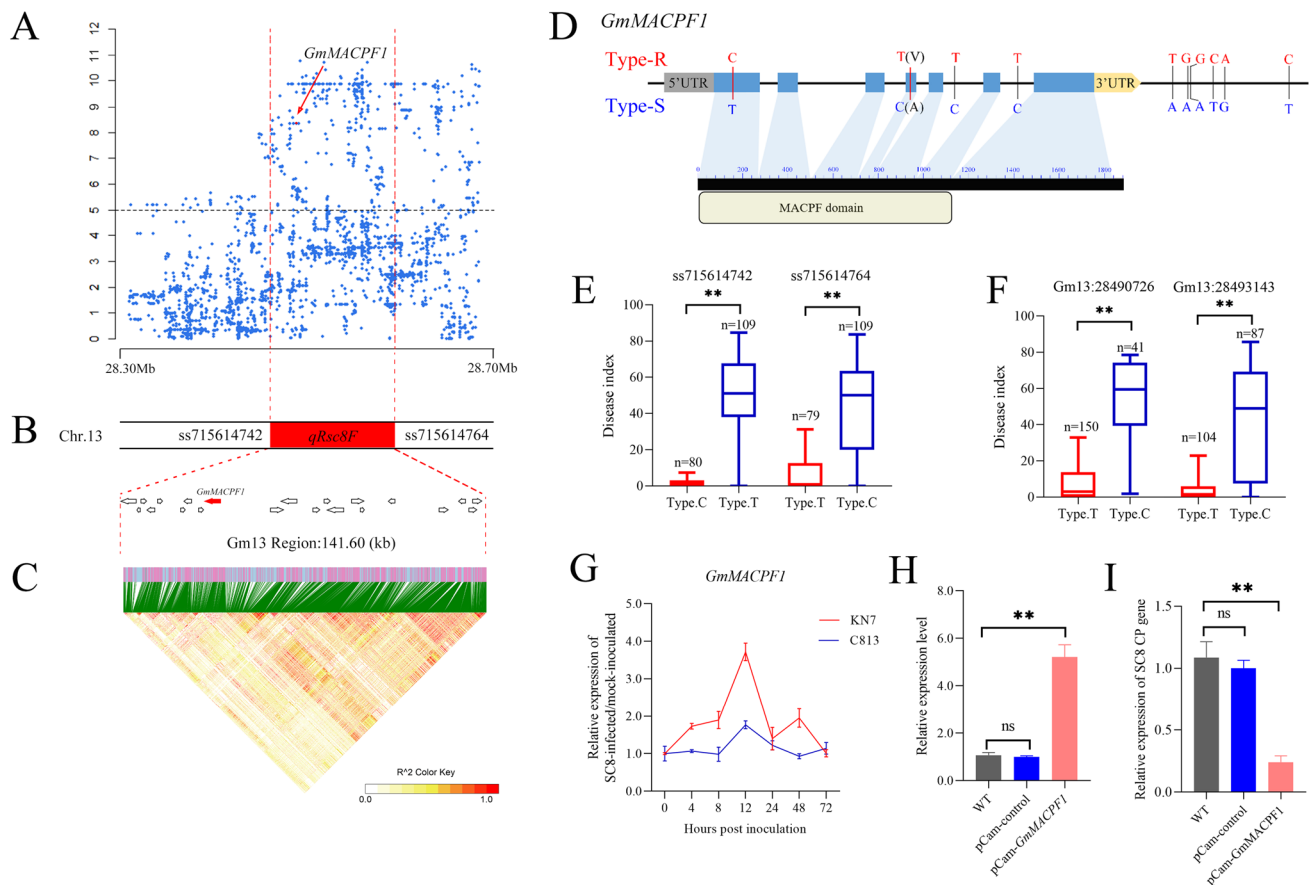
Fig. 3 Identification and function analysis of causal gene *GmMACPF1*. **A** Regional GWAS signals on chromosome 13 for disease index to SC8. The red dashed line shows the consistent region of GWAS and linkage mapping. Red arrow indicates the position of causal gene *GmMACPF1*. **B** QTL *qRsc8H* and its containing genes in the consistent region. Red arrow indicates the causal gene *GmMACPF1*. **C** LD heat-map for the consistent region of GWAS and linkage mapping. **D** Gene structure of *GmMACPF1*. Type-R indicates the resistant soybean accessions, Type-S indicates the susceptible accessions. Blue rectangles indicate exons. Vertical lines indicate different SNP variations. **E** Box plots for disease index of RIL population, based on the classify results of QTL *qRsc8H* flanking markers ss715614742 and ss715614764. The significance of difference was analyzed by two-tailed *t*-test. **F** Box plots for disease index of natu-

