



Genome-wide association study for soybean mosaic virus SC3 resistance in soybean

Zhijun Che · Honglang Yan · Hailun Liu · Hui Yang ·
Haiping Du · Yuming Yang · Baohui Liu · Deyue Yu

Received: 20 April 2020 / Accepted: 3 July 2020 / Published online: 15 July 2020
© Springer Nature B.V. 2020

Abstract Soybean mosaic virus (SMV) threatens soybean (*Glycine max* (L.) Merr.) production and seed quality worldwide. Host plant resistance is the most environmentally friendly and economically effective method to deduce the damages of SMV. However, the use of limited resistant loci in SMV management will be conquered due to SMV strain shifts caused by evolutionary arms races. Thus, exploration of novel germplasms of SMV resistance and identification of novel resistant genes are necessary. In this study, we inoculated 219 soybean accessions with a SC3 virus

strain in three environments. A genome-wide association study was conducted with 355K SoySNP array to dissect the genetic basis of SMV resistance. A total of 24 SNPs significantly associated with resistance to SC3 were identified, and these SNPs explained 25.54–33.60% of the phenotypic variations. Many significant SNPs were located near to the known SMV resistance loci *Rsv1*, *Rsv4*, and *Rsv5*. Furthermore, one SNP on chromosome 8 and four SNPs on chromosome 20 were novel loci. Three genes underlying these two novel loci with different genetic variation between resistant and susceptible accessions and with high expression levels in leaf were identified as promising candidate genes. Therefore, the novel loci and candidate genes were identified as new genetic sources for marker-assisted selection breeding in soybean.

Zhijun Che, Honglang Yan and Hailun Liu contributed equally to this work.

Z. Che · H. Yang · H. Du · B. Liu
Innovative Center of Molecular Genetics and Evolution, School of Life Sciences, Guangzhou University, Guangzhou, China

Z. Che (✉) · Y. Yang · D. Yu (✉)
National Center for Soybean Improvement, National Key Laboratory of Crop Genetics and Germplasm Enhancement, Jiangsu Collaborative Innovation Center for Modern Crop Production, Nanjing Agricultural University, Nanjing 210095, China
e-mail: chezj@gzhu.edu.cn
e-mail: dyyu@njau.edu.cn

H. Yan
Jiangsu Yanjiang Institute of Agricultural Sciences, Nantong, China

H. Liu
Key Laboratory of South China Agricultural Plant Molecular Analysis and Genetic Improvement, and Guangdong Provincial Key Laboratory of Applied Botany, South China Botanical Garden, Chinese Academy of Sciences, Guangzhou, China

Keywords Soybean · Soybean mosaic virus · SC3 strain · Association analysis · Disease rate

Introduction

Soybean mosaic virus (SMV) is a single-stranded, positive-sense RNA virus of the *Potyvirus* genus (Adams et al. 2005). In SMV-infected soybean, an obvious mosaic symptom was observed in leaves, causing severe yield loss. Based on disease symptoms in specific soybean cultivars, SMV isolates were divided into seven strains (G1 to G7) in America (Cho and Goodman 1979) and there were twenty-two strains (SC1 to SC22) in China (Wang et al. 2003; Guo et al.

2005; Zhan et al. 2015; Li et al. 2010). However, the relationship among different SMV strains between America and China has not been reported yet. Breeding and planting soybean cultivars with broad-spectrum resistance were effective and eco-friendly methods to decrease the yield loss caused by SMV. Four resistant loci, *Rsv1*, *Rsv3*, *Rsv4*, and *Rsv5* (Resistance to SMV), have been reported to be effectively resistant to several SMV strains (Kiihl and Hartwig 1979; Buss et al. 1997; Hayes et al. 2000; Klepadlo et al. 2017). However, a different naming method has been used in China. Some resistance loci on chromosome 2 containing *R_{SC7}* and *R_{SC8}* (Resistance to SMV in China) derived from Kefeng No. 1 were resistant to SC1-SC13, SC16, SC18-SC21, and G1-G7 (Li et al. 2010; Yan et al. 2015; Wang et al. 2011b; Cai et al. 2014). Resistance loci on chromosome 13 involving *R_{SC3}* and *R_{SC14}* derived from Qihuang No. 1 were resistant to SC1-SC8, SC11-SC14, SC16-SC18, SC20-SC22, and G1-G7 (Li et al. 2010; Cai et al. 2014; Zheng et al. 2014; Ma et al. 2011). Resistance loci *R_{SC4}* on chromosome 14 derived from Dabaima was resistant to SC4 and SC10 (Wang et al. 2011a; Li et al. 2012). *R_{SC15}* was a newfound locus on chromosome 6 derived from RN-9, which could resist the most virulent strain SC15 (Ren et al. 2017). To compare the different loci, clusters of resistant genes containing nucleotide-binding sites (NBS) and leucine-rich repeat (LRR) domains were identified in *Rsv1* and *Rsv3* (Wang et al. 2011a; Hayes et al. 2004; Yang et al. 2013; Ma et al. 2016). Recently, *Rsv4* gene encoding an RNase H family protein was map-based cloned in Peking; *Rsv4* protein inhibited virus multiplication by degrading dsRNA, thus conferring a broader resistance to potyvirus (Ishibashi et al. 2019). Silencing a chaperone *GmHSP40.1* gene near *Rsv4* enhanced the susceptibility of soybean plants to SMV (Liu and Whitham 2013). Taken together, it is of great significance to explore resistant genes by using resistant sources for breeding broad-spectrum-resistant soybean varieties.

A lot of work has been done in the field of resistant loci mapping and candidate genes prediction. However, few of *Rsv* or *Rsc* genes were cloned and functionally identified. The reasons for these results may be that many resistant loci were mapped to large regions by adopting biparental mapping populations with low genetic diversity and low mapping resolution. Therefore, high-density markers and new methods are required for the identification of resistant genes. Many putative genes have been identified in plant via genome-wide

association study (GWAS) based on broad heritable diversity populations and high-density single-nucleotide polymorphisms (SNPs) markers. Three causal genes, *LOC_Os01g62780*, *LOC_Os11g08410*, and *LOC_Os08g37890*, which are associated, respectively, with days to heading, plant height, and panicle length, were identified in rice by GWAS and thus are important for agronomic traits improvement (Yano et al. 2016). Soybean *GmMYB29*, an R2R3-type MYB transcription factor, was identified by GWAS and obtained the function in isoflavone biosynthesis (Chu et al. 2017). Two homeologous genes *Tof12* and *Tof11* were identified by GWAS and provided the evidence that flower phenology was a key domestication trait in soybean (Lu et al. 2020; Gong 2020). Che et al. (2017) identified 104 SNPs significantly associated with resistance to SMV strain of SC7 in a soybean mutant panel (Che et al. 2017).

Few studies have been reported on the excavation of SMV resistance loci in soybean natural panel by GWAS. The object of this study was to screen some SMV-resistant germplasms in a soybean natural population and to identify the important SMV-resistant genetic loci in multiple environments using GWAS. This information could be used for marker-assisted selection in soybean-resistant breeding programs.

Material and methods

Plant materials and SC3 strain

A natural population consisting of 219 soybean accessions (including 195 landraces and 24 elite varieties) collected from 26 diverse provinces and six diverse ecological regions in China was used for GWAS (Wang et al. 2016). The viruliferous leaves of SMV strain of SC3 were preserved in a refrigerator at $-80\text{ }^{\circ}\text{C}$. All soybean accessions and SMV strain were provided by the National Center for Soybean Improvement (Nanjing Agricultural University, Nanjing, China). The resistance evaluation experiments were conducted in 2010, 2011, and 2012 in an aphid-free greenhouse at the Jiangpu Experimental station of Nanjing Agricultural University ($32^{\circ} 12' \text{N}$ and $118^{\circ} 37' 48'' \text{E}$), and these three environments were named E1, E2, and E3, respectively. In each environment, all the accessions were planted in a randomized complete block design.

Resistance evaluation and phenotype investigation

About thirty seedlings of each accessions were planted in round pots filled with sand. The unifoliate leaves were inoculated with a SC3 inoculum, which were viruliferous leaves grinded by using a mortar and pestle in 0.1 M sodium phosphate buffer (PH 7.2–7.4), and finally rinsed with tap water. Foliar symptoms to SC3 were monitored weekly after inoculation, compared with continuous counting 2 to 4 weeks post inoculation, then disease rate (DR) phenotype was investigated according to previously reported standard (Pu et al. 1983). The rate of the number of systematic mosaic plants to total plants was defined as DR. Based on this standard, DR under 10% was defined as resistant accession and DR over 10% was susceptible accession.

