Identification of drought-tolerance genes in the germination stage of soybean

Xingzhen Zhao

CAAS ICS: Chinese Academy of Agricultural Sciences Institute of Crop Sciences

Zhangxiong Liu

CAAS ICS: Chinese Academy of Agricultural Sciences Institute of Crop Sciences

Huihui Li

CAAS ICS: Chinese Academy of Agricultural Sciences Institute of Crop Sciences

Yanjun Zhang

Gansu Academy of Agricultural Sciences

Lili Yu

CAAS ICS: Chinese Academy of Agricultural Sciences Institute of Crop Sciences

Xusheng Qi

Gansu Academy of Agricultural Sciences

Huawei Gao

CAAS ICS: Chinese Academy of Agricultural Sciences Institute of Crop Sciences

Yinghui Li

CAAS ICS: Chinese Academy of Agricultural Sciences Institute of Crop Sciences

Lijuan Qiu (✉ qiulijuan@caas.cn)

Chinese Academy of Agricultural Sciences Institute of Crop Sciences <https://orcid.org/0000-0001-5777-3344>

Research article

Keywords: Soybean, Germination, Drought tolerance, Genome-wide association analysis (GWAS)

Posted Date: November 25th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-112822/v1>

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

Abstract

Background

Drought stress influences the vigor of plant seeds and inhibits seed germination, making it one of the primary environmental factors adversely affecting food security. The seed germination stage is critical to ensuring the growth and productivity of soybeans in soils prone to drought conditions. We here examined the genetic diversity and drought-tolerance phenotypes of 410 accessions of a germplasm diversity panel for soybean and conducted quantitative genetics analyses to identify loci associated with the drought-tolerance of seedlings.

Results

We uncovered significant differences among the diverse genotypes for 4 growth indices and 5 drought-tolerance indices, which revealed abundant variation among genotypes, upon drought stress, and for genotype $\times$ treatment effects. We also used 158,327 SNP markers, and performed GWAS for the drought-related traits. Our data met the conditions (PCA + K) for using a mixed linear model in TASSEL, and we thusly identified 26 SNPs associated with drought tolerance indices for germination stage soybean plants. These were distributed across 10 soybean chromosomes, and these explain 5.19–9.66% of the observed phenotypic variation. Nine SNP sites, including for example Gm20_34956219 and Gm20_34956219, were associated with two or more phenotypic indices, and there were nine SNP markers located in or adjacent to (within 500 kb) previously reported drought tolerance QTLs. These SNP led to our identification of 41 candidate genes related to drought tolerance in the germination stage.

Conclusion

The results of our study contribute to a deeper understanding of the genetic mechanisms underlying drought tolerance in soybeans at the germination stage, thereby providing a molecular basis for identifying useful soybean germplasm for breeding new drought-tolerant varieties.

Background

Drought is one of the most impactful environmental factors affecting agricultural production [1, 2]. In recent years, global warming and decreased precipitation have caused frequent droughts [3], resulting in substantial % reductions in global crop yields [4]. Irrigation is often limited in certain regions and increases production costs [1, 5]. Therefore, much current agricultural research focuses on understanding drought tolerance in crops, seeking to accelerate the development of suitably drought-tolerant cultivars [2, 6].

Soybean is an important global food source and a valuable economic crop: it is one of the main sources of plant fat and protein for humans, possesses medicinal value, and is widely used as a raw material in industry [7, 8]. Of the legumes, soybean is the most sensitive to water [9]; indeed, drought conditions have reduced annual global soybean production by approximately 40% [10, 11]. Seed germination is an extremely important growth stage for plants, and drought stress during germination can reduce the overall number of seedlings by 20% and in serious cases can reduce yields by more than 50% [4, 12]. This makes the identification of drought-tolerant germplasm and the cultivation of drought-resistant cultivars particularly important in soybean.

Extensive research has been conducted to study drought tolerance in soybean. The main drought tolerance traits identified to date include germination rate [2, 13, 14], germination energy [2], drought index [15], leaf wilting [1, 16, 17], water use efficiency [18, 19], canopy wilt [7, 20], fibrous roots [20], and yield under water deficit [21]. Research investigating drought tolerance at the germination stage of soybean has been mainly based on PEG treatment to simulated drought conditions. Vijay *et al.* [2] employed different concentrations of PEG-6000 to study the drought tolerance of soybean accessions, demonstrating that germination rate, vigor index, and stress tolerance index can be used to evaluate the drought tolerance of soybeans at the germination stage. Dantas *et al.* [22] found that the optimum concentration of PEG-6000 to study drought tolerance during soybean germination was -0.2 MPa.