ral population, based on the classify results of nonsynonymous SNP mutation (T/C) and synonymous SNP mutation (C/T) in the exons of *GmMACPF1*. The significance of difference was analyzed by two-tailed *t*-test. **G** Expression analysis of *GmMACPF1* by qRT-PCR. Y-axis indicates the relative expression ratio of samples inoculated with SC8 and mock at different time-points (0~72 hpi), X-axis indicates different time-points after inoculation. Asterisks represents statistical significance at 0.01 level. **H** Expression level of *GmMACPF1* in transgenic soybean hairy roots, compared with the empty-vector and wild-type controls. **I** Expression level of coat protein (CP) gene of SMV-SC8 in transgenic soybean hairy roots with over-expressed *GmMACPF1*, compared with the empty-vector and wild-type controls. Asterisks represents significance at 0.01 level

indicating a region of colocalization between the association and linkage analyses (Fig. 3).

Screening of the candidate genes for SC8 resistance

There were 21 genes in the overlapping ~141.6-kb region on chromosome 13 according to the Williams82 reference genome (Wm82.a4.v1), and the types of variants (SNPs and Indels) in these genes were analysed (Tables S7 and S8). Thereafter, the expression of these genes in different soybean organs and the positions of these SNPs and Indels in the corresponding protein-encoding regions were analysed.

Genes that were expressed in at least one organ or showed amino acid alterations in conserved domains were screened out (Fig. S4). Next, the positions of SNPs and Indels in the 5'UTR and the upstream region were analysed. Finally, by combining the gene annotations and SMV resistance-related reports, two candidate genes, *Glyma.13G176600* and *Glyma.13G176100*, were screened out for further study (Fig. 3, Table S7).

Among these two candidate genes, *Glyma.13G176600* possessed a synonymous mutation (C-T) and a nonsynonymous mutation (T-C) in the first and fourth exons (Table S7), which resulted in an amino acid change (Val-Ala) in the
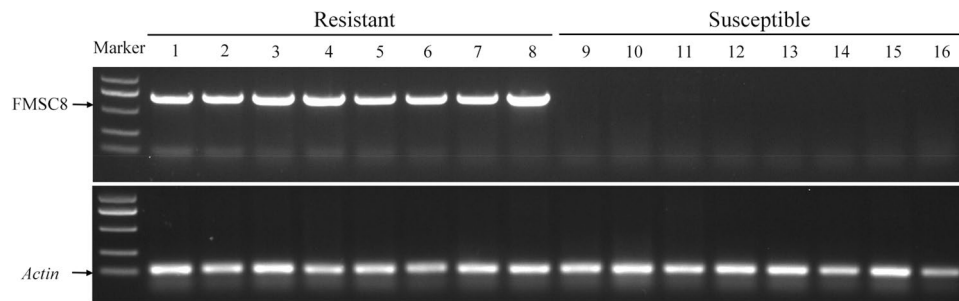
Fig. 4 PCR amplification of resistant and susceptible soybean accessions via the functional marker FMSC8. Marker DNA marker DL2000; Samples 1~8 indicate resistant accessions (1: KN7, DI = 9.69; 2: Shiheidou2, DI = 0.00; 3: Handou7, DI = 6.30; 4: Cangdou10, DI = 2.17; 5: Yudou19, DI = 0.00; 6: Fendou95, DI = 4.17; 7: Jian7, DI = 0.00; 8: Shanda1, DI = 0.00); Samples 9~16 indicate susceptible accessions (9: C813, DI = 65.91; 10: Kexuan13, DI = 46.80; 11: Weidou267, DI = 75.00; 12: Yudou17, DI = 51.79; 13: Ji17-J6, DI = 79.98; 14: Cangdou13, DI = 58.56; 15: Huangaizao, DI = 68.39; 16: Tiejiaqing, DI = 64.62)

conserved domain of the encoding membrane attack complex component/perforin (MACPF) protein (Fig. 3); this gene was designated *GmMACPF1*. Moreover, based on the two associated SNPs, the 209 accessions could be divided into two types, and the disease index of the resistant type (11.94 and 4.86) was extremely significantly lower than that of the susceptible type (52.57 and 41.85) (Fig. 3).