Phenotypic data analysis

Phenotypic data including descriptive statistics, the best linear unbiased prediction (BLUP), analysis of variance (ANOVA), and broad heritability for disease rates were performed by R software (R Core Team 2018). To minimize the effects of variation among the three different environments, BLUP was computed based on the phenotypic data according to the R software with lme4 package (Merk et al. 2012): $Y_{ik} = \mu + G_i + Y_k + GY_{ik} + \varepsilon_{ik}$, where Y_{ik} is the phenotype evaluate, μ is the overall mean, G_i is i th's genotypic effect, Y_k is k th's year effect, GY_{ik} is the interaction effect of genotype \times year, and ε_{ik} is the residual error. The broad-sense heritability (h^2) of phenotype data was estimated according to the formula: $h^2 = \sigma_g^2 / (\sigma_g^2 + \sigma_{gv}^2/n + \sigma_e^2/nr)$, where σ_g^2 is the genotype variance, σ_{gv}^2 is the genotype and environment interaction variance, σ_e^2 is the residual error, n is the number of environments, and r represents the replications (Knapp et al. 1985).

GWAS analysis

The 219 soybean accessions were genotyped using NJAU 355 K SoySNP array as previously reported (Wang et al. 2016). A total of 292,053 high-quality SNP markers were identified and provided approximately one SNP per 3.3 kb among the 20 soybean chromosomes. In this study, we screened out 207,608 SNPs with minor allele frequency (MAF) $\geq 5\%$ and three environmental phenotypes and BLUP values were used for association analysis. GWAS was performed by

using a compressed mixed linear model (CMLM) based on an R package named the Genome Association and Prediction Integrated Tool (Lipka et al. 2012). The Bonferroni threshold of significantly associated SNPs with DR was set as $P \leq 1/207,608 = 4.82e-06$ or $-\text{Log}_{10}P \geq 5.32$. The linkage disequilibrium (LD) block and haplotype analysis were performed using Haploview 4.2 software (Barrett et al. 2005).

Putative gene prediction and expression analysis

To explore the putative genes underlying significant association signals, the predicted genes in the LD decay (130 kb) genomic regions were searched based on the annotation of the soybean reference genome Glyma.Wm82.a1.v1.1 (<http://phytozome.net>). Genes with annotations related to biotic stress or disease resistance were selected as putative genes. Meanwhile, a BLAST-P analysis was performed in *Arabidopsis* and rice genome using the amino acid sequences of soybean putative genes.

Two RNA-seq data downloaded from public databases were used for the expression analysis of putative genes. One database including seed development, vegetative tissues (leaves, roots, stems, and seedlings), and reproductive tissues (floral buds) was downloaded from the Plant Expression Database (<https://www.plexdb.org>, GEO Accession number: GSE29163). Another containing six tissues with different development stages was downloaded from the SoyBase (<https://www.soybase.org/soyseq/>). The multi-array comparative analysis method is the most popular preprocessing algorithm for gene expression microarrays. The normalized values from microarray were log₂-transformed. To visualize the expression data, two heat maps were constructed by HemI software (Deng et al. 2014).

To evaluate whether these three candidate genes were induced by SC3 strain, resistant accession Kefeng No.1 and susceptible accession Nannong 1138-2 were planted in an aphid-free greenhouse, and leaves with three biological replications were collected in liquid N₂ at 0, 1, 2, and 3 days after SC3 inoculating or mock inoculating with sodium phosphate buffer. Total RNA was isolated with Ultrapure RNA Kit (CW BIO). Synthesis of complementary DNA and removal of DNA were performed with first strand cDNA synthesis kit (Takara). Quantitative real-time PCR (qRT-PCR) was performed using the LightCycler 480 (Roche). The PCR cycling conditions were 95 °C for 30 s; 40 cycles: 95 °C

for 10 s and 60 °C for 60 s. The soybean tubulin gene (*Glyma.05G157300*) as an internal control for calculating the relative expression levels of candidate genes.

Results

Significant phenotypic variation in soybean natural population

The disease rates of SMV strain of SC3 in three environments were used as phenotypic traits, which were analyzed by using descriptive statistics, ANOVA, and broad heritability to determine the resistance diversities in a soybean natural population. A total of 219 accessions, representing a broad level of genetic variation of soybean cultivation in China, were collected from three ecological habitats. As expected, a great level of genetic variation in DR was observed in this panel (Table 1). The DR was ranging from 0 to 1 with average values ranged from 0.64 to 0.70 in three environments. Moreover, these phenotypic values exhibited large coefficients of variation (CV), and mean CV in three environments were 52.94, 39.39, and 52.83, respectively. The frequency distribution of DR among three environments showed a skewed normal distribution (Fig. 1), suggesting DR was a complex quantitative trait controlled by multiple genes. The ANOVA showed that genotype, and genotype and environment interaction dramatically influenced the DR. To explore whether DR could stably be inherited to the progeny, the broad-sense heritability was calculated in this population. Unsurprisingly, the broad-sense heritability was 79.02%, which was a high value. Collectively, these results indicated that DR with high variation diversity and heritability was stable in this population, especially affected by a genetic factor.

Significant loci associated with DR

We screened out 207,608 polymorphic SNPs with $MAF \geq 0.05$ for GWAS analysis. By using a CMLM model to correct the population structure and kinship, as shown in the quantile-quantile (QQ) plots (Fig. 2), in the left corner of scatter diagram were low significant loci that were not associated with disease rate trait, so the observed $-\log_{10}(P)$ should be consistent with the expected $-\log_{10}(P)$. The fact that these loci were indeed on the diagonal line indicates that the CMLM was suitable for disease rate trait in GWAS. A total of twenty-four SNPs exceeding a significant threshold ($-\log_{10}P \geq 5.32$) in all three environments and BLUP (Table 2). Additionally, ten SNPs on chromosome 13 significantly associated with DR were repetitively detected at least twice. The other fourteen significant SNPs identified only in a single environment were located on chromosomes 2, 8, 13, and 20 (Table 2, Fig. 2).

Four SNPs, which were identified by GWAS within or near previously fine-mapped QTLs associated with resistance to SMV on chromosome 2 (Yan et al. 2015; Wang et al. 2011b; Karthikeyan et al. 2017), closed to each other and explained almost 31.48% of the phenotypic variation. On chromosome 13, a total of fifteen SNPs were divided into two SNP clusters. One cluster included four SNPs (AX-93813964, AX-93813987, AX-93814000, and AX-93939503) that were significantly detected by GWAS within or near the previously identified resistant QTLs, such as *Rsv1*, *Rsc-pm*, and *R_{SC11}* (Yang et al. 2013; Shi et al. 2008; Bai et al. 2009); the other cluster containing eleven SNPs were located near several resistant QTLs, such as *R_{SC18B}*, *Rsc-ps*, *Rsv5* (Klepadlo et al. 2017; Yang et al. 2013; Li et al. 2015). Remarkably, the phenotypic variation of these 15 adjacent SNPs ranged from 25.54 to 36.27%, suggesting that these loci were major QTLs with large effects. In

Table 1 Descriptive statistics, ANOVA and broad heritability of disease rates to SC3 in three environments

| Env | Mean | SD | Median | CV(%) | Min | Max | Skew | kurtosis | G | G × E | h ² (%) |
|-----|------|------|--------|-------|-----|-----|-------|----------|----|-------|--------------------|
| E1 | 0.68 | 0.36 | 0.85 | 52.94 | 0 | 1 | -0.83 | -0.93 | ** | ** | 79.02 |
| E2 | 0.66 | 0.26 | 0.72 | 39.39 | 0 | 1 | -0.51 | -0.8 | ** | ** | |
| E3 | 0.53 | 0.28 | 0.52 | 52.83 | 0 | 1 | -0.07 | -1.14 | ** | ** | |

Env environment, SD standard deviation, Min minimum, Max maximum, G genotype, G × E genotype × environment, h² broad heritability

