

# Association mapping of seed vigor in spring soybean (*Glycine max* (L.) Merr.) in northeast China

Caijin Wang

Heilongjiang University

Huiyan Zhao

Heilongjiang University

Yang Wang (✉ [2004084@hlju.edu.cn](mailto:2004084@hlju.edu.cn))

Heilongjiang University

---

## Research Article

**Keywords:** Soybean, association analysis, simple sequence repeat, linkage disequilibrium, elite allele, seed vigor

**Posted Date:** March 15th, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-288706/v1>

**License:**   This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

---

# Abstract

High-vigor soybean seeds can result in the rapid and uniform emergence of seeds under a wide range of field conditions. We aimed to explore the quantitative trait loci of seed vigor traits and superior allelic and carrier materials from the soybean in northeast China. A total of 257 improved lines and 104 soybean landraces were used to evaluate seed vigor, using normal seedling fresh weight, normal seedling length, main root length, hypocotyl length, and germination percentage, in 2018 and 2019. In this study, the different degrees of LD was detected not only among syntenic markers but also among nonsyntenic ones, suggesting that there had been historical recombination among linkage groups. A total of 19 and 10 simple sequence repeat (SSR) markers associated with seed vigor were detected in improved lines and soybean landraces in 2018 and 2019, respectively. Our results also showed five SSR markers co-associated with two or three vigor traits. In addition, 41 superior alleles and their carrier materials were mined. Based on the elite alleles detected, the best cross combinations for improving seed vigor traits were proposed. The results demonstrate that association mapping is a valuable foundation for molecular breeding for soybean seed vigor.

## Introduction

Soybean (*Glycine Max* (L.) Merr.) is an important oil crop that provides 69 % of plant protein and 30 % of plant oil worldwide<sup>1,2</sup>. Good production levels are essential, and high seed vigor varieties can increase the crop yield of soybeans by 20–30 %<sup>3,4</sup>. Seed vigor reflects in seed germination potential, field emergence rate, and seed storage tolerance under a range of conditions<sup>5</sup>. Seeds with high vigor show increase in seed reserve capacity during germination, which provides a better guarantee for seedling development<sup>6</sup>. Furthermore, seeds with high vigor have a strong competitive advantage over weeds; seeds with low vigor have a low germination rate, weak resistance, and low yield<sup>7,8</sup>. Therefore, clarifying the genetic structure of soybeans and exploring elite alleles for the seed germination stage will promote genetic improvement of soybean seed vigor.

Seed vigor is a complex trait that is mainly manifested by factors such as seed germination rate, germination percentage (GP), seedling length (SL), root length (RL), seedling fresh weight (FW), and seed longevity<sup>5</sup>. The genetic mechanisms of seed vigor have been examined in crops such as rice<sup>9,10</sup>, maize<sup>11,12</sup>, and *Arabidopsis*<sup>13</sup>. Studies have shown that seed vigor is controlled by polygens, and many quantitative trait loci (QTL) for seed vigor have been identified<sup>12,14</sup>. The ability of seeds to resist aging is often used to measure seed vigor. Very little is known about the genetic control of seed vigor in soybeans. Zhang et al<sup>15</sup> used two recombinant inbred line (RIL) populations to evaluate three related seed germination traits by aging seed treatment to uncover the genetic mechanisms of soybeans. Thirty four QTL were identified on 11 chromosomes; Twenty one of these QTL were clustered in five QTL-rich regions on Chromosome 3 (Chr3), Chr5, Chr17, and Chr18. Dargahi et al<sup>16</sup> constructed the F2:3 population developed from a cross between poor-longevity soybeans and a landrace cultivar with higher levels of longevity. Three QTL and 13 simple sequence repeat (SSR) loci associated with the relative germination rate and storability of seeds were detected. Singh et al<sup>17</sup> identified QTL for seed longevity in soybeans, using aged seeds of 153 F2:3 lines collected from replicated trails. There were four SSR markers associated with seed longevity, which explained the phenotypic variation that ranged from 6.3 % (Satt285) to 7.5 % (Satt434). All QTL were examined using family based linkage mapping as the main method to study the genotypes of the quantitative traits of plants. However, the results of this method often have limitations. Allelic

variation could not be detected if there was no separation or recombination at some sites between the two parents<sup>18</sup>.

Association mapping identifies QTL based on linkage disequilibrium (LD) and examines the marker-trait associations in natural populations. Association mapping is an effective method to mine elite alleles because of a short research time and high mapping resolutions; it also facilitates allele discovery<sup>19,20</sup>. QTL detected by association mapping in soybeans include some important quantitative traits such as yield<sup>21,22</sup>, plant type<sup>23–25</sup>, content of oil<sup>26</sup>, and resistance<sup>27,28</sup>.

However, there is no study that examines the genetic mechanisms for seed vigor in the soybean, using association mapping. In this study, 361 soybean accessions were used to conduct association mapping for seed vigor by combining information from 175 SSR markers. FW, SL, RL, HL, and GP were measured in 2018 and 2019, to indicate seed vigor. The purpose of this study was (1) to investigate phenotypic and genetic diversities; (2) to detect the extent of LD in improved lines populations and landraces soybean populations; (3) to detect SSR loci controlling seed vigor and to mine elite alleles and corresponding frontal vector material; and (4) to propose a cross combination to improve seed vigor traits.

## Materials And Methods

### Plant materials and field planting

Of the 361 soybean accessions, 257 belonged to improved lines and 104 belonged to landrace accessions. They were all from four provinces in the northeast of China (Tables S1 and S2). The field experiments were carried out at the Hulan Experimental Station at Heilongjiang University (Harbin), Heilongjiang Province, China. Seed was sown on May 10, in 2018 and in 2019. Each variety was planted in two rows, 0.5 m apart, with 10 plants in a row in a random design. The harvested seeds of the 361 accessions were dried to a seed moisture of approximately 13 % and then used to evaluate seed vigor.

### Investigation Of Traits

Fifty soybean seeds of each genotype were tested in our study. We used the methods outlined in Wang et al<sup>29</sup> to treat seeds before germination and to manage the conditions during germination. Seeds were cultivated in a 13×19×12 cm germination box in a germination room for 11 days, using the between paper (BP) method. This process was repeated twice. Five vigor-related traits were measured in the five seedlings that grew the fastest in each genotype, as follows: GP, normal SL, main RL, HL, and normal seedling FW. The average measurements of the five seedlings were used in the study. GP was calculated as the percentage of seeds that germinated after 3 days ( $GP (\%) = 3 \text{ days of germinated seeds} / \text{total number of seeds} \times 100$ ), according to the standard of seed germination identified by Wang et al<sup>30</sup>.

### DNA Extraction And SSR Marker

Genomic DNA was extracted from the primary leaf of soybean seedling, using the cetyl trimethylammonium bromide (CTAB) method<sup>31</sup>. A total of 175 SSR markers were synthesized by Shanghai Genaray Biotech Co. Ltd., Shanghai, China, following the sequences published on the SoyBase website (<https://www.soybase.org/>). The polymerase chain reaction (PCR) was a 10 µL mixture, consisting of 1.5 µL of DNA, 1.5 µL of forward and

reverse primers, 5.7  $\mu$ L of ddH<sub>2</sub>O, 0.2  $\mu$ L of dNTP, 1.0  $\mu$ L of PCR buffer, and 0.1  $\mu$ L of Taq DNA polymerase. The following steps were performed in the PCR process: pre-denaturation at 94°C for 5 min and denaturation at 94°C for 30 s. The cycle was repeated 35 times, followed by annealing at 42–61°C for 30 s, extension at 72°C for 30 s, and a final extension at 72°C for 5 min. The PCR products were separated in 8 % polyacrylamide gels, and silver staining was used to enable visualization<sup>32</sup>.