By predicting the tertiary structure of the *GmMACPF1*, we found that its peptide chain folding exhibited obvious alterations near the nonsynonymous mutation site, which might affect protein function (https://swissmodel.expasy.org/interactive) (Fig. S5). Subsequently, the expression of *GmMACPF1* was analysed via qRT−PCR after SC8 inoculation. Compared to the mock control, *GmMACPF1* showed significantly higher expression at 4, 8, 12 and 48 hpi in the resistant accession KN7 but presented no obvious change at the four time points in the susceptible accession C813 (Fig. 3). Additionally, the Rad60-SLD gene *Glyma.13G176100* had one associated SNP in its 5'UTR and three associated SNPs in the downstream region and was designated *GmRad60* (Table S7). Compared to the mock control, *GmRad60* showed significantly higher expression at several time points in the resistant accession KN7 and the susceptible accession C813 (Fig. S6). In addition, some SNPs identified in the promoters of these two genes could be used to divide the soybean accessions into two groups with significant differences in disease index of SC8 infection (Fig. S7).

GmMACPF1 overexpression inhibits the accumulation of SC8 in soybean hairy roots

The functions of the two candidate genes, *GmMACPF1* and *GmRad60*, were analysed via the "*Agrobacterium rhizogenes*-induced soybean hairy roots versus Soybean mosaic virus" method as described in a previous report (Jiang et al. 2019). First, the comparison of the ORF

sequences of *GmMACPF1* in the RIL parents KN7 and C813 revealed the same synonymous mutation (C-T) and nonsynonymous mutation (T-C) identified in the 209 resequenced accessions, while the ORF sequences of *GmRad60* showed no difference between the RIL parents.

Subsequently, we transformed C813 with these two genes and analysed the expression of the genes and the SC8 coat protein (CP) gene in transgenic soybean hairy roots. The results showed that the expression levels of *GmMACPF1* and *GmRad60* were significantly higher than those in the empty-vector control and wild-type (WT) (Fig. 3, Fig. S6). More importantly, the SC8 CP gene expression level in transgenic hairy roots overexpressing *GmMACPF1* was significantly decreased 3.3- and 3.1-fold compared to the empty-vector control and WT, respectively, while there were no significant differences in the *GmRad60*-overexpressing hairy roots, indicating that *GmMACPF1* significantly decreased the SC8 virus content and positively regulated resistance to SC8 in soybean (Fig. 3, Fig. S6).

Development of the functional gene marker FMSC8

The functional gene marker FMSC8 was developed based on the nonsynonymous mutation of *GmMACPF1*, and the corresponding primers could amplify a 671-bp fragment in genotypes similar to KN7 but produced no PCR product in genotypes similar to C813 (Fig. 4). By using the FMSC8 marker to evaluate the 419 accessions (including 209 resequencing accessions, two RIL parents and 208 other accessions), 327 accessions were attributed to the KN7 genotype, among which 263 had a disease index lower than 15 (0~14.55, resistant to SC8), while the remaining 92 accessions were attributed to the genotype C813, among which 73 had a disease index higher than 15 (21.23~89.68, susceptible to SC8). Accordingly, the detection rate of the

FMSC8 functional marker was calculated as 80.19% in the 419 soybean accessions (Table S9, Fig. 4).