**Significant at $P \leq 0.01$

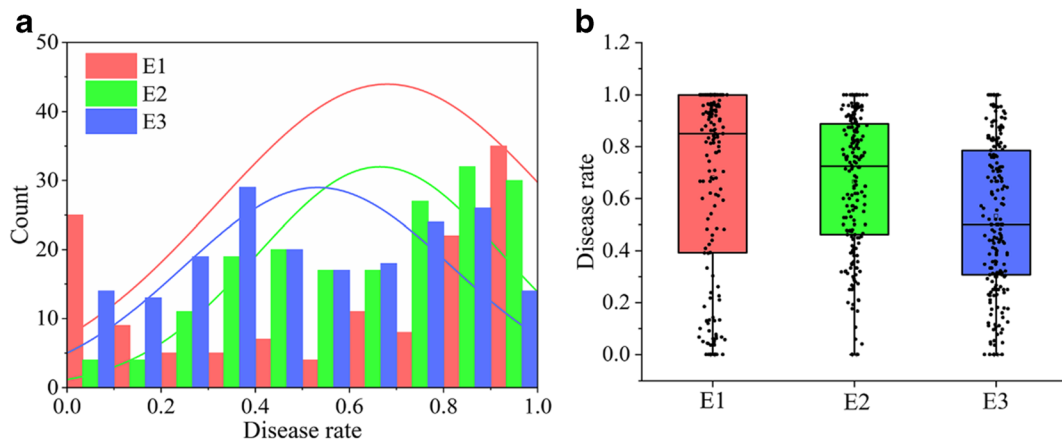


Fig. 1 Phenotypic analysis of disease rates to SC3 in three individual environments. **a** Frequency distribution of DR in three environments, E, environment. **b** Box plot of DR in three environments. Bold line is median value, black dots are DR values of 219 soybean accessions

addition, five SNPs significantly associated with resistance to SC3 located on chromosomes 8 and 20 were not reported in previous reports and explained 31.50–32.65% of the phenotypic variation, possibly representing novel major loci underlying soybean resistance to SC3 (Fig. 3a, b). A new SNP locus AX-93753793 on chromosome 8 was significantly identified in E1 (Fig. 3a–c). In comparison with allelic variation, the superior alleles of this locus may decrease the average DR from 0.79 to 0.32 (Fig. 4a). Furthermore, resistant accession Kefeng No.1 and susceptible accession Nannong 1138-2 were also divided into different groups based on allelic variation of AX-93753793. On chromosome 20, four SNPs loci were significantly detected in E1 (Fig. 3d–f). Based on these four alleles, the genotype of natural population used in this study was classified into three haplotypes (Hap1–Hap3): Hap1 (ACCA), Hap2 (CTTG), and Hap3 (ATTG). Among them, Hap2 contained 19 soybean accessions with average DR less than 0.1, suggesting Hap2 was the favorable haplotype for resistance to SC3 (Fig. 4b). These results indicated that the genetic diversity association panel and abundant genome-wide SNP markers are both important factors for mapping resolution, which would enable to map previously identified SMV-resistant QTLs and detect novel loci via GWAS.

Prediction of candidate genes and expression patterns analysis

A total of 24 SNPs significantly associated with DR detected in three environments and BLUP. More than two-thirds (19/24) of SNPs were located within or

closed to previously reported loci. Fifteen SNPs on chromosome 13 associated with resistance to SC3 flanked with known major resistance loci *Rsv1* and *Rsv5*. Previous studies reported that the resistant gene in *Rsv1* loci contained typical nucleotide-binding sites and leucine-rich repeat domain. Four genes (*Glyma13g25920*, *Glyma13g25950*, *Glyma13g25970*, and *Glyma13g26000*) homologous to those reported resistant genes *5gG3* and *3gG2* were predicted as candidate genes (Li et al. 2017). This co-location of previously reported SC3 resistance loci with SNPs significantly identified in this study supplied a powerful evidence showing the credible of the GWAS results.

For high-density of SNP markers used in this panel, providing the possibility to predict the candidate genes. A new SNP locus AX-93753793 on chromosome 8 explained 31.69% of the DR variation. We found a total of 28 genes in the LD region (260 kb) of SNP AX-93753793 in the Williams 82 reference genome, then compared the different genetic variation of these genes between Kefeng No.1 and Nannong 1138-2 based on the whole genome re-sequencing data previously reported (Ma et al. 2019). Results showed that AX-93753793 was located in an LD block harboring two candidate genes encoding aldolase-type TIM barrel family protein (*Glyma.08g175800*) and TCP-1/cpn60 chaperonin family protein (*Glyma.08g175900*), respectively. An indel variation in 3'-UTR of *Glyma.08g175800* and fourteen bases of frameshift insertion in the exonic region of *Glyma.08g175900* were found (Table 3). Another unreported SC3 resistance loci was identified to be associated with DR on chromosome 20. In this region, *Glyma.20g190000* encoding 3'-phosphoinositide-

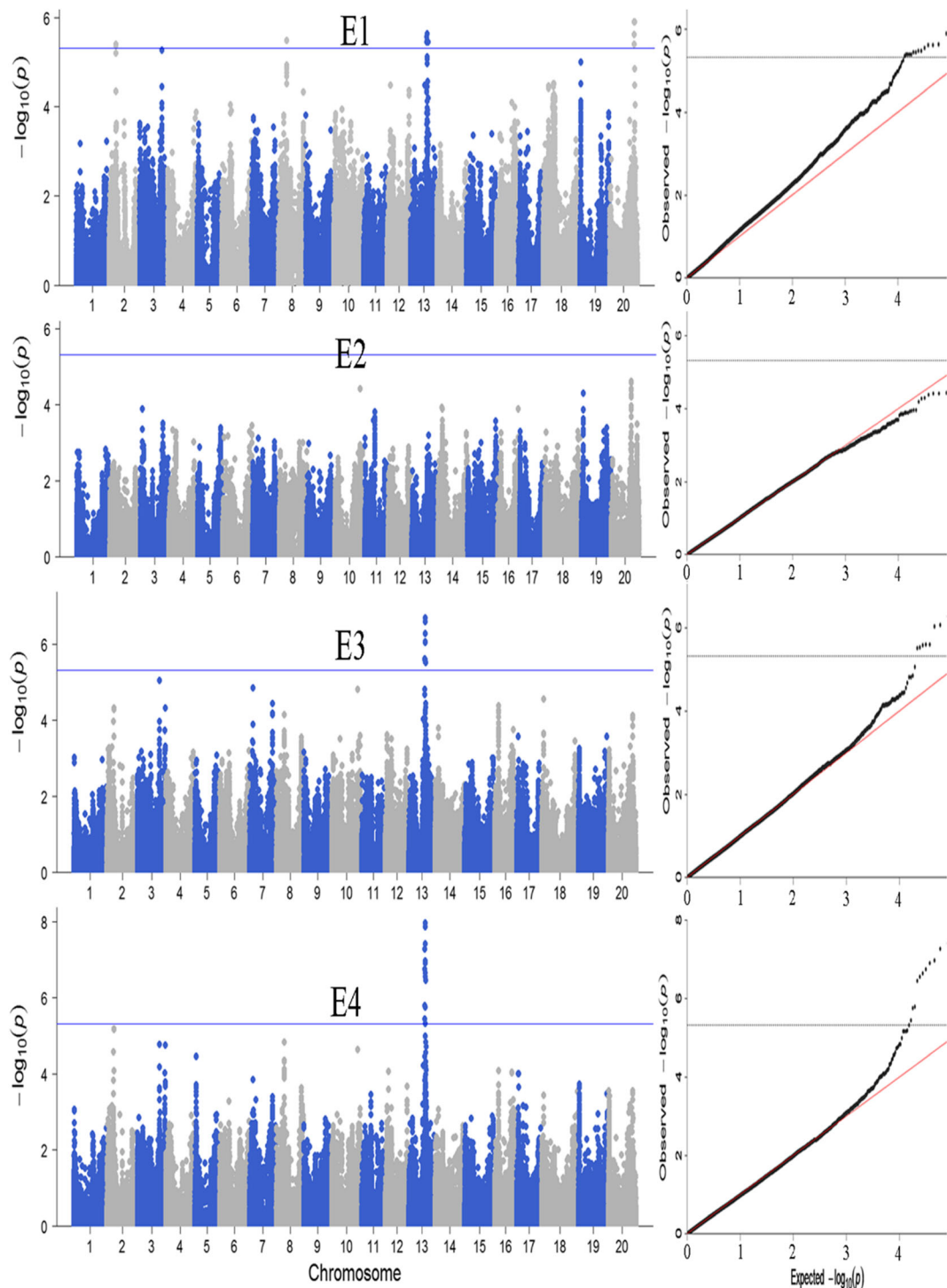


Fig. 2 Manhattan plots and quantile-quantile plots for three environments and BLUP in GWAS results. E, environment, BLUP, best linear unbiased prediction. Horizontal lines indicates Bonferroni-adjusted significance threshold $-\log_{10}(P) \geq 5.32$

dependent protein kinase 1 was regarded as a candidate gene. Two different indel variations in the 5'-UTR region

of *Glyma.20g190000* was identified between Kefeng No.1 and Nannong 1138-2 (Table 3). *Glyma.20g190000* was