Drought tolerance in plants is a quantitative trait controlled by multiple genes [23]. QTL (quantitative trait locus) is related to drought tolerance and can be used as a molecular marker to improve selection efficiency [24]. There are currently 120 drought-tolerant QTLs represented in SoyBase (www.soybase.org); these are primarily distributed on chromosomes 2, 5, 17, and 19 [7, 15, 20]. Specht *et al.* [25] constructed a recommended inbred line (RIL) population with 236 lines from a Minsoy × Noir1 cross, and performed six different levels of water stress experiments over two years; this work identified three QTL loci related to yield under drought stress. Hwang *et al.* [7] used five RIL populations to study the QTLs controlling leaf wilting, identifying seven stable QTLs. Kaler *et al.* [24] used 31,260 SNPs to conduct genome-wide association analysis on 373 soybean accessions and identified 47 SNP markers related to WUE (water use efficiency). Liu *et al.* [14] used 4616 SNPs to conduct genome-wide association analysis on 259 soybean accessions at the germination stage and identified 15 QTLs related to drought tolerance indices during germination.

It is notable that the results from previous studies often differ, owing in part to their various use of diverse different mapping groups, molecular markers, experimental environments, and calculation methods. There are therefore relatively few drought-tolerance QTLs that have been separately verified by separate research groups. Due to drought localization (which is limited by the number of parents) the identified drought tolerance QTL is easily overlooked [26]. As such, the drought tolerance of soybeans during the germination stage requires further study.

Here, we examined a soybean germplasm diversity panel comprising 410 accessions, which we screened with drought conditions using PEG-6000. We used the data from the germination stage growth assays to calculate the relative germination rate (RGR), relative germination energy (RGE), germination drought tolerance index (GDTI), germination stress index (GSI), and membership function value (MFV) for each of the soybean accessions. We used to genotype the 410 accessions using 158,327 SNPs, and then conducted a whole-genome association analysis. Our GWAS identified 26 SNPs associated with drought tolerance indices at the germination stage, among which nine were associated with two or more indices. There were nine SNP markers located in or adjacent to previously reported drought tolerance QTLs, and 41 candidate genes related to drought tolerance in the germination stage were identified. Our results will provide a molecular basis for identifying drought-tolerant germplasm and will help develop new soybean cultivars that exhibit strong drought tolerance at the germination stage.

Results

Selection of the optimal concentration of PEG-6000

Seeds of the 6 variously drought tolerance soybean accessions were treated with 5 different concentrations of a PEG-6000 solution (Fig. 1). As expected, the GR and GE values for all of the accessions decreased as the concentration of PEG-6000 increased. At 25% PEG-6000, the soybeans did not germinate. When treated with 10% PEG-6000 and 15% PEG-6000, there were no significant differences among the genotypes for GR, GE, GI, or GDI compared to the control samples (0% PEG-6000). When treated with 20% PEG-6000, the GR, GE, GI, and GDI of all 6 accessions were significantly reduced (to varying degrees) compared to the control, and the differences were significant between the 20% PEG-6000 treatment and the control. Therefore, we selected the 20% PEG-6000 concentration for simulating the drought-stress condition for our large-scale, germination-stage screen of the 410 accessions of our soybean germplasm diversity panel.

Phenotype analysis of soybean germplasm at the germination stage

Descriptive analysis of four germination-related traits and drought tolerance traits

We measured four germination-related traits (GI, GDI, GE, and GR) for germinating seeds of the 410 soybean accessions under 20% PEG-6000 (T) or 0% PEG-6000 (C). Table S3 displays the calculated mean for each trait, the ranges, the standard deviations, and the coefficients of variation. The mean values for the whole drought-treated panel for the GI, GDI, GE, and GR traits were 0.68, 17.49, 13.58%, and 15.07%, respectively, whereas the means for the controls were 7.50, 189.20, 96.33%, and 96.92%. The results of a MIXED model procedure ANOVA (Table 1) identified significant differences among genotypes, treatment, and genotypes × treatment ($P < 0.001$). The treatment mean square was the largest, suggesting that the drought treatment was the most impactful factor.

Table 1
Analysis of variance (ANOVA) of four germination-related traits under 0% PEG and 20% PEG conditions for seeds of the 410 soybean accessions

Trait	Source	DF	Sum of Square	Mean Square	F Value	Pr > F
GR	Geno	409	251929.70	615.97	7.29	< .0001
	Treatment	1	4374687.00	4374687.00	51778.10	< .0001
	Block/Treat	2	357.14	178.57	2.11	0.1211
	Geno × Treat	409	246113.60	601.75	7.12	< .0001
GE	Geno	409	236174.00	577.44	7.74	< .0001
	Treatment	1	4470355.00	4470355.00	59943.60	< .0001
	Block/Treat	2	359.92	179.96	2.41	0.0898
	Geno × Treat	409	226055.60	552.70	7.41	< .0001
GDI	Geno	409	1658797.00	4055.74	10.23	< .0001
	Treatment	1	19290378.00	19290378.00	48679.80	< .0001
	Block/Treat	2	2222.70	1111.35	2.80	0.0608
	Geno × Treat	409	613890.90	1500.96	3.79	< .0001
GI	Geno	409	3440.47	8.41	8.82	< .0001
	Treatment	1	30333.58	30333.58	31790.80	< .0001
	Block/Treat	2	9.94	4.97	5.21	0.0055
	Geno × Treat	409	1756.14	4.29	4.50	< .0001
GR, germination rate; GE, germination energy; GDI, germination drought index; GI, germination index; DF, degree of freedom.						

