Molecular Plant Breeding 2021, Vol.12, No.12, 1-22
http://genbreedpublisher.com/index.php/mpb

Research Article Open Access

Association Analysis of Resistance to Soybean Mosaic Virus and Variation in Resistance Allelic Excavation

Weiwei Wang, Lai Wei, Zhiyuan Yu, Guixing Zhao, Weiwei Bi, Lijun Liu
Soybean Research Institute of Heilongjiang Academy of Agricultural Sciences, Harbin 150086, China
Corresponding author Email: nkyssbd@126.com

Molecular Plant Breeding, 2021, Vol.12, No.12 doi: 10.5376/mpb.2021.12.0012
Received: 17 Dec., 2020
Accepted: 31 Mar., 2021
Published: 11 Apr., 2021

Preferred citation for this article:
Wang W.W., Wei L., Yu Z.Y., Zhao G.X., Bi W.W., and Liu L.J., 2021, Association analysis of resistance to soybean mosaic virus and variation in resistance allelic excavation, Molecular Plant Breeding, 12(12): 1-22 (doi: 10.5376/mpb.2021.12.0012)

Abstract Soybean Mosaic Virus (SMV) is one of the most important soybean diseases in the world, and it seriously affects soybean yield and quality. In this paper, 405 SSR molecular markers were used in genome-wide scan to analysis 327 soybean varieties in the North of China. The purpose was to search the genetic loci and excellent resistance genes related to SMV1 and SMV3, and to reveal the differences between the genetic models of varieties. The results confirmed that the genetic differences among the tested varieties were large. Through the association analysis of the population genome-wide association (GWA) and linkage group association (LGA), 14 highly significant sites related to SN1 and SN3 were found. There are three sites related to SN1 (Sat-297, sat-296, sat-154) which are coincident with or adjacent to the resistance sites found by predecessors; while there are two at SN3 (Sat-154 and sat-726) which are coincident with previous studies, and sat-154 is a resistance site shared by SN1 and SN3. 7 genes related to resistance were obtained, of which Glyma13g25420 and Glycma13g25440, Glyma14g08700 and Glyma14g08710 has resistance genes on the NBS-LRR domain; Glyma14g08900, Glyma14g08910, and Glyma14g08920 have PLP2 (Patation-like Protein) resistance genes. This study provides an important reference for SMV resistance research and technical support for soybean resistance breeding.

Keywords Soybean Mosaic Virus; Association analysis; Resistance genes

Background

SMV is one of the major diseases affecting the world's main soybean-producing areas. Soybeans infected with SMV have wrinkled leaves; are dwarfed; have fewer pods, which are wrinkled and speckled; have a significantly decreased number of root nodules, 100-seed weight and yield per plant; and exhibit decreased protein and oil levels. As a result, SMV suppresses soybean yield and quality and can even result in a failed harvest (Hill et al., 1987; Steinlage et al., 2002; Zhi et al., 2005; Liu, 2007; Hui et al., 2019; Gao and Shu, 2020; Chen et al., 2020).

SMV is divided into different strains based on the host in various incidence regions. Cho and Goodman divided American SMV materials into 7 strains (G1-G7) (Cho and Goodman, 1979). Wang divided common SMV in southern and northern China into 6 strains (Sa-Sf) and 3 strains (N1-N3), respectively (Wang et al., 2004. Northeastern China is the main production area for quality soybeans. Kai Li found that no two strains from these studies were the same; strains from different regions were specific (Li et al., 2014). SMV accounts for more than 9% of the total incidence of disease in some parts of this area. The main pathogens are SMV strain N1 (SN1) and SMV strain N3 (SN3). At present, it is believed that the cultivation of SMV-resistant material is a relatively effective method for preventing and controlling the virus. In early research into SN1 and SN3, scholars had proposed different genetic models based on the resistance separation ratio of offspring in genetic groups. Sun et al. (1990) demonstrated that SN1 resistance was controlled by 2 pairs of complementary recessive genes and SN3 resistance by 1 pair of recessive genes. Zheng et al. (2000a) demonstrated that SN3 resistance was controlled by 1 pair of recessive genes. The limited data on SMV genetics have contributed to the variation in scholars' views on genetic models of the virus. Therefore, it is difficult to devise an accurate genetic model simply by analyzing the separation ratio, as the heredity of soybean SMV resistance is complex.

As research methods developed, linkage mapping and association analysis have been applied in searching for SMV resistance genes and linked markers, and many resistance-related molecular markers have been identified.

By utilizing linkage mapping, Teng et al. (2006) found 2 SSR markers closely linked to SN3 resistance and 6 SSR molecular markers closely linked to SN1 resistance. By utilizing association analysis, Kumar et al. (2015) found 2 SSR molecular markers that were significantly correlated with soybean yellow mosaic virus resistance. By utilizing both linkage mapping and association analysis, Yan et al. (2015) revealed a correlation between 19 SNPs and soybean SMV resistance and located gene Rsc7 on chromosome 2 (D1b linkage group).

SMV has different strains corresponding to different resistance genes. The resistance genes are primarily found in linkage groups B2, D1b and F and include Rsv1, Rsv3 and Rsv4 in strains G1-G7 (Yu et al., 1996; Hayes et al., 2000; Jeong et al., 2002); Rsa (Wang et al., 2004), Rsc4 (Wang et al., 2011), Rsc7 (Fu et al., 2006), Rsc8, Rsc9 (Wang et al., 2004), Rsc11 (Bai et al., 2009) and Rsc13 (Guo et al., 2007) in strains Sa-Sf; and Rsmv1 (Teng et al., 2011) and Rsmv3 (Teng et al., 2006) in strains SN1 and SN3.

The present research covered 327 soybean genetic resources as materials. By utilizing two association analysis strategies, genome-wide association (GWA) and linkage group association (LGA) analysis, of linkage groups B2, D1b, and F, the research explored the SSR markers associated with SN1 and SN3 resistance in soybeans in northeastern China. The results not only lay the foundation for SMV-resistant soybean breeding but are also valuable for practical application.

1 Results and Analysis

1.1 Virus resistance analysis of soybean germplasm resources

The resistances of the materials, that is, the phenotypic data, were detected by an artificial inoculation identification method under an insect-proof net, which could effectively reduce interference by the environment. The identification results showed that the plants inoculated with SN1 exhibited substantial variation in their resistance, with a variation coefficient of 0.28, and the disease index varied from 0.09 to 0.55. There was one material with a DI of 9%, 20 materials with a DI of 12%~19%, 160 materials with a DI of 20%~30%, 81 materials with a DI of 30%~39%, 33 materials with a DI of 40%~50%, 23 materials with a DI of 50%~59%, and 3 materials with a DI of 62%~63%.

After the inoculation with SN3, the resistance variation range was also large, with a variation coefficient of 0.26 and a disease index from 0.20 to 0.69. There were 62 materials with a DI of 20%~29%, 85 materials with a DI of 30%~39%, 111 materials with a DI of 40%~49%, 49 materials with a DI of 50%~59%, and 14 materials with a DI of 60%~69% (Figure 1).

Figure 1 The classification standard of SMV

Note: Changes in soybean plants infected with soybean mosaic virus(SMV), that were divided into five levels: (A)Level one, Leaves contain mild light yellow mottled; (B) Level two, yellow and green leaf mottled, leaf blade protruding leaf curl backward; (C) Level three, follicular protrusions, slightly distorted distortion; (D) Level four, light mosaic and wrinkled mosaic mixed with yellow mottled; (E)Level five, diseased plants atrophy curly withered buds

The descriptive statistics indicated that the experimental materials exhibit a wide variation in resistance to SN1 and SN3, and the DI was normally distributed (Figure 2).

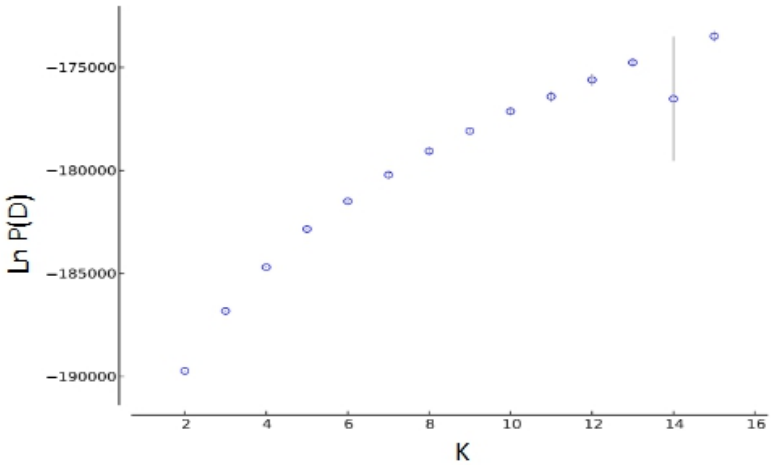

Figure 2 Change of LnP(D) according to K-value

1.2 Diversity of population genetics

In this paper, 1616 allelic variations were detected by 405 SSR markers; each locus had four allelic variations on average with a range of 2~9, among which Sct_033 had the most alleles (9). The Shannon index distribution of SSR markers was 0.089-4.136, with an average of 1.76. The PIC value of the SSR markers varied from 0.01 to 0.84, with an average of 0.51. The Shannon index and the PIC value (Table 1) showed that the SSR markers used in this experiment were sufficiently abundant for association analysis.