## Data Analysis

### Descriptive statistics

Basic statistics were calculated using SPSS 22.0 software (<http://www.spss.com>). Heritability in the broad sense ( $H_B^2$ ) was determined using the formula:  $H_B^2 = Vg/(Vg + Ve/n) \times 100$  %, where  $Vg$  is genetic variance,  $Ve$  is error variance, and  $n$  is number of replicates.

### Genetic Diversity And Population Structure

Genetic diversity was analyzed using POWERMARKER 3.25 software<sup>33</sup>. Population structure analysis was performed using STRUCTURE V 2.3.4 software<sup>34</sup>. In the STRUCTURE setting, a burn-in period of 10,000 and a Markov Chain Monte Carlo (MCMC) run length of 100,000 was set. We implemented six independent runs for each subpopulation (K), with the setting varying from 2 to 10. Evanno's method was used to determine the best Delta K estimate<sup>35</sup>. Relative kinship was estimated using SPAGeDi software<sup>36</sup>.

### Linkage Disequilibrium And Association Analysis

LD was calculated using TASSEL V 2.1 software<sup>37</sup>. Pairwise estimates of standardized disequilibrium coefficients ( $D'$ ) were used to measure the level of LD between pairs of SSR markers<sup>38,39</sup>. The LD decay graph was plotted with genetic distance (cM) versus  $D'$  for intra-chromosomal combinations, and the background of LD decay was measured as  $D'$  dropped to half its maximum value<sup>22</sup>. The general linear model (GLM) method was used to test the associations between the phenotypes and markers. Markers associated with seed vigor traits at a significance level of  $P < 0.01$  were selected. The "null allele" was used to estimate the phenotypic allele effect<sup>40,41</sup>.

## Results

### Phenotypic evaluation and correlation

In both groups, the absolute values of kurtosis and skewness were less than 1 (Table 1), and the five traits conformed to a normal distribution (Figures S1 and S2). The HB2 of the five traits, averaged across 2 years, ranged from 62.19 % to 93.48 %. The coefficient of genetic variation of the five traits ranged from 13.21 % to 43.16 % (Table 1). In general, four of the vigor traits (excluding HL) under the improved lines had higher values than those under soybean landraces.

Table 1

Analysis on the variation of vigor traits of 257 improved lines and 104 landraces soybean accessions

Traits	Years	257 improved lines accessions					104 landrace accessions				
		Mean $\pm$ SD	K	S	CV (%)	$H_B^2$ (%)	Mean $\pm$ SD	K	S	CV (%)	$H_B^2$ (%)
FW (g)	2018	2.06 $\pm$ 0.37	0.12	0.12	18.08	82.04	1.82 $\pm$ 0.33	2.05	0.41	18.19	93.48
	2019	2.04 $\pm$ 0.37	0.10	0.17	18.05		1.80 $\pm$ 0.32	2.25	0.62	17.92	
SL (cm)	2018	8.88 $\pm$ 1.76	0.01	-0.15	19.76	90.81	8.61 $\pm$ 1.63	0.18	-0.27	18.95	87.35
	2019	8.54 $\pm$ 1.56	0.47	-0.26	18.32		8.41 $\pm$ 1.49	0.57	-0.27	17.67	
RL (cm)	2018	19.20 $\pm$ 4.06	-0.30	0.24	21.17	90.71	17.89 $\pm$ 3.82	0.53	0.53	21.37	93.05
	2019	18.01 $\pm$ 4.07	-0.07	0.24	22.61		17.50 $\pm$ 3.78	1.33	0.58	21.60	
HL (cm)	2018	9.13 $\pm$ 1.21	-0.31	-0.11	13.21	84.17	9.22 $\pm$ 1.70	-0.55	-0.18	18.43	87.88
	2019	9.23 $\pm$ 1.36	0.16	0.02	14.71		9.42 $\pm$ 1.82	-0.65	-0.10	19.32	
GP (%)	2018	61.24 $\pm$ 21.76	-0.57	-0.39	35.54	74.40	59.71 $\pm$ 21.18	-0.22	-0.54	35.47	62.19
	2019	58.38 $\pm$ 16.83	-0.38	-0.27	28.83		42.95 $\pm$ 18.53	-0.44	0.23	43.16	

Abbreviations: FW, seedling fresh weight; SL, seedling length; RL, main root length; HL, hypocotyl length; GP, germination percentage; SD, standard deviation; K, Kurtosis; S, Skewness; CV, coefficient of variation;  $H_B^2$ , heritability in the broad sense

The Pearson correlation of the five traits showed there was a significant positive and negative correlation between the five traits of improved lines and soybean landraces (Table 2). There was no significant correlation between HL and FW, SL, whereas HL and RL were significantly negatively correlated (Table 2).

Table 2  
Pearson correlation of five traits of improved lines and landraces soybean accessions

Populations	Traits	FW	SL	RL	HL
Improved lines	SL	0.378**			
	RL	0.373**	0.373**		
	HL	-0.101	-0.023	-0.155*	
	GP	0.436**	0.258**	0.446**	-0.131*
Landraces	SL	0.336**			
	RL	0.385**	0.220*		
	HL	-0.057	0.109	-0.218*	
	GP	0.340**	0.462**	0.306**	0.079

Note: \* and \*\* denote significant differences at the 0.05 and 0.01 probability levels, respectively

## Molecular Markers And Genetic Structure

Of the 175 SSR markers that were polymorphic, a total of 844 alleles were detected. The number of alleles ranged from 2 to 12, with a mean of 4.82 alleles per marker. Genetic diversity ranged from 0.01 to 0.85, with a mean of 0.50. Polymorphic information content (PIC) ranged from 0.01 to 0.85, with a mean of 0.45 (Table S3). In comparison, higher values were observed for the numbers of alleles, genetic diversity, and PIC in soybean landraces than those in improved lines (Table S4).

STRUCTURE software analysis showed that the 175 markers were relatively independent. The log-likelihood function [ $\ln P(D)$ ] has no obvious inflection point (Fig. 1A). Therefore, the suitable value of K was determined using Delta K; here, three subpopulations were detected (Fig. 1B). For improved lines, the Pop<sub>1</sub>, Pop<sub>2</sub>, and Pop<sub>3</sub> subpopulations contained 62, 81, and 114 accessions (Fig. 2A), respectively. For soybean landraces, the Pop<sub>1</sub>, Pop<sub>2</sub>, and Pop<sub>3</sub> subpopulations contained 41, 38, and 25 accessions (Fig. 2B), respectively.

## Linkage Disequilibrium

LD of allelic variation between SSR loci is the basis of association analysis. Considering that group mixing leads to the enhancement of LD, an LD analysis of both the improved lines and the soybean landraces was carried out. For the improved lines, a total of 14,992 pairs were detected, which included both inter-and intra-chromosomal combinations (Fig. 3A). The SSR pairs of loci with LD supported by  $P < 0.01$  occupied 19.66 % of the total population (Fig. 3A). The  $D'$  of linked marker pairs, supported by the  $P < 0.01$  mean, was 0.38 (Table 3). For the soybean landraces, a total of 14,834 pairs were detected, which included both inter-and intra-chromosomal combinations (Fig. 3B). The SSR pairs of loci with LD, supported by  $P < 0.01$ , occupied 6.38 % of the total population (Fig. 3A). The  $D'$  of linked marker pairs, supported by the  $P < 0.01$  mean, was 0.44 (Table 3). The study revealed that the improved lines had more LD locus pairs than soybean landraces, whereas soybean

landraces had higher D' values than improved lines. This suggests that the history of soybean landraces has undergone more reorganization, retaining higher LD.

