

Genome-wide Association of Single Nucleotide Polymorphism Loci and Candidate Genes for Frog-eye Leaf Spot (*Cercospora Sojina*) Resistance in Soybean

Xin Gu

Heilongjiang Academy of Agricultural Sciences

Shanshan Huang

Beidahuang Kenfeng Seed Co., Ltd

Yansong Ma

Beidahuang Kenfeng Seed Co., Ltd

Xiaohe Yang

Heilongjiang Academy of Agricultural Sciences

Liangliang Yao

Heilongjiang Academy of Agricultural Sciences

Xuedong Gao

Heilongjiang Academy of Agricultural Sciences

Maoming Zhang

Heilongjiang Academy of Agricultural Sciences

Wei Liu

Heilongjiang Academy of Agricultural Sciences

Lei Qiu

Heilongjiang Academy of Agricultural Sciences

Haihong Zhao

Heilongjiang Academy of Agricultural Sciences

Qingsheng Wang

Heilongjiang Academy of Agricultural Sciences

Zengjie Li

Heilongjiang Academy of Agricultural Sciences

Zhimin Li

Heilongjiang Academy of Agricultural Sciences

Qingying Meng

Heilongjiang Academy of Agricultural Sciences

Shuai Yang

Heilongjiang Academy of Agricultural Sciences

Chao Wang

Beidahuang Kenfeng Seed Co., Ltd

Xiping Hu

Beidahuang Kenfeng Seed Co., Ltd

Junjie Ding (✉ me999@126.com)

Heilongjiang Academy of Agricultural Sciences

Research Article

Keywords: Soybean, Cercospora sojina, Genome-wide association study, resistant haplotype, frog-eye leaf spot resistant genes

Posted Date: June 1st, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-564949/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published at BMC Plant Biology on December 1st, 2021. See the published version at <https://doi.org/10.1186/s12870-021-03366-y>.

Abstract

Background: Frogeye leaf spot (FLS) is a destructive fungal disease in soybean production. The most economical and effective strategy to control FLS is to use resistant cultivars. However, the use of a limited number of resistant loci in FLS management will be countered by the emergence of new, and more virulent, *Cercospora sojina* races. Thus, we identified quantitative trait loci (QTL) that control resistance to FLS and identified novel resistant genes using a genome-wide association study (GWAS) on 234 Chinese soybean cultivars.

Results: A total of 30,890 single nucleotide polymorphism (SNP) markers were used to estimate linkage disequilibrium (LD) and population structure. The GWAS results identified six loci ($p < 0.001$) distributed over chromosomes (Chr.) 2, 5, 20 comprising 69 SNPs significantly associated with FLS resistance. No previous studies have reported resistance loci in these regions. Subsequently, 55 genes in three resistance-related haplotype blocks were annotated. *Glyma20g31630* encoding pyruvate dehydrogenase (PDH), *Glyma05g28980* encoding mitogen-activated protein kinase 7 (MPK7), and *Glyma20g31510*, *Glyma20g31520* encoding the calcium-dependent protein kinase 4 (CDPK4) in the haplotype blocks were found to be associated with FLS resistance.

Conclusions: The proteins encoded by these four genes directly, or indirectly participate in the biological pathway of salicylic acid (SA) and jasmonic acid (JA). These two plant hormones can induce the expression of disease resistance related genes and are essential for plant systemic acquired resistance (SAR). Therefore, the novel loci and candidate genes provide genomic location information for further advances in FLS resistance gene cloning and molecular breeding.

Background

In most soybean-growing countries, soybean is prone to many plant diseases. Among these, frogeye leaf spot (FLS) caused by the fungus *Cercospora sojina* Hara (*C. sojina*) is one of the most economically harmful [1]. In the main soybean-producing areas of Northeast China [2], the local temperature and leaf wetness periods are very suitable for the occurrence of FLS and for epidemics [3]. Because the character of the disease is polycyclic, a suitable environment and improper control measures can result in outbreaks [4]. Although FLS can be controlled by fungicides, there are challenges, such as fungicide resistance and environmental pollution [5]. Thus, the development of resistant cultivars is a preferred disease management strategy. However, the disadvantage of this strategy is that the resistance of cultivars may be quickly lost. The main reason for this is that resistance mechanisms can be overcome by the emergence of new *C. sojina* pathotypes [6]. Therefore, high-density markers and methods to counter new pathotype races are key for disease resistance breeding.

Rcs1 [7] (Resistant to *C. sojina* Race 1), *Rcs2* [8], and *Rcs3* [9] have been recognized by the Soybean Committee [1]. Although the selection pressures on *C. sojina* populations caused by planting resistant cultivars have produced FLS strains that have overcome the *Rcs1* and *Rcs2* genes [9], the *Rcs3* gene

continues to confer resistance against most races of *C. sojina* in the U.S.A. [10]. Quantitative trait loci (QTL) were mapped close to a known resistance gene cluster on soybean linkage group (LG) J by restriction fragment length polymorphism (RFLP) and simple sequence repeats (SSR) [11]. The gene was mapped in the 2-centimorgan (cM) interval between SSR markers Satt244 and Satt547. Several *Phialophora gregata* resistance genes, such as *Rbs1*, *Rbs2*, and *Rbs3*, conferring resistance against brown stem rot have also been mapped within 10 cM of the *Rcs3* gene [12-13]. Two SSR markers associated with *Rcs3* were confirmed in 64 soybean cultivars and breeding lines of "Davis" ancestors and progenies [14]. However, SSR markers are less suitable for high-throughput and association studies because they are based on repeat length variants, electrophoresis detection methods, and the occurrence of SSR alleles of identical size but different evolutionary origins [15]. Several single nucleotide polymorphism (SNP) and insertion or deletion (Indel) markers associated with *Rcs3* have been evaluated. A total of 19 SNPs/Indels were identified and verified, but only 11 SNPs/Indels were located close to *Rcs3* in an F₂ population of "Davis" × "Blackhawk." The *Rcs3* gene was located close to Satt244 and 0.50 cM from SNPs AZ573TA150 and AZ573CA393. Neither SNPs nor Indels had a direct effect on the phenotype of *Rcs3*, but 11 SNPs were located in the 3-cM interval around *Rcs3* [16]. Subsequently, five plant introductions (PIs) from China (PI594619, PI594661, PI594662A, PI594774, and PI594891) were found to be resistant to FLS in a broad soybean spectrum [17]. Bulk segregant analysis (BSA) results showed that the resistance genes in PI594619 and PI594662A were located near Satt501 on chromosome (Chr.) 18 (LG-G) and near Satt547 and Satt244 on Chr. 16 (LG-J), respectively. A resistance gene in PI594661 was found near Satt244 on Chr. 16 (LG-J). In addition, the resistance of PI594891 and PI594774 to FLS is controlled by two single dominant genes on Chr. 13 near Satt114, which is different from the *Rcs3* allele on Chr. 18. They had a high level of resistance similar to the *Rcs3* gene in "Davis" among the reported FLS resistance genes [17]. The analysis of lines with key recombination events was used to narrow down the FLS-resistance genomic region of PI594891 from 3.3 Mb to 72.6 kb with five annotated genes. The resistance gene of PI594774 was fine-mapped into a 540-kb region, including the 72.6 kb of PI594891. Five candidate genes of PI594891 were sequenced, and multiple mutations in the promoter, intron, and 5' and 3' UTR regions were found [18].