Discussion

First application of deep resequencing of populations to study soybean SMV resistance

A high density of molecular markers is critical for improving the accuracy of QTL mapping (Song et al. 2016; Cao et al. 2017). Different molecular marker systems have been developed and used to map resistance genes in soybean. In recent years, the SNP array technique has been increasingly applied for the mining of genetic loci and candidate genes resistant to SMV in soybean. Chu et al. (2021) identified a major QTL related to resistance to soybean mosaic virus in diverse genetic populations based on the SoySNP6K array, and 5046 high-performing SNPs with an MAF > 0.05 and call rate > 80% were used for GWAS. Che et al. (2017) revealed a novel locus for SC7 resistance in a soybean mutant panel based on a 355-K SoySNP array, and a total of 61,543 SNPs with an MAF > 0.05 were used for GWAS.

With the development of resequencing technology, a large number of SNP markers have been used to find SNPs associated with important agronomic and economic traits and to mine functional genes in rice, corn, cotton and other crops (Ma et al. 2018; Yang et al. 2021). However, there has been no previously report of the application of GWAS to investigate resistance to SMV based on the resequencing of soybean populations. In the present study, a natural population composed of 209 soybean accessions subjected to deep resequencing with an average sequencing depth of 18.44 × was first used to identify genetic loci and candidate genes for resistance to SMV-SC8. More than 5.0 M SNPs were applied in the GWAS, which laid a solid foundation for identifying genetic loci for disease resistance.

Genetic loci associated with SC8 resistance on chromosome 13

Plants have gradually developed diverse defence mechanisms against viruses and other pathogens over time, but the underlying genetic basis of plant resistance still needs to be further dissected. In soybean, three independent loci (*Rsv1*, *Rsv3* and *Rsv4*) have been characterized as conferring resistance to SMV strains G1 ~ G7 (Maroof et al. 2008; Shakiba et al. 2012), and several loci (*Rsc4*, *Rsc5*, *Rsc7*, *Rsc8*, *Rsc15* and *Rsc20*) has been reported to confer resistance against SMV-SC strains (Wang et al. 2011a, 2011b, 2017; Karthikeyan et al. 2017, 2018; Rui et al.

2017). Eight of these loci (R_{sv4}, R_{SC5}, R_{SC6}, R_{SC7}, R_{SC8}, etc.) were located on chromosome 2, ten loci (*Rsv1*, *Rsv5*, R_{SC3Q}, R_{SC11}, R_{SC12}, etc.) were located on chromosome 13, and two loci were located on each of chromosome 6 (R_{SC15} and R_{SC18}) and chromosome 14 (*Rsv3* and R_{SC4}).

In China, SMV-SC8 has become a main epidemic virus strain in many soybean-growing regions, but only a few studies with a small number of molecular markers have focused on the mining of relevant genetic loci and candidate genes via a linkage mapping strategy to date (Wang et al. 2011b; Zhao et al. 2016; Ren et al. 2021). Therefore, more than 5.0 M SNPs were used for GWAS in the present study, and 3035 SNPs were identified as significantly associated with resistance to SC8. Additionally, a major QTL, *qRsc8F*, was shown to control SC8 resistance in one genetic population across multiple environments. More importantly, 327 SNPs on chromosome 13 obtained from the association analysis were located in the region of the *qRsc8F*, which was first reported to confer resistance to SMV-SC8.

Moreover, to analyse the relationship of *qRsc8F* and previously reported QTLs on chromosome 13 with other SMV strains (such as the relationship of the *Rsv1* locus with strains G1 ~ G7) (Table S4), we conducted a haplotype analysis based on the SNPs in the 28.42 ~ 30.30 Mb region (the region including *qSC8F*, *Rsv1* and ten other resistance QTLs) (Fig. S8). The results showed that nine of the eleven previous QTLs (including *Rsv1*) were attributed to different LD blocks relative to *qSC8F*; the two remaining QTLs, *Rsc18A* and *Rsc15ZH*, were shown to confer resistance to the SMV-SC18 and SMV-SC15 strains, partially overlapped with *qSC8F*, indicating a broad-spectrum resistance of this genomic region for diverse SMV strains, including the SC8 resistance identified in the present study.