Table 1 405 SSR markers information

ID	SSR loci	Linkage group	Locus	Alleles		Mean		Mean
1	Sat_137	A1	0	4	0.66418	0.53367	2.1574	1.797
2	Sat_217	A1	3.542	3	0.57144		1.8961	
3	sat_265	A1	14.369	3	0.35661		1.0721	
4	Sat_368	A1	17.162	4	0.35701		1.299	
5	Sat_385	A1	26.42	3	0.38775		1.3884	
6	Satt050	A1	27.784	6	0.67954		2.7902	
7	satt073	A1	28.955	3	0.38016		0.1711	
8	satt155	A1	30.926	7	0.68273		2.7379	
9	Satt174	A1	31.074	4	0.44371		1.5653	
10	Satt211	A1	31.144	4	0.61364		2.132	
11	Satt225	A1	32.675	3	0.58976		1.9004	
12	Satt236	A1	33.275	3	0.52467		1.7418	
13	Satt276	A1	33.971	5	0.71037		2.3672	
14	Satt300	A1	46.454	4	0.48873		1.7056	
15	Satt364	A1	88.577	4	0.57794		1.87	
16	Satt382	A1	93.229	5	0.72226		2.3757	
17	Satt449	A1	95.16	4	0.57655		1.9277	
18	Satt591	A1	95.962	3	0.22013		0.862	
19	Satt684	A1	101.565	4	0.43609		1.5191	
20	Satt619	A1	69.206	5	0.69007		2.4609	
21	Sat_040	A2	9.143	3	0.17585	0.48137	0.7249	1.77202
22	Sat_294	A2	14.991	5	0.72385		2.3817	
23	Sat_377	A2	36.77	2	0.0454		0.2368	

Continued Table 1

ID	SSR loci	Linkage group	Locus	Alleles		Mean		Mean
24	sat_382	A2	51.505	3	0.42453		1.4788	
25	Sat_392	A2	54.917	3	0.52444		1.7439	
26	Satt133	A2	90.84	3	0.51973		1.739	
27	Satt177	A2	96.974	4	0.51995		1.8375	
28	Satt187	A2	106.29	4	0.67053		2.3479	
29	Satt209	A2	107.051	4	0.65051		2.3934	
30	Satt228	A2	108.782	3	0.31845		1.1712	
31	Satt333	A2	116.407	4	0.39054		1.4367	
32	Satt377	A2	116.64	4	0.43849		1.1407	
33	Satt378	A2	116.731	6	0.75217		3.3464	
34	Satt390	A2	118.639	6	0.75541		3.4549	
35	Satt409	A2	119.59	4	0.66436		2.2407	
36	Satt437	A2	125.382	4	0.44867		1.5887	
37	satt470	A2	128.441	2	0.01492		0.089	
38	satt508	A2	131.97	4	0.45731		1.3902	
39	Satt525	A2	145.565	2	0.37221		1.3752	
40	Satt538	A2	154.114	3	0.42782		1.5213	
41	Satt632	A2	159.63	6	0.63027		2.3245	
42	Sct_067	A2	165.73	5	0.66473		2.5837	
43	Sat_123	B1	32.51	4	0.65731	0.52652	2.4664	1.80566
44	Sat_247	B1	36.479	3	0.34423		1.2626	
45	sat_364	B1	37.799	4	0.5958		2.0016	
46	Satt197	B1	46.393	4	0.66539		2.2601	
47	Satt251	B1	49.731	4	0.60214		2.1245	
48	Satt444	B1	73.767	3	0.12438		0.5307	
49	Satt453	B1	78.127	3	0.55079		1.8088	
50	Satt484	B1	84.249	4	0.61994		2.266	
51	Satt509	B1	85.916	3	0.13033		0.5132	
52	satt597	B1	100.877	4	0.4742		1.0236	
53	satt638	B1	118.531	6	0.41666		0.5452	
54	Sct_026	B1	123.962	4	0.63851		2.5354	
55	Satt583	B1	84.189	8	0.82014		4.1355	
56	Sat_177	B2	6.047	5	0.63974	0.48972	2.427	1.68146
57	Sat_189	B2	7.843	5	0.50208		1.8369	
58	Sat_264	B2	12.558	6	0.74193		2.5932	
59	Satt020	B2	17.774	3	0.34415		1.1858	
60	Satt063	B2	27.632	6	0.5446		1.911	
61	Satt066	B2	55.198	6	0.4091		1.5201	
62	Satt070	B2	65.555	3	0.4945		1.6703	
63	Satt126	B2	70.56	5	0.51031		1.4147	
64	Satt168	B2	72.127	4	0.3981		1.4298	
65	Satt304	B2	72.808	3	0.30559		1.0782	
66	Satt467	B2	72.918	3	0.14433		0.5047	
67	Satt474	B2	73.213	4	0.5756		1.8524	
68	Satt534	B2	75.346	5	0.69325		2.2851	
69	Satt556	B2	78.844	4	0.51182		1.7937	

Continued Table 1

ID	SSR loci	Linkage group	Locus	Alleles		Mean		Mean
70	Satt560	B2	87.592	3	0.29027		1.077	
71	Satt577	B2	93.488	5	0.75295		2.5198	
72	Satt687	B2	97.924	6	0.55295		1.6956	
73	Satt726	B2	100.548	3	0.58308		1.8882	
74	SCT_094	B2	113.614	4	0.53694		1.8539	
75	Sat_009	B2	78.661	3	0.34816		1.2615	
76	Sct_064	B2	89.307	4	0.40459		0.3976	
77	Sat_042	C1	0	4	0.49936	0.45782	1.7251	1.74335
78	Sat_331	C1	9.02	5	0.73141		3.1082	
79	Satt136	C1	10.336	4	0.49529		1.2304	
80	Satt161	C1	24.106	5	0.53724		1.932	
81	satt164	C1	26.353	6	0.72674		3.0401	
82	Satt180	C1	65.078	6	0.76578		3.456	
83	Satt190	C1	67.025	5	0.57877		1.971	
84	Satt194	C1	73.324	3	0.43655		1.5533	
85	Satt361	C1	73.393	2	0.1501		0.6042	
86	Satt396	C1	73.788	7	0.5673		2.0189	
87	Satt399	C1	75.105	3	0.16995		0.6916	
88	Satt565	C1	75.52	3	0.45954		1.6044	
89	Satt578	C1	76.228	4	0.59683		2.0092	
90	Satt607	C1	82.506	3	0.40206		1.4497	
91	Satt718	C1	125.742	3	0.40073		1.409	
92	Sct_186	C1	127.774	2	0.32579		1.1998	
93	SOYGPATR	C1	132.458	3	0.52258		1.7564	
94	AW277661	C1	74.79	3	0.14836		0.621	
95	Sat_252	C2	4.223	4	0.67703	0.50503	2.1905	1.78045
96	AW734043	C2	26.649	4	0.57751		2.0119	
97	GmAC7L	C2	30.467	3	0.34065		1.2401	
98	Sat_062	C2	30.801	7	0.72902		2.7081	
99	Sat_076	C2	40.298	7	0.65605		2.3387	
100	Sat_213	C2	42.365	4	0.53636		1.866	
101	Sat_246	C2	44.662	3	0.34486		1.0976	
102	Sat_263	C2	45.757	4	0.44933		1.6095	
103	Sat_402	C2	69.665	5	0.72106		2.3724	
104	satt079	C2	75.431	3	0.40888		1.4982	
105	Satt100	C2	82.23	6	0.77417		2.5837	
106	Satt202	C2	89.305	3	0.43812		1.5599	
107	Satt227	C2	90.93	4	0.36553		1.3433	
108	Satt277	C2	91.805	4	0.53812		1.8298	
109	Satt281	C2	97.829	5	0.73145		2.9852	
110	Satt286	C2	99.181	4	0.5609		1.9134	
111	Satt289	C2	101.754	4	0.59767		2.149	
112	satt291	C2	103.328	6	0.52538		1.6086	
113	Satt305	C2	107.592	3	0.12689		0.5473	
114	Satt307	C2	111.682	5	0.55333		1.918	
115	Satt316	C2	112.189	3	0.481		1.6699	

Continued Table 1

ID	SSR loci	Linkage group	Locus	Alleles		Mean		Mean
116	Satt322	C2	112.352	3	0.2107		0.8225	
117	Satt365	C2	113.392	3	0.37498		1.3852	
118	Satt376	C2	113.96	3	0.55489		1.8288	
119	satt422	C2	117.766	4	0.59916		2.0337	
120	Satt450	C2	117.868	2	0.04163		0.2094	
121	Satt460	C2	118.782	4	0.70104		2.7692	
122	Satt489	C2	121.271	4	0.59163		1.9221	
123	Satt520	C2	126.236	3	0.50022		1.6812	
124	Satt557	C2	127.003	3	0.3774		1.3886	
125	Satt640	C2	127.668	5	0.57102		2.1123	
126	Sat_036	D1a	49.835	5	0.6428	0.53279	2.1988	1.88814
127	Sat_160	D1a	53.671	5	0.72672		2.7615	
128	Sat_201	D1a	54.499	4	0.5715		1.9121	
129	Sat_345	D1a	55.221	5	0.67411		2.4955	
130	Sat_346	D1a	55.681	4	0.49333		1.7195	
131	Sat_414	D1a	56.08	4	0.65992		2.1453	
132	Satt077	D1a	56.434	3	0.39902		1.473	
133	Satt198	D1a	56.566	2	0.36566		1.3611	
134	Satt254	D1a	60.969	4	0.62417		2.208	
135	Satt283	D1a	62.367	5	0.68417		2.7191	
136	Satt295	D1a	64.522	4	0.46467		1.6394	
137	Satt383	D1a	65.63	5	0.68049		2.3518	
138	Satt436	D1a	68.619	3	0.57747		1.8764	
139	Satt439	D1a	70.691	3	0.3723		1.3512	
140	Satt502	D1a	72.274	3	0.21249		0.6888	
141	satt507	D1a	75.251	4	0.45258		1.5982	
142	Satt515	D1a	77.49	4	0.59465		1.993	
143	Satt580	D1a	85.484	4	0.6291		2.2812	
144	Satt603	D1a	104.283	2	0.29789		1.1008	
145	Sat_089	D1b	0	4	0.34196	0.48464	1.2667	1.72289
146	Sat_096	D1b	9.8	5	0.69027		2.3084	
147	Sat_135	D1b	20.614	4	0.48443		1.6904	
148	Sat_254	D1b	37.073	5	0.66162		2.2098	
149	Sat_283	D1b	40.041	5	0.6833		2.7191	
150	Sat_289	D1b	46.617	3	0.58994		2.0354	
151	Sat_351	D1b	46.92	4	0.56439		1.8967	
152	Sat_423	D1b	52.613	3	0.42133		1.5148	
153	Satt005	D1b	59.61	4	0.57483		1.9302	
154	Satt041	D1b	67.625	3	0.34253		1.2432	
155	Satt141	D1b	70.653	5	0.43661		1.574	
156	Satt157	D1b	72.887	6	0.69531		2.3421	
157	Satt189	D1b	74.786	3	0.385		1.4267	
158	satt216	D1b	75.292	3	0.28127		1.0244	
159	Satt266	D1b	75.412	3	0.5888		1.901	
160	Satt274	D1b	75.674	3	0.49348		1.664	
161	Satt296	D1b	75.939	4	0.28062		1.054	