Genome-wide association studies (GWASs) can mine alleles using phenotypic variation and recombination in population evolution. This is significantly related to the variation of target traits at the whole genome level without constructing a mapping population [19]. GWASs have been widely used to identify resistance genes in soybean. Che [20] performed a GWAS on a collection of 219 soybean breeding lines for resistance to the SC3 strain of soybean mosaic virus (SMV). A total of 24 SNPs were identified, which accounted for 25.54%-33.60% of the phenotypic variation. Many SNPs were found close to *Rsv1*, *Rsv4*, and *Rsv5*. To study soybean resistance to *Sclerotinia* Stem Rot (SSR) by GWASs, Wei [21] genotyped and sequenced 420 soybean lines. Two computational models (Compressed Mixed Linear Model with Genome Association and Prediction Integrated Tool, the Fixed and Random Model Circulating Probability Unification) were used to identify SNPs with significant association to disease response. These similar, but different, results indicated that the use of different computational models can affect the SNP association to SSR resistance. A total of 125 genes located in the linkage disequilibrium (LD) of

Chr. 1, Chr. 11, and Chr. 18 were identified by the two models. The candidate genes in these LD blocks can encode isochorismate synthase and regulate host cell death pathways. GWASs can investigate a set of genetic unstructured genotypes. More precise QTLs can be detected in case of sufficiently used genetic markers. Vuong [22] studied a set of 553 soybean accessions and genotyped and identified 14 loci distributed on different chromosomes, including 60 SNPs significantly associated with soybean cyst nematode (SCN) resistance. GWAS results also confirmed six QTLs and eight novel QTLs that were previously mapped using bi-parental populations. Many studies have shown that GWASs based on LD can be used to examine the genetic variation for soybean disease resistance traits, and it is a main tool to identify candidate resistance genes by associating the phenotypic trait to its underlying genetics [23]. However, there are no reports of GWASs being used to study resistance to *C. sojina* in soybean.

In this study, we analyzed 234 soybean accessions, using the genotype data of SoySNP180K BeadChips. We used a mixed linear model to evaluate the genetic effects of resistance sites and to identify resistance-related SNP loci and haplotypes. The possible candidate resistance genes were annotated. This is an effective approach for molecular breeding and identification of disease resistance mechanisms.

Methods

Plant material and phenotypic evaluation

A population of 234 soybean accessions collected from three different breeding departments in northeast China was used for GWAS. The accessions are public and available for non-commercial purposes. The FLS strain used was *C. sojina* Race15 which was identified by Chinese differential cultivars stored in liquid nitrogen. All soybean accessions were provided by the Beidahuang Kenfeng Seed Co., Ltd. and *C. sojina* strains were provided by the Jiamusi Branch of Heilongjiang Academy of Agricultural Sciences.

The valuation experiments of FLS incidence were conducted during the growth season from 2017 to 2019. Before spray inoculation, the strain was purified by single-spore isolation and then cultured on V8 juice agar (V8 juice 200 mL/L, CaCO₃ 3.0 g/L, agar 15.0 g/L, sterile water to 1 L) [24], After gently washing down from the medium by using sterile water and filtering through sterilized multi-layer gauze, the conidial suspensions were adjusted to a concentration of 6×10^4 conidia mL⁻¹. At the V2~V3 growth stage, one trifoliolate leaf per soybean seedling was inoculated with 0.3 mL of the conidial suspension on the upper leaf surface. The inoculated seedlings were then transferred to a humidity chamber at 26 to 28°C for 72 h. At 14 and 21 d after inoculation, resistance assessment was evaluated primarily on leaf spot size and the number of spots formed on the most severely affected trifoliolate leaves. The disease evaluation scale was as follows: spots ≤ 2 mm diameter, or ≤ 10 spots of 2~3 mm diameter per trifoliolate leaf were categorized as a resistant response (R); 10~20 spots of ≤ 3 mm diameter per trifoliolate leaf was categorized as a moderately resistant response (MR); > 20 spots of 3 mm was categorized as a susceptible response (S); Spots that were connected into large groups and most of the leaves dying early

due to the disease were categorized as a highly susceptible response (HS) [25]. All inoculation tests were repeated at least 3 times and 30 plants of each soybean accession were inoculated. Phenotypic data descriptive statistics were performed by R software [26].

GWAS analysis and Linkage disequilibrium analysis (LD)

The 234 soybean accessions were genotyped using SoySNP180K BeadChips. To determine the association site, the mixed linear model (MLM) in TASSEL 5.0 software was used. The Q value and the kinship K value after the Structure calculation were used as covariates, combined with the SNP chip genotype data, and the traits correlation analysis with tags. The threshold of the SNP locus that is significantly associated with the trait was set to $P < 0.001$. The remaining GWAS analyses were also performed using TASSEL 5.0 including the descriptive analysis, the genetic distance matrix of soybean varieties of genotype data, the principal component analysis and the kinship matrix [27]. Cluster analysis was performed by the UPGMA method and the phylogenetic tree was drawn by Mega4 software. The Quantile-Quantile (Q-Q) plots and Manhattan were drawn by TASSEL5.0 software. The degree of linkage disequilibrium (LD) between each pair of SNPs was estimated with the correlation coefficient r^2 by Haploview5.0 (minor allele freq: 0.001. Fraction of strong LD informative comparisons must be at least 0.97 and all other parameters used were default parameters).

Candidate genes annotation

According to the results of GWAS, SNPs related to FLS resistance were located when $p < 0.001$. Selection of the genes within 200 kb on both sides of the significant SNPs, were performed by BLAST searching against NR [28], GO [29] and KEGG databases [30] to find candidate genes related to FLS resistance.

Results

Phenotyping

We initially analyzed the genotypes of 382 soybean accessions using the SoySNP180K BeadChips, which were divided into four panels. The total number of SNP markers was 180961. Of these, 58,388 of the SNP markers were of the Poly High Resolution type and accounted for 32.24% of the total markers. The selection of SNP markers was carried out in the next step. After comparing data over three years, 277 accessions with stable results were selected as the candidate population for the GWAS of the resistance genes. Finally, 234 accessions with both genotype and phenotype data were selected as the GWAS population. All the tested accessions had different degrees of lesions, including the most resistant genotypes, which indicated that the infection process was successful. There were seven resistant accessions (R), 119 moderate resistant accessions (MR), 87 susceptible accessions (S), and 21 highly susceptible accessions (HS). Additional file 1: Table S1).

Genotyping

A total of 58,388 SNP genotypes of 234 accessions were performed by descriptive analysis. The results showed that 27,302 SNPs had no polymorphism in the association analysis population. After filtering, 30,890 SNPs on 20 soybean chromosomes were selected as the genotype data sources for the GWAS. The number of SNPs on Chr. 18 was the largest (2051), and the numbers of SNPs on Chr. 1 and Chr. 12 were the least (1198) (Fig. 1). On average, there were 1544.5 SNPs on each chromosome, and each SNP covered 31.301 kb of chromosome (Table 1).

Population structure analysis of 234 accessions

A total of 30,890 SNPs from 234 accessions were analyzed by Principal component analysis (PCA). The first and second principal components explained 6.44% and 4.60% of the variance, respectively, and they explained 11.04% of the phenotypic variation. A scatter-plot of the first and second principal components showed that the soybean genotypes collected from different sources were closer to each other. A subpopulation structure was not observed in this population (Fig. 2). Cluster analysis of 234 soybean accessions based on UPGMA was conducted using 30,890 SNP marker genotypes. There was no obvious classification of the accessions, which was consistent with the PCA results (Additional file 2: Figure S1).

GWAS of genes resistant to *C. soja*

The mixed linear model (MLM) and the "Q+K" model were used for the GWAS (Additional file 3: Figure S2). The K matrix of 234 accessions was completed by TASSEL5.0, and the Q matrix was constructed by selecting the first three principal components. A total of six SNPs were associated with FLS resistance ($p \leq 0.001$), of which, one SNP, three SNPs, and two SNPs were located on Chr. 2, Chr. 5, and Chr. 20, respectively (Fig. 3). The phenotypic variation explained by each peak ranged from 6.17% to 9.20%, and the highest peak was for Affx-89062122 on Chr. 5 (Table 2).