Analysis of drought tolerance

The respective mean values for RGR, RGE, GDTI, GSI, and MFV were 0.16, 0.14, 0.09, 0.08, and 0.15 for the 410 soybean germplasm accessions (Table 2, Fig. 2). The coefficient of variation ranges for RGR, RGE, and MFV were 0–1, the coefficient of variation ranges for GDTI and GSI were 0-0.57 and 0-0.48, respectively, and the maximum coefficient of variation for RGE was 136.51. The minimum coefficient of variation for MFV was 121.81. The results of variance analysis of RGR, RGE, GDTI, and GSI demonstrated that there were significant differences among genotypes for all examined indices ($P < 0.001$), but no significant differences between repeats (Table 3). The results of the correlation analysis demonstrated that there were significant positive correlations for all indices ($P < 0.001$) (Table 4), which could be because the phenotypic values of five drought tolerance indices were calculated on the basis of the germination number. The general heritability of MFI, GSI, GDTI, RGE, and RGR was high: 90.41%, 87.78%, 88.90%, 90.86%, and 91.43%, respectively, which aid in early selection of offspring.

Table 2
Descriptive statistics of five drought tolerance indices

Traits	Mean	SD	Skewness	Kurtosis	Range	CV
RGR	0.16	0.20	1.55	2.03	0.01	129.88
RGE	0.14	0.19	1.63	2.40	0.01	136.51
GDTI	0.09	0.12	1.60	2.21	0.0057	135.45
GSI	0.08	0.11	1.43	1.44	0.0048	129.00
MFV	0.15	0.19	1.41	1.60	0.01	121.81
RGR, relative germination rate; RGE, relative germination energy; GDTI, germination drought tolerant index; GSI, germination stress index; MFV, Membership function value; SD standard deviation; CV coefficient of variation.						

Table 3
Analysis of variance of four drought tolerance indices of 410 accessions at 0% PEG and 20% PEG

Trait	Source	DF	Sum of Square	Mean Square	FValue	Pr > F
RGR	Geno	409	44.94	0.11	11.67	< .0001
	Block	2	0.01	0.00	0.28	0.75
RGE	Geno	409	39.22	0.10	10.94	< .0001
	Block	2	0.01	0.00	0.47	0.63
GDTI	Geno	409	13.88	0.03	9.01	< .0001
	Block	2	0.00	0.00	0.28	0.75
GSI	Geno	409	12.41	0.03	8.19	< .0001
	Block	2	0.00	0.00	0.16	0.86
RGR, relative germination rate; RGE, relative germination energy; GDTI, germination drought tolerant index; GSI, germination stress index; DF, degree of freedom.						

Table 4
Phenotypic correlations between five drought tolerance indices in the 410 soybean accessions

Trait	RGR	RGE	GDTI	GSI
RGE	0.9840***			
GDTI	0.9640***	0.9766***		
GSI	0.9490***	0.9524***	0.9729***	
MFV	0.9862***	0.9902***	0.9909***	0.9819***
*** significant level under 0.0001 for Pearson correlation test.				
RGR, relative germination rate; RGE, relative germination energy; GDTI, germination drought tolerant index; GSI, germination stress index; MFV, Membership function value				

Analysis of soybean genetic diversity

Analysis of genetic diversity and linkage disequilibrium

The results of a PowerMarker analysis demonstrated that the mean MAF value for 117,811 SNPs among the 410 accessions of the diversity panel was 0.2228 (ranging from 0 to 0.5030) and the proportion of SNPs with MAF greater than 0.2228 was approximately 46.9%. The mean values of genetic diversity, heterozygosity, and PIC were 0.3043, 0.0237, and 0.2548, respectively, and the ranges were 0-0.5061, 0-0.4070, and 0-0.3843 (Fig. 3, Table S4). The results of an LD analysis demonstrated that the

whole-genome mean LD for the diversity panel was $r^2 = 0.3440$. The r^2 value decreased to approximately half of its maximum level once the LD decay distance reached approximately 75 kb (Fig. 4). This suggests that LD decayed relatively quickly among the accession of the panel.

Population genetic structural analysis

To avoid false-positive associations due to population stratification, three calculations were executed to study population structure: principal component analysis (PCA), phylogenetic-tree construction, and population-structure analysis with ADMIXTURE. Our PCA analysis based on the SNP data for the whole panel indicated as expected that eigenvalues decreased as the number of PCs increased (Fig. 5B and C). With fewer than 4 PCs, the eigenvalues decreased gradually; this suggests that accessions in the diversity panel can be plausibly divided into four subgroups. To better understand the genetic diversity of the soybean germplasm panel, we built a neighbor-joining tree based on the incidence of common alleles between the accessions. This analysis also divided the accessions of the panel into four subgroups (Fig. 5A), findings consistent our results from the PCA.