Continued Table 1

ID	SSR loci	Linkage group	Locus	Alleles		Mean		Mean
162	Satt350	D1b	76.283	3	0.36753		1.3269	
163	Satt428	D1b	76.323	3	0.57739		1.8763	
164	Satt459	D1b	76.596	3	0.36776		1.3078	
165	Satt506	D1b	77.349	3	0.38516		1.3652	
166	Satt537	D1b	84.04	5	0.54867		1.9121	
167	Satt546	D1b	87.204	3	0.38513		1.4272	
168	Satt579	D1b	98.745	8	0.73951		3.4182	
169	Satt600	D1b	116.35	4	0.48097		1.7092	
170	Satt634	D1b	118.618	4	0.25372		0.9555	
171	Satt701	D1b	131.917	4	0.63507		2.2868	
172	Satt703	D1b	136.659	3	0.41114		1.4344	
173	BE475343	D1b	30.743	4	0.38682		1.1434	
174	Sat_022	D2	16.756	6	0.77691	0.51852	3.526	1.90828
175	Sat_086	D2	26.049	4	0.59255		1.9966	
176	Sat_092	D2	47.726	6	0.76408		3.4678	
177	Sat_194	D2	57.07	4	0.63918		2.2046	
178	Sat_300	D2	57.505	7	0.72535		2.4703	
179	satt002	D2	67.712	3	0.47512		1.6324	
180	Satt031	D2	79.231	4	0.51217		1.8108	
181	Satt135	D2	85.148	4	0.261		1.0066	
182	Satt154	D2	85.686	3	0.53446		1.7829	
183	Satt186	D2	86.69	4	0.51316		1.77	
184	Satt226	D2	87.557	4	0.26224		0.9468	
185	Satt301	D2	87.666	6	0.67965		2.2944	
186	Satt328	D2	87.876	2	0.05159		0.25	
187	Satt386	D2	88.02	3	0.24551		0.8694	
188	Satt389	D2	93.709	8	0.80579		3.9422	
189	Satt514	D2	105.451	4	0.64713		2.1477	
190	Satt543	D2	114.966	3	0.56265		1.8458	
191	Satt574	D2	115.703	3	0.42995		1.5049	
192	Satt662	D2	118.662	3	0.53615		1.7877	
193	Satt669	D2	120.299	5	0.51649		1.8405	
194	Satt672	D2	124.998	4	0.59377		1.9389	
195	sct_137	D2	128.95	3	0.28245		0.9459	
196	sat_136	E	3.721	6	0.58268	0.43551	2.062	1.55723
197	Sat_273	E	12.919	3	0.39009		1.3722	
198	satt045	E	19.296	7	0.62122		2.4326	
199	Satt117	E	32.101	5	0.57102		2.1123	
200	satt151	E	32.272	4	0.5288		1.4972	
201	Satt185	E	34.197	3	0.56462		1.8449	
202	Satt212	E	39.162	2	0.34902		1.2832	
203	satt213	E	41.241	3	0.01507		0.0947	
204	Satt263	E	41.683	4	0.62908		2.0694	
205	Satt369	E	43.363	4	0.47681		1.6335	
206	Satt384	E	44.763	3	0.25448		0.97	
207	satt411	E	44.925	3	0.32224		1.1799	

Continued Table 1

ID	SSR loci	Linkage group	Locus	Alleles		Mean		Mean
208	satt452	E	44.976	3	0.33001		1.1896	
209	satt483	E	45.096	4	0.31921		1.1402	
210	satt598	E	45.397	2	0.33407		1.2094	
211	Satt602	E	45.627	4	0.47119		1.0434	
212	Satt651	E	45.782	6	0.77479		3.548	
213	Satt699	E	46.646	4	0.67403		2.5664	
214	satt706	E	47.501	3	0.34605		1.2723	
215	satt716	E	56.272	2	0.1558		0.6234	
216	AW186493	F	1.918	4	0.66191	0.48225	2.1548	1.6988
217	BE806387	F	2.233	4	0.70043		2.7536	
218	Sat_074	F	3.043	4	0.61974		2.0469	
219	Sat_090	F	3.634	3	0.57781		1.8804	
220	Sat_120	F	3.952	2	0.20053		0.7552	
221	Sat_133	F	5.361	3	0.39318		1.4246	
222	Sat_154	F	10.653	5	0.62944		2.1592	
223	Sat_240	F	15.291	4	0.66176		2.1482	
224	Sat_297	F	16.084	4	0.61957		2.0842	
225	Sat_309	F	18.125	3	0.29696		1.1082	
226	Satt030	F	20.564	4	0.61876		2.0712	
227	Satt114	F	21.036	7	0.72416		2.638	
228	Satt144	F	22.965	3	0.41667		0.6306	
229	Satt145	F	25.584	3	0.48368		1.6322	
230	Satt146	F	26.706	3	0.46761		1.5885	
231	Satt149	F	33.185	4	0.63749		2.104	
232	Satt160	F	41.471	6	0.731		3.152	
233	Satt252	F	43.013	4	0.28207		1.0523	
234	Satt325	F	43.442	3	0.42832		1.5072	
235	Satt334	F	44.419	3	0.47542		1.6441	
236	Satt335	F	50.78	3	0.28166		1.0259	
237	Satt343	F	59.597	6	0.72484		2.5494	
238	satt348	F	63.689	3	0.21669		0.8511	
239	Satt362	F	68.908	3	0.33121		1.2093	
240	Satt374	F	71.412	4	0.50786		1.7515	
241	satt423	F	74.128	4	0.48305		1.6775	
242	satt425	F	75.972	4	0.49299		1.7039	
243	Satt490	F	77.704	3	0.19772		0.7847	
244	Satt510	F	78.055	5	0.43859		1.594	
245	Satt516	F	82.834	3	0.27834		1.0461	
246	Satt522	F	97.966	3	0.20966		0.9418	
247	Satt554	F	102.083	4	0.58081		1.8278	
248	satt586	F	111.891	5	0.55435		1.9475	
249	Satt649	F	119.186	3	0.4324		1.5202	
250	Satt656	F	130.643	4	0.22059		0.4022	
251	Satt659	F	135.121	4	0.43027		1.5311	
252	Sct_033	F	142.348	9	0.83581		2.9673	
253	Sat_064	G	0	7	0.79074	0.49924	2.6896	1.73685

Continued Table 1

ID	SSR loci	Linkage group	Locus	Alleles		Mean		Mean
254	Sat_141	G	1.84	5	0.58342		1.9671	
255	Sat_164	G	2.201	6	0.63159		2.1843	
256	Sat_185	G	9.183	3	0.30636		1.1272	
257	Sat_203	G	10.923	3	0.34431		1.2448	
258	Sat_223	G	12.542	4	0.65692		2.4555	
259	Sat_372	G	21.889	4	0.56165		1.9112	
260	Satt012	G	33.263	4	0.25797		1.0081	
261	Satt038	G	43.378	4	0.39912		1.4132	
262	Satt163	G	52.938	4	0.44134		0.8757	
263	Satt191	G	56.523	4	0.58568		1.9787	
264	Satt235	G	60.371	4	0.40275		1.4828	
265	Satt275	G	61.644	4	0.60693		2.0781	
266	Satt288	G	62.077	5	0.61077		2.0769	
267	Satt324	G	63.003	4	0.70004		2.7508	
268	Satt394	G	66.551	4	0.50967		1.7501	
269	Satt472	G	68.673	5	0.5813		1.9976	
270	Satt505	G	69.869	3	0.21249		0.6888	
271	Satt517	G	76.77	4	0.50367		1.7187	
272	Satt533	G	94.403	4	0.65572		2.437	
273	Satt594	G	94.844	5	0.29722		1.1441	
274	Satt610	G	96.572	3	0.36513		1.3325	
275	Satt688	G	107.747	4	0.66884		2.5284	
276	sct_199	G	108.702	3	0.39485		1.415	
277	Satt131	G	48.923	2	0.26648		0.7982	
278	Sat_290	G	29.027	4	0.64518		2.1036	
279	Sat_127	H	0.594	3	0.51196	0.46458	1.738	1.37557
280	Sat_205	H	2.851	4	0.47771		1.6892	
281	Sat_214	H	4.882	5	0.6208		2.0608	
282	Sat_401	H	8.48	4	0.36984		1.3491	
283	Satt052	H	28.803	4	0.45098		1.601	
284	Satt181	H	38.894	4	0.61942		2.043	
285	Satt192	H	44.035	5	0.40313		0.4031	
286	Satt222	H	46.947	3	0.38577		1.324	
287	Satt314	H	53.35	2	0.1323		0.5436	
288	Satt353	H	58.906	6	0.53485		1.8389	
289	Satt442	H	64.101	3	0.51824		1.7243	
290	Satt469	H	66.21	6	0.48885		1.1473	
291	Satt541	H	68.082	4	0.55234		1.8857	
292	Satt635	H	68.182	6	0.5049		1.3944	
293	Satt666	H	69.115	4	0.43283		0.7022	
294	sctt009	H	91.116	3	0.42929		0.5645	
295	GMGLPSI2	I	18.5	3	0.48991	0.50744	1.6607	1.8014
296	Sat_268	I	27.979	4	0.62778		2.1139	
297	Sat_299	I	31.494	3	0.43634		1.5025	
298	Sat_418	I	33.984	6	0.67397		2.3073	
299	Sat_419	I	35.347	5	0.67085		2.6366	