Haplotype analysis for FLS resistance gene

The linkage disequilibrium (LD) between pairs of SNPs on Chr. 2, Chr. 5, and Chr. 20 was analyzed by Haploview Software 5.0. *C. soja* resistance-related SNPs and adjacent SNPs form different haplotype blocks. Seven significant SNPs on Chr. 2 were located within an adjacent haplotype block of 37 kb (Fig. 4a). The r^2 values of all SNPs in the LD block were close to 1, indicating that these seven SNPs were highly associated and might have the same causal site (s) with FLS resistance. A total of 27 SNPs on Chr. 5 were located in one adjacent haplotype block forming six haplotypes (Fig. 5a). A total of 35 SNPs on Chr. 20 were located in one adjacent haplotype block forming six haplotypes (Fig. 6a). Through the T test, we found a significant difference in the resistance and susceptibility index of Chr. 2. Compared with the Hap C (haplotype block C) genotypes, Hap A was significantly more resistant to FLS ($p=0.016$, less than 0.038). Hap A is the resistance genotype, and Hap C is the susceptible genotype (Fig. 4b). The Hap A and Hap D genotypes on Chr. 5 showed a significant difference in resistance in the susceptibility index ($p=0.025$, less than 0.05). Hap A is the resistant genotype, and Hap D is the susceptible genotype (Fig. 5b). On Chr. 20, Hap A and Hap C genotypes showed significant differences in the resistance to disease index ($p=0.016$, less than 0.05). Hap B and Hap C also showed significant differences ($p=0.019$, less than

0.05). Hap A and Hap B are the resistance genotypes, and Hap C is the susceptible genotype (Fig. 6b). The gene information distributed in three haplotype blocks was extracted, and the positions on the chromosomes were indicated (Fig. 7).

Candidate genes for FLS resistance at GWAS loci

A total of 55 genes within the three haplotype blocks on Chr. 2, Chr. 5, and Chr. 20 were annotated with Glyma1.0 in NR, GO, and KEGG databases (Additional file 4: Table S2). These genes were separated into 29 GO terms, mainly including mitochondrial outer membrane (GO:0005741), calcium-dependent protein serine/threonine kinase activity (GO:0009931), calcium-dependent protein kinase activity (GO:0010857), MAP kinase activity (GO:0004707), protoxylem development (GO:0090059), and xylan metabolic process (GO:0045491) (Additional file 5: Table S3). The enriched KEGG pathway is involved in plant–pathogen interaction (gmx04626), MAPK signaling pathway–plant interaction (gmx04016), and biosynthesis of secondary metabolites (gmx01110) (Additional file 6: Table S4). Among these genes, *Glyma05g28980* encodes mitogen-activated protein kinase 7 (MPK7). *Glyma20g31510* and *Glyma20g31520* encode the calcium-dependent protein kinase (CDPK4) family proteins, and *Glyma20g31630* encodes pyruvate dehydrogenase (PDH), which may be involved in plant disease resistance. These genes were predicted to be candidate resistance genes.

Discussion

GWASs have made great progress in soybean genetics research, but the current research scope is limited to a few characters [31]. Compared with traditional QTL mapping, whole genome association analysis has the advantages of wide detection range [32], high resolution [33-34], and more material sources [35]. However, GWAS analysis has its own limitations. Association mapping is complementary compared to traditional bi-parental linkage mapping, but bi-parental linkage mapping cannot be replaced. The interaction effects of genes and the environment will affect GWAS analysis results; so, rigorous phenotyping is required. Due to the scale of the study, more complex traits controlled by multiple loci with relatively small phenotypic effects will require large populations [36]. Population structure is a very important factor in the correlation study; so, it is necessary to carefully select germplasm resources and evaluate the population structure. This study used more than three replications to increase the accuracy of phenotype identification. In addition, GWAS is only a prediction of candidate genetic sites, and further research is needed in combination with other experimental methods to explore their biological functions. Previous studies have shown that there is a significant QTL close to Satt319 on LG C2 (Chr. 7; LOD 3.8; r^2 52%) and a significant QTL near Satt632 on LG A2 (Chr. 8; r^2 was 15%; LOD was 3.6) associated with resistance to *C. sojina* race 2 [37]. Additional minor loci were also found, which were detected on LGs A1, B1, F, G, H, I, J, K, L, M, and O. Among the loci, one was associated with Satt440 on LG I (Chr. 20) at 112.7 cM. The locus explained 15% of the variation in FLS at 42 dai ($p=0.001$) with the “Essex” allele reducing disease severity by up to 0.95 units, but it does not overlap with our loci. Therefore, the six SNPs associated with FLS resistance may represent new loci that require further verification.

In this study, one encoding PDH gene, one encoding MPK7 gene, and two encoding CDPK4 genes in the haplotype block were associated with FLS resistance. They are all related to the biological pathway of salicylic acid (SA). Plant systemic acquired resistance (SAR) is an inducible immune system. The cells in the infected parts of plants can produce signal molecules, such as salicylic acid (SA), lipids, peptides, and nitric oxide. These signal molecules diffuse to the normal tissues and cells of plants through the vascular system and then activate the expression of stress resistance genes and the regulation of physiological metabolism in the normal cells, so as to enhance the immune ability of cells and effectively restrict disease spread [38]. SA can induce the expression of a variety of pathogenesis-related (PR) genes and help plants resist the invasion of disease organisms, such as viruses, bacteria, and fungi. SA is involved in the formation of plant innate immunity and effector-triggered immunity in plants. SA can also activate the suicide process of plant cells and form necrotic spots in the infected parts to prevent the invasion and spread of pathogens [39]. A gene encoding MPK7 (GhMPK7) cloned from cotton belongs to the C MAPK group. It plays an important role in the broad-spectrum resistance to fungi and viruses regulated by SA and is also involved in the regulation of plant growth and development [40]. MPK7 is co-expressed with MKK3 and promotes the strong expression of *Pseudomonas syringae* resistance genes in plants [41]. This study also proved that MKK3 pathway plays a role in pathogen defense and further underscores the importance and complexity of MAPK signaling in plant stress responses. MKK3 plays a role in jasmonate (JA)-mediated developmental signaling and generates H₂O₂ to activate MPK7, which acts as a secondary signal to activate defense genes [42]. Among phytohormones, JA plays an important role in resisting biological stress [43]. However, reverse genetic studies indicated that MAPKs, SA-induced protein kinase (SIPK), and wound-induced protein kinase (WIPK), are rapidly activated by fatty acid–amino acid conjugates. Then, MAPKs and calcium-dependent protein kinase (CDPK) are necessary for the induction of JA in response to biological stress [44]. When CDPK4 and CDPK5 in *Nicotiana attenuata* were simultaneously silenced, transgenic plants (IRcdpk4/5) induced high levels of defense metabolites. Yang [43] found that CDPK4 and CDPK5 affect plant resistance to biological stress in a JA- and JA-signaling-dependent manner. Transgenic plants showed over activation of SIPK, a MAPK involved in various stress responses, and genetic analysis indicated that the increased SIPK activity in IRcdpk4/5 plants leads to the exceptionally high JA levels. Some studies have found that CPK4/11 and CPK5/6 play an important role in the resistance of *Arabidopsis* to *Pseudomonas syringae* in a MAPK-independent manner [44]. The effector in *Heterodera avenae* can interact specifically with an *Arabidopsis* pyruvate dehydrogenase subunit, which might interfere with the SA signaling pathway and suppress plant defense responses [45]. We located candidate genes for soybean resistance to FLS using association analysis [17]. The effective use of these QTLs will require functional verification combined with proteomics when effective markers are identified for use in resistance breeding.

Conclusion

Based on the GWAS results of a mixed linear model, a total of six SNPs ($p < 0.001$) significantly related to FLS resistance were detected on Chr. 2, Chr. 5, and Chr. 20, which formed three haplotype blocks. *Glyma20g31630* encoding pyruvate dehydrogenase (PDH), *Glyma05g28980* encoding mitogen-activated

protein kinase 7 (MPK7), and *Glyma20g31510* and *Glyma20g31520* encoding the calcium-dependent protein kinase 4(CDPK4) in the haplotype blocks were associated with FLS resistance. The proteins encoded by these four genes directly or indirectly participate in the biological pathways of salicylic acid (SA) and jasmonic acid (JA). These two plant hormones can induce the expression of disease resistance related genes and are essential for plant SAR. This study provides a resource for elucidating the mechanism underlying FLS resistance and for breeding resistant soybean cultivars by molecular-assisted selection.