Table 2 SNPs significantly associated with disease rate trait in soybean

| SNP | Chr | Position (Williams 82.a1.v1.1) | -Log ₁₀ P | | | | R ² (%) | Related QTLs |
|-------------|-----|--------------------------------|----------------------|----|------|------|--------------------|---|
| | | | E1 | E2 | E3 | BLUP | | |
| AX-93681616 | 2 | 11,779,855 | 5.39 | ns | ns | ns | 31.48 | <i>R_{SC7}</i> (Yan et al. 2015) |
| AX-93977811 | 2 | 11,780,847 | 5.36 | ns | ns | ns | 31.43 | <i>R_{SC8}</i> (Wang et al. 2011b) |
| AX-93681618 | 2 | 11,784,615 | 5.39 | ns | ns | ns | 31.48 | <i>Rsv4</i> (Ishibashi et al. 2019) |
| AX-93681619 | 2 | 11,786,644 | 5.39 | ns | ns | ns | 31.48 | |
| AX-93753793 | 8 | 14,196,133 | 5.48 | ns | ns | ns | 31.69 | Fusarium lesion length 1-1 (Ellis et al. 2012) |
| AX-93813964 | 13 | 27,261,629 | ns | ns | ns | 5.80 | 27.05 | <i>Rsc-pm</i> (Yang et al. 2013) |
| AX-93813987 | 13 | 27,325,009 | 5.56 | ns | 5.59 | 7.27 | 25.58–31.86 | <i>Rsv1</i> (Shi et al. 2008) |
| AX-93814000 | 13 | 27,358,186 | ns | ns | ns | 5.45 | 27.03 | |
| AX-93939503 | 13 | 27,475,067 | 5.44 | ns | 5.60 | 6.75 | 25.58–31.60 | <i>R_{SC11}</i> (Bai et al. 2009) |
| AX-93814205 | 13 | 27,893,804 | ns | ns | 6.07 | 6.97 | 25.54–27.34 | |
| AX-93814208 | 13 | 27,900,033 | ns | ns | 5.53 | 6.65 | 25.55–27.17 | |
| AX-93814209 | 13 | 27,901,549 | 5.63 | ns | 6.69 | 7.87 | 25.54–32.03 | <i>R_{SC18B}</i> (Li et al. 2015) |
| AX-94110993 | 13 | 27,906,097 | 5.47 | ns | 6.27 | 7.41 | 25.55–31.67 | |
| AX-94110994 | 13 | 27,908,496 | 6.31 | ns | 6.59 | 7.96 | 25.69–33.60 | |
| AX-94111015 | 13 | 27,948,177 | ns | ns | 5.60 | 6.55 | 25.74–27.10 | <i>Rsc-ps</i> (Yang et al. 2013) |
| AX-93814235 | 13 | 27,957,401 | ns | ns | ns | 5.32 | 27.25 | |
| AX-93814296 | 13 | 28,101,481 | ns | ns | ns | 5.76 | 27.25 | |
| AX-93643473 | 13 | 28,266,587 | 5.62 | ns | 6.04 | 6.91 | 25.69–32.02 | |
| AX-94111196 | 13 | 28,363,355 | ns | ns | 5.51 | 6.45 | 25.64–27.37 | |
| AX-94111786 | 13 | 29,865,665 | 5.45 | ns | ns | ns | 36.27 | <i>Rsv5</i> (Klepadlo et al. 2017) |
| AX-93957459 | 20 | 41,544,070 | 5.61 | ns | ns | ns | 32.00 | SCN 34-4 (Winter et al. 2007) |
| AX-93661259 | 20 | 41,622,297 | 5.39 | ns | ns | ns | 31.50 | SCN43-6, SCN43-8, SCN 44-12 and SCN 44-8 (Jiao et al. 2015) |
| AX-94207740 | 20 | 41,631,762 | 5.90 | ns | ns | ns | 32.65 | |
| AX-93957473 | 20 | 41,680,482 | 5.90 | ns | ns | ns | 32.65 | |

Chr chromosome, E1 – E3 environment 1 – environment 3, ns not significant, SNP significant at $P \leq 1/207,608 = 4.82e-06$ or $-\log_{10} P \geq 5.32$, R² phenotypic variation explained with the SNP

homologous to *Ospdk1* in rice. *OsPdk1* was reported to regulate basal disease resistance, overexpression of *Ospdk1*-enhanced basal resistance against both fungus, *Magnaporthe oryzae*, and bacterial pathogen, *Xanthomonas oryzae* pv. *oryzae* (Matsui et al. 2010). Therefore, these three genes (*Glyma.08g175800*, *Glyma.08g175900*, and *Glyma.20g190000*) located under the peak SNPs of newly found SC3-resistant loci were identified as candidate genes.

Moreover, elucidating the expression patterns of candidate genes during different soybean tissues would be favorable for understanding the potential functions. The GEO RNA-seq data in public databases was used for expression analysis. Three genes *Glyma.08g175800*, *Glyma.08g175900*, and *Glyma.20g190000* were

expressed in all soybean tissues, including leaf, root, stem, flower, seeding, mature embryo, dry seed, and cotyledon (Fig. 5a). However, they showed different expression patterns: *Glyma.08g175800* had high expression in leaf, flower, and seeding but low in root; *Glyma.08g175900* was expressed in all tissues and highest in leaf; the expression level of *Glyma.20g190000* in flower and cotyledon was twice as high as that in other tissues, and it also showed medium expression level in leaf and seeding. To verify the reliability of expression patterns of candidate genes, another RNA-seq data downloaded from Soybase was examined and acquired similar results (Fig. 5b).

The expression analysis in leaves of three candidate genes with SC3 or mock inoculation was shown in Fig. 6. Comparing with mock inoculation, the

Fig. 3 Regional plots of novel significant association loci on chromosomes 8 and 20. **a** A 260-kb region of local Manhattan plot harboring the significant SNP on chromosome 8. The red dot is the most significant SNP, AX-93753793. **b** The soybean reference genome (Williams 82 a1.V1.1) in a 260-kb region on chromosome 8. Candidate genes were showed in red color. The linkage disequilibrium heat map shown in **c**. **d** Local Manhattan plot in 396 kb region on chromosome 20, four significant SNPs are showed in red dots. The soybean reference genome (e) and linkage disequilibrium heat map (f) in a 396-kb region on chromosome 20, and the candidate gene was showed in red color

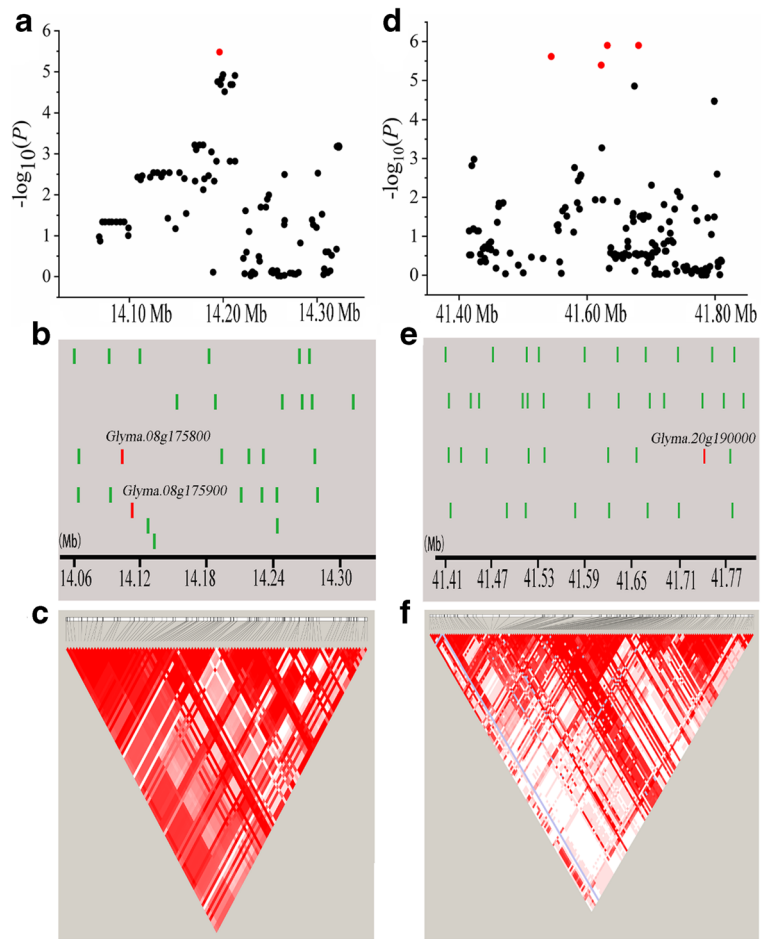


Fig. 4 Superior allele analysis and haplotype analysis of SNPs significantly identified in GWAS. **a** Superior allele analysis on chromosome 8. **b** Haplotype analysis on chromosome 20. Box edges represent 25% quantile (top) and 75% quantile (bottom), bold line is median value, square dot is mean value, and black dot is abnormal value. **, significant at $P \leq 0.01$