Continued Table 1

ID	SSR loci	Linkage group	Locus	Alleles		Mean		Mean
300	Sat_421	I	36.935	4	0.28901		1.1023	
301	Satt127	I	46.22	2	0.33848		1.2462	
302	satt162	I	50.108	3	0.51703		1.725	
303	Satt239	I	55.097	3	0.55406		1.8165	
304	Satt270	I	74.263	6	0.59453		2.2092	
305	Satt354	I	81.399	3	0.26727		1.0106	
306	Satt367	I	86.741	5	0.70494		2.8303	
307	Satt440	I	97.047	3	0.45525		1.5599	
308	Satt571	I	98.113	4	0.5163		1.8032	
309	Satt587	I	99.827	4	0.59155		1.9762	
310	Sct_189	I	112.7	4	0.58683		1.9745	
311	Sctt012	I	113.768	3	0.31238		1.1489	
312	Sat_151	J	11.741	4	0.54032	0.56124	1.8385	1.93968
313	Sat_224	J	12.409	4	0.65526		2.133	
314	Sat_395	J	15.688	4	0.59092		1.9708	
315	Satt132	J	25.513	3	0.25827		0.9826	
316	Satt183	J	37.042	4	0.6954		2.231	
317	Satt215	J	38.097	3	0.31264		1.1401	
318	Satt244	J	38.192	4	0.64135		2.0991	
319	Satt249	J	39.181	4	0.61686		2.2405	
320	Satt285	J	39.643	2	0.27839		1.0336	
321	Satt287	J	41.354	3	0.43416		1.5104	
322	Satt405	J	41.903	5	0.70494		2.8324	
323	Satt406	J	42.253	5	0.7318		2.6077	
324	Satt414	J	42.505	4	0.50152		1.7259	
325	Satt431	J	44.084	4	0.61755		2.2012	
326	Satt529	J	65.037	3	0.57592		1.956	
327	Satt547	J	67.786	4	0.64071		2.0982	
328	Satt596	J	75.132	5	0.74567		2.4768	
329	satt622	J	78.569	4	0.50833		1.7885	
330	Satt654	J	89.478	6	0.69638		2.8364	
331	Sctt011	J	62.885	4	0.47847		1.0908	
332	Sat_043	K	0.912	3	0.45163	0.52361	1.5486	1.83761
333	Sat_044	K	1.798	6	0.55547		1.9699	
334	Sat_167	K	14.354	4	0.63199		2.2716	
335	Sat_293	K	32.955	4	0.52374		1.7818	
336	Satt046	K	36.991	3	0.54748		1.836	
337	Satt055	K	40.864	4	0.38119		1.3853	
338	Satt137	K	42.392	4	0.12898		0.552	
339	Satt178	K	42.709	4	0.46309		1.6255	
340	Satt242	K	43.042	4	0.61698		2.0388	
341	Satt260	K	43.347	5	0.64042		2.1599	
342	Satt349	K	45.594	4	0.41849		1.5226	
343	Satt475	K	46.796	6	0.67197		2.7733	
344	Satt499	K	50.804	4	0.64149		2.0997	
345	satt539	K	58.005	2	0.36173		1.3271	

Continued Table 1

ID	SSR loci	Linkage group	Locus	Alleles		Mean		Mean
346	Satt544	K	61.674	5	0.56242		1.982	
347	Satt555	K	71.005	4	0.64636		2.3663	
348	Satt673	K	78.683	3	0.38707		1.418	
349	Satt715	K	80.119	3	0.51103		1.7121	
350	Satt727	K	85.188	4	0.68762		2.2123	
351	Sct_196	K	99.103	5	0.64298		2.1694	
352	Sat_150	L	0	3	0.42025	0.53029	1.5182	1.92074
353	Satt076	L	10.345	4	0.53176		1.7922	
354	Satt156	L	14.027	4	0.58681		1.9667	
355	Satt166	L	19.928	4	0.70162		2.762	
356	Satt182	L	27.922	4	0.70298		2.7714	
357	Satt229	L	34.54	5	0.63933		2.1431	
358	Satt232	L	53.671	4	0.49251		1.4448	
359	Satt238	L	54.574	2	0.19296		0.7474	
360	Satt313	L	56.142	6	0.80409		3.7604	
361	Satt373	L	61.349	4	0.60927		2.0166	
362	Satt448	L	64.661	4	0.64269		2.1033	
363	Satt481	L	66.51	4	0.59154		1.9654	
364	Satt495	L	70.364	2	0.36887		1.362	
365	Satt513	L	93.885	5	0.48611		1.7374	
366	Satt523	L	106.366	4	0.34262		1.2697	
367	Satt527	L	107.24	2	0.37118		1.3712	
368	Sat_226	M	7.841	4	0.5971	0.57091	1.993	1.91304
369	Satt150	M	18.582	3	0.39846		1.464	
370	Satt175	M	33.468	6	0.66982		2.2757	
371	Satt245	M	35.854	3	0.54631		1.8073	
372	Satt308	M	50.097	4	0.66961		2.2189	
373	Satt463	M	53.538	5	0.66713		2.2036	
374	Satt536	M	58.595	4	0.5509		1.8287	
375	Satt540	M	61.043	4	0.52401		1.8139	
376	Satt567	M	62.14	3	0.58253		1.9093	
377	Satt590	M	65.791	4	0.58024		1.93	
378	Satt626	M	66.988	4	0.5715		1.822	
379	Satt702	M	130.756	4	0.49334		1.6901	
380	Sat_084	N	28.52	4	0.49743	3.98272	1.6647	2.09624
381	Sat_208	N	32.848	5	0.59481		2.2084	
382	Satt009	N	34.649	5	0.62673		2.0724	
383	Satt022	N	35.32	6	0.80461		3.766	
384	Satt530	N	36.863	5	0.72424		2.4437	
385	Satt584	N	37.976	3	0.31794		1.1583	
386	Satt624	N	39.351	4	0.40296		1.4389	
387	Satt675	N	102.055	4	0.60887		2.0175	
388	BE801128	O	5.44	4	0.65455	3.98272	2.4379	1.63525
389	Sat_108	O	8.748	5	0.63254		2.104	
390	Sat_132	O	20.428	4	0.55059		1.9132	
391	Sat_231	O	39.817	3	0.16469		0.6767	

Continued Table 1

ID	SSR loci	Linkage group	Locus	Alleles	Mean	Mean
392	Sat_242	O	42.287	2	0.35703	1.3155
393	Sat_274	O	49.711	3	0.35487	1.3095
394	Sat_291	O	51.897	6	0.66662	2.383
395	Satt153	O	58.398	3	0.553	1.8169
396	Satt173	O	59.489	6	0.75699	2.8402
397	Satt241	O	68.97	3	0.36182	1.3121
398	Satt243	O	74.049	3	0.49466	1.6912
399	Satt259	O	100.376	3	0.45629	1.5549
400	Satt347	O	107.577	4	0.40434	0.4241
401	Satt358	O	118.137	4	0.44498	1.5297
402	Satt420	O	119.504	3	0.46921	1.6098
403	Satt445	O	128.443	3	0.49435	1.7002
404	Satt592	O	129.296	3	0.4251	1.4884
405	Sat_221	O	51.009	3	0.37486	1.3272

The genetic similarity coefficient (genetic similarity, GS) of 327 materials varied from 0.668 to 0.926, with an average of 0.76 (Figure 3), and the GS of the experimental materials showed a large genetic difference and high genetic diversity.