Table 3

Comparison of D' of LD for pairwise SSR loci between improved lines and landraces soybean accessions

Populations	No. of LD <sup>a</sup> locus pairs	Frequency distribution of D' <sup>b</sup> (P < 0.01)					Mean of D' <sup>b</sup>
		0-0.2	0.2-0.4	0.4-0.6	0.6-0.8	0.8-1	
Improved lines	208(29.71 %)	15	127	38	21	7	0.38
Landraces soybean	83(11.93 %)	1	34	36	9	3	0.44

Note: <sup>a</sup> LD means linkage disequilibrium; <sup>b</sup> D' means standardized disequilibrium coefficients

Regression analysis of the D' value and the genetic distance shows that the attenuation of the D' value follows the  $Y=b \ln(x) + c$  derivative. The D' between SSR loci decreased with increasing genetic distance for improved lines and soybean landraces (Figs. 4A and 4B). The minimum (0.5) distance of LD decay for improved lines and landraces was 2.53 cM and 1.46 cM, respectively. It can thus be seen that the LD decay rate of soybean landraces was slow, whereas that of improved lines was fast.

### Association Mapping

Using the GLM program of marker-trait association, and based on the threshold ( $p < 0.01$ ), 19 and 10 SSR loci were revealed in the two populations in 2018 and 2019, respectively (Table S5). Two years were co-detected. The SSR markers for improved lines were seven for FW, two for SL, two for RL, one for HL, and two for GP, and the SSR markers for soybean landraces were three for FW, two for SL, two for RL, one for HL and two GP (Table 3).

Association mapping identified seven markers for FW in the improved lines. The range of phenotypic variation explained (PVE) was from 5.95 % to 12.08 %, of which Sat\_256\_Chr7 accounted for maximum phenotypic variations: 12.08 % in 2018 and 11.32 % in 2019 (Table 4). There were two markers associated with mapping for SL; the range of PVE was from 6.35 % to 13.35 %, of which Satt441\_Chr7 accounted for the maximum phenotypic variations, viz. 13.35 % in 2018 and 2019 (Table 4). There were two markers of association mapping for RL. The range of PVE was from 7.09 % to 12.11 %, of which Satt329\_Chr8 accounted for maximum phenotypic variations: 12.11 % in 2018 and 11.32 % in 2019 (Table 4). There was one marker of association mapping with HL, located on nine chromosomes, which accounted for phenotypic variations of 8.58 % in 2018 and 10.22 % in 2019 (Table 4). There were two markers of association mapping for GP. The range of PVE was from 6.98 % to 10.79 %, of which Satt303\_Chr18 accounted for maximum phenotypic variations: 10.79 % in 2018 and 6.98 % in 2019 (Table 4). Three markers were co-associated with two or three vigor traits: Aw277661, Satt441, and Satt606 (Table 4).

Table 4  
Simple sequence repeat (SSR) marker significantly associated with five traits (P < 0.01)

Traits	SSR marker	Chr.	Position (cM)	Start position (bp)	End position (bp)	2018		2019	
						P value	PVE(%)	P value	PVE(%)
Improved lines									
FW	Sat_266	3	47.27	36046708	36046763	0.0085	5.95	0.0031	6.57
	Aw277661	3	74.49	32803245	32803313	0.0004	6.25	0.0005	6.37
	Sat_256	7	74.52	17606889	17606932	0.0001	12.08	0.0008	11.32
	Satt441	9	46.2	11271676	11271771	0.0001	8.59	0.0001	8.17
	Satt577	14	6.05	668802	668837	0.0001	9.07	0.0001	9.41
	Satt606	15	39.77	14918728	14918821	0.0001	8.34	0.0001	9.23
	Sat_228	16	23.91	3049971	3050050	0.0019	7.96	0.0011	8.32
SL	Aw277661	4	74.49	32803245	32803313	0.0003	7.44	0.0016	6.35
	Satt441	9	46.2	11271676	11271771	0.0001	13.35	0.0001	13.35
RL	Satt329	8	110.94	21161155	21161226	0.0001	12.11	0.0001	11.32
	Satt606	15	39.77	14918728	14918821	0.0005	7.09	0.0001	11.74
HL	Satt588	9	117.01	45420963	45421065	0.0059	8.58	0.0011	10.22
GP	Satt606	15	39.77	14918728	14918821	0.0001	10.79	0.0005	6.98
	Satt413	17	113.61	40043052	40043061	0.0029	8.34	0.0031	7.84
Landraces soybean									
FW	Sat_378	8	116.62	23605177	23605222	0.0001	27.21	0.0003	23.11
	Satt262	10	57.02	35510863	35510922	0.0004	18.38	0.0005	18.44
	Satt509	11	32.5	6206957	6207049	0.0072	14.18	0.0040	14.96
SL	Satt291	6	45.75	7320677	7320718	0.0061	12.74	0.0037	13.31
	Sat_378	8	116.62	23605177	23605222	0.0059	16.19	0.0093	16.47
RL	Sat_337	4	32.09	4172809	4172860	0.0070	16.27	0.0031	18.2
	Satt571	20	18.5	1291809	1291850	0.0013	12.37	0.0011	12.74
HL	Satt304	14	65.55	13284499	13284588	0.0013	15.36	0.0049	12.6
GP	Sat_378	8	116.62	23605177	23605222	0.0009	21.08	0.0001	28.13

Abbreviations: FW, seedling fresh weight; SL, seedling length; RL, main root length; HL, hypocotyl length; GP, germination percentage; Chr., chromosome; PVE, phenotypic variations explained

Traits	SSR marker	Chr.	Position (cM)	Start position (bp)	End position (bp)	2018		2019	
						P value	PVE(%)	P value	PVE(%)
	Satt509	11	32.5	6206957	6207049	0.0022	15.07	0.0021	13.8

Abbreviations: FW, seedling fresh weight; SL, seedling length; RL, main root length; HL, hypocotyl length; GP, germination percentage; Chr., chromosome; PVE, phenotypic variations explained

In the soybean landraces, there were three markers associated with mapping for FW; the range of PVE was from 14.18 % to 27.21 %, of which Sat\_378\_Chr8 accounted for maximum phenotypic variations of 27.21 % in 2018 and 23.11 % in 2019 (Table 4). There were two markers of association mapping for SL. The range of PVE was from 12.74 % to 16.47 %, and Sat\_378\_Chr8 accounted for maximum phenotypic variations: 16.19 % in 2018 and 16.47 % in 2019 (Table 4). There were two markers of association mapping for RL. The range of PVE was from 12.37 % to 18.2 %, and Sat\_337\_Chr4 accounted maximum phenotypic variations of 16.27 % in 2018 and 18.2 % in 2019 (Table 4). There was one marker of association mapping for HL, located on 14 chromosomes, which accounted for phenotypic variations: 15.36 % in 2018 and 12.6 % in 2019 (Table 4). There were two markers of association mapping for GP. The range of PVE was from 13.8 % to 28.13 %, and Sat\_378\_Chr8 accounted for maximum phenotypic variations: 21.08 % in 2018 and 28.13 % in 2019 (Table 4). Two markers were co-associated with two or three vigor traits, namely Sat\_378 and Satt509 (Table 4).

### Mining Of Elite Alleles And Carrier Materials

The alleles and carrier materials for the loci associated with the five vigor traits were identified (Table S6). In the improved lines, Sat\_256-236bp had the greatest positive effect (+ 0.32 g) for FW. The carrier material was Hefeng 48 (Table S6). Satt441-281bp had the greatest positive effect (+ 1.21 cm) for SL, and the carrier material was Heilong 66 (Table S6). Satt329-262bp had the greatest positive effect (+ 3.58 cm) for RL, and the carrier material was Heinong 44 (Table S6). Satt588-128bp had the greatest positive effect (+ 0.99 cm) for HL, and the carrier material was Jiunong 66 (Table S6). Satt413-196bp had the greatest positive effect (+ 12.94 %) for GP, and the carrier material was Heinong 44 (Table S6). Some of the phenotypic effect values of locus-allelic viz., Aw277661-266bp, Satt588-185bp, and Satt441-294bp had positive effects, and the elite allele frequencies were 78.21 %, 74.30 %, and 73.05 %, respectively. This may be because of strong artificial selection.