We again detected a 4 subgroup structure for the accessions of the panel when we conducted a population structure analysis using the ADMIXTURE program (Fig. 5D). For the 4 groups, the G1 accessions were primarily from the United States, Japan, and other countries outside China, the G2 accessions were primarily from Northern China, G3 accessions were from diverse locations within China, and the G4 accessions were primarily from Southern China.

GWAS to identify SNPs associated with drought tolerance

Our MLM GWAS analysis identified a total of twenty-six SNP loci that were significantly associated with drought-tolerance traits (Table 5; Fig. 6); these were distributed on 15 chromosomes. There were 8, 8, 22, 5, and 8 SNPs associated with RGR, RGE, GDRI, GSI, and MFV, respectively. There were 5 significantly drought-tolerance trait related SNPs on both chromosomes 1 and 20, 4 SNPs on chromosome 8, 2 SNPs each on chromosomes 4, 9, and 15, as well as 1 significant SNP each for chromosomes 2, 3, 5, 6, 7, 11, 13, 14, and 19. Notably 9 loci were associated with two or more drought-related traits. The amount of phenotypic variation explained by these SNPs was 5.19–9.66%, with an average of 6.99%. It was also highly notable that 21 of the QTLs identified in our GWAS are located within 500 kb of previously identified loci in quantitative genetics analyses of soybean, with 9 of these near loci previously associated with drought-related traits. The remaining 16 were for yield-related traits.

Table 5 SNPs significantly associated with five drought tolerance indices (-logP>4.5)

Marker	Chr	Position	Associated traits(R ²)	Reported QTLs/genes
Gm01_35877607	1	35877607	RGR(7.20), RGE(7.53), GDTI(6.93), MFV(6.95)	Seed set (Ning et al., 2018); Seed weight (Panthee et al., 2005; Han et al., 2012)
Gm01_38948188	1	38948188	GDTI(6.73)	Pod wall weight (Guo et al., 2011)
Gm01_47042336	1	47042336	GSI(6.31), GDTI(6.51)	Drought index (Du et al., 2009a); Root area (Wu et al., 2012); Root length (Wu et al., 2012)
Gm01_48619013	1	48619013	GDTI(6.16)	Drought index(Du et al., 2009a)
Gm02_6357585	2	6357585	GDTI(7.11)	Canopy wilt (Hwang et al., 2015)
Gm03_39037	3	39037	GDTI(6.08)	
Gm04_4484515	4	4484515	RGR(6.84)RGE(7.22)GDTI(6.99)MFV(6.74)	Canopy wilt (Hwang et al., 2015); Seed set (Ning et al., 2018); Seed weight (Li et al., 2010)
Gm04_50945875	4	50945875	GDTI(6.69)	Seed number (Li et al., 2010); WUE (Dhanapal et al., 2015)
Gm05_38540838	5	38540838	RGR(7.60)	Cellwall polysacch composition (Stombaugh et al., 2004)
Gm06_9791913	6	9791913	GDTI(6.17)	Seed weight (Moongkanna et al., 2011); Shoot weight (Vieira et al., 2006)
Gm07_24735482	7	24735482	GDTI(6.59)	
Gm08_1438457	8	1438457	RGE(6.08)	Seed weight per plant (Liu et al., 2011)
Gm08_4052111	8	4052111	GDTI(6.91)	Canopy wilt (Hwang et al., 2015); Seed weight (Teng et al., 2009)
Gm08_7972856	8	7972856	RGR(8.03)	Root density, lateral (Manavalan et al., 2015); Seed set (Ning et al., 2018)
Gm09_11414508	9	11414508	RGE(5.19)GDTI(6.00)MFV(5.32)	Seed yield (Kabelka et al., 2004; Guzman et al., 2007)
Gm09_18023730	9	18023730	GSI(6.46)GDTI(6.77)MFV(6.03)	Seed yield (Yuan et al., 2002; Guzman et al., 2007)
Gm11_30280479	11	30280479	RGR(8.33)RGE(7.04)GDTI(8.08)MFV(7.53)	Seed set (Ning et al., 2018)
Gm13_35517964	13	35517964	GDTI(6.55)	
Gm14_46603856	14	46603856	GDTI(6.19)	Canopy wilt (Abdel-Haleem et al., 2012)
Gm15_11950665	15	11950665	GDTI(6.52)	Seed weight (Liu et al., 2011)
Gm15_47429024	15	47429024	GSI(6.45)	
Gm19_49449499	19	49449499	GDTI(6.46)	Canopy wilt (Hwang et al., 2015); Drought tolerance(Zhang et al., 2012)
Gm20_4618170	20	4618170	GDTI(6.28)	
Gm20_13921498	20	13921498	RGR(6.66)RGE(7.26)GDTI(7.01)MFV(6.64)	Seed weight (Han et al., 2012)
Gm20_34956219	20	34956219	RGR(6.88)RGE(7.87)GSI(7.22)GDTI(8.33)MFV(7.81)	Canopy wilt(Abdel-Haleem et al., 2012); Root density, lateral (Liang et al., 2014); Seed set (Han et al., 2012); WUE (Kaler et al., 2017)
Gm20_36902659	20	36902659	RGR(7.69)RGE(8.03)GSI(8.19)GDTI(9.66)MFV(8.57)	Root density, lateral (Liang et al., 2014)