Abbreviations

GWAS: genome-wide association analysis; FLS: Frogeye leaf spot; QTL: quantitative trait loci; SNP: single nucleotide polymorphisms; LD: linkage disequilibrium; PCA: Principal component analysis; MLM: mixed linear model

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and material

Data generated or analyzed during this study are included in this published article and its supplementary information files.

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by the Applied Technology Research and Development Plan of Heilongjiang Province (GA20B104) and the Agricultural Science and Technology Innovation Leaping Project of Heilongjiang Academy of Agricultural Sciences (HNK2019CX14). The funding bodies played no role in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript.

Author contributions

GX and HSS both participated in the conception and design of the study, performed the experimental work and data analysis, and drafted the manuscript. DJJ and HXP supervised the study and wrote the

commentary. LW, WC, MYS, WQS, and YXH assisted with data analysis and biological interpretation, and revised the manuscript. LZM, LZJ, YL, and GXD participated in the experimental design. ZMM and YS participated in the conception and experimental design of the study and assisted in biological interpretation. All of the authors reviewed and accepted the final version of the manuscript.

Acknowledgements

We thank LetPub (www.letpub.com) for linguistic assistance during manuscript preparation.

References

1. Mian MA, Missaoui AM, Walker DR, Phillips DV, Boerma HR. Frogeye leaf spot of soybean: A review and proposed race designations for isolates of *Cercospora sojina* Hara. *Crop Sci.* 2008;48:14–24.
2. Gu X, Ding JJ, Liu W, Yang XH, Yao LL, Gao XD, et al. Comparative genomics and association analysis identifies virulence genes of *Cercospora sojina* in soybean. *BMC Genomics*, 2020;21:172.
3. Kim H, Newell AD, Cota-Sieckmeyer RG, Rupe JC, Fakhoury AM, Bluhm BH. Mating-type distribution and genetic diversity of *Cercospora sojina* populations on soybean from Arkansas: evidence for potential sexual reproduction. *Phytopathology.* 2013;103(10):1045–51.
4. Camera JN, Ghissi VC, Reis EM, Deuner CC. The combined effects of temperature and leaf wetness periods on soybean frogeye leaf spot intensity. *Semin Cienc Agrar.* 2016;37(1):77–84.
5. Shrestha SK, Cochran A, Mengistu A, Lamour K, Castro-Rocha A, Young-Kelly H. Genetic diversity, QoI fungicide resistance, and mating type distribution of *Cercospora sojina*—Implications for the disease dynamics of frogeye leaf spot on soybean. *PLoS One.* 2017;12(5):e0177220.
6. Gu X, Yang S, Yang XH, Yao LL, Gao XD, Liu Wet, et al. Comparative transcriptome analysis of two *Cercospora sojina* strains reveals differences in virulence under nitrogen starvation stress. *BMC Microbiology*, 2020, 20:166.
7. Athow K, Probst AH. The inheritance of resistance to frogeye leaf spot of soybeans. *Phytopathology.*1952; 42(12):660–2 pp.
8. Athow KL, Probst AH, Kurtzman CP, Laviolette FA. A newly identified physiological race of *Cercospora sojina* on soybean. *Phytopathology.* 1962;52(7):712–4 pp.
9. Phillips D, Boerma H. Two genes for resistance to race 5 of *Cercospora sojina* in soybeans. *Phytopathology.* 1982; 72(7):764–6.
10. Baker, WA, Weaver D.B, Qui J, Pace PF. Genetic analysis of frogeye leaf spot resistance in PI54610 and Peking soybean. *Crop Sci.* 1999;39:1021–5.
11. Mian, MAR, Wang T, Phillips DV, Alvernaz J, and Boerma HR. Molecular mapping of the Rcs3 gene for resistance to frogeye leaf spot of soybean. *Crop Sci.* 1999. 39:1687–91.
12. Bachman MS, Tamulonis JP, Nickell CD, Bent AF. Molecular markers linked to brown stem rot resistance genes Rbs1 and Rbs2 in soybean. *Crop Sci.* 2001;41:527–35.

13. Webb, D.M. Brown stem rot resistance in soybeans. Pioneer Hi-Bred Int., Inc., Johnston, IA. U.S. 1997. Patent No. 5689035
14. Missaoui AM, Phillips DV, Boerma H. DNA Marker Analysis of 'Davis' Soybean and Its Descendants for the Rcs3 Gene Conferring Resistance to *Cercospora sojina*. *Crop Sci.* 2007;47:1263–70.
15. Viard F, Franck P, Dubois MP, Estoup A, Jarne P. Variation of microsatellite size homoplasy across electromorphs, loci, and populations in three invertebrate species. *J.Mol. Evol.* 1998;47:42–51.
16. Missaoui AM, Ha BK, Phillips DV, Boerma HR. Single nucleotide polymorphism detection of the Rcs3 gene for resistance to frogeye leaf spot in soybean. *Crop Sci.* 2007;47:1681–90.
17. Hoskin A. Genetic mapping of soybean resistance genes to frogeye leaf spot in five Chinese Plant Introductions and efficiency of early generation selection for low phytate soybean lines. Institute of Plant Breeding, Genetics, and Genomics. University of Georgia. 2011.
18. Anh-Tung P, Harris DK, James B, Aaron H, Jonathan S, Hussein AH, et al. Fine Mapping and Characterization of Candidate Genes that Control Resistance to *Cercospora sojina* K. Hara in two soybean germplasm accessions. *Plos One*, 2015; 10(5):e0126753.
19. Nordborg M, Weigel D. Next-generation genetics in plants. *Nature.* 2008;456: 720–3.
20. Che Z, Yan H, Liu H, Yang H, Yu D. Genome-wide association study for soybean mosaic virus SC3 resistance in soybean. *Molecular Breeding.* 2020; 40:69.
21. Wei W, Oliveira MAC, Figueiró Adriana de A, Xing Wet, Shilpa M, Wickland DP, et al. Genome-wide association mapping of resistance to a Brazilian isolate of *Sclerotinia sclerotiorum* in soybean genotypes mostly from Brazil. *BMC Genomics.* 2017;18(1):849.
22. Vuong TD, Sonah H, Meinhardt CG, Deshmukh R, Kadam S, Nelson JG, et al. Genetic architecture of cyst nematode resistance revealed by genome-wide association study in soybean. *BMC Genomics.* 2015;16:593.
23. Wei L, Jian H, Lu K, Filardo F, Yin N, Liu L, et al. Genome-wide association analysis and differential expression analysis of resistance to *Sclerotinia* stem rot in *Brassica napus*. *Plant Biotechnology Journal.* 2015;pp 1–13.
24. Kim JS, Lee YS, Kim SK, Kim KD, Kim KW. Differential responses of soybean cultivars to *Cercospora sojina* isolates, the causal agent of frogeye leaf spot in Korea. *Plant Pathology Journal.* 2011;27(2):183–6.
25. Mian MAR, Missaoui AM, Walker DR, Phillips DV, Boerma HR. Frogeye leaf spot of soybean: a review and proposed race designations for isolates of *Cercospora sojina* Hara. *Crop Sci.* 2008;48(1):14–24.
26. R Core Team. R: a language and environment for statistical computing. In: R Foundation for statistical computing. 2018; Vienna <https://www.R-project.org/>
27. Yu J, Pressoir G, Briggs WH, Vroh BI, Yamasaki M, Doebley JF, et al. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nature Genetics.* 2006;38(2):203–208.