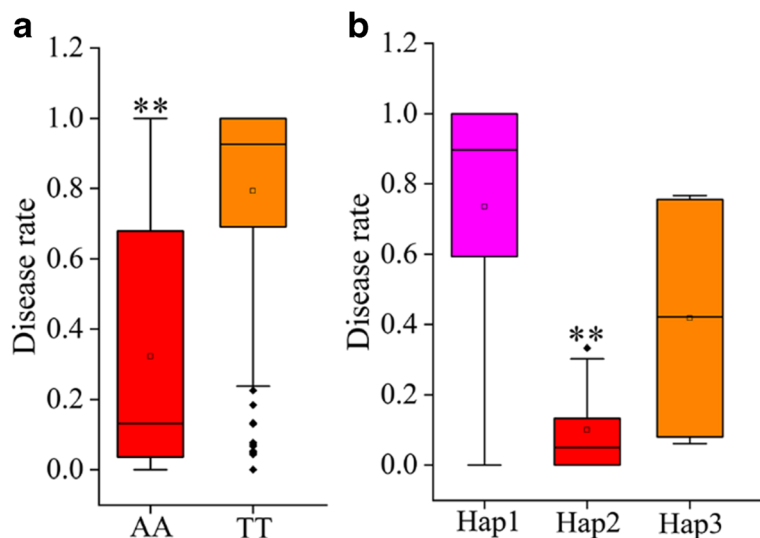


Table 3 Different variations of candidate genes in Kefeng No.1 and Nannong 1138-2

| Gene | Ref | Alt | Region | Variation type |
|------------------------|-------|----------------|--------|----------------------|
| <i>Glyma.08g175800</i> | TGACA | – | 3'-UTR | Deletion |
| <i>Glyma.08g175900</i> | – | GAGCAACGAGCAAC | Exonic | Frameshift insertion |
| <i>Glyma.20g190000</i> | – | GA | 5'-UTR | Insertion |
| | T | – | 5'-UTR | Deletion |

Ref Williams 82 reference genome, Alt Alter variation type

expression level of *Glyma.08g175800* was induced up-regulated in Kefeng No.1 at 1, 2, and 3 days after SC3 inoculating and significantly upregulated in Nannong 1138-2 at 3 days after SC3 inoculating. On the contrary, the expression levels of *Glyma.08g175900* and *Glyma.20g190000* were induced downregulated in both Nannong 1138-2 and Kefeng No.1 at 1, 2, and 3 days with SC3 inoculation than mock inoculation. These three candidate genes were all induced by SC3 virus strain which may suggest some special functions in SMV resistance. The expression analysis was an important basis for further estimating the functional study of SC3 candidate genes.

Discussion

Factors determining association analysis

To improve the resolution of GWAS for the rapid identification of new loci associated with DR to SC3, we employed high-density SNP markers in this study. SNPs have been extensively applied in genetic mapping research for their wide and uniform distribution in genome, reducing the possibility of missing important QTLs. A SoySNP50 K array was used for GWAS to detect QTLs related to soybean seed protein and oil content (Hwang et al. 2014). A total of 62,423 SNPs

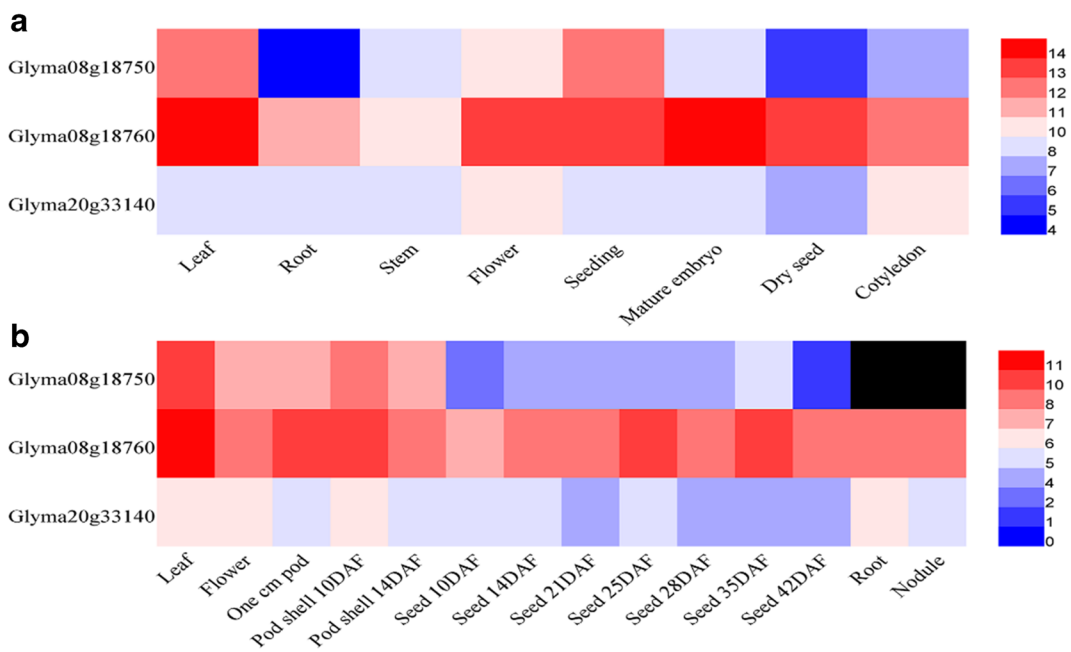


Fig. 5 Digital expression profiles of three candidate genes in different tissues. **a** Expression levels of three genes in eight different tissues from microarray data (GEO Accession number: GSE29163). **b** Expression levels of three genes in different tissues and at different stages based on RNA-seq data. The values used in

heatmaps from the microarray data were \log_2 -transformed (*Glyma08g18750*, *Glyma08g18760* and *Glyma20g33140* were corresponding to *Glyma.08g175800*, *Glyma.08g175900* and *Glyma.20g190000*, respectively, in the a2 version of the soybean Williams 82 reference genome)

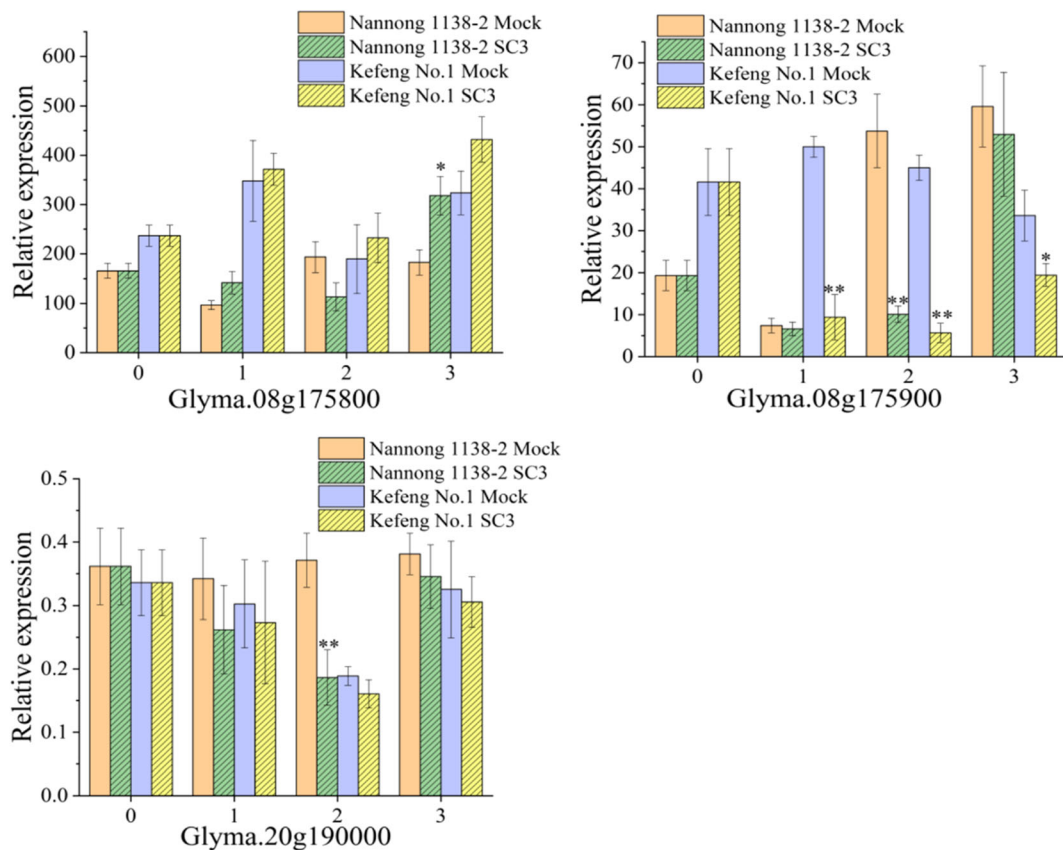


Fig. 6 Expression analysis of three candidate genes in leaves after SC3 inoculating. Error bar indicate the standard deviation. *Significant at $P \leq 0.05$; **Significant at $P \leq 0.01$