The causal gene regulating SC8 resistance

The colocalization interval of ss715614742 ~ ss715614764 on chromosome 13 was obtained from linkage mapping and association analysis in the present study; this interval is close to the physical positions of genetic loci related to other SMV strains reported previously. There are a large number of genes in this region, two of which have been predicted to be related to the resistance of the SC15 strain (Rui et al. 2017). To further determine the causal gene related to the SC8 strain in this region, two previously predicted genes (*GmRad60* and *Glyma.13g176300*) and another gene, *GmMACPF1*, were transformed into soybean hairy roots, and the results demonstrated that only *GmMACPF1* overexpression significantly decreased the SC8 virus content of the transgenic soybean hairy roots, which indicated that *GmMACPF1* was the causal gene of SC8 resistance in this consistent genomic region (Fig. 3, S6).

Some studies have reported that the functions of the MACPF protein are involved in plant resistance to different pathogens via the salicylic acid (SA) pathway. *AtMACP2* can regulate resistance to diverse diseases through the SA pathway (Zhang et al. 2022). Another MACPF protein-encoding gene, *CAD1*, plays an important role in immunity-induced programmed cell death (PCD) (Tsutsui et al. 2006). Similarly, the mutation of necrotic spotted lesion 1 resulted in obviously spotted necrotic lesions, retarded plant growth and high accumulation of SA (Noutoshi et al. 2006). In soybean, SA has already been reported to affect resistance to different diseases (Baebler et al. 2014; Alazem et al. 2015; Zhang et al. 2012), such as *Rsv1*-mediated resistance to most SMV strains. The silencing of *GmEDS1a/GmEDS1b* or *GmPAD4* reduced SA accumulation and enhanced susceptibility to SMV-G5 in soybean (Wang et al. 2014). Thus, we deduced that the functional gene *GmMACPF1* might confer resistance to SC8 via the SA pathway.

The causal gene *GmMACPF1* identified in the present study encoded a homologous protein to AtMACP2, and a nonsynonymous mutation in this gene induced an alteration in peptide chain folding between the resistant and susceptible soybean accessions. Previous studies showed that *GmMACPF1* was highly expressed in soybean leaf and root, and was induced by the inoculation of bean pod mottle virus and SMV-G7 (Fig. S9). In the present study, the expression of *GmMACPF1* was much higher at 4, 8, 12 and 48 hpi after SC8 inoculation in the resistant accession KN7 but presented no obvious change in the susceptible accession C813 (Fig. 3). Further analysis showed that there were some variations in the promoter of *GmMACPF1*, and several of them altered regulatory elements of the promoter, suggesting that the difference in *GmMACPF1* expression might be due to allelic variations in the promoter region. To further verify this, we cloned the promoter of *GmMACPF1* from the two parents KN7 and C813, and the much higher expression of *GUS* and β-galactosidase activity were observed in *GmMACPF1*[KN7]-Pro::GUS, compared to *GmMACPF1*[C813]-Pro::GUS (Fig. S10). Furthermore, a functional gene marker of *GmMACPF1* was developed based on the nonsynonymous mutation (T/C), which could divide the soybean accessions into resistant and sensitive types (Table S9). This also demonstrated the function of the SC8 resistance gene from another point of view.

Conclusion

A deeply resequenced natural soybean population and a RIL population were used to explore genetic loci and genes for SMV-SC8 resistance. In total, 3030 SNPs on chromosome 13 were identified as significantly associated with SC8 resistance in the natural population, among which 327 were located within an ~ 0.14 Mb region of the major QTL *qRsc8F* in the RIL population. One candidate gene, *GmMACPF1*, screened in the region of consistent linkage and association showed different expression levels between resistant and susceptible accessions after SC8 inoculation, and a significant decrease in virus content was observed in *GmMACPF1*- overexpressing soybean hairy roots. Furthermore, a functional gene marker, FMSC8, was developed based on the variant in *GmMACPF1*, and a high coincidence rate of 80.19% between the disease index and marker genotype was determined in 419 soybean accessions. Thus, these results provide valuable resources for studies of the molecular mechanism of SMV resistance and genetic improvement in soybean.