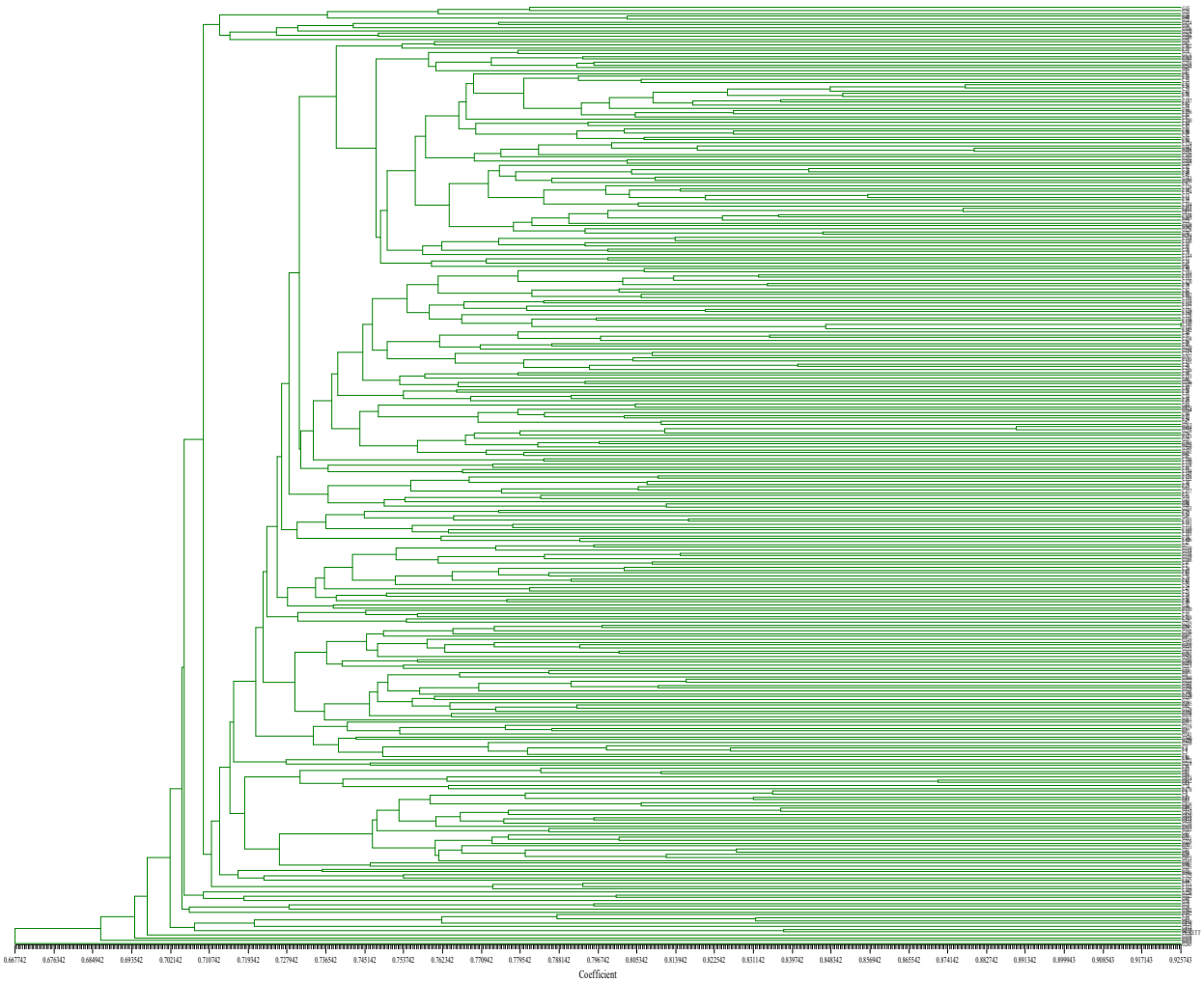

Figure 3 GC analysis of population

1.3 The group structure and the linkage disequilibrium analysis

The experimental materials were partitioned by the population genetic structure based on the mathematic model with the help of STRUCTURE 2.3.4, and the Q-value was calculated. The maximum likelihood 1n(D) was found to increase with the increase in K-value. Therefore, the ΔK value must be calculated to confirm the ideal K value (G. EVANNO, 2005) (Figure 4).

Using this method, the K-value was confirmed as 3 (Figure 4) and the experimental materials were divided into 3 sub-populations.

The numbers of materials in the 3 sub-groups, respectively, were 123, 124, and 54, with rates of the total group of 38%, 38% and 17%. The other 26 materials belonged to the mixed subpopulation (Figure 5). In the LGA analysis, the experimental materials were divided into two sub-populations, with material numbers of 170 and 157.

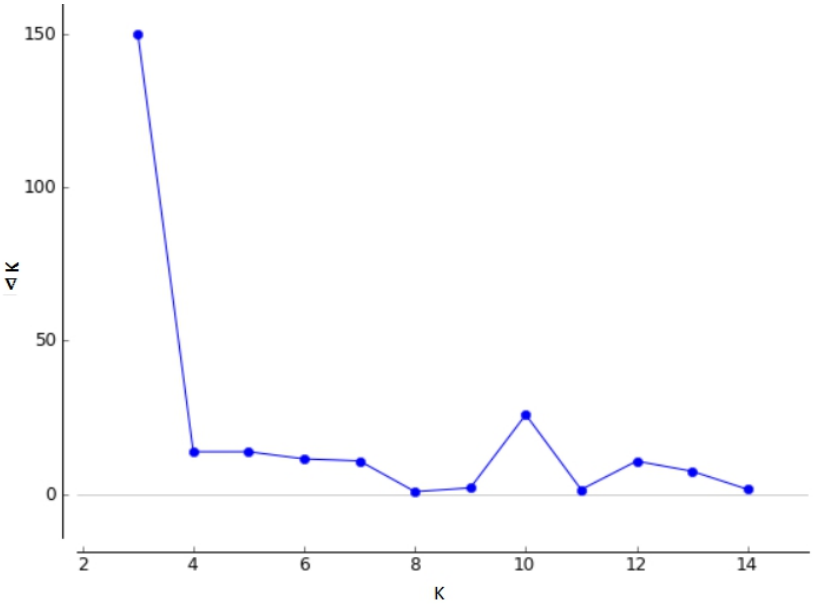

Figure 4 Change of ΔK according to K-value

The software Structure Harvester was used to calculate the ΔK value and provide the ideal K value:

$L'(K) = L(K) - L(K-1)$

$|L''(K)| = |L'(K+1) - L'(K)|$

$\Delta K = \text{mean}(|L''(K)|) / \text{sd}(L(K))$

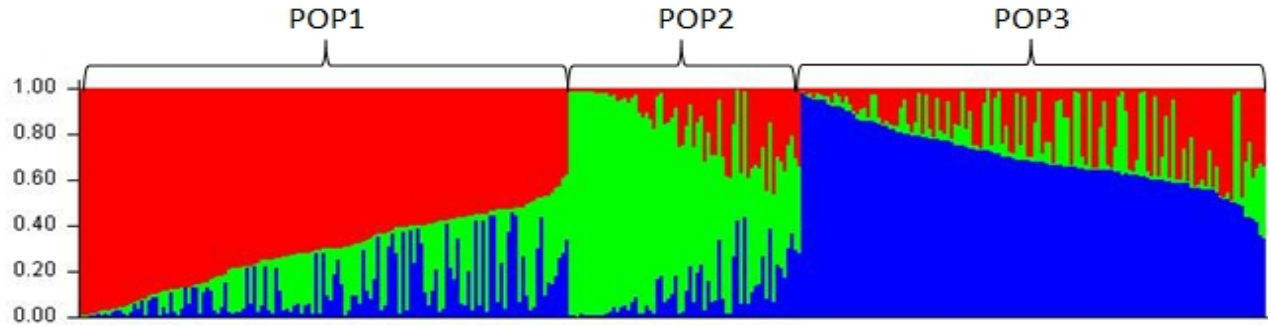

Figure 5 The population structure of 326 soybean materials. The population structure of 326 soybean materials were calculated by the Software STRUCTURE. The association mapping population (AMP) was divided into three subgroups, which were represented by Arabic numerals (1-3). The vertical axis shows the probability that the material belongs to the population

The linkage disequilibrium analysis was the basis of the association analysis. In the experiment, a linkage disequilibrium existed between collinear SSR markers (in the same chromosome) and non-collinear (in different chromosomes) SSR markers to a different extent (Table 2). Among them, the linkage groups A1 and C1 had the highest linkage disequilibrium (D`>0.60), while the linkage group F had the lowest.

Table 2 Distribution of D`values of pairwise SSR loci

Significant LD locus pairs	Frequency of D`value (p<0.001)					Mean of D`
	0-0.20	0.21-0.40	0.41-0.60	0.61-0.80	0.81-1.00	
7972	6	4614	2669	438	218	0.408

1.4 The analysis of resistance genes correlated with the soybean locus

GWA and LGA were conducted under the MLM model to map the loci related to SN1 resistance. Seven extremely significantly (p<0.01) associated loci and 24 significantly associated loci were obtained from GWA. Nine (p<0.05) significantly associated loci were obtained from LGA. There were five significantly associated loci common to both the GWA and LGA results.

In GWA, the accumulated phenotypic variation explanation ratio of seven extremely significantly associated SSR loci was 32.1%. Sat_150 in linkage group L exhibited the maximal explanation (6.8%), while Satt150 in linkage group M exhibited the minimal explanation (3%). All 7 loci were obtained for the first time (Table 3).

Table 3 SSR loci association with SMV1

Maker	Linkage Group	Locus	P-Value	R2
Sat_150	L	53.67	0.0000278	0.068
Satt377	A2	90.84	0.000709	0.04653
Satt624	N	35.32	0.001908	0.04005
Sct_196	K	43.04	0.002352	0.03869
Satt376	C2	97.82	0.003133	0.03682
Satt202	C2	126.23	0.006015	0.03259
Satt150	M	18.58	0.008892	0.03006

In LGA, the accumulated phenotypic variation explanation ratio of 9 significant associated SSR loci was 24.4%. Satt296 exhibited the maximal explanation (3.34%), while Satt149 exhibited the minimal explanation (2.3%).

Candidate genes were mined from 3 cM sections on both sides of 15 associated SSR loci obtained from GWA and LGA. As shown in table 4, associated genes can generally be categorized into four types. In the first type, most resistance genes have NB-ARC and LRR domains, which identifies them as R resistance genes with the NBS-LRR domain. In the second type, most genes have the S_TKc domain, which identifies them as receptor-like protein kinase or MAPK (mitogen-activated protein kinase). The third type is related to jasmonate (JA) and the synthesis process of JAs, such as linoleate 9S-lipoxygenase and glyceraldehyde-3-phosphate dehydrogenase. The forth type of genes is related to biotic stress response proteins, such as SACPD (stearoyl-ACP desaturase) and cysteine protease (Table 4).

Taking Sat_150 (P=2.78E-05) positioned at 53.67 cM of the L Linkage group from the GWA analysis as an example, 9 resistance-related genes were obtained by gene mining. Among them, Glyma19g31950, Glyma19g32080, Glyma19g32110, and Glyma19g32150 are R resistance genes with NB-ARC and LRR domains. Glyma19g33290, Glyma19g33300, Glyma19g33310, Glyma19g33320, and Glyma19g33330 are the DIR (Dirigent-like) genes that participate in the synthesis of lignin and the defense response to fungi (Fristensky et al., 1985; Riggleman et al., 1985).