RGR, relative germination rate; RGE, relative germination energy; GDTI, germination drought tolerant index; GSI, germination stress index; MFV, Membership function value

Discussion

Identifying germplasm resources in order to evaluate drought tolerance in soybeans is necessary for drought-tolerant breeding, the study of drought tolerance mechanisms, and the detection of molecular markers [14, 24]. Previous results demonstrated that germplasm with high drought tolerance had high rates and uniformity of germination [22]. Germination speed, uniformity, and elongation of young roots were then used to explore the drought tolerance of germplasm, and RGR, RGE, GDTI, and GSI data were used to evaluate drought tolerance [28, 40]. In the present study, we first used six accessions with different drought tolerance levels and conducted treatments with different concentrations of PEG-6000. By comparatively analyzing the GR, GE, GI, and GDI, we determined that the optimal concentration of PEG-6000 for a larger scale screen was 20%. We then subjected the 410 accessions of our germplasm diversity panel to drought-stress. Analysis with a linear ANOVA model indicated significant ($P < 0.001$) variation for drought tolerance among genotypes (Table 3). The phenotypic coefficient of variation of traits related to drought tolerance was large (121.8-136.5) (Table 2), suggesting significant phenotypic variation; while drought-related traits displayed highly generalized heritability ($\geq 85\%$).

The effect of drought on crops is multifaceted. The membership function method be used to synthesize multiple evaluation indices, to avoid the bias of a single index [27, 29], and to better evaluate the drought tolerance of soybean [27]. In this study, a total of 26 drought-tolerant loci were identified. Among these loci, the number for detected based on the MFV data was 8; and there were also loci detected based on MFV and the other indices, for example 8 by both MFV and GDRI, 7 by both MFV and RGE, 6 by both MFV and RGR, and 3 by both MFV and GSI. Of the SNPs associated with the loci of the five drought tolerance indices, nine were associated with two or more traits, among which eight were detected by MFV. As such, MFV is excellent for evaluating drought tolerance, highlighting its utility for evaluating drought tolerance in the germination stage of soybean.

Drought tolerance is a quantitative trait controlled by multiple genes [14, 23]. In this study, the distribution of phenotypic index values demonstrated significant variation in the 410 accession panel (Fig. 3), and a genome-wide association analysis identified 26 QTLs from the 117,811 SNPs that were related to drought tolerance (Table 5). These results reinforce that drought tolerance during the germination stage is controlled by multiple genes.

Nine of the 26 significantly association drought-tolerance associated loci were associated with two or more drought-related traits. Two loci (Gm20_34956219 and Gm20_349602658) were associated with five drought tolerance indices; four loci (Gm01_35877607, Gm04_4484515, Gm11_30280479, and Gm20_13921498) were associated with four drought tolerance traits; two loci (Gm09_11414508 and Gm09_18023730) were associated with three drought tolerance traits; and Gm_01_47042336 was associated with two drought tolerance traits. These results are in agreement with previously reported results about the involvement of multiple loci in drought-tolerance responses in soybean [41–43]. Moreover, as these drought-tolerance associated loci have been detected several times, our study reinforces that these are relatively stable QTL loci.

We compared the results of the association mapping of drought tolerance with previously studied QTLs within a 500 kb range using Soybase (<http://www.soybase.org>). Of the 26 significant SNPs in this study, nine were located in or near the reported QTLs related to drought tolerance (Table 5). Of these, six were related to canopy wilt [7, 20], two were related to drought index [15], two were related to WUE [23, 24], and one was related to drought tolerance in the germination stage [44]. Gm19_49449499 is located downstream of the QTL satt513 [13], which is reportedly related to drought tolerance in the germination stage; at the same time, a canopy wilting QTL exists near Gm19_49449499 [7]. Gm20_34956219 was associated with data for five drought tolerance indices in our study; this is located near a wilting canopy QTL [20] and within 14 Kb of the WUE marker ss715637488 [24]. Additionally, a QTL locus reported to control root density [45] is located near Gm20_34956219. Wang *et al.* [6] reported that the more lateral roots soybean accessions have, the stronger the water absorption and thus drought resistance. These are stable QTL loci detected by different research materials and using different research methods, so it is likely that there are drought-tolerant genes within their genomic regions. These drought-tolerance associated markers should be useful for identifying causal genes that can be used to improve drought tolerance in soybeans.