28. Li W, Jaroszewski L, Godzik A. Tolerating some redundancy significantly speeds up clustering of large protein databases. *Bioinformatics*. 2002;18(1):77–82.
29. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Geneontology: tool for the unification of biology. *Nature Genetics*. 2000;25:25–9.
30. Kanehisa M, Goto S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Research*. 2000;28(1):27–30.
31. Contreras R, Mora F, Mar O, Higashi W, Schuster I. A Genome-wide association study for agronomic traits in soybean using SNP markers and SNP-based haplotype analysis. *PLoS One*. 2017;12(2): e0171105.
32. Yu J, Holland JB, McMullen MD, Buckler ES. Genetic design and statistical power of nested association mapping in maize. *Genetics*, 2008, 178(1); 539–51.
33. Yang J, Jiang H, Yeh CT, Yu J, Jeddloh JA, Nettleton D, et al. Extreme-phenotype genome-wide association study (XP-GWAS): A method for identifying trait-associated variants by sequencing pools of individuals selected from a diversity panel. *Plant Journal*. 2015;84(3):587–96.
34. Huang X, Wei X, Sang T, Zhao Q, Feng Q, Zhao Y, et al. Genome-wide association studies of 14 agronomic traits in rice landraces. *Nature Genetics*. 2010;42(11):961–7.
35. Winham SJ, Biernacka, JM. Gene-environment interactions in genome-wide association studies: Current approaches and new directions. *Journal of Child Psychology and Psychiatry*. 2013;54(10):1120–34.
36. Yang Q, Wu HS, Guo CY, Fox CS. Analyze multivariate phenotypes in genetic association studies by combining univariate association tests. *Genetic Epidemiology*. 2010;34 (5): 444–54.
37. Sharma H, Lightfoot DA. Quantitative trait loci underlying partial resistance to *Cercospora sojina* Race 2 detected in Soybean Seedlings in greenhouse assays. *Atlas Journal of Biology*. 2014;3(1):175–82.
38. Park SW, Kaimoyo E, Kumar D, Mosher S, Klessig DF. Methyl salicylate is a critical mobile signal for plant systemic acquired resistance. *Science*. 2007;318(5487):113–6.
39. Bari R, Jones JDG. Role of plant hormones in plant defence responses. *Plant Mol Biol*. 2009;69:473–88.
40. Shi J, An HL, Zhang L, Go Z, Guo XQ. *GhMPK7*, a novel multiple stress-responsive cotton group C MAPK gene, has a role in broad spectrum disease resistance and plant development. *Plant Molecular Biology*. 2010;74:1–17.
41. Do'czi, Ro'bert, Brader, Günter, Pettkó-Szandtner, Aladár, et al. The Arabidopsis mitogen-activated protein kinase kinase MKK3 is upstream of group C mitogen-activated protein kinases and participates in pathogen signaling. *American Society of Plant Biologists*. 2007;19:3266–79.
42. Orozco-Cardenas M, Ryan, C.A. Hydrogen peroxide is generated systemically in plant leaves by wounding and systemin via the octadecanoid pathway. *Proc. Natl. Acad. Sci. USA*. 1999;96:6553–7.

43. Zhu SY, Yu XC, Wang XJ, Zhao R, Li Y, Fan RC, et al. Two calcium-dependent protein kinases, CPK4 and CPK11, regulate abscisic acid signal transduction in Arabidopsis. *Plant Cell*. 2007;19:3019–36.
44. Yang DH, Hettenhausen C, Baldwin IT, Wu JQ. Silencing *Nicotiana attenuata* Calcium-Dependent Protein Kinases, CDPK4 and CDPK5, Strongly Up-Regulates Wound- and Herbivory-Induced Jasmonic Acid Accumulations. *Plant physiology*. 2012;159(4):1591–1607.
45. Yang SS, Pan LL, Chen YP, Yang D, Liu Q and Jian H. *Heterodera avenae* GLAND5 Effector Interacts With Pyruvate Dehydrogenase Subunit of Plant to Promote Nematode Parasitism. *Frontiers in Microbiology*. 2019; 10:1241.

Tables

Table 1 Distribution of 30890 SNPs on 20 chromosomes of soybean

Chromosome	No. of SNP markers	Starting position (kb)	End position (kb)	SNP coverage (kb)	Single SNP coverage (kb/SNP)
1	1198	46.209	55901.759	55855.55	46.624
2	1930	59.603	51642.224	51582.621	26.727
3	1436	18.316	47723.675	47705.359	33.221
4	1606	30.661	49194.977	49164.316	30.613
5	1274	69.827	41932.26	41862.433	32.859
6	1665	10.299	50630.237	50619.938	30.402
7	1389	7.129	44605.694	44598.565	32.108
8	1777	57.34	46925.494	46868.154	26.375
9	1719	18.921	46832.712	46813.791	27.233
10	1259	98.244	50913.805	50815.561	40.362
11	1434	16.255	39163.086	39146.831	27.299
12	1198	46.664	40103.384	40056.72	33.436
13	1714	9.149	44339.594	44330.445	25.864
14	1243	14.558	49706.317	49691.759	39.977
15	1903	9.353	50860.462	50851.109	26.722
16	1540	17.629	37327.774	37310.145	24.227
17	1551	252.274	41878.591	41626.317	26.838
18	2051	7.193	62264.999	62257.806	30.355
19	1504	18.1	50558.55	50540.45	33.604
20	1499	31.395	46769.259	46737.864	31.179

Table 2 SNPs responsible for *C. sojina* resistance traits by GWAS

SNP	Chromosome	Position	p -value	SNP r^2 (%)
Affx-89212224	2	50738780	8.8898E-4	6.26
Affx-89062085	5	34589138	9.9358E-4	6.17
Affx-89062094	5	34596356	2.7756E-4	7.47
Affx-89062122	5	34658974	3.7851E-5	9.20
Affx-89163211	20	40257935	2.7861E-4	7.37
Affx-89163218	20	40262553	1.7863E-4	7.76

Figures

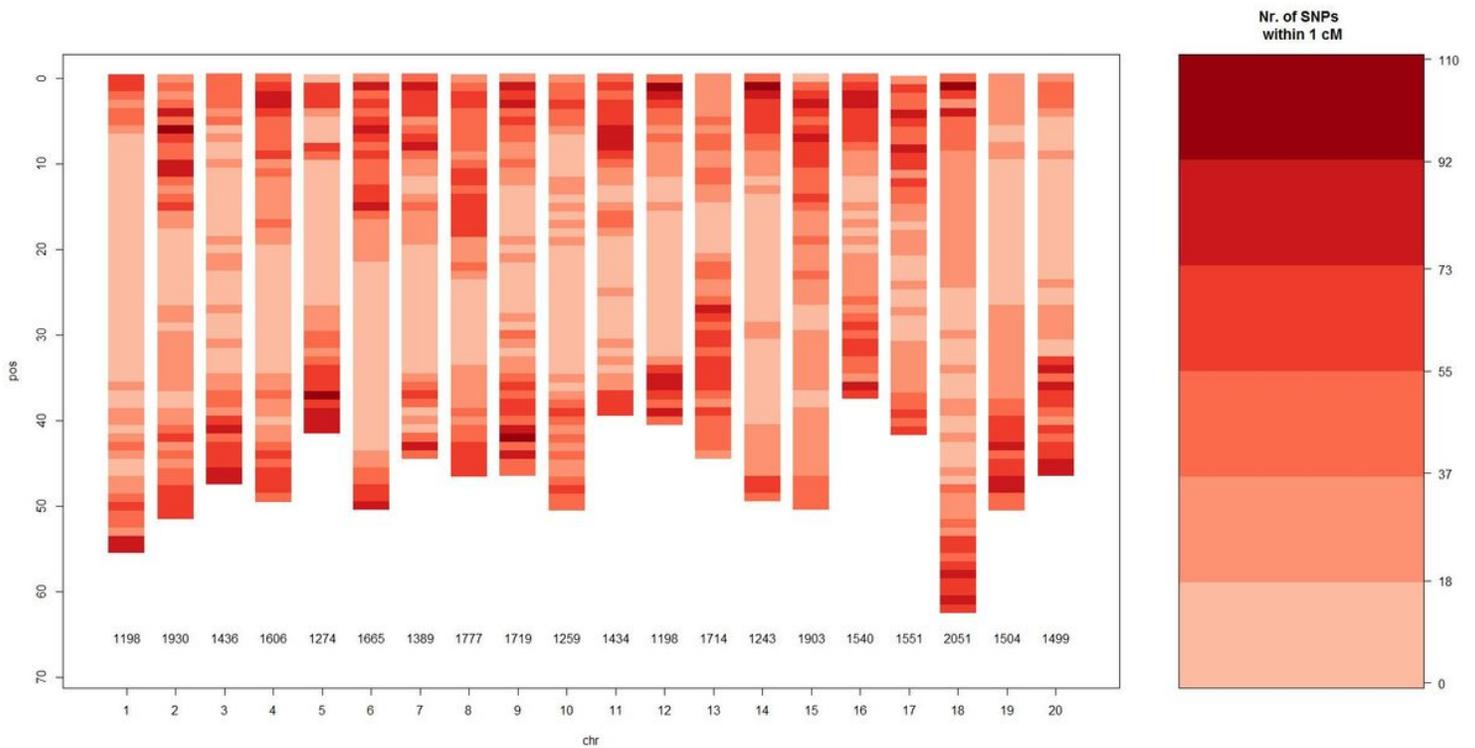


Figure 1

Distribution of the 30890 SNPs on the 20 soybean chromosomes. The graph shows the number of SNPs with a $MAF \geq 0.05$ on each chromosome. The red bar from light to dark on the right represents the number of SNPs within 1 cM.