Declarations

References

Alazem M, Lin NS (2015) Roles of plant hormones in the regulation of host–virus interactions. Mol Plant Pathol 16:529–540

Baebler Š, Witek K, Petek M, Stare K, Tušek-Žnidarič M, Pompe-Novak M, Renaut J, Szajko K, Strzelczyk-Żyta D, Marczewski W, Morgiewicz K, Gruden K, Hennig J (2014) Salicylic acid is an indispensable component of the *Ny-1* resistance-gene-mediated response against *Potato virus Y* infection in potato. J Exp Bot 65:1095–1109

Cao YC, Li SG, He XH, Chang FG, Kong JJ, Gai JY, Zhao TJ (2017) Mapping QTLs for plant height and flowering time in a Chinese summer planting soybean RIL population. Euphytica 213:39

Che ZJ, Liu HL, Yi FL, Cheng H, Yang YM, Wang L, Du JY, Zhang PP, Wang J, Yu DY (2017) Genome-wide association study reveals novel loci for SC7 resistance in a soybean mutant panel. Front Plant Sci 8:1771–1783

Chen P, Buss GR, Roane CW, Tolin SA (1991) Allelism among genes for resistance to Soybean mosaic virus in strain-differential soybean cultivars. Crop Sci 31:305–309

Cho EK, Goodman RM (1979) Strains of Soybean mosaic virus: classification based on virulence in resistant soybean cultivars. Phytopathology 69:467–470

Cho EK, Choi SH, Cho WT (1983) Newly recognized Soybean mosaic virus mutants and sources of resistance in soybeans. Res Rep off Rural Dev 25:18–22

Chu JH, Li WL, Piao DR, Lin F, Huo XB, Zhang H, Du H, Kong YB, Jin Y, Li XH, Zhang CY (2021) Identification of a major QTL related to resistance to soybean mosaic virus in diverse soybean genetic populations. Euphytica 217:176

Galvez LC, Banerjee J, Pinar H, Mitra A (2014) Engineered plant virus resistance. Plant Sci 228:11–25

Guo DQ, Zhi HJ, Wang YW, Gai JY, Zhou XA, Yang CL (2005) Identification and distribution of strains of soybean mosaic virus in middle and northern of Huang Huai Region of China. Soybean Sci 27:64–68

Hartman GL, West ED, Herman TK (2011) Crops that feed the World 2. Soybean-Worldwide production, use, and constraints caused by pathogens and pests. Food Secur 3:5–17

Huo XB, Li XH, Du H, Kong YB, Tian R, Li WL, Zhang CY (2019) Genetic loci and candidate genes of symbiotic nitrogen fixation-related characteristics revealed by a genome-wide association study in soybean. Mol Breed 39(9):127

Jiang H, Li K, Gai JY (2019) *Agrobacterium rhizogenes*-induced soybean hairy roots versus *Soybean mosaic virus* (ARISHR-SMV) is an efficiency pathosystem for studying soybean–virus interactions. Plant Methods 15:1–11

Karthikeyan A, Li K, Jiang H, Ren R, Li C, Zhi H, Chen S, Gai J (2017) Inheritance, fine-mapping, and candidate gene analyses of resistance to soybean mosaic virus strain SC5 in soybean. Mol Genet Genomics 292:811–822

Karthikeyan A, Li K, Li C, Yin J, Li N, Yang Y, Song Y, Ren R, Zhi H, Gai J (2018) Fine-mapping and identifying candidate genes conferring resistance to *Soybean mosaic virus* strain SC20 in soybean. Theor Appl Genet 131:461–476

Kim JS (1991) A new virulent strain of Soybean mosaic virus infecting SMV resistant soybean cultivar Deogyou. Plant Pathol J 7:37–41

Kim YH, Kim O, Lee BC, Moon JK, Lee SC, Lee JY (2003) G7H, a new *Soybean mosaic virus* strain: its virulence and nucleotide sequence of CI gene. Plant Dis 87:1372–1375

Klepadlo M, Chen P, Shi A, Mason RE, Korth KL, Srivastava V (2017) Two tightly linked genes for soybean mosaic virus resistance in soybean. Crop Sci 57:1844–1853

Kong YB, Li XH, Wang B, Du H, Zhang CY (2018) The soybean purple acid phosphatase GmPAP14 predominantly enhances external phytate utilization in plants. Front Plant Sci 9:292