In the soybean landraces, Sat\_378-168bp had the greatest positive effect (+ 0.18 g) for FW, and the carrier material was Keshandajinhuang (Table S6). Sat\_378-156bp had the greatest positive effect (+ 0.62 cm) for SL, and the carrier material was (Table S6). Sat\_337-281bp had the greatest positive effect (+ 3.48 cm) for RL, and the carrier material was Longyoutai (Table S6). Satt304-169bp had the greatest positive effect (+ 0.41 cm) for HL, and the carrier material was Keshandajinhuang (Table S6). Sat\_378-156bp had the greatest positive effect (+ 7.96 %) for GP, and the carrier material was Fangzhengbailudou (Table S6). Some of the phenotypic effect values of locus-allelic viz., Satt571-135bp, Satt509-189bp, Satt262-256bp, and Satt291-223bp had positive effects, and the elite allele frequencies were 94.23 %, 81.55 %, 76.92 %, and 70.19 %. This may be because of the long-term natural selection.

### Design For The Positive Elite Crosses Combinations

The two best cross combinations for improving seed vigor traits were identified (Table 5), based on the elite alleles detected. The positive effect of elite alleles on five seed vigor traits, namely Suinong 26, Hefeng 48, Heilong 44, Dalihuang and Keshandajinhuang emerged repeatedly (Table S6), suggesting that these materials carry superior fragments and are more likely to improve the corresponding characteristics in practical breeding if multiple vigor traits are improved simultaneously, and all superior genes are polymerized in one variety to the greatest extent.

Table 5  
The positive elite crosses predicted by association mapping with five traits

Traits	Parental combinations	No. of elite alleles predicted
Fresh weight	Suinong 26 × Dalijuang	10
	Suinong 26 × Keshandajinhuang	8
Seedling length	Heimong 66 × Tiejiasilihuang	3
	Heimong 66 × Ninganxiaoheidou	3
Root length	Heinong 44 × Helongyoutai	4
	Heinong 37 × Helongyoutai	3
Hypocotyl length	Jiunong 1 × Keshandajinhuang	2
	Dengke 1 × Keshandajinhuang	2
Germination percentage	Heinong 44 × Fangzhengbailudou	2
	Hefeng 24 × Xiaoheiqi	2
<b>Association mapping of seed vigor in spring soybean (<i>Glycine max</i> (L.) Merr.) in northeast China</b>		

## Discussion

The vigor traits of soybean seeds during the germination period are the basis of seedling establishment. This study explored the elite alleles of the correlation loci for traits associated with seed vigor in the soybean. Seed vigor traits were identified for the spring soybean in northeast China in 2018 and 2019. Our data showed that the vigor traits of the two groups were normally distributed and that the genetic variation was abundant.

Analysis of genetic diversity of germplasm resources is an important basis for exploring favorable genes<sup>42-44</sup>. In this study, 175 SSR markers were polymorphic, with a mean of 4.82 alleles per marker, indicating that this material is more genetically diverse. In addition, higher values were observed for the alleles, gene diversity, and PIC value for soybean landraces than for soybeans in improved lines. This may result from a reduction in the genetic diversity of the improved lines because of crossbreeding and the long-term selection process.

The relative kinship estimates based on the 175 SSR markers indicated that more than 74.67 % and 74.48 % of the pairwise kinship estimates in improved lines and soybean landraces accessions were within the range of 0 to 0.05, whereas 17.43 % and 18.93 % of the improved lines and soybean landraces accessions kinship

estimates ranged from 0.05 to 0.15. The remaining pairs of accessions with the higher estimated categories continued to decrease in number, revealing that the pairwise accessions had a null or weak relationship (Fig. 5). Association analysis required non-structural groups and no direct relatives between the materials. The materials in this test met the basic requirements for association analysis.

Association analysis with the natural population as the object of study is more likely to cause false positives than traditional linkage analysis, as the mixing of subsets can increase the LD in a population. Understanding the population structure is important to avoid spurious associations in association mapping<sup>34,45</sup>. There have been many reports on the genetic structure of soybean<sup>26,46</sup>. In this study, based on the Bayesian clustering method, the population of improved lines and landrace accessions was clustered into three main subpopulations. It is noteworthy that the germplasm resources of Heilongjiang Province of China were distributed in three subpopulations, indicating some exchange of genes with those from the other three provinces.

Since LD between QTL and marker sites is the premise and foundation of association analysis, it is necessary to establish the LD status of the genomic region of the population in the study before association analysis. In this study, the value of  $D'$  was higher in the landraces than in the improved soybean lines, suggesting that there had been historical recombination among linkage groups in the former population. Artificial selection further leads to different LD levels among soybean gene groups. Some studies suggest that the LD decay distance of soybean is more than 50–420 kb<sup>2,47,48</sup>, whereas other studies found a LD decay distance of more than 10–12 cM<sup>26,49</sup>. This indicates the inconsistency of LD decay levels in improved lines and soybean landraces, and highlights the need to determine the range of LD variation in improved lines and soybean landraces. In this study, the improved soybean lines had more LD locus pairs than soybean landraces, and the background of LD decay was from 1.46 cM for landraces to 2.53 cM for improved lines. Therefore, when performing association mapping, different genetic populations will require different marker densities, and to some extent, lower LD decay rates may limit the resolution of association analysis.

A total of ten and seven SSR markers associated with the vigor traits were screened out from cultivated and landraces of soybean in 2018 and 2019, respectively. We also compared the SSR markers associated with seed vigor with findings from previous studies. Our results were different from those reported by Zhang et al<sup>15</sup>, Singh et al<sup>16</sup> and Dargahi et al<sup>17</sup>. These differences in the results obtained could be attributed to the use of different materials and markers in these studies. Moreover, the improved lines and landraces of soybean identified loci were not consistent. The elite alleles that led to the control of the same trait may be located at different loci. This may result from the hitchhiking effect of the selection pressure of the two groups, resulting in great differences in genetic structure and traits between the two groups<sup>50</sup>.

In the improved soybean lines, two SSR markers were co-associated with FW and SL, namely on the alleles Aw277661-246bp, Aw277661-266bp, Aw277661-268bp, Satt441-281bp, Satt441-287bp, and Satt441-294bp. One SSR marker was co-associated with FW, SL, and GP, in which the direction of the phenotypic effects of alleles Satt606-279bp, Satt606-301bp, and Satt606-329bp was consistent. In the soybean landraces, two SSR markers were co-associated with FW and GP, on the alleles Sat\_378-156bp, Sat\_378-168bp, Sat\_378-173bp, Sat\_378-178bp, Sat\_378-301bp, Satt509-181bp, Satt509-189bp, and Satt509-197bp. The direction of phenotypic effects was highly consistent. In addition, the results of the study revealed five loci associated with two or more

traits simultaneously, and these multi-effect loci may be the genetic basis for their phenotypic correlation, which can be applied to the synergistic improvement of multiple traits<sup>41</sup>.

Through sexual hybridization, more favorable elite alleles can be aggregated to improve the seed vigor traits of the variety. There were 16 positive alleles for the FW trait, for which the phenotypic effects varied from 0.04 g to 0.32 g, and these positive alleles were carried by six varieties. There were three elite alleles for Hefeng 48 and Dalihuang, and seven elite alleles for Suinong 26. There were seven positive alleles for the SL trait, with phenotypic effects varying from 0.02 cm to 1.21 cm, and these positive alleles were carried by six varieties. There were two elite alleles for Heinong 46. There were six positive alleles for the RL trait, with phenotypic effects varying from 0.35 cm to 3.58 cm, and these positive alleles were carried by four varieties. For Heinong 44 and Helongyoutai, there were two elite alleles. There were four positive alleles for the HL trait, for which the phenotypic effects varied from 0.16 cm to 0.99 cm, and these positive alleles were carried by four varieties. There were eight positive alleles for the GP traits, with phenotypic effects varying from 1.78 % to 12.94 %, and these positive alleles were carried by eight varieties.