Drought is one of the primary abiotic stresses affecting crop production and severely restricts soybean yield [4, 12]. There are QTLs related to yield traits near the significant SNPs detected in this study, which are located on chromosomes 1, 4, 5, 6, 8, 9, 11, 15, and 20 (Table. 5) [5, 42, 44, 46–53]. Gm08_4052111 is adjacent to the canopy wilt marker [7], and is located between regions (ranging from satt390-satt424) related to seed weight [44]. We found that Gm09_11414508 is related to three drought-related traits, and is

positioned within the seed yield marker range satt518-BARC-041991-08155 [49]. Gm08_1438457 is located between Sat_383 - BARC-010329-00586, which reportedly controls single seed weight QTL [51]. Genomic regions with multiple associated traits suggest pleiotropy of a single causal gene or the close association of multiple causal genes. Using MAS diagrams, these markers can in theory be used for molecular marker-assisted selection to help improve both drought tolerance and yield in soybeans.

Of the SNPs with significant associations detected in this study, and in addition to Gm20_34956219, there are three SNPs located close to reported QTLs associated with root traits. For example, Gm20_36902659 is adjacent to a reported QTL related for lateral root density [45]. Near Gm01_47042336, there are root area locus [54] and root length locus [54]. There are also five drought-associated QTLs that we identified which have not been reported in previous studies. These are new loci and will thus require verification by additional studies.

We used the SoyBase database to identify candidate genes directly associated with the SNPs of our detected QTLs or in nearby genes. We identified 26 candidate regions containing 41 genes. Of these, 12 SNPs related to the drought tolerance indices detected in this study were located within genes, including SNPs causing four non-synonymous mutations, three synonymous mutations; there were also three SNPs in the 3'-UTR of genes and two SNPs positioned in gene introns. Fourteen of the significantly associated SNPs were located in intergenic regions (Table S5). The results of functionally annotating 41 phytozome genes (<https://phytozome.jgi.doe.gov>) suggested that these candidate genes may have functions related to bidirectional sugar transporter sweet, transferase, exportin, and hydrolase. The gene *Glyma.01g106000* (adjacent site Gm01_35877607) and *Glyma.08g103900* (Gm08_7972856 is located on the gene) regulate root morphology and the expression of a transferase in soybean [55, 56], which could be related to drought tolerance in soybean. The gene *Glyma.08g017800* (Gm08_1438457 is located on the gene) improves drought tolerance by regulating the rise and fall of glucose under drought conditions in soybean [6]. Gm01_48619013 is located in an exon of the *Glyma.01g149300* gene (*Glyma.01g35220* in v1.1) (Fig. S2), which encodes a methyltransferase PMT21-related protein (Table S5), while the *Glyma.01g149300* gene improves drought tolerance in soybeans by regulating protein synthesis under drought conditions [57]. The consistency of the associations was tested by comparing the drought tolerance of particular genotypes of Gm01_48619013 (Fig. S2) SNP sites as defined by this study. Drought tolerance in accessions that carry Gm01_48619013-GG genotypes was significantly higher within populations than for genotypes homozygous to alternate alleles.

Conclusion

In this study, 410 soybean accessions were tested for drought tolerance by simulating drought conditions with 20% PEG-6000. Variance analysis demonstrated that there were significant differences among the genotypes in five drought tolerance indices: RGR, RGE, GDRI, GSI, and MFV. A whole-genome association analysis was performed, using 158,327 SNP markers. Twenty-six SNP loci related to drought tolerance during the germination stage were detected. Of these, nine SNP loci were significantly related to two or more drought-tolerance traits, nine loci were near QTL loci reportedly related to drought tolerance, and two SNP-related genes were associated with drought tolerance in soybeans. It is extremely important to continue studying drought-tolerance genes and markers to assist with the selection and development of drought-tolerant soybean accessions.

Methods

Plant materials

410 soybean accessions were obtained from the Chinese National Soybean GeneBank (CNSGB), including 110 Non-Chinese accessions (from 8 countries, including the United States and Russia) and 300 Chinese domestic accessions (from 27 Chinese provinces, with many accessions from the Northeastern provinces of Heilongjiang and Liaoning, in the known center of soybean domestication) (Table S1). Pilot screening identified the set of 6 accessions (Table S2)—each with different levels of drought tolerance—that we used to initially optimize the PEG-6000 concentration we used for the larger-scale screen of the full 410 accession germplasm diversity panel.