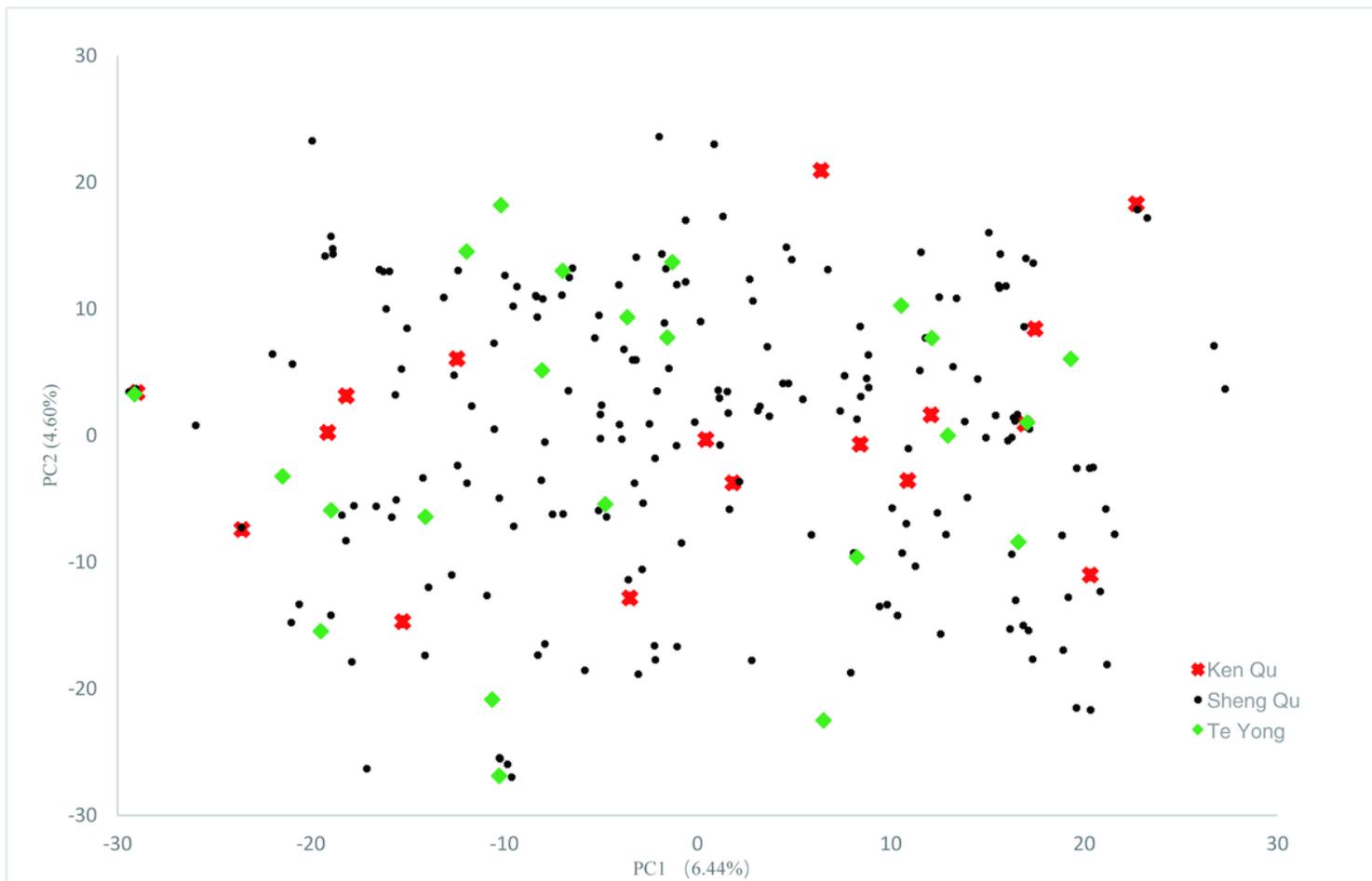


Figure 2

Principal Component Analysis (PCA) of a diverse set of 243 soybean cultivars. PCA scatter plot showing the two main principal components. Dots with different colors and shapes represent different sources.

P-Values by Chromosome for FLS

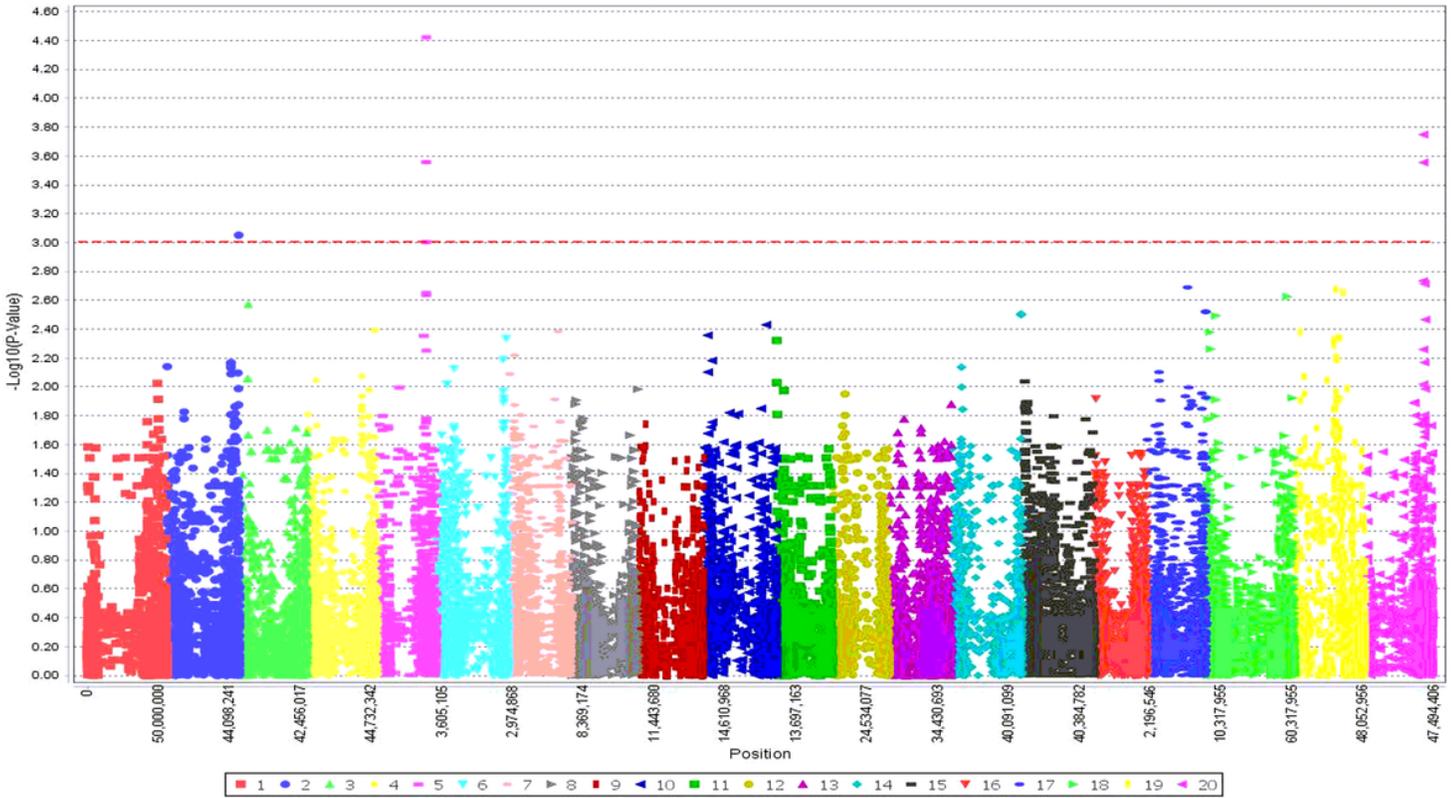


Figure 3

Manhattan plots of the association of SNPs with FLS resistance in soybean identified by GWAS. The chromosomal SNPs are differentiated by various colors. The red dashed line represents the significance threshold $-\log_{10}(p) = 3.00$.