## Conclusions

In conclusion, seed vigor traits are of great importance for the breeding of the soybean. However, these are quantitative traits controlled by multiple genes. Association mapping has become the main method for exploring the genetic architecture of complex quantitative traits. Improving the detection efficiency of genetic loci and revealing the mechanism of genetic correlation and interaction between genes and the environment is a major challenge in analyzing plant quantitative traits. This study determined that more LD loci were found in improved lines than in landraces, which had a higher degree of LD. We identified 17 SSR marker traits associated with five vigor traits. Association mapping information can compensate for the lack of family linkage QTL location information. Information on allelic variations can be used to select parents, cross design for pyramiding novel alleles, and perform marker-assisted selection (MAS) in soybean breeding at the germination stage.

## Declarations

### Data availability

All data generated or analyzed during this study are included in this published article (and its supplementary information files). The improved lines are approved by the National Crop Variety Approval Committee. Soybean landraces are commonly used by breeders as the base material for variety selection and breeding. They meet the requirements for cultivation in Heilongjiang Province. The seeds are provided by the relevant breeding units and can be planted for use.

### Compliance with ethical standards

The authors declare to do not conflict of interest among authors and in the research work. The above manuscript publication is approved by all authors and by the responsible authorities where the work was carried out.

### Funding information

### Conflict of Interest

No conflict of interest among authors and in the research work.

### Authors Contribution

Caijin Wang wrote the manuscript, performed association analysis, linkage disequilibrium. Caijin Wang and Huiyan Zhao evaluated the five vigor-related traits; Yang Wang was involved in all aspects of the study and reviewed and edited the manuscript.

### Acknowledgments

The authors are grateful to professor Yang Wang for critical review of the manuscript. Financially supported by the Grant/Award Number: 2018NWB036-12; Master Innovation Research Project of Heilongjiang University, Grant/Award Number: YJSCX2020-216HLJU

### Additional Information

Additional supporting information may be found in the supplementary information files.

## References

1. Wilson, R. F. & Soybean Market Driven Research Needs in Genetics and Genomics of Soybean. *Springer*, 3–15. doi:10.1007/978-0-387-72299-3\_1 (2008)
2. Lam, H. M. *et al.* Resequencing of 31 wild and cultivated soybean genomes identifies patterns of genetic diversity and selection. *Nature Genetics*. **42**, 1053–1059 <https://doi.org/10.1038/ng.715> (2010).
3. Stevan, S. *et al.* Influence of seeds vigor in the attributes of soybean yield. *Revista de Ciências Agrárias*. **42**, 698–703 <https://doi.org/10.19084/rca.16389> (2019).
4. Kolchinski, E. M., Schuch, L. O. B. & Peske, S. T. Vigor de sementes e competição intra-específica em soja Seeds vigor and intra-specific competition in soybean. *Ciência Rural*. **35**, 1248–1256 <https://doi.org/10.1590/s0103-84782005000600004> (2005).
5. Sun, Q., Wang, J. H. & Sun, B. Q. Advances on Seed Vigor Physiological and Genetic Mechanisms., 6,1060–1066. doi:10.1016/s1671-2927(07)60147-3 (2007)
6. Henning, F. A. *et al.* Composição química e mobilização de reservas em sementes de soja de alto e baixo vigor. *Bragantia*. **69**, 727–734 <https://doi.org/10.1590/s0006-87052010000300026> (2010).
7. Zhao, D. L., Bastiaans, L., Atlin, G. N. & Spiertz, J. H. J. Interaction of genotype management on vegetative growth and weed suppression of aerobic rice. *Field Crops Research*. **100**, 327–340 <https://doi.org/10.1016/j.fcr.2006.08.007> (2007).
8. Leandro, D. C. *et al.* Variabilidade de plantas de soja originadas de sementes de diferentes níveis de qualidade fisiológica. *Acta Agron.* **64** (3), 234–238 (2015).

9. Xie, L. X. *et al.* Identification and fine mapping of quantitative trait loci for seed vigor in germination and seedling establishment in rice. *Journal of Integrative Plant Biology*. **56**, 749–759 <https://doi.org/10.1111/jipb.12190> (2014).
10. Lai, Y. Y. *et al.* Identification of QTLs with Additive, Epistatic, and QTL × Seed Maturity Interaction Effects for Seed Vigor in Rice. *Plant Molecular Biology Reporter*. **34**, 160–171 <https://doi.org/10.1007/s11105-015-0913-7> (2016).
11. Liu, J. B. *et al.* Identification of QTLs for maize seed vigor at three stages of seed maturity using a RIL population. *Euphytica*. **178**, 127–135 <https://doi.org/10.1007/s10681-010-0282-0> (2011).
12. Fu, Z. Y. *et al.* Functional marker related to germination vigor of maize seed. *Mol. Breeding*. **36**, 159 <https://doi.org/10.1007/s11032-016-0581-9> (2016).
13. Clerkx, E. J. M., Vries, H. B. D., Ruys, G. J., Steven, P. C. G. & Maarten, K. Genetic differences in seed longevity of various Arabidopsis mutants. *Physiologia Plantarum*. **121**, 448–461 <https://doi.org/10.1111/j.0031-9317.2004.00339.x> (2004).
14. Dang, X. J. *et al.* Genetic diversity and association mapping of seed vigor in rice (*Oryza sativa*L.). *Planta*. **239**, 1309–1319 <https://doi.org/10.1007/s00425-014-2060-z> (2014).
15. Zhang, X. *et al.* Whole-genome mapping identified novel “QTL hotspots regions” for seed storability in soybean (*Glycine max* L.). *BMC Genomics*. **20**, 499 <https://doi.org/10.1186/s12864-019-5897-5> (2019).
16. Dargahi, H., Tanya, P. & Srinives, P. Mapping of the genomic regions controlling seed storability in soybean (*Glycine max*L.). *Journal of Genetics*. **93**, 365–370 <https://doi.org/10.1007/s12041-014-0381-0> (2014).
17. Singh, R. K. *et al.* SSR markers associated with seed longevity in soybean. *Seed Science and Technology*. **36**, 162–167 <https://doi.org/10.15258/sst.2008.36.1.17> (2008).
18. Zhang, W. J. *et al.* Epistatic Association Mapping for Alkaline and Salinity Tolerance Traits in the Soybean Germination Stage. *PLoS ONE*. **9** (1), e84750 <https://doi.org/10.1371/journal.pone.0084750> (2014).
19. Mackay, L. & Powell, W. Methods for linkage disequilibrium mapping in crops. *Trends in Plant Science*. **12**, 57–63 <https://doi.org/10.1016/j.tplants.2006.12.001> (2007).
20. Yu, J. M. & Buckler, E. S. Genetic association mapping and genome organization of maize. *Current opinion in biotechnology*. **17** (2), 155–160 <https://doi.org/10.1016/j.copbio.2006.02.003> (2006).
21. Hu, Z. B. *et al.* Association mapping of yield-related traits and SSR markers in wild soybean (*Glycine soja* Sieb. and Zucc.). *Breed. Sci.* **63**, 441–449 <https://doi.org/10.1270/jsbbs.63.441> (2014).
22. Wen, Z. X. *et al.* Genome-wide association mapping of quantitative resistance to sudden death syndrome in soybean. *BMC Genomics*. **15**, 809 <https://doi.org/10.1186/1471-2164-15-809> (2014).
23. Assefa, T. *et al.* Genome-wide associations and epistatic interactions for internode number, plant height, seed weight and seed yield in soybean. *BMC Genomics*. **20**, 527 <https://doi.org/10.1186/s12864-019-5907-7> (2019).
24. Chang, F. G. *et al.* Genome-Wide Association Studies for Dynamic Plant Height and Number of Nodes on the Main Stem in Summer Sowing Soybeans. *Frontiers in Plant Science*. **9**, 184 <https://doi.org/10.3389/fpls.2018.01184> (2018).
25. Lü, H. Y. *et al.* Genome-wide association study of dynamic developmental plant height in soybean. *Canadian Journal of Plant Science*. **97**, 308–315 <https://doi.org/10.1139/CJPS-2016-0152> (2017).