Li K, Yang QH, Zhi HJ, Gai JY, Yu DY (2010) Identification and distribution of *soybean mosaic virus* strains in southern China. Plant Dis 94:351–357

Li CY, Chang HL, Wang JH, Zou JN, Shi Y, Zhang YJ, Wei W, Xin DW (2018) Genetic analysis of nodule traits in soybean via wild soybean background population and high generation recombination inbred lines. Int Jagric Bio 20:2521–2528

Ma G, Chen P, Buss GR, Tolin SA (2002) Complementary action of two independent dominant genes in Columbia soybean for resistance to Soybean mosaic virus. J Hered 93:179–184

Ma ZY, He SP, Wang XF, Sun JL, Zhang Y, Zhang GY, Wu LQ, Li ZK, Liu ZH, Sun GF, Yan YY, Jia YH, Yang J, Pan ZE, Gu QS, Li XY, Sun ZW, Dai PH, Liu ZW, Gong WF, Wu JH, Wang M, Liu HW, Feng KY, Ke H, Wang JD, Lan HY, Wang GN, Peng J, Wang N, Wang LR, Pang BY, Peng Z, Li RQ, Tian SL, Du XM (2018) Resequencing a core collection of upland cotton identifies genomic variation and loci influencing fiber quality and yield. Nat Genet 50:803–813

Malapi-Nelson M, Wen RH, Ownley BH, Hajimorad MR (2009) Co-infection of soybean with *Soybean mosaic virus* and *Alfalfa mosaic virus* results in disease synergism and alteration in accumulation level of both viruses. Plant Dis 93(12):1259–1264

Maroof MAS, Jeong SC, Gunduz I, Tucker DM, Buss GR, Tolin SA (2008) Pyramiding of soybean mosaic virus resistance genes by marker-assisted selection. Crop Sci 48:517–526

Noutoshi Y, Kuromori T, Wada T, Hirayama T, Kamiya A, Imura Y, Yasuda M, Nakashita H, Shirasu K, Shinozaki K (2006) Loss of *Necrotic Spotted Lesions 1* associates with cell death and defense responses in *Arabidopsis thaliana*. Plant Mol Biol 62:29–42

Nyquist WE, Baker RJ (1991) Estimation of heritability and prediction of selection response in plant populations. Crit Rev Plant Sci 10:235–322

Ren QY, Jiang H, Xiang WY, Nie Y, Xue S, Zhi HJ, Li K, Gai JY (2021) A MADS-box gene is involved in soybean resistance to multiple *Soybean mosaic virus* strains. Crop J 21:802–808

Rui R, Liu SC, Karthikeyan A, Wang T, Niu HP, Yin JL, Yang YH, Wang LQ, Yang QH, Zhi HJ (2017) Fine-mapping and identification of a novel locus *Rsc15* underlying soybean resistance to *Soybean mosaic virus*. Theor Appl Genet 130:2395–2410

Shakiba E, Chen P, Shi A, Li D, Dong D, Brye K (2012) Two novel alleles at the *Rsv3* locus for resistance to *Soybean mosaic virus* in PI399091 and PI61947 soybeans. Crop Sci 52:2587–2594

Shao ZQ, Shao JB, Huo XB, Li WL, Kong YB, Du H, Li XH, Zhang CY (2022) Identification of closely associated SNPs and candidate genes with seed size and shape via deep re-sequencing GWAS in soybean. Theor Appl Genet 135(7):2341–2351

Shi AN, Chen PY, Zheng CM, Hou AF, Zhang B (2008) A PCR-based marker for the *Rsv1* locus conferring resistance to *Soybean mosaic virus*. Crop Sci 48:262–268

Shi AN, Chen PY, Vierling R, Li DX, Zheng CM (2013) Identification of Soybean mosaic virus resistance alleles in Jindou 1 soybean. Euphytica 192:181–187

Song QJ, Jenkins J, Jia GF, Hyten DL, Pantalone V, Jackson SA, Schmutz J, Cregan PB (2016) Construction of high resolution genetic linkage maps to improve the soybean genome sequence assembly Glyma1.01. BMC Genom 17:33