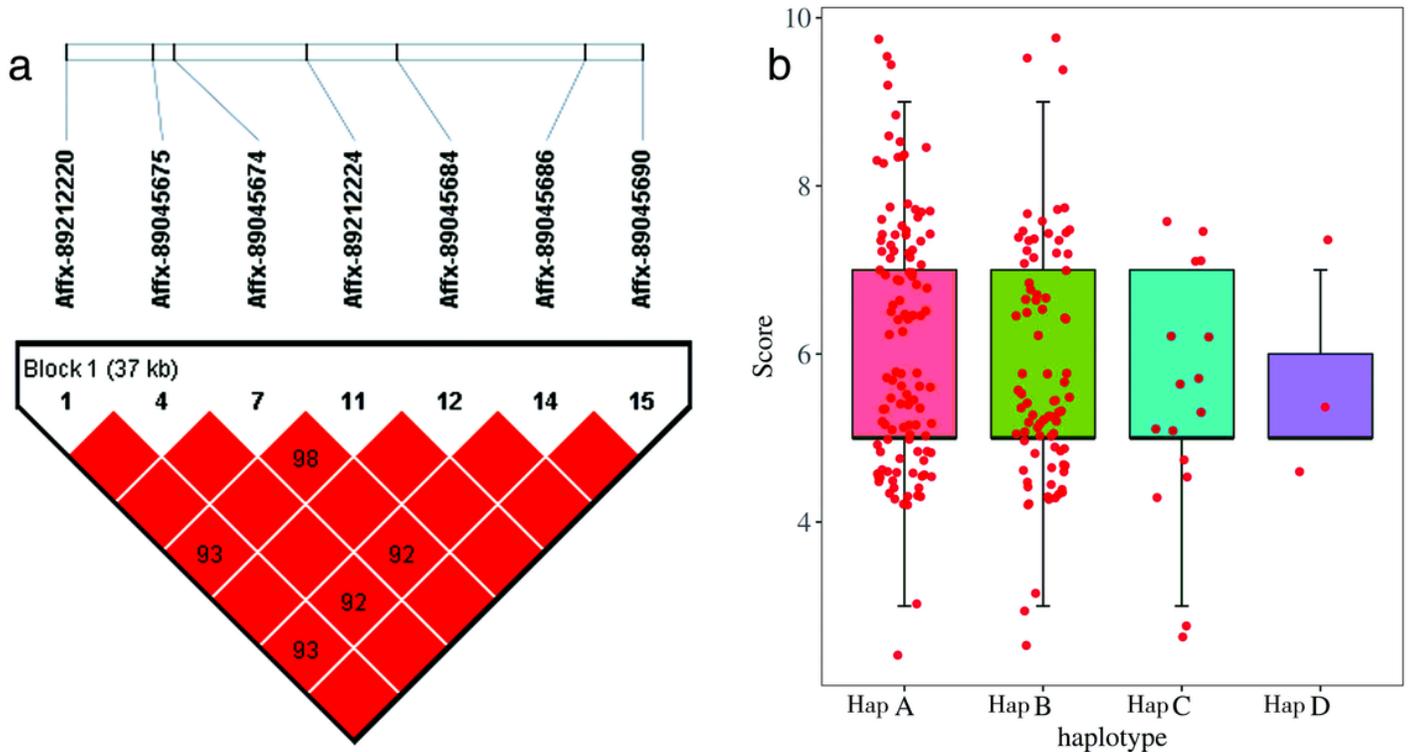


Figure 4

Haplotype block for SNPs significantly associated with FLS resistance on Chr. 2. a. Numbers in squares indicate 100-fold r^2 values of each pair of SNPs. The bars above LD plots represent the physical positions of SNPs. LD blocks are marked with black triangles. b. Box plots showing the phenotypic values of four haplotypes and the distribution of four haplotypes in different subspecies. Different letters indicate that $P < 0.05$.

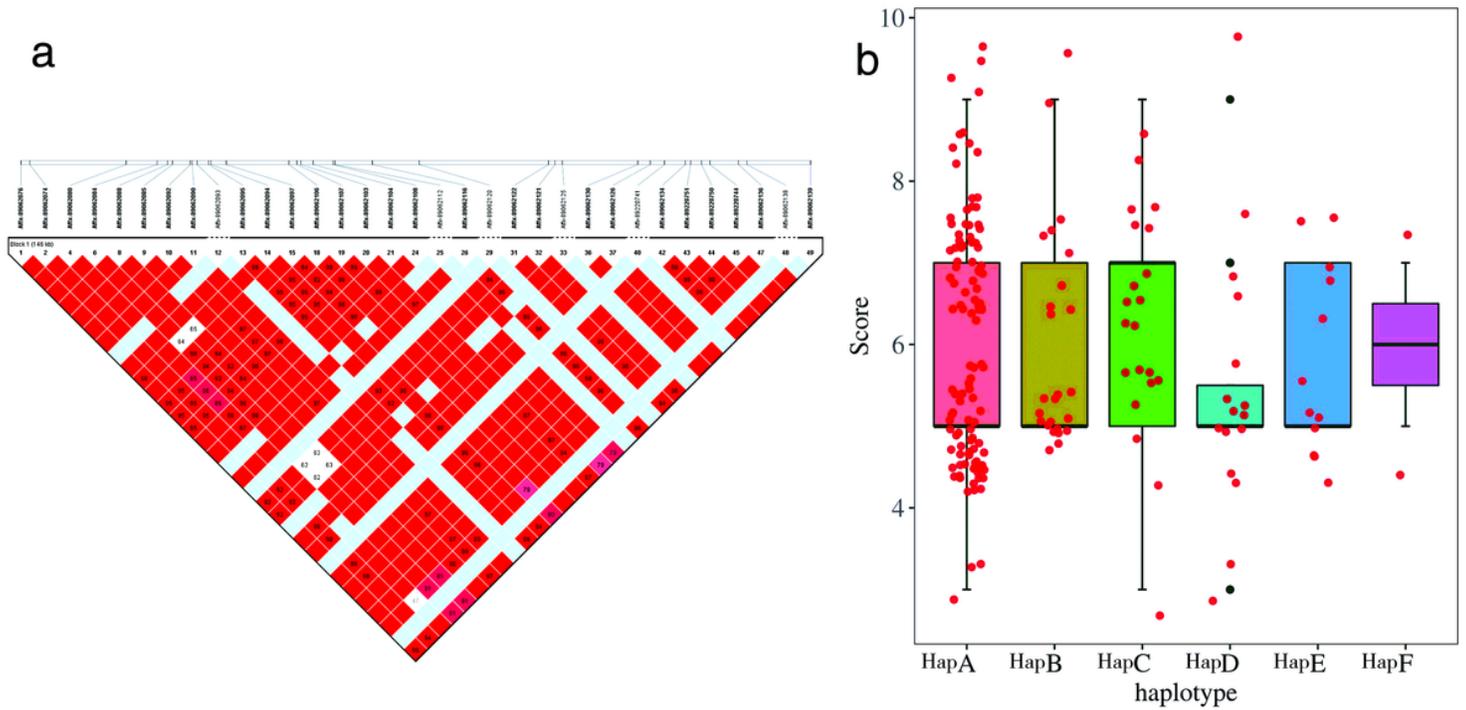


Figure 5

Haplotype block for SNPs significantly associated with FLS resistance on Chr. 5. a. Numbers in squares indicate 100-fold r^2 values of each pair of SNPs. The bars above LD plots represent the physical positions of SNPs. LD blocks are marked with black triangles. b. Box plots showing the phenotypic values of six haplotypes and the distribution of six haplotypes in different subspecies. Different letters indicate that $P < 0.05$.