Table 4 SSR loci association with SMV 1

Maker	Linkage Group	Locus	P-Value	R2
Satt296	D1b	52.613	0.005119	0.033575
Satt216	D1b	9.8	0.012206	0.027967
Satt157	D1b	37.073	0.04301	0.019893
Sat_189	B2	72.918	0.044479	0.019678
Satt168	B2	55.198	0.038399	0.020617
Sat_154	F	68.908	0.013676	0.027236
Sat_297	F	59.597	0.015139	0.026582
Sat_240	F	25.584	0.02188	0.024218
Satt149	F	18.125	0.026754	0.022928

Taking Satt296 (P=5.12E-03) positioned at 52.61 cM of the D1b linkage group from LGA analysis as an example, 6 resistance-related genes were obtained by gene mining (Table 4). Among them, Glyma02g13460 and Glyma02g13470 are receptor-like protein kinase genes. Gene Glyma02g15600 encodes for SACPD (stearoyl-ACP desaturase), which may participate in the plant defense response to viruses. Silencing SACPD-encoded genes could increase the content of stearic acid (SA) and the gene expression of constitutive pathogenicity-related genes, which leads to the improvement of resistance (Kachroo et al., 2008). The gene Glyma02g14940 encodes for a type of ethylene-responsive transcription factor (ERF), which could regulate the expression of the RP gene by combining the ethylene-responsive transcription factor GCC-box with the RP gene promoter. Others have found that overexpressing ERF in *Arabidopsis* and tobacco can induce the expression of RP genes and improve the resistance to fungi, bacteria and viruses (Park and Twell, 2001; Yi, 2004; Zuo et al., 2007). The gene Glyma02g15690 encodes for a type of MAPK protein that belongs to the last protein kinase in the MAPK cascade, which has an important role in plant systemic defense responses (Meng et al., 2012). The gene Glyma02g15831 encodes for a type of cysteine protease that participates in the hypersensitivity caused by the invasion of germs and the PCD (programed cell death) of organ failure (Chen et al., 2010).

1.5 Related locus analysis of soybean SN3 resistance

GWA and LGA were conducted under the MLM model to map the loci related to SN3 resistance. Seven extremely significantly (p<0.01) associated loci and 39 significantly associated loci were obtained from GWA. Twelve significantly (p<0.05) associated loci were obtained from LGA. There were 5 significantly associated loci common to the GWA and LGA results.

According to GWA, the accumulated phenotypic variation explanation ratio of 7 extremely significantly associated SSR loci was 24.27%. Satt252 in linkage group F had the maximal explanation (4.61%), and Satt020 had the minimal explanation (2.96%) (Table 5).

According to LGA, the accumulated phenotypic variation explanation ratio of 12 significantly associated SSR loci was 35.4%. Satt252 had the maximal explanation (4.7%), and Satt726 had the minimal explanation (2.0%) (Table 6).

Candidate genes from a 3 cM section on both sides of 15 associated SSR loci obtained from GWA and LGA were mined. According to Table 8, the associated genes can generally be categorized into three types. In the first type, most resistance genes have NB-ARC and LRR domains, which identifies them as R resistance genes with an NBS-LRR domain. The second type is related to JA and the synthesis process of JAs, such as linoleate 13S-lipoxygenase, glyceraldehyde-3-phosphate dehydrogenase and 12-oxo-phytodienoic acid reductase. The third type of gene contains the S_TKc domain, which identifies them as MAPKKK (mitogen-activated protein kinase kinase kinase).

Taking Satt726 (P=4.20E-02) positioned at 100.55 cM of the B2 linkage group from the LGA analysis as an example, 9 resistance-related genes were obtained by gene mining. Among them, Glyma14g37860 belongs to the

RPP13-like gene, which inhibits the growth of bacteria and viruses. The other 8 genes (Glyma14g36511, Glyma14g38500, Glyma14g38516, Glyma14g38533, Glyma14g38561, Glyma14g38586, Glyma14g38700, and Glyma14g38741) are all R resistance genes with an NBS-LRR domain.

Table 5 SSR loci association with SMV3

Make	Linkage Group	Locus	P-Value	R2
Satt252	F	16.084	0.000744	0.0461
Satt376	C2	97.829	0.00282	0.0374
Satt168	B2	55.198	0.005748	0.0328
Sat_074	F	142.348	0.005845	0.0327
Sat_418	I	74.263	0.006344	0.0322
Satt304	B2	65.555	0.006414	0.0321
Satt020	B2	72.127	0.009489	0.0296

Table 6 SSR loci association with SMV strain N3

Maker	Linkage Group	Locus	P-Value	R2
Satt252	F	16.084	5.77E-04	0.04782
Sat_074	F	142.348	0.001459	0.041753
Satt020	B2	72.127	0.002694	0.037758
Satt304	B2	65.555	0.003224	0.036591
Satt168	B2	55.198	0.006616	0.031933
Satt516	F	44.419	0.01391	0.02714
Satt428	D1b	77.349	0.022254	0.024121
Satt474	B2	75.346	0.026966	0.02289
Satt141	D1b	72.887	0.027278	0.022816
SCT_094	B2	70.56	0.035557	0.021119
Sat_154	F	68.908	0.038924	0.020541
Satt726	B2	100.548	0.041964	0.02006

Taking Satt252 positioned at 43.01 cM of the F Linkage group from the GWA and LGA analyses as an example, 6 resistance-related genes, Glyma13g04200, Glyma13g04230, Glyma13g04410, Glyma13g04540, Glyma13g03790 and Glyma13g06010, were obtained by gene mining. Among them, Glyma13g04200 and Glyma13g04230 encode two RPP13-like antiviral proteins with NB-ARC and LRR domains, respectively. Glyma13g04410 encodes an alpha-dioxygenase that belongs to the peroxidase superfamily, which catalyzes the oxidation of fatty acids to form 2-hydroperoxy fatty acid. Its expression levels in different plants were associated with plant resistance to the invasion of pathogens (Hamberg et al., 2003). Glyma13g03790 encodes for a linoleate 13S-lipoxygenase, which is a key enzyme of the fatty acid metabolic pathway. Its activity enhancement and rapid peroxidation of fatty acid play a role in the stress response mechanism for plants in response to biotic and abiotic stress (Hwang and Hwang 2010; Vicente et al., 2012). Glyma13g06010 and Glyma13g04540 encode histone deacetylase and carboxylesterase, respectively.

1.6 Comprehensive analysis of soybean SN1 and SN3 resistance

The comprehensive analysis of soybean SN1 and SN3 resistance revealed that three loci are associated with two characters simultaneously. These loci are Sat_154 from the F linkage group, Satt168 from the B2 linkage group and Satt376 from the C2 linkage group.

Two resistance-related genes, Glyma13g25420 and Glyma13g25440, were obtained by gene mining around Sat_154. These genes are resistance genes with an NBS-LRR domain.

Locus Satt168 at 68.91 cM of the B2 linkage group is related to anti-fungal activity (Wang and Ghabrial, 2002) and *Heterodera glycines* (Yue et al., 2001) resistance in soybean. Soybeans with this locus exhibit broad-spectrum

resistance that is related to the pest control response. Five resistance-related genes were obtained by gene mining. Among them, Glyma14g08700 and Glyma14g08710 are resistance genes with an NBS-LRR domain. Glyma14g08900, Glyma14g08910 and Glyma14g08920 are PLP2 (patatin-like protein) resistance genes.

Locus Satt376 is related to the tocopherol content of soybean. The tocopherol cyclase (TC) gene, which controls the synthesis of tocopherol, could be over-expressed to strengthen the resistance of plants (Li and Shi, 2010). Four genes related to resistance were obtained by gene mining. Among them, Glym06g19890 and Glyma06g19920 encode an EDS1K-like protein and regulate a resistant protein. Glyma06g18110 and Glyma06g18120 encodes glyceraldehyde-3-phosphate dehydrogenase (GAPDH), which may play a role in the viral infection process.

2 Discussion

2.1 Analysis and comparison of GWA and LGA

This experiment adopted two association analysis strategies, genome-wide association (GWA) and linkage group association (LGA). GWA used SSR loci distributed in the whole genome to mine soybean SMV resistance-related loci with a threshold of 0.001. LGA used SSR loci, distributed mainly on the linkage groups (B1, D1b, F) gathering known SMV resistance-related loci, to mine related loci with a threshold of 0.05.

In association analysis, the P value (threshold value) was selected to meet the requirements for finding significantly associated loci and to reduce false-positive associations. In GWA, when the P value was increased from 0.001 to 0.05, the corresponding associated loci number for SN1 and SN3 increased from 7 and 7 to 24 and 39, respectively. In addition, the degree of significant increase in false positives was greater than that of association. Therefore, to ensure the accuracy of associated loci finding, we selected the threshold value of 0.001.

Compared with the GWA, LGA had fewer molecular data, but three linkage groups contain most of the known loci related to SMV resistance. Setting the P value to 0.05 ensured valid mining of correlated loci when there were fewer molecular data, but the determination of final candidate-associated loci should be combined with the bioinformatics analysis of loci.

Comparing the results from GWA and LGA revealed that 5 loci from LGA (Satt296, Sat_154, Sat_240, Satt149, and Satt157) related to SN1 resistance showed significant but not extremely significant correlations in GWA ($0.001<p<0.05$). In the association analysis of SN3, there were 5 loci with significant correlation and extremely significant correlation by LGA and GWA separately. Therefore, the results obtained by the mutual authentication of two methods can prevent the omission of less intensely associated loci and provide secondary verification of a portion of the results. However, the LGA method is applicable to associated traits with a certain basis of molecular research, and related loci should distribute to a small number of linkage groups or chromosomes.

2.2 The associated loci and known RSV gene locations

Among the loci associated with SN1 resistance, 3 (Sat_297, Satt296 and Sat_154) were the same as or adjoining to those found in previous studies. Two relevant loci (Sat_154 and Satt726) associated with SN3 resistance that were the same as found in previous studies. Locus Sat_154 was associated with both SN1 and SN3.

Among those resistance-associated loci, Sat_297 (59.597) was mapped by Li et al. (2010). Shi et al. (2008) found that Satt726 was linked closely to the Rsv3 gene of cultivar J05, with a 2.0 cM genetic distance.

The locus Satt296, related to SN1 resistance according to Luan, was associated with SN3 resistance in our research (Luan et al., 2006). In the public genetic map, this locus has distances of 0.41 cM and 6 cM from Satt542 (53.02) related to gene Rsv4 from Hayes' research and Satt634 (46.61) related to gene Rsc7 from Fu's research (Hayes et al., 2000; Fu et al., 2006).