Sun YQ, Chen SL, Chu JH, Li XH, Zhang CY (2022) Mining of QTLs and candidate genes for pod and seed traits via combining BSA-seq and linkage mapping in soybean. J Agric Sci Technol. https://doi.org/10.13304/j.nykjdb.2022.0089

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

Tsutsui T, Morita-Yamamuro C, Asada Y, Minami E, Shibuya N, Ikeda A, Yamaguchi J (2006) Salicylic acid and a Chitin Elicitor both control expression of the *CAD1* gene involved in the plant immunity of *Arabidopsis*. Biosci Biotech Bioch 70(9):2042–2048

Wang DG, Ma Y, Yang YQ, Liu N, Li CY, Song YP, Zhi HJ (2011a) Fine mapping and analyses of *Rsc8* resistance candidate genes to soybean mosaic virus in soybean. Theor Appl Genet 122:555–565

Wang DG, Ma Y, Liu N, Yang ZL, Zheng GJ, Zhi HJ (2011b) Fine mapping and identification of the soybean *Rsc4* resistance candidate gene to soybean mosaic virus. Plant Breed 130:653–659

Wang JL, Shine MB, Gao QM, Navarre D, Jiang W, Liu CY, Chen QS, Hu GH, Kachroo A (2014) Enhanced disease susceptibility mediates pathogen resistance and virulence function of a bacterial effector in soybean. Plant Physiol 165:1269–1284

Wang DG, Zhao L, Li K, Ma Y, Wang LQ, Yang YQ, Zhi HJ (2017) Marker-assisted pyramiding of soybean resistance genes *RSC4*, *RSC8*, and *RSC14Q* to soybean mosaic virus. J Integr Agric 16:2413–2420

Wu LZ, Wang XF, Li XH, Ma ZY (2014) Construction and function identification of universal plant expression vector pCamE. J Agric Biotechnol 22:661–671

Yan L, Hofmann N, Li SX, Ferreira ME, Song QJ, Jiang JL, Ren SX, Quigley C, Fickus E, Cregan P, Song QJ (2017) Identification of QTL with large effect on seed weight in a selective population of soybean with genome-wide association and fixation index analyses. BMC Genomics 18:529

Yang YQ, Hou WH, Bing QL, Yan L, Zhang MS, Meng XM, Liu LJ, Lin J, Zhi HJ, Zhang MC (2014) Composition and distribution of SMV strains in Hebei. Soybean Sci 33:87–90

Yang CM, Yan J, Jiang SQ, Li X, Min HW, Wang XF, Hao DY (2021) Resequencing 250 soybean accessions: new insights into genes associated with agronomic traits and genetic networks. Genom Proteom Bioinf 24:1672–2229

Zhang CQ, Grosic S, Whitham SA, Hill JH (2012) The requirement of multiple defense genes in soybean *Rsv1*–mediated extreme resistance to *Soybean mosaic virus*. Mol Plant Microbe Interact 25:1307–1313

Zhang J, Wen ZX, Li W, Zhang YW, Zhang LF, Dai HY, Wang DC, Xu R (2017) Genome-wide association study for soybean cyst nematode resistance in Chinese elite soybean cultivars. Mol Breed 37:60

Zhang JP, Wang XZ, Lu YM, Bhusal SJ, Song QJ, Cregan PB, Yen Y, Brown M, Jiang GL (2018) Genome-wide scan for seed composition provides insights into soybean quality improvement and the impacts of domestication and breeding. Mol Plant 11:460–472

Zhang X, Dai YS, Wang YX, Su ZZ, Yu LJ, Zhang ZF, Xiao S, Chen QF (2022) Overexpression of the *Arabidopsis* MACPF protein *AtMACP2* promotes pathogen resistance by activating SA signaling. Int J Mol Sci 23(15):8784

Zhao L, Wang DG, Zhang HY, Shen YC, Yang YQ (2016) Fine mapping of the *Rsc8* locus and expression analysis of candidate SMV resistance genes in soybean. Plant Breed 135:701–706

Zhi HJ, Gai JY, He XH (2005) Inheritance of resistance in infection and resistance in development to soybean mosaic virus in soybeans. Acta Agron Sin 31:1260–1264