Methods

Optimum PEG-6000 concentration screening

We used six soybean accessions with different levels of drought tolerance at the germination stage for this drought treatment optimization process. Seeds of uniform size were selected from each accession; these were sterilized with 0.1% HgCl₂ for 30 s, washed with sterile water 2-3 times, and dried with filter paper. Twenty seeds per genotype were used in each of three replications. The seeds were placed on wetted filter paper in 9-cm-diameter Petri dishes to evaluate the growth performance and phenotypic variation in seedlings. We then added 15 ml of PEG-6000 solution to each dish at the following concentrations: 0% (CK), 10%, 15%, 20%, and 25% (W/W). The culture dish was then placed in an artificial climate incubator at constant temperature and humidity (25°C), and the appropriate PEG-6000 concentrations were added to each of the treatments each two days to keep the germinating bed moist. Germination was assessed at 24-hour intervals for 8 consecutive days. The germination rate (GR), germination energy (GE), germination index (GI), and germination drought index (GDI) of different PEG-6000 concentrations for each accession were compared to determine the optimal concentration of PEG-6000, according to the method of Ku *et al.* [27] and Thabet *et al.* [28].

$$GR(\%) = n/N \times 100;$$

$$GE(\%) = m/N \times 100;$$

$$GI = \sum(DG/DT);$$

$$GDI = (1.00)nd_2 + (0.75)nd_4 + (0.50)nd_6 + (0.25)nd_8.$$

In these formulas, “n” is the number of germinated seeds on the eighth day, “m” is the number of seeds germinated on the sixth day, “N” is the total number of seeds, “DG” is the number of seeds germinated every two days, “DT” is the germination days corresponding to DG, “nd₂”, “nd₄”, “nd₆”, and “nd₈” are the germination rates of seeds on the second day, the fourth day, the sixth day, and the eighth day, respectively, and “1.00”, “0.75”, “0.50”, and “0.25” are the drought tolerance coefficients assigned by the corresponding germination days, respectively.

Phenotype identification and drought tolerance evaluation in the germination stage

Using the optimum PEG-6000 concentration (20%), 410 germinating plants from each accession of the panel were tested for traits including GR, GE, GI, and GDI. The RGR, RGE, GDTI, GSI [28], and MFV [27, 29] traits were used as evaluation indices to examine the drought tolerance of the materials during the germination stage. The calculation method was as follows.

$$RGR = GR_T / GR_C;$$

$$RGE = GE_T / GE_C;$$

$$GDTI = GDI_T / GDI_C;$$

$$GSI = GI_T / GI_C;$$

$$MFV = (\sum X_u) / M, X_u = (X - X_{Min}) / (X_{Max} - X_{Min}).$$

In these formulas, “T” is the treatment, “C” is the control (water), is the subordinate function of an indicator of the accessions, “X” is the measured value of an indicator of the accessions, “X_{Min}” and “X_{Max}” are the minimum and maximum values within the measured value of an indicator of all accessions, and “M” is the number of measured indicators.

Phenotypic data analysis

Statistical analysis of all phenotypic data across the four germination-related traits and five drought tolerance indices was conducted using the software SAS PROC GLM. (SAS Institute 1999). The broad-sense heritability (h²) [30] of each trait was estimated using the variance components. All of the above variance values can be calculated using the REML method for the SAS VARCOMP procedure.

$$H^2 = \frac{\sigma_{\text{Geno}}^2}{\sigma_{\text{Geno}}^2 + \sigma_{\varepsilon}^2}$$

Genotype identification and analysis

Genotype identification

Genomic DNA was extracted from soybean seedling leaves according to the methods used by Kisha *et al.* [31], and DNA quality was detected by 1% agarose gel electrophoresis and a spectrophotometer. A genome-wide genotyping array containing 158,327 SNPs was applied to genotype the 410 accessions using the Illumina Infinium platform according to the manufacturer's protocol (Illumina) [32, 33]. All SNP genotype data were treated with raw data normalization, clustering, and genotype calling using Illumina Genome Studio Genotyping Module (Illumina). The SNPs with a minor allele frequency (MAF) <0.05 and missing rates < 0.25 were removed to avoid problems of spurious LD and false positive associations. Finally, 117,811 high-quality SNPs were used for GWAS analysis. The SNPs were distributed relatively evenly across the 20 soybean chromosomes (Fig. S1).

Analysis of gene diversity, linkage disequilibrium, and population structure

We used PowerMarker v3.25 software to analyze MAF, PIC, heterozygosity, and gene diversity [34]; PLINK software to analyze the attenuation distance of linkage disequilibrium (LD) of the related population [35], and the R language for mapping [32]. We used half of the maximum distance for LD attenuation to identify LD blocks; this was the support interval we used for identifying significant SNPs related to a particular trait. We used multivariate analysis to classify the soybean accessions into subgroups, including cluster analysis with a neighbor-joining algorithm, model-based population structure analysis, and principal component analysis (PCA). Cluster analysis and the PCA were performed in TASSEL 5.0. When the eigenvalue is flat the subgroup structure is determined (after PC4 in our model), population structural analysis was performed using the admixture program [36].