with a $MAF \geq 5\%$ were used for GWAS to identify major and minor loci associated with dynamic plant height and number of node traits (Chang et al. 2018). In this study, a higher-density NJAU 355 K SoySNP array containing 292,053 SNPs was used for GWAS. These high-quality SNPs were strictly screened from SNP database which represents the population diversity of wild and cultivated soybean. The 93.6% of SNPs span less than 9 kb, and the average SNP spacing was approximately 3.3 kb among 20 chromosomes (Wang et al. 2016). Based on high-density SNP array, the resolution is much higher compared with 1536 SNPs used in previous study (Liu et al. 2016). Genotyping with high-density markers and phenotyping in multiple environments can improve the ability of GWAS to detect potential SC3 resistance loci.

LD is a theoretical basis for mapping candidate loci or genes in GWAS. A much smaller LD value in out-pollination crop than self-pollination crop, because the genome recombination rate was increasing and

accumulating by generation in cross-pollinated crop. As out-pollination crop, the LD was evaluated only 2 kb in maize lines and such small LD value improved the localization accuracy to the gene level (Gore et al. 2009). Soybean is a strictly self-pollination crop so that its genome exhibits a strong LD, which was estimated about 500 kb in previous report (Hyten et al. 2007). Based on NJAU 355 K SoySNP array, LD analysis was performed in a large population containing 367 soybean accessions, the LD which decays to half of its maximum value was 130 kb in cultivated soybeans (Wang et al. 2016). In this study, the relatively small LD value enhances mapping resolution and significantly shrinks genomic region to effectively identify candidate gene underlying SNP loci. Population structure is another important factor that determines the power of GWAS. A complicated population structure can lead to unequal allele frequencies in the subpopulations, which may result in false SNP loci associated with target traits. In this study, we selected a panel with no high population

structure (Wang et al. 2016). The same population have been used for identifying genetic loci of water-soluble protein content trait via GWAS (Zhang et al. 2017). As shown in the quantile-quantile and Manhattan plots (Fig. 2) for BLUP and three individual environments, we identified significant SNP signals for DR after employing a CMLM model in accounting for population structure and genetic relatedness.

Novel SNP loci for DR

Development of soybean accessions with broad-spectrum resistance to diverse SMV strains, as well as identification of novel resistant loci and genes, is necessary for reducing damages caused by SMV. Recently, the cloning of a major gene *Rsv4* (Ishibashi et al. 2019) with broad resistance to potyvirus has provided increasing knowledge to understand the genetic variation and resistance mechanism in soybean defense response against SMV. The genomic regions of *Rsv4* or *R_{SC8}*, *R_{SC7}* significantly associated with SMV resistance, have been previously fine-mapped in a closed region on chromosome 2 (Ishibashi et al. 2019; Wang et al. 2011b; Yan et al. 2015). Thus, it is not surprising that four SNPs located in the *Rsv4*-linked regions were also identified in our study. In addition, fifteen SNPs located in a 2.6-Mb region identified in our study are close to *Rsv1* (Shi et al. 2008), *Rsv5* (Klepadlo et al. 2017), *Rsc-pm* (Yang et al. 2013), and *Rsc-ps* (Yang et al. 2013), and these loci also implied a contribution to broad SMV resistance. However, studies focusing on *Rsv1* loci are difficult due to many *R* genes (contained NBS-LRR domains) flock to each other or to multiple alleles of the same gene (Yang et al. 2013; Ma et al. 2016). Previous reports indicate that *5gG3* and *3gG2* were strong candidate genes tightly linked to *Rsv1* loci in PI 96983 (Hayes et al. 2004; Gore et al. 2002). A recombinant line only containing *3gG2* gene inherited from PI 96983 conferred resistance to certain SMV strains as well as cultivars Marshall and Ogden; thereby, it is speculated that diverse *Rsv1* alleles in differ cultivars were due to recombination of major R genes (Hayes et al. 2004). Therefore, understanding structural variations caused by recombination in *Rsv1* region with the help of sequencing technique may figure out the genetic architecture in this genomic region.

In addition, loci located on chromosomes 8 and 20 significantly identified in this study were novel loci with average R^2 values as high as 32.10%. These five SNPs

were located in the previous reported biotic stress-related QTLs of *Fusarium lesion length 1-1* (Ellis et al. 2012), *SCN 34-4* (Winter et al. 2007), *SCN 43-6*, *SCN 43-8*, *SCN 44-12*, and *SCN 44-8* (Jiao et al. 2015), indicating that the two novel loci on chromosomes 8 and 20 were critical in the resistance of pathogens and soybean cyst nematode. We identified favorable allele and haplotypes in these two novel loci (Fig. 4). A few accessions in Hap1 have low DR, perhaps possessing other resistant loci. The average DR in Hap1 is 0.73, which suggested that most accessions in Hap1 were susceptible to SMV (Fig. 4b). Nineteen accessions in Hap2 were identified as resistant germplasm with average disease rates less than 0.1. These results will provide insights for accelerating the development of molecular marker and breeding for SMV resistance.

Candidate resistance genes for SC3

Most researchers reported resistant loci such as *Rsv1*, which are acquired from a limited number of resistant soybean cultivars (Kiihl and Hartwig 1979), whereas exploration in large population and identification of novel genes are lagging. Hence, novel resistant genes identified via GWAS in soybean germplasm resource population are necessary for SMV management. In contrast with *Rsv1* and *Rsv5* loci, no NBS-LRR genes were found in the region of two novel loci identified in our study, suggesting these loci likely contained new classes of resistance genes. Three candidate genes were identified underlying the two novel loci. *Glyma.08g175800* is annotated as an aldolase in hydrogen peroxide (H_2O_2) biosynthetic process and highly expressed in leaf. A lethal systemic hypersensitive response caused by reactive oxygen lead to necrosis in leaf (Hernández et al. 2016). Local cell death caused by reactive oxygen can not only prevent SMV from spreading in adjacent cells, but also as a diffusible signal to induce antiviral production and pathogenesis-related defense responses in adjacent plant tissues (Chen et al. 2017). In addition, *Glyma.08g175900*, a member of TCP-1/cpn60 chaperonin family protein, has been reported to play important role in cell apoptosis and the immune responses of human cells (Tsan and Gao 2009). Furthermore, its homologous proteins in *Arabidopsis* and maize were significant upregulated after pathogen infection (Hwang et al. 2012; Wu et al. 2013). High expression level of *Glyma.08g175900* in soybean leaf might suggest a potential role in SMV infection defense. Another

gene *Glyma.20g190000* annotated as a 3'-phosphoinositide-dependent protein kinase has not been studied in soybean, but its homologous gene *PDK1* in tomato is involved in the regulation of cell death. Silencing *PDK1* induces whole plant death in tomato, and further research showed that *PDK1* phosphorylates AGC kinase negatively regulates cell death (Devarenne et al. 2006). Moreover, overexpression of *Ospdk1* in rice enhanced defensive reaction against pathogens (Matsui et al. 2010). These novel candidate genes identified in this study provided valuable gene resources to reveal the resistance mechanism to SMV.

Author contributions DY designed the experiment. HY performed the experiment. ZC performed data analysis. ZC wrote the manuscript, and HL, HY, HD, YY, BL, and DY revised it. All authors read and approved the final manuscript.