26. Priolli, R. H. G., Campos, J. B., Stabellini, N. S., Pinheiro, J. B. & Vello, N. A. Association mapping of oil content and fatty acid components in soybean. *Euphytica*. **203**, 83–96 <https://doi.org/10.1007/s10681-014-1264-4> (2015).
27. Ali, G., Babak, R., Siamak, R. & Seyed, A. K. Association Analysis of Charcoal Rot Disease Resistance in Soybean. *Plant pathology journal*. **35**, 189–199 <https://doi.org/10.5423/PPJ.OA.12.2018.0283> (2019).
28. Zhao, X. *et al.* Loci and candidate gene identification for resistance to *Sclerotinia sclerotiorum* in soybean (*Glycine max* L. Merr.) via association and linkage maps. *Plant J*. **82**, 245–255 <https://doi.org/10.1111/tpj.12810> (2015).
29. Wang, Z. F., Wang, J. F., Bao, Y. M., Wu, Y. Y. & Zhang, H. S. Quantitative trait loci controlling rice seed germination under salt stress. *Euphytica*. **178**, 297–307 <https://doi.org/10.1007/s10681-010-0287-8> (2011).
30. Wang, Z. F., Wang, F. H., Zhou, R., Wang, J. F. & Zhang, H. S. Identification of quantitative trait loci for cold tolerance during the germination and seedling stages in rice (*Oryza sativa* L.). *Euphytica*. **181**, 405–413 <https://doi.org/10.1007/s10681-011-0469-z> (2011).
31. Murray, M. G. & Thompson, W. F. Rapid isolation of high molecular weight plant DNA. *Nucleic Acids research*. **8**, 4321–4326 <https://doi.org/10.1093/nar/8.19.4321> (1980).
32. Sanguinetti, C. J., Dias, N. E. & Simpson, A. J. Rapid silver staining and recovery of PCR products separated on polyacrylamide gels. *Biotechniques*. **17**, 915–919 (1994).
33. Liu, K. & Muse, S. V. POWERMARKER: an integrated analysis environment for genetic marker data. *Bioinformatics*. **21**, 2128–2129 <https://doi.org/10.1093/bioinformatics/bti282> (2005).
34. Pritchard, J. K., Stephens, M. & Donnelly, P. Inference of population structure using multilocus genotype data. *Genetics*, 155, 945–959. [doi.org/10.1111/j.1471-8286.2007.01758.x](https://doi.org/10.1111/j.1471-8286.2007.01758.x) (2000)
35. Evanno, G., Regnaut, S. & Goudet, J. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol. Ecol.* **14**, 2611–2620 <https://doi.org/10.1111/j.1365-294X.2005.02553.x> (2005).
36. Hardy, O. J. & Vekemans, X. SPAGEDi: a versatile computer program to analyse spatial genetic structure at the individual or population levels. *Mol. Ecol. Notes*. **2**, 618–620 <https://doi.org/10.1046/j.1471-8286.2002.00305.x> (2002).
37. Bradbury, P. J. *et al.* TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics*. **23**, 2633–2635 <https://doi.org/10.1093/bioinformatics/btm308> (2007).
38. Flint-Garcia, S. A., Thornsberry, J. M. & Buckler, E. S. Structure of linkage disequilibrium in plants. *Annual Review of Plant Biology*. **54**, 357–374 <https://doi.org/10.1146/annurev.arplant.54.031902.134907> (2003).
39. Farnir, F. *et al.* Extensive genome-wide linkage disequilibrium in cattle. *Genome Res*. **10**, 220–227 <https://doi.org/10.1101/gr.10.2.220> (2000).
40. Breseghello, F. & Sorrells, M. S. Association mapping of kernel size and milling quality in wheat (*Triticum aestivum* L.) cultivars. *Genetics*. **172**, 1165–1177 (2006).
41. Wen, Z. X. *et al.* Association analysis of agronomic and quality traits with SSR markers in *Glycine max* and *Glycine soja* in China: II. exploration of elite alleles. *Acta Agronomica Sinica*. **34**, 1339–1349 <https://doi.org/10.3724/SP.J.1006.2008.01339> (2008).

42. Maja, Å. M. *et al.* Genetic diversity of European commercial soybean [*Glycine max* (L.) Merr.] germplasm revealed by SSR markers. *Genet. Resour. Crop Evol.* **67**, 1587–1600 <https://doi.org/10.1007/s10722-020-00934-3> (2020).
43. Wen, Z. X., Ding, Y. L., Zhao, T. J. & Gai, J. Y. Genetic diversity and peculiarity of annual wild soybean (*G. soja* Sieb. et Zucc.) from various eco-regions in China. *Theoretical and Applied Genetics.* **119**, 371–381 <https://doi.org/10.1007/s00122-009-1045-y> (2009).
44. Wang, L. X. *et al.* Genetic diversity center of cultivated soybean (*Glycine max*) in China-New insight and evidence for the diversity center of Chinese cultivated soybean. *Journal of Integrative Agriculture.* **15**, 2481–2487 [https://doi.org/10.1016/S2095-3119\(15\)61289-8](https://doi.org/10.1016/S2095-3119(15)61289-8) (2016).
45. Yu, J. M. *et al.* A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nature Genetics.* **38**, 203–208 <https://doi.org/10.1038/ng1702> (2006).
46. Kumar, B. *et al.* Population structure and association mapping studies for important agronomic traits in soybean. *Journal of Genetics.* **93**, 775–784 <https://doi.org/10.1007/s12041-014-0454-0> (2014).
47. Zhu, Y. L. *et al.* Single-nucleotide polymorphisms in soybean. *Genetics.* **163**, 1123–1134 (2003).
48. Zhou, Z. K. *et al.* Resequencing 302 wild and cultivated accessions identifies genes related to domestication and improvement in soybean. *Nature Biotechnology.* **33**, 408–414 <https://doi.org/10.1038/nbt.3096> (2015).
49. Jun, T. H., Van, K., Kim, M. Y., Lee, S. H. & Walker, D. R. Association analysis using SSR markers to find QTL for seed protein content in soybean. *Euphytica.* **162**, 179–191 <https://doi.org/10.1007/s10681-007-9491-6> (2008).
50. Zhang, X. Y. *et al.* Hitchhiking Effect Mapping: A New Approach for Discovering Agronomic Important Genes. *Scientia Agricultura Sinica.* **39**, 1526–1535 (2006).