The locus Sat_154 (68.908) from the F linkage group located between the SN1-resistant section Satt114 (63.689)-Satt510 (71.412) was mapped by Teng et al. (2011), and that between the Rsc11 resistance section Satt114 (63.689)- Sct_033 (74.12) was mapped by Bai et al. (2009). From the public genetic map, Sat_154 and Sat_297 were both located upstream of gene Rsv1 with genetic distances to gene Rsv1 of 2.3 cM and 11.6 cM,

respectively. Additionally, Sat_154 was located within the SN3 resistance section Satt114 (63.689)-Satt362 (82.834), mapped by Teng et al. (2005).

The research results showed that there were one or more genes related to SMV resistance surrounding these 4 associated loci. However, fine mapping and further gene mining are still required to specify these genes. The same and adjoining loci also showed that our research method has reliability and reference value.

2.3 The influence of populations and strains used in this research

In previous studies, mining the SMV resistance locus was uncommon; most studies used recombinant inbred lines as experimental materials, which led to fewer linkage loci identifications and limited application to the corresponding genetic population. Compared with the genetic population used in composite interval mapping, the resource population accumulated a large amount of recombinant information during long-term evolution. This research successfully structured a large resource population containing 327 breeding materials commonly used in northeastern China and took SMV resistance as the target trait. The primary popular strains N1 and N3 of SMV in northeastern China were chosen as the pathogenic strains. Because northeastern China is the main production area and source land for Chinese soybean and local cultivars and SMV pathogenic strains are regional, the resistance studies on other strains from different areas are difficult to apply. The present research focused on studying SMV resistance in northeastern China among local breeding materials, and the associated loci identified in the research may have limited application. The present findings also provide a partial molecular background for all the materials in this resource population, which would benefit breeding materials selection and SMV resistance marker breeding in the northeast region.

In the survey of SN1 infection, the DI of the entire population was mainly in the low acceptance area of 0.20-0.30, accounting for 49.84%, and its variance range was 0.1, defined as hypovirulence. For SN3 infection, the DI was mainly in the mid-acceptance area of 0.40-0.50, accounting for 34.58%, and its variance range was 0.1 with strong pathogenicity. The results of the surveys on inflection were the same as previously reported (Zheng et al., 2000b).

The resource population containing 327 materials used in this research covered most of the breeding materials in northeastern China. These materials were all susceptible to SMV but with different degrees of resistance, which means that SMV is a common disease in the northeast with a high incidence rate, leading to yield decrease. Soybean materials with full resistance were not found. Because the materials collected in the population were mainly breeding materials, which had all undergone resistance selection for SMV acceptance during their cultivation, material exhibiting extreme acceptance did not appear in the population. However, cultivars with obviously extreme resistance were also not found in this study. This result highlights the importance of identifying key resistance genes and cultivating resistance cultivars.

The association analysis method has a greater ability to explain the recombinant information in the resource population. In the genome-wide scan, a number of molecular markers were needed to obtain more precise and associated loci for the target trait. This research used SSR markers with the convenience of obtaining data and higher repeatability, but as its marker density was lower than for SNP markers, the precision of our research was somewhat diminished. In later studies, based on the current results, we could specifically develop a large number of SNP markers to improve the density. Combining these results with soybean genomic sequencing results, we could further define the molecular markers that are closely associated with SMV resistance genes and clone and validate the related candidate genes, providing a corresponding theoretical basis for SMV resistance breeding and the selection of breeding materials with higher resistance to SMV.

3 Materials and Methods

3.1 Test materials

This research used 327 soybean cultivars and lines as materials, which were the common breeding materials in northeastern China collected by our laboratory.

The soybean mosaic virus SN1 and SN3 strains were provided by the Jilin Academy of Agricultural Sciences, China.

3.2 Identification of resistance to soybean diseases SN1 and SN3

The disease index detection for SN1 and SN3 infection was performed by Jia Liu from the Jilin Academy of Agricultural Sciences. The procedure was as follows: 150 plants of each cultivar or line were planted in a 6-meter-long ridge with 3 repetitions, and evaluations were made after artificial inoculation under an anti-aphid net. After inoculation for 20 to 30 days, when the plant was diseased, the states of illness were divided into five grades according to leaf color changes and the extent of shrinkage. The incidence grade for each plant was determined and recorded, and the disease index was calculated using the formula below, which was used as the phenotype of correlation analysis.

$$\text{Disease index (DI)} = \frac{\sum(\text{plant number for each grade} \times \text{corresponding grades})}{\text{total plant number} \times 5} \times 100\%$$

3.3 SSR markers genome scans

Soybean leaf DNA was extracted using the AxyPrep kit (Axygen Scientific Inc, Silicon Valley, USA). The DNA concentrations were 100~200 ng/μL as determined with a NANODROP 2000C instrument. The SSR primer information was obtained from the USDA soybean genetic database (http://www.soybase.org/), and 1015 pairs of primers were prepared by Sangon Biotech Co., Ltd. Ten random soybean materials were used to screen the primers. Four hundred and five pairs of SSR primers distributed evenly on 20 chromosomes with good polymorphisms were selected in this research. The PCR amplification products were detected by 6% polyacrylamide gel electrophoresis and developed by silver staining.

3.4 Data analysis

The software package Popgene32 was for the statistical analysis of allele variation number and the Shannon index. The software PIC6.0 was used to calculate and obtain the PIC (polymorphism information content) value of the SSR markers. The software NTSYS was used to determine the genetic similarity coefficient. The software STRUCTURE 2.3.4 was used to analyze the group structure of the material. The calculation assumed that the sub-group number K was between 2 to 15 and the calculation repetition number was from 10,000 to 100,000, adding 10,000 repetition times for each calculation, with 5 iterations. The LnP (D) value of each assumed population was obtained for each calculation, and the K value was confirmed through ΔK. The software SPAGeDi-1.3d was used to process SSR marker data to obtain a kinship matrix between different materials. The software TASSEL3.0 was used with a Q value as the covariate and a mixed linear model (MLM, Q+K) for regression analysis of SSR marker data and phenotypic data.

3.5 Gene mining around the SSR loci

Based on the whole-genome information for the soybean cultivar Williams82 provided by the SoyBase database (http://www.soybase.org/), the genomic positions of relevant SSR loci were obtained. In this research, the candidate-related genes were searched for in a 3 cM section on both sides of associated SSR loci and the candidate genes were sorted by GO (Gene ontology) enrichment analysis. Based on the GO enrichment results, the genes relevant to the plant-resistance response were preliminarily screened; these genes may be involved in defense reactions against bacteria, fungi and viruses or may be involved in the plant hormone synthetic process related to anti-pathogenic infection and systematic acquired disease resistance. By comparing the gene-encoded protein sequence through the NCBI BlastP online server, the structural domain information was obtained, and a second screening was conducted. Based on the enrichment information, structural domain information and previous research, the candidate genes were defined if they were homologous to known anti-disease genes, exhibited typical structural features of anti-disease genes, or participated in the synthesis process of plant hormones related to anti-phytopathogenic infection.

Authors' contributions

Lai Wei, Zhiyuan Yu, Guixing Zhao, Weiwei Bi, Lijun Liu, Weiwei Wang and Zhiyuan Yu carried out the molecular genetic studies, participated in the sequence alignment and drafted the manuscript. Lai Wei participated in the sequence alignment. Guixing Zhao and Weiwei Bi participated in the design of the study and performed the statistical analysis. Lijun Liu conceived of the study, and participated in its design and coordination and helped to draft the manuscript. All authors read and approved the final manuscript.

Acknowledgements

This research is supported by the National Soybean Industry Technology System (CARS-O4-PSO5), the cultivation of new varieties of resistant genetically modified soybeans (2016ZX08004002), the Agricultural Science and Technology Innovation Leapfrog Project (HNK2019CX01), the Harbin Science and Technology Bureau Project (2017RAQYJ034), and the Postdoctoral Fund of Heilongjiang Province (LBH-Z16184) and Heilongjiang Academy of Agricultural Sciences (2017BZ14).

References

Bai L., Li H.C., Ma Y., Wang D. G., Liu N., and Zhi H.J., 2009, Inheritance and gene mapping of resistance to soybean mosaic virus strain SC-11 in soybean, Soybean Science, 28(1): 1-6

Kumar B., Talukdar A., Verma K., Bala I., Harish G.D., Gowda S., Lal S.K., Sapra R.L., and Singh K.P., 2015, Mapping of yellow mosaic virus (YMV) resistance in soybean (*Glycine max* L. Merr.) through association mapping approach, Genetica, 143(1): 1-10
https://doi.org/10.1007/s10709-014-9801-6
PMid:25475043

Chen A.G, Wang Y., Meng W.L., Li Z.B., Cui X. G., and Wu Y., 2020, Evaluation cluster analysis for glycine soja of different habitats resistant to soybean Mosaic Virus (SMV), Liaoning Agricultural Sciences, 1: 7-13

Chen X.B., Zhang Z.L., Liu D.M., Zhang K., Li A., and Mao L., 2010, SQUAMOSA promoter-binding protein-like transcription factors: star players for plant growth and development, J Journal of Integrative Plant Biology, 52(11): 946-951
https://doi.org/10.1111/j.1744-7909.2010.00987.x
PMid:20977652

Cho E.K., and Goodman R.M., 1979, Strains of soybean mosaic virus: classification based on virulence in resistant soybean cultivars, Phytopathology, 69(5): 467-470
https://doi.org/10.1094/Phyto-69-467

Fristensky B., Riggleman R.C., Wagoner W., and Hadwiger L.A., 1985, Gene expression in susceptible and disease resistant interactions of peas induced with fusarium solani pathogens and chitosan, Physiologial Plant Pathology, 27(1): 15-28
https://doi.org/10.1016/0048-4059(85)90053-0

Fu S., Zhan Y., Zhi H., Gai J., and Yu D., 2006, Mapping of smv resistance gene rsc-7 by ssr markers in soybean., Genetica, 128(1-3): 63-69
https://doi.org/10.1007/s10709-005-5535-9
PMid:17028941