Funding This work was supported in part by Ministry of Science and Technology (2017YFE0111000), Ministry of Agriculture (2016ZX08004-003), National Natural Science Foundation of China (31871649), and China Postdoctoral Science Foundation (2019M662842)

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

- Adams MJ, Antoniw JF, Fauquet CM (2005) Molecular criteria for genus and species discrimination within the family Potyviridae. *Arch Virol* 150:459–479. <https://doi.org/10.1007/s00705-004-0440-6>
- Bai L, Li HC, Ma Y et al (2009) Inheritance and gene mapping of resistance to soybean mosaic virus strain SC-11 in soybean. *Soybean Sci* 28:1–6. <https://doi.org/10.1111/j.1744-7909.2006.00365.x>
- Barrett JC, Fry B, Maller J, Daly MJ (2005) Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 21:263–265. <https://doi.org/10.1093/bioinformatics/bth457>
- Buss GR, Ma G, Chen P, Tolin SA (1997) Registration of LN92-12033 and LN92-12054 soybean germplasm lines near-isogenic for brown stem rot resistance gene Rbs2. *Crop Sci* 37:1987–1987. <https://doi.org/10.2135/cropsci1997.0011183x003700060067x>
- Cai C, Jiang X, Zhao C, Ma J (2014) Sequence analysis of the coat protein gene of Chinese soybean mosaic virus strain SC7 and comparison with those of SMV strains from the USA. *Bing Du Xue Bao* 30:489–494. <https://doi.org/10.13242/j.cnki.bingduxuebao.002537>
- Chang F, Guo C, Sun F, Zhang J, Wang Z, Kong J, He Q, Sharmin RA, Zhao T (2018) Genome-wide association studies for dynamic plant height and number of nodes on the main stem in summer sowing soybeans. *Front Plant Sci* 9:1–13. <https://doi.org/10.3389/fpls.2018.01184>
- Che Z, Liu H, Yi F, Cheng H, Yang Y, Wang L, du J, Zhang P, Wang J, Yu D (2017) Genome-wide association study reveals novel loci for SC7 resistance in a soybean mutant panel. *Front Plant Sci* 8:1–12. <https://doi.org/10.3389/fpls.2017.01771>
- Chen H, Adam Arsovski A, Yu K, Wang A (2017) 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* 18:391–404. <https://doi.org/10.1111/mpp.12407>
- Cho E-K, Goodman RM (1979) Strains of soybean mosaic virus: classification based on virulence in resistant soybean cultivars. *Phytopathology* 69:467. <https://doi.org/10.1094/Phyto-69-467>
- Chu S, Wang J, Zhu Y, Liu S, Zhou X, Zhang H, Wang CE, Yang W, Tian Z, Cheng H, Yu D (2017) An R2R3-type MYB transcription factor, GmMYB29, regulates isoflavone biosynthesis in soybean. *PLoS Genet* 13:1–26. <https://doi.org/10.1371/journal.pgen.1006770>
- Deng W, Wang Y, Liu Z, Cheng H, Xue Y (2014) HemI: a toolkit for illustrating heatmaps. *PLoS One* 9:9–13. <https://doi.org/10.1371/journal.pone.0111988>
- Devarenne TP, Ekengren SK, Pedley KF, Martin GB (2006) Adi3 is a Pdk1-interacting AGC kinase that negatively regulates plant cell death. *EMBO J* 25:255–265. <https://doi.org/10.1038/sj.emboj.7600910>
- Ellis M, Wang H, Paul P et al (2012) Identification of soybean genotypes resistant to and genetic mapping of resistance quantitative trait loci in the cultivar Conrad. *Crop Sci* 52:2224–2233. <https://doi.org/10.2135/cropsci2011.11.0624>
- Gong ZZ (2020) Flowering phenology as a core domestication trait in soybean. *J Integr Plant Biol* 62:546–549. <https://doi.org/10.1111/jipb.12934>
- Gore MA, Hayes AJ, Jeong SC, Yue YG, Buss GR, Maroof MAS (2002) Mapping tightly linked genes controlling potyvirus infection at the Rsv1 and Rpv1 region in soybean. *Genome* 45:592–599. <https://doi.org/10.1139/g02-009>
- Gore MA, Chia JM, Elshire RJ, Sun Q, Ersoz ES, Hurwitz BL, Peiffer JA, McMullen MD, Grills GS, Ross-Ibarra J, Ware DH, Buckler ES (2009) A first-generation haplotype map of maize. *Science* 326:1115–1117. <https://doi.org/10.1126/science.1177837>
- Guo D, Zhi H, Wang Y et al (2005) Identification and distribution of strains of soybean mosaic virus in middle and northern of HuangHuai Region of China. *Chin J Oil Crop Sci* 27:64–68. <https://doi.org/10.3321/j.issn:1007-9084.2005.04.015>
- Hayes AJ, Ma G, Buss GR, Maroof MAS (2000) Molecular marker mapping of RSV4, a gene conferring resistance to all known strains of soybean mosaic virus. *Crop Sci* 40:1434–1437. <https://doi.org/10.2135/cropsci2000.4051434x>
- Hayes AJ, Jeong SC, Gore MA, Yu YG, Buss GR, Tolin SA, Maroof MAS (2004) Recombination within a nucleotide-binding-site/leucine-rich-repeat gene cluster produces new variants conditioning resistance to soybean mosaic virus in soybeans. *Genetics* 166:493–503. <https://doi.org/10.1534/genetics.166.1.493>
- Hernández JA, Gullner G, Clemente-Moreno MJ, Künstler A, Juhász C, Díaz-Vivancos P, Király L (2016) Oxidative stress