## Figures

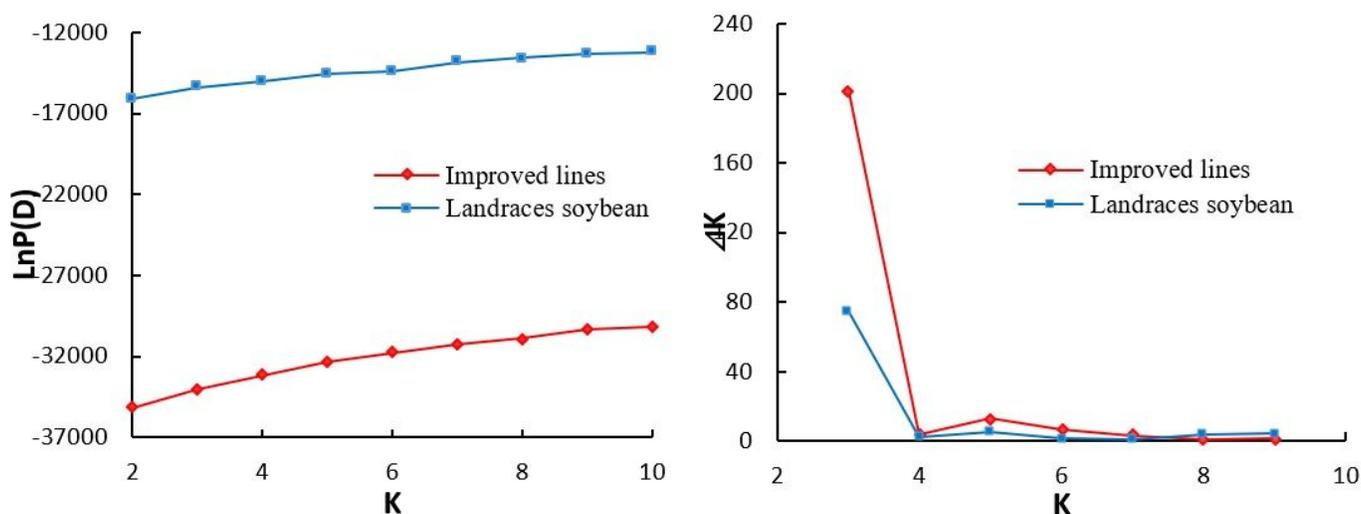
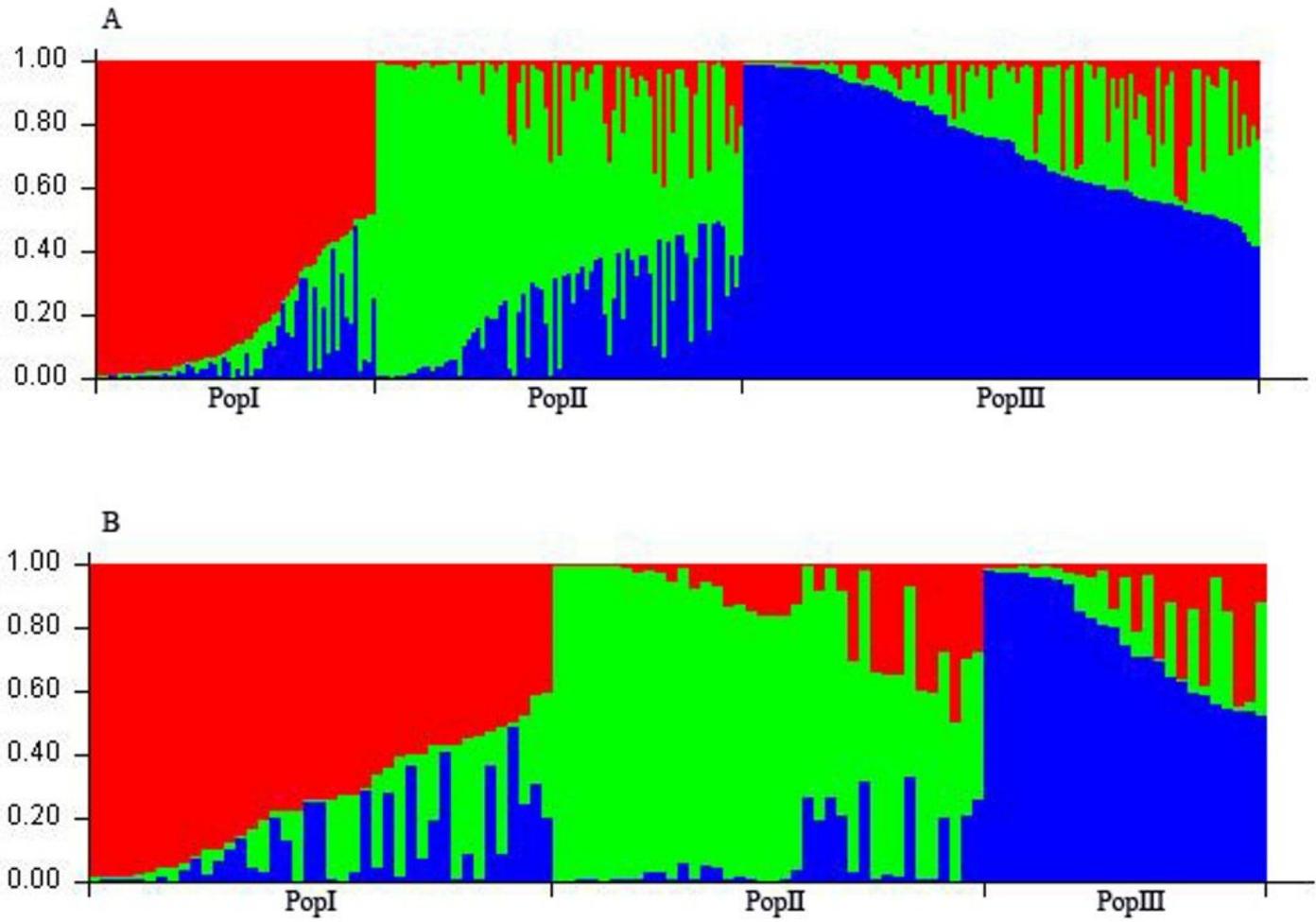