Gao Y, and Shu S. S., 2020, Survey of Soybean Virus Disease Vectors and Insects, Soybean Science and Technology, 1: 49-54

Guo D.Q., Wang Y.W., Zhi H.J., Gai J.Y., and Li K., 2007, Inheritance and gene mapping of resistance to smv strain group sc-13 in soybean, Soybean Science, 1: 21-24

Hamberg M., Sanz A., Rodriguez M.J., Calvo A.P., and Castresana C., 2003, Activation of the fatty acid alpha-dioxygenase pathway during bacterial infection of tobacco leaves. formation of oxylipins protecting against cell death, Journal of Biological Chemistry, 278(51): 51796-51805
https://doi.org/10.1074/jbc.M310514200
PMid:14522973

Hayes A.J., Ma G.R., Buss G.R., and Maroof M.A.S., 2000, Molecular marker mapping of rsv4, a gene conferring resistance to all known strains of soybean, Crop Science. 40(5): 1434-1437
https://doi.org/10.2135/cropsci2000.4051434x

Hill J.H., Bailey T.B., Benner H.I., Tachibana H., Durand D.P., 1987, Soybean Mosaic Virus: effects of primary disease incidence of yield and seed quality, Plant Dis, 71: 237-239
https://doi.org/10.1094/PD-71-0237

Hui J. T, Chen B, and Zhang Y., 2019, Field Screening of Soybean Virus Resistance Varieties, Shaanxi Journal of Agricultural Sciences, 65(9): 7-8+17

Hwang I.S., and Hwang B.K., 2010, The pepper 9-lipoxygenase gene calox1 functions in defense and cell death responses to microbial pathogens, Plant Physiology, 152(2): 948-967
https://doi.org/10.1104/pp.109.147827
PMid:19939946 PMCid:PMC2815858

Jeong S.C., Kristipati S., Hayes A.J., Maughan P. J., and Maroof M.A.S., 2002, Genetic and sequence analysis of markers tightly linked to the soybean mosaic virus resistance gene, rsv3, Crop Science, 42(1): 265-270
https://doi.org/10.2135/cropsci2002.2650
PMid:11756284

Kachroo AI., Fu D.Q., Havens W., Navanrre D., and Kachroo P., 2008, Anoleic acid-mediatedpathway induces constitutive defense signaling and enhanced resistance to multiple pathogens in soybean, Molplant Microbe Interact, 5.21(5): 564-575
https://doi.org/10.1094/MPMI-21-5-0564
PMid:18393616

Kumar B., Talukdar A., Verma K., Girmilla V., and Sapra R.L., 2014, Screening of soybean [glycine max (l.) merr.] genotypes for yellow mosaic virus (ymv) disease resistance and their molecular characterization using rga and ssrs markers, Australian Journal of Crop ence, 8(1): 27-34

Li K., Xia Y.C., Wang D.G., Yang Y.Q., Ren R., Gao L., Zhang K., and Zhi H.J., 2014, Analysis of dynamic change of soybean Mosaic Virus Strains in Heilongjiang Province of China, Soybean Science, 33(6): 880-884

Li M.J., and Shi Z.L., 2010, Sequence Analysis of Plant Acyl Carrier Protein., Acta Agriculturae Boreall-Sinica, 25(S1): 1-6

Li W.F., Zhu X.S., Wang X.F., Wang K., Liu C.Y., Chen Q.S., and Hu G.H., 2010, Identification of the smv adult-plant and seed coat mottling resistance in soybean germplasms using ssr markers, Journal of Plant Genetic Resources, 11(2): 239-243

Liu L.J., 2007, High quality soybean in Northeast China, Heilongjiang Science and Technology Press, Harbin, China

Luan X.Y., Li Z.F., Man W.Q., Bai Y.N., Ma Y.S., Liu X.L., Zheng B.X., and Du W.G., 2006, Identification of Molecular Markers Linked to Resistance for SMV3 in Soybean, Molecular Plant Breeding, (06),841-845.

Meng X.Z., Wang H.C., He Y.X., Liu Y.D., Walker J. C., Torii K.U., and Zhang S.Q., 2012, A mapk cascade downstream of erecta receptor-like protein kinase regulates arabidopsis inflorescence architecture by promoting localized cell proliferation, Plant Cell, 24(12): 4948-4960
https://doi.org/10.1105/tpc.112.104695
PMid:23263767 PMCid:PMC3556968

Park S.K., and Twell D., 2001, Novel patterns of ectopic cell plate growth and lipid body distribution in the arabidopsis gemini pollen1 mutant, Plant Physiology, 126(2): 899-909
https://doi.org/10.1104/pp.126.2.899
PMid:11402217 PMCid:PMC111179

Riggleman R.C., Fristensky B., Hadwiger L.A., 1985, The disease resistance response in pea is associated with increased levels of specific mrnas, Plant Molecular Biology, 4(2-3): 81-86
https://doi.org/10.1007/BF02418753
PMid:24310743

Shi A.N., Chen P.Y., Li D.X., Zheng C.M., Hou A.F., and Zhang B., 2008, Genetic confirmation of 2 independent genes for resistance to soybean mosaic virus in j05 soybean using ssr markers, Journal of Heredity, 99(6): 598-603
https://doi.org/10.1093/jhered/esn035
PMid:18492653

Steinlage T.A., Hill J.H., and Nutter Jr F.W., 2002, Temporal and spatial spread of soybean mosaic virus (smv) in soybeans transformed with the coat protein gene of smv, Phytopathology, 92(5): 478
https://doi.org/10.1094/PHYTO.2002.92.5.478
PMid:18943021

Sun Z. Q., Liu Y. Z., Sun D. M., and Liao L., 1990, Inheritance of resistance to SMV strain 1.2.3 in soybeans, China Oils, 2: 20-24

Teng W.L., Li W.B., Han Y.P., Zhao G.Y., Qiu L.J., and Chang R.Z., 2005, SSR markers of disease resistance genes in soybean SMV3, Collection of abstracts from the National Symposium on Crop Biotechnology and Mutagenesis Technology, 79, Wuyi Mountain, Fujian, China

Teng W.L., Li W.B., Qiu L.J., Han Y.P., Zhao G.Y., Guan R.X., and Chang R.Z, 2006, Identification of SSR marker linked to the resistance gene of SMV3 in soybean, Soybean Science, 3: 244-249

Teng W.L., Lu S.Y., Gao Y., Sun M.M., Han Y.P., Li W.B., Zhu D., Cheng Z., Hu H.B., and Qiu L.J., 2011, Identification of smv1 and smv3 resistance in soybean cultivars(lines) from different provinces, Journal of Northeast Agricultural University, 42(10): 16-20

Vicente J., Cascón T., Vicedo B., García-Agustín P., Hamberg M., and Castresana C., 2012, Role of 9-lipoxygenase and α-dioxygenase oxylipin pathways as modulators of local and systemic defense, Molecular Plant, 5(4): 914-928
https://doi.org/10.1093/mp/ssr105
PMid:22199234

Wang D., Ma Y., Liu N., Yang Z., Zheng G., Zhi H., 2011, Fine mapping and identification of the soybean rsc4 resistance candidate gene to soybean mosaic virus, Plant Breeding, 130(6): 653-659
https://doi.org/10.1111/j.1439-0523.2011.01888.x

Wang R.Y., and Ghabrial S.A., 2002, Effect of aphid behavior on efficiency of transmission of soybean mosaic virus by the soybean-colonizing aphid, aphis glycines, Plant Disease, 86(11): 1260-1264
https://doi.org/10.1094/PDIS.2002.86.11.1260
PMid:30818478

Wang Y.J., Yang D.F., Wang X.Q., Yang Y. L., Yu D.Y., Gai J.Y., Wu X.L., He C.Y., Zhang J.S., and Chen S.Y., 2004, Mapping of five genes resistant to smv strains in soybean, Acta Genetica Sinica, 31(1): 87-90

Yan H.L., Wang H., Cheng H., Hu Z.B., Chu S.S., and Zhang G.Z., 2015, Detection and fine-mapping of sc7 resistance genes via linkage and association analysis in soybean, Journal of Integrative Plant Biology, 57(8): 722-729
https://doi.org/10.1111/jipb.12323

Yi S.Y., Kim J.H., Joung Y.H., Lee S., Kim W.T., Yu S.H., and Choi D., 2004, The pepper transcription factor capf1 confers pathogen and freezing tolerance in Arabidopsis, Plant Physiology,136(1): 2862-2874
https://doi.org/10.1104/pp.104.042903
PMid:15347795 PMCid:PMC523348

Yu Y.G., Maroof M.A.S., and Buss G.R., 1996, Divergence and allelomorphic relationship of a soybean virus resistance gene based on tightly linked dna microsatellite and rflp markers, Theoretical & Applied Genetics, 92(1): 64-69
https://doi.org/10.1007/s001220050093
PMid:24166117

Yue P., Arelli P.R., and Sleper D.A., 2001, Molecular characterization of resistance to heterodera glycines in soybean pi 438489b, Theoretical & Applied Genetics, 102(6-7): 921-928
https://doi.org/10.1007/s001220000453

Zheng C.M., Chang R.Z., and Qiu L.J., 2000a, Progress on the disease of soybean mosaic virus, Acta Phytopathologica Sinica, 30(2): 97-105

Zheng C.M., Chang R.Z., Qiu L.J., Wu Z.P., and Gao F.L., 2000b, Identification the resistance of soybean germplasm to smv3, Soybean Science, 19(4): 299-306

Zhi H.J., Gai J.Y., and He X.H., 2005, Study on methods of classification quantitative resistance to soybean mosaic virus in soybean, Soybean Science, 24(1): 5-11

Zuo K.J., Qin J., Zhao J.Y., Ling H., Zhang L.D., Cao Y.F., and Tang K.X., 2007, Over-expression gberf2 transcription factor in tobacco enhances brown spots disease resistance by activating expression of downstream genes, Gene, 391(1-2): 80-90
https://doi.org/10.1016/j.gene.2006.12.019
PMid:17321073