- and antioxidative responses in plant–virus interactions. *Physiol Mol Plant Pathol* 94:134–148. <https://doi.org/10.1016/j.pmp.2015.09.001>
- Hwang IS, Kim NH, Choi DS, Hwang BK (2012) Overexpression of *Xanthomonas campestris* pv. *vesicatoria* effector AvrBsT in *Arabidopsis* triggers plant cell death, disease and defense responses. *Planta* 236:1191–1204. <https://doi.org/10.1007/s00425-012-1672-4>
- Hwang EY, Song Q, Jia G, Specht JE, Hyten DL, Costa J, Cregan PB (2014) A genome-wide association study of seed protein and oil content in soybean. *BMC Genomics* 15:1–12. <https://doi.org/10.1186/1471-2164-15-1>
- Hyten DL, Choi IY, Song Q, Shoemaker RC, Nelson RL, Costa JM, Specht JE, Cregan PB (2007) Highly variable patterns of linkage disequilibrium in multiple soybean populations. *Genetics* 175:1937–1944. <https://doi.org/10.1534/genetics.106.069740>
- Ishibashi K, Saruta M, Shimizu T, Shu M, Anai T, Komatsu K, Yamada N, Katayose Y, Ishikawa M, Ishimoto M, Kaga A (2019) Soybean antiviral immunity conferred by dsRNase targets the viral replication complex. *Nat Commun* 10:1–3. <https://doi.org/10.1038/s41467-019-12052-5>
- Jiao Y, Vuong TD, Liu Y, Meinhardt C, Liu Y, Joshi T, Cregan PB, Xu D, Shannon JG, Nguyen HT (2015) Identification and evaluation of quantitative trait loci underlying resistance to multiple HG types of soybean cyst nematode in soybean PI 437655. *Theor Appl Genet* 128:15–23. <https://doi.org/10.1007/s00122-014-2409-5>
- 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 Gen Genomics* 292:811–822. <https://doi.org/10.1007/s00438-017-1310-8>
- Kiihl RAS, Hartwig EE (1979) Inheritance of reaction to soybean mosaic virus in soybeans 1. *Crop Sci* 19:372–375. <https://doi.org/10.2135/cropsci1979.0011183x001900030024x>
- Klepadlo M, Chen P, Shi A, Mason RE, Korth KL, Srivastava V, Wu C (2017) Two tightly linked genes for soybean mosaic virus resistance in soybean. *Crop Sci* 57:1844–1853. <https://doi.org/10.2135/cropsci2016.05.0290>
- Knapp SJ, Stroup WW, Ross WM (1985) Exact confidence intervals for heritability on a progeny mean basis 1. *Crop Sci* 25: 192–194. <https://doi.org/10.2135/cropsci1985.0011183x002500010046x>
- Li K, Yang QH, Zhi HJ, Gai JY (2010) Identification and distribution of soybean mosaic virus strains in southern China. *Plant Dis* 94: 351–357. <https://doi.org/10.1094/PDIS-94-3-0351>
- Li C, Yang Y, Wang D et al (2012) Studies on mapping and inheritance of resistance genes to SMV strain SC10 in soybean. *Sci Agric Sin* 45:4335–4342. <https://doi.org/10.3864/j.issn.0578-1752.2012.21.001>
- Li K, Ren R, Adhimoolum K, Gao L, Yuan Y, Liu Z, Zhong Y, Zhi H (2015) Genetic analysis and identification of two soybean mosaic virus resistance genes in soybean [*Glycine max* (L.) Merr]. *Plant Breed* 134:684–695. <https://doi.org/10.1111/pbr.12315>
- Li C, Adhimoolum K, Yuan Y, Yin J, Ren R, Yang Y, Zhi H (2017) Identification of candidate genes for resistance to soybean mosaic virus strain SC3 by using fine mapping and transcriptome analyses. *Crop Pasture Sci* 68:156–166. <https://doi.org/10.1071/CP16353>
- Lipka AE, Tian F, Wang Q, Peiffer J, Li M, Bradbury PJ, Gore MA, Buckler ES, Zhang Z (2012) GAPIT: genome association and prediction integrated tool. *Bioinformatics* 28:2397–2399. <https://doi.org/10.1093/bioinformatics/bts444>
- Liu JZ, Whitham SA (2013) Overexpression of a soybean nuclear localized type-III DnaJ domain-containing HSP40 reveals its roles in cell death and disease resistance. *Plant J* 74:110–121. <https://doi.org/10.1111/tj.12108>
- Liu H, Che Z, Zeng X, Zhang G, Wang H, Yu D (2016) Identification of single nucleotide polymorphisms in soybean associated with resistance to common cutworm (*Spodoptera litura* Fabricius). *Euphytica* 209:49–62. <https://doi.org/10.1007/s10681-016-1631-4>
- Lu SJ, Dong LD, Fang C, Liu S, Kong L, Cheng Q, Chen L, Su T, Nan H, Zhang D, Zhang L, Wang Z, Yang Y, Yu D, Liu X, Yang Q, Lin X, Tang Y, Zhao X, Yang X, Tian C, Xie Q, Li X, Yuan X, Tian Z, Liu B, Weller JL, Kong F (2020) Stepwise selection on homeologous PRR genes controlling flowering and maturity during soybean domestication. *Nat Genet* 52:428–436. <https://doi.org/10.1038/s41588-020-0604-7>
- Ma Y, Wang DG, Li HC, Zheng GJ, Yang YQ, Li HW, Zhi HJ (2011) Fine mapping of the RSC14Q locus for resistance to soybean mosaic virus in soybean. *Euphytica* 181:127–135. <https://doi.org/10.1007/s10681-011-0457-3>
- Ma FF, Wu XY, Chen YX, Liu YN, Shao ZQ, Wu P, Wu M, Liu CC, Wu WP, Yang JY, Li DX, Chen JQ, Wang B (2016) 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* 129:2227–2236. <https://doi.org/10.1007/s00122-016-2769-0>
- Ma Y, Ma W, Hu D, Zhang X, Yuan W, He X, Kan G, Yu D (2019) QTL mapping for protein and sulfur-containing amino acid contents using a high-density bin-map in soybean (*Glycine max* L. Merr.). *J Agric Food Chem* 67:12313–12321. <https://doi.org/10.1021/acs.jafc.9b04497>
- Matsui H, Miyao A, Takahashi A, Hirochika H (2010) Pdk1 kinase regulates basal disease resistance through the OsOx1l-OsPti1a phosphorylation cascade in rice. *Plant Cell Physiol* 51:2082–2091. <https://doi.org/10.1093/pcp/pcq167>
- Merk HL, Yames SC, Van Deynze A et al (2012) Trait diversity and potential for selection indices based on variation among regionally adapted processing tomato Germplasm. *J Am Soc Hortic Sci* 137:427–437. <https://doi.org/10.21273/jashs.137.6.427>
- Pu Z, Cao Q, Fang D et al (1983) Resistance in soybean varieties to six strains of soybean mosaic virus. *J Nanjing Agric Univ* 3:41–45. <https://doi.org/10.7685/j.issn.1000-2030.1983.03.006>
- R Core Team (2018) R: a language and environment for statistical computing. In: R Foundation for statistical computing. Austria. Available online at, Vienna <https://www.R-project.org/>
- Ren R, Liu S, Karthikeyan A et al (2017) Fine-mapping and identification of a novel locus Rsc15 underlying soybean resistance to soybean mosaic virus. *Theor Appl Genet* 130: 2395–2410. <https://doi.org/10.1007/s00122-017-2966-5>
- Shi A, Chen P, Li DX, Zheng C, Hou A, Zhang B (2008) Genetic confirmation of 2 independent genes for resistance to soybean mosaic virus in J05 soybean using SSR markers. *J Hered* 99:598–603. <https://doi.org/10.1093/jhered/esn035>
- Tsan M-F, Gao B (2009) Heat shock proteins and immune system. *J Leukoc Biol* 85:905–910. <https://doi.org/10.1189/jlb.0109005>

- Wang X, Gai J, Pu Z (2003) Classification and distribution of strains of soybean mosaic virus in middle and lower Huanghuai and Changjiang river valleys. *Soybean Sci* 22:102–107. <https://doi.org/10.3969/j.issn.1000-9841.2003.04.001>
- Wang D, Ma Y, Liu N, Yang Z, Zheng G, Zhi H (2011a) Fine mapping and identification of the soybean RSC4 resistance candidate gene to soybean mosaic virus. *Plant Breed* 130: 653–659. <https://doi.org/10.1111/j.1439-0523.2011.01888.x>
- Wang D, Ma Y, Yang Y, Liu N, Li C, Song Y, Zhi H (2011b) Fine mapping and analyses of RSC8 resistance candidate genes to soybean mosaic virus in soybean. *Theor Appl Genet* 122: 555–565. <https://doi.org/10.1007/s00122-010-1469-4>
- Wang J, Chu S, Zhang H, Zhu Y, Cheng H, Yu D (2016) Development and application of a novel genome-wide SNP array reveals domestication history in soybean. *Sci Rep* 6:1–10. <https://doi.org/10.1038/srep20728>
- Winter SMJ, Shelp BJ, Anderson TR, Welacky TW, Rajcan I (2007) QTL associated with horizontal resistance to soybean cyst nematode in Glycine soja PI464925B. *Theor Appl Genet* 114:461–472. <https://doi.org/10.1007/s00122-006-0446-4>
- Wu L, Han Z, Wang S, Wang X, Sun A, Zu X, Chen Y (2013) Comparative proteomic analysis of the plant-virus interaction in resistant and susceptible ecotypes of maize infected with sugarcane mosaic virus. *J Proteome* 89:124–140. <https://doi.org/10.1016/j.jprot.2013.06.005>
- Yan H, Wang H, Cheng H, Hu Z, Chu S, Zhang G, Yu D (2015) Detection and fine-mapping of SC7 resistance genes via linkage and association analysis in soybean. *J Integr Plant Biol* 57:722–729. <https://doi.org/10.1111/jipb.12323>
- Yang Y, Zheng G, Han L, Dagang W, Yang X, Yuan Y, Huang S, Zhi H (2013) 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* 126: 1783–1791. <https://doi.org/10.1007/s00122-013-2092-y>
- Yano K, Yamamoto E, Aya K, Takeuchi H, Lo PC, Hu L, Yamasaki M, Yoshida S, Kitano H, Hirano K, Matsuoka M (2016) Genome-wide association study using whole-genome sequencing rapidly identifies new genes influencing agronomic traits in rice. *Nat Genet* 48:927–934. <https://doi.org/10.1038/ng.3596>
- Zhan Y, Zhi H-j, Yu D, Gai J (2015) Identification and distribution of SMV strains in Huang-Huai Valleys. *Sci Agric Sin* 39: 2009–2015. <https://doi.org/10.3321/j.issn:0578-1752.2006.10.009>
- Zhang D, Lü H, Chu S, Zhang H, Zhang H, Yang Y, Li H, Yu D (2017) The genetic architecture of water-soluble protein content and its genetic relationship to total protein content in soybean. *Sci Rep* 7:1–13. <https://doi.org/10.1038/s41598-017-04685-7>
- Zheng G-j, Yang Y q, Ma Y et al (2014) Fine mapping and candidate gene analysis of resistance gene RSC3Q to soybean mosaic virus in Qihuang 1. *J Integr Agric* 13:2608–2615. [https://doi.org/10.1016/S2095-3119\(13\)60738-8](https://doi.org/10.1016/S2095-3119(13)60738-8)

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.