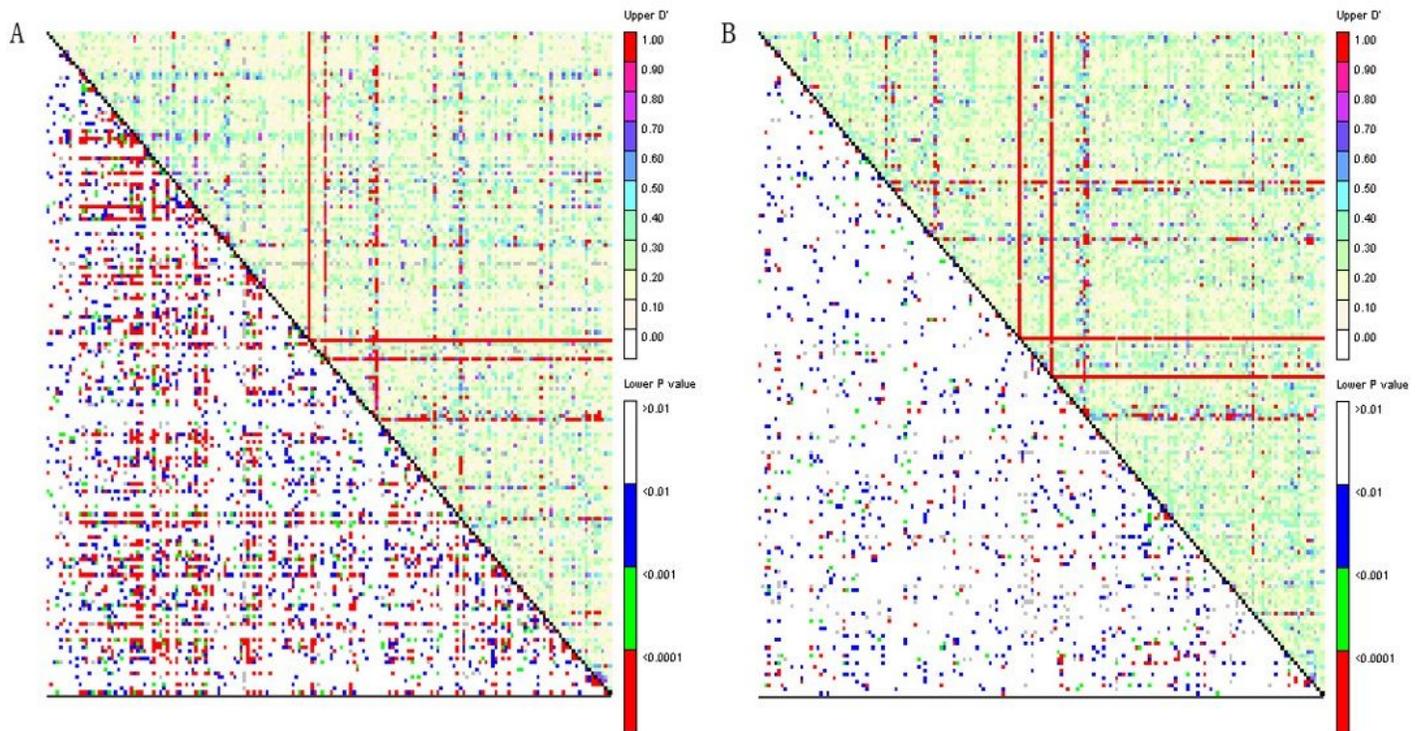
Figure 1

Magnitude of K and Delta K of population structure analysis based on improved lines and landraces soybean



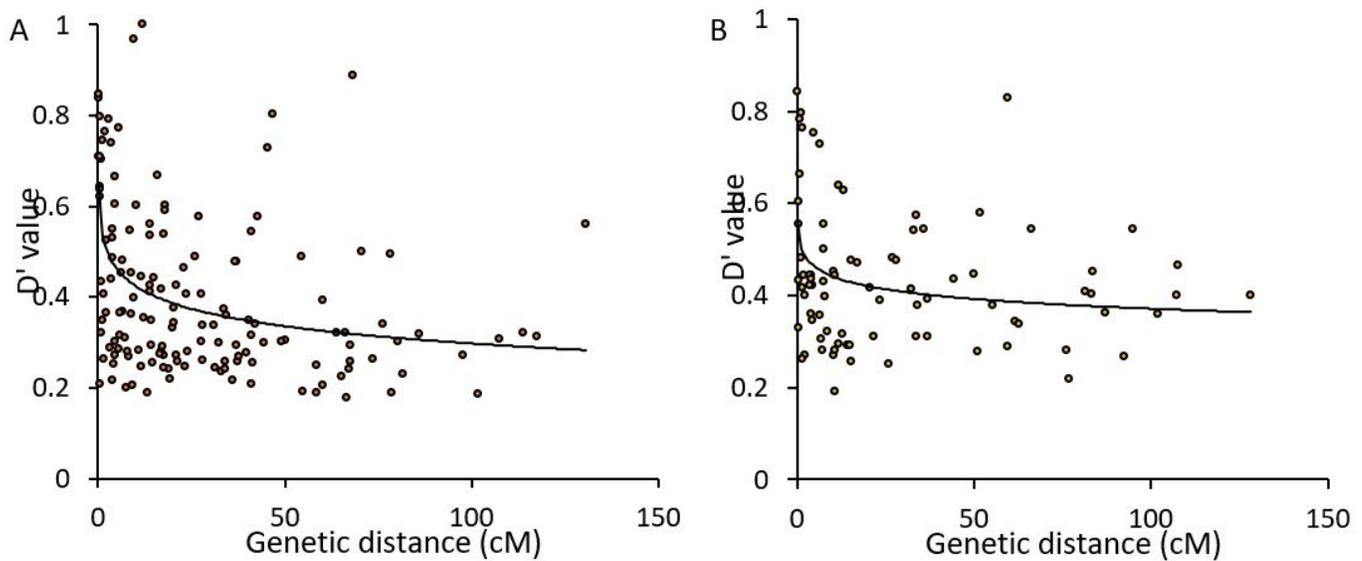
**Figure 2**

Population structure of improved lines and landraces soybean. Identified subpopulations are PopI (red color), PopII (green color), PopIII (blue color). (A, improved lines; B, landraces soybean)



**Figure 3**

Distribution of linkage disequilibrium among SSR loci (A, improved lines; B, landraces soybean). Note: Each pixel above the diagonal indicates the  $D'$  size of the corresponding marker pair as shown in the color code at the upper right, and each pixel below the diagonal indicates the P-value size of the testing LD of the corresponding marker pairs as shown in the color at the lower right



**Figure 4**

Scatter plot of the  $D'$ -value of marker genetic distance (A, improved lines; B, landraces soybean)

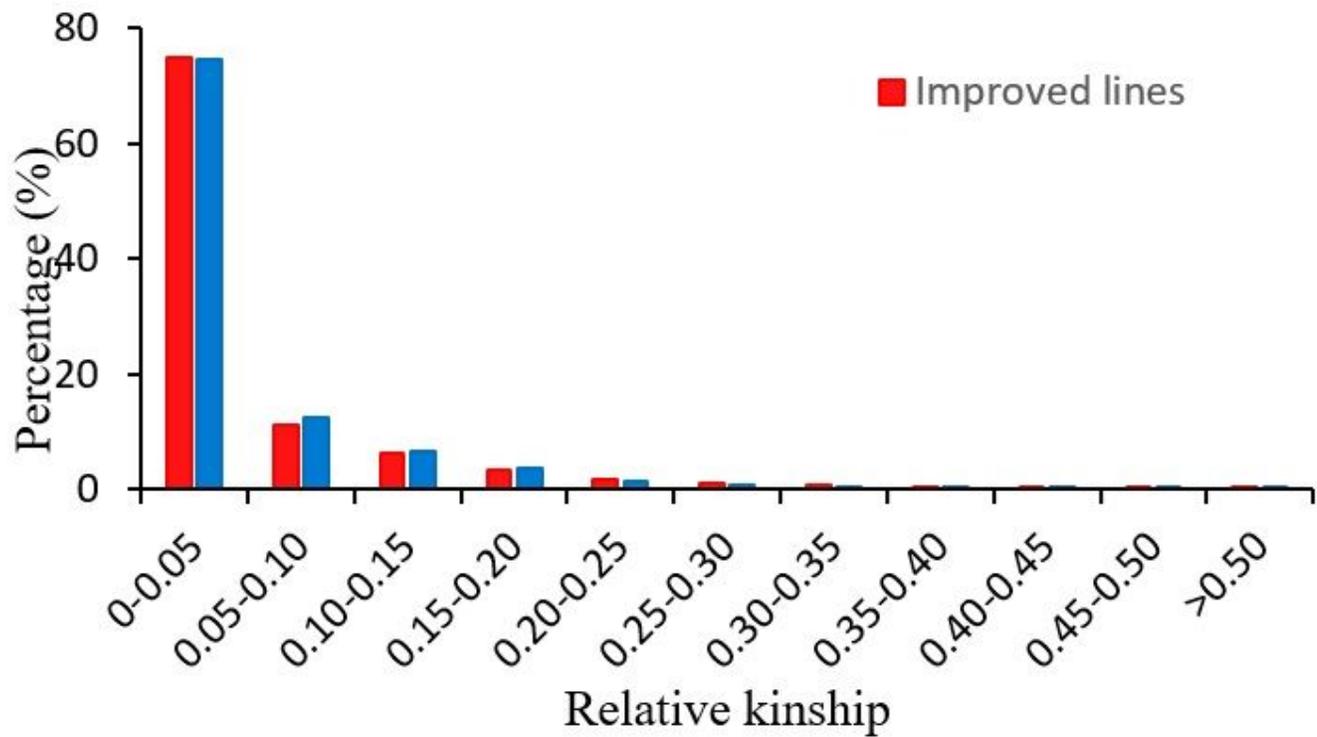


Figure 5

Distribution of pair-wise kinship coefficients among improved lines and landraces soybean

## Supplementary Files

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

- [SupplementaryInformation.docx](#)