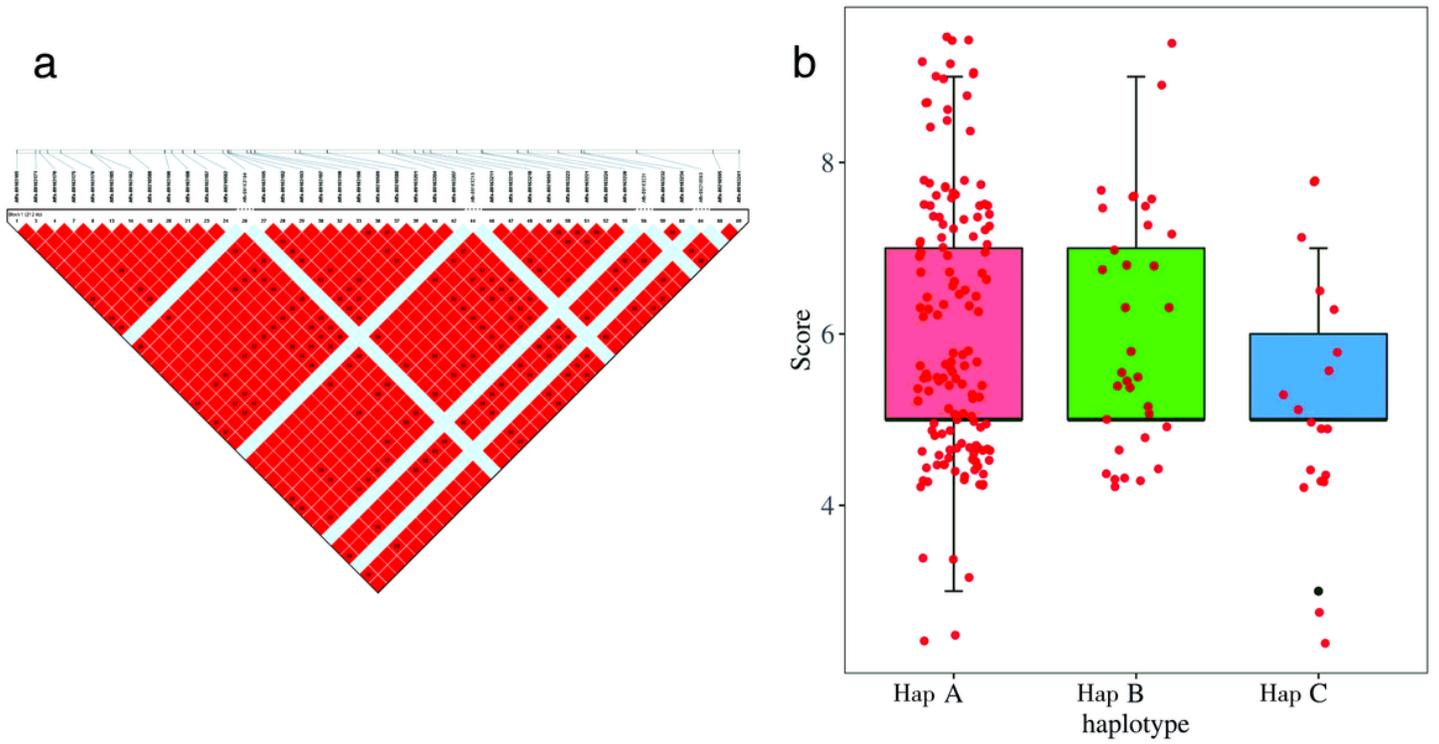


Figure 6

Haplotype block for SNPs significantly associated with FLS resistance on Chr. 20. a. Numbers in squares indicate 100-fold r^2 values of each pair of SNPs. The bars above LD plots represent the physical positions of SNPs. LD blocks are marked with black triangles. b. Box plots showing the phenotypic values of three haplotypes and the distribution of three haplotypes in different subspecies. Different letters indicate that $P < 0.05$.

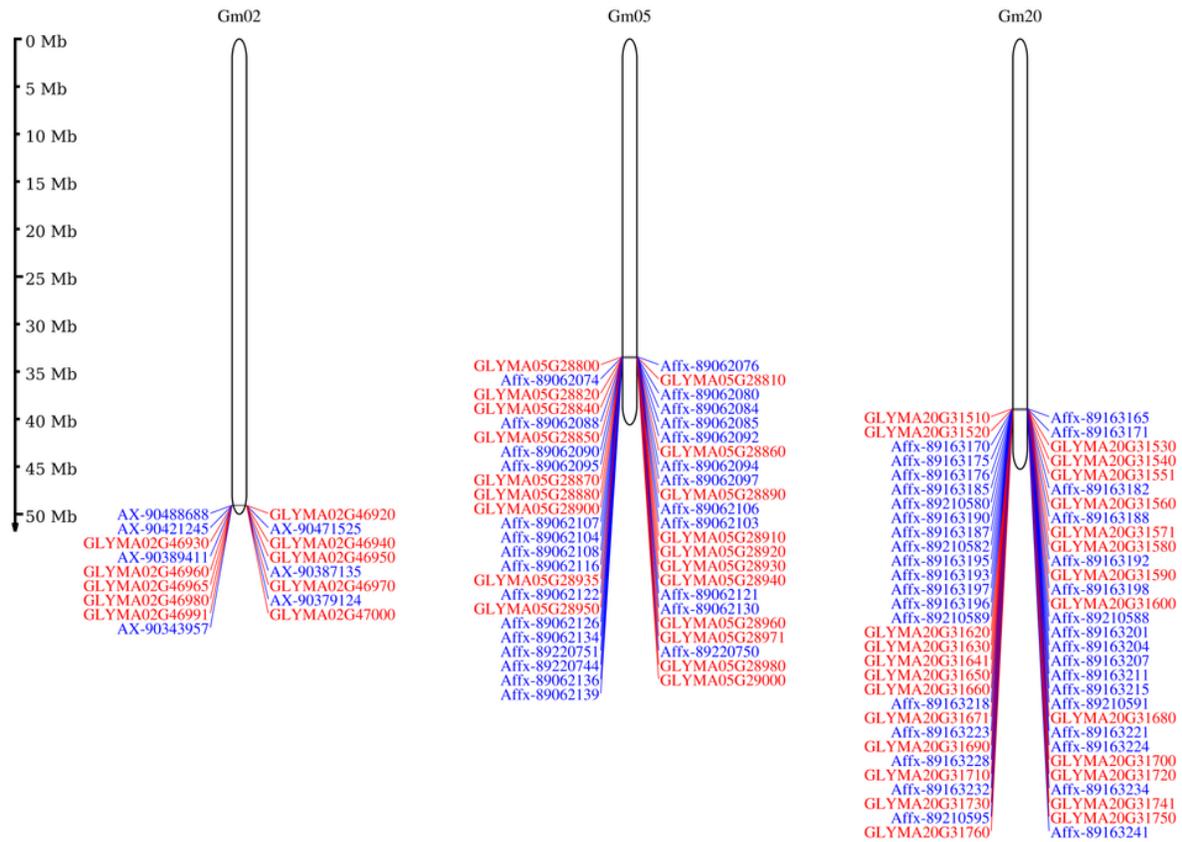


Figure 7

Location map of the gene information in the three haplotype blocks on chromosomes 2, 5, and 20.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Additionalfile1TableS1.xls](#)
- [Additionalfile3FigureS2.png](#)
- [Additionalfile4TableS2.xls](#)
- [Additionalfile5TableS3.xls](#)
- [Additionalfile3FigureS2.png](#)
- [Additionalfile6TableS4.xlsx](#)