Genome-wide association analysis

We performed a genome-wide association analysis using a mixed linear model (MLM) that accounted for kinship (K matrix) and population structure (PAC matrix) in TASSEL 5.0 [37]. The Loiselle algorithm [38, 39] was used to approximate the kinship coefficient between each pair of accessions in TASSEL 5.0. Significant SNPs were those with $-\log(P) > 4.5$ in MLM. Any significant markers positioned within a single LD block were viewed as one QTL region.

Abbreviations

GWAS: genome-wide association mapping; QTL: quantitative trait loci; SNP; RIL: recombinant inbred lines; GR: germination rate; GE: germination energy; GI: germination index; GDI, germination drought index; RGR, relative germination rate; RGE: relative germination energy; GDTI: germination drought tolerant index; GSI: germination stress index; MFV: Membership function value.

Declarations

Acknowledgements

No applicable.

Funding:

This work was supported by the National Key R & D Program for Crop Breeding (grant 2016YFD0100304), the Agricultural Science and Technology Innovation Program of the Chinese Academy of Agricultural Sciences (CAAS), and the Improvement of Soybean Abiotic Stress Tolerance to Address the Climate Change (grant PJ0121092018).

Availability of data and material

The datasets used and/or analysed during the current study available from the corresponding author on reasonable request.

Author contributions

LJQ, ZXL, and XZZ conceived and designed the experiments. XZZ and ZXL performed the experiments. XZZ, ZXL, HHL, YJZ, LLY, XSQ, HWG, and YHL analyzed data, and XZZ and ZXL wrote the manuscript. All authors read and approved the manuscript.

Competing interest

The authors have declared that no competing interests exist.

Consent for Publication

No applicable.

Ethics approval and consent to participate

This article does not contain any studies with human participants or animals performed by any of authors.

References

1. Du W, Yu D, Fu S. Detection of quantitative trait loci for yield and drought tolerance traits in soybean using a recombinant inbred line population. *Journal of Integrative Plant Biology*, 2009a, 51: 868-878.
2. Vijay R, Ravichandran V, Boominathan P. Assessment of soybean genotypes for PEG induced drought tolerance at germination and seedling level. *Madras Agricultural Journal*, 2018, 105: 1.
3. Lesk C, Rowhani P, Ramankutty N. Influence of extreme weather disasters on global crop production. *Nature*, 2016, 529: 84-87.
4. Zhao C, Liu B, Piao S, Wang X, Lobell DB, Huang Y, et al. Temperature increase reduces global yields of major crops in four independent estimates. *Proceedings of the National Academy of Sciences*, 2017, 114: 9326-9331.
5. Ning H, Yuan J, Dong Q, Li W, Xue H, Wang Y, et al. Identification of QTLs related to the vertical distribution and seed-set of pod number in soybean [*Glycine max* Merr]. *PLoS ONE*, 2018, 13: e0195830.
6. Wang X, Oh M, Sakata K, Komatsu S. Gel-free/label-free proteomic analysis of root tip of soybean over time under flooding and drought stresses. *Journal of Proteomics*, 2016, 130: 42-55.
7. Hwang S, King CA, Ray JD, Cregan PB, Chen P, Carter TE, et al. Confirmation of delayed canopy wilting QTLs from multiple soybean mapping populations. *Theoretical and Applied Genetics*, 2015, 128: 2047-2065.
8. Ye H, Roorkiwal M, Valliyodan B, Zhou L, Chen P, Varshney RK, et al. Genetic diversity of root system architecture in response to drought stress in grain legumes. *Journal of Experimental Botany*, 2018, 69: 3267-3277.
9. Condon AG, Richards RA, Rebetzks GJ, Farquhar GD. Breeding for high water-use efficiency. *Journal of Experimental Botany*, 2004, 55: 2447-2460.
10. Frederick JR, Camp CR, Bauer PJ. Drought-stress effects on branch and mainstem seed yield and yield components of determinate soybean. *Crop Science*, 2001, 41: 759-763.
11. Mishra V, Cherkauer KA. Retrospective droughts in the crop growing season: Implications to corn and soybean yield in the Midwestern United States. *Agricultural and Forest Meteorology*, 2010, 150: 1030-1045.
12. Devi JM, Sinclair TR, Chen P, Carter TE. Evaluation of elite southern maturity soybean breeding lines for drought tolerant traits. *Agronomy Journal*, 2014, 106: 1947-1954.
13. Zhang WB, Qiu PC, Jiang HW, Liu CY, Li CD, Hu GH, et al. Dissection of genetic overlap of drought and low-temperature tolerance QTLs at the germination stage using backcross introgression lines in soybean. *Molecular Biology Reports*, 2012, 39: 6087-6094.
14. Liu ZX, Li HH, Gou ZW, Zhang YJ, Wang XR, Ren HL, et al. Genome-wide association study of soybean seed germination under drought stress. *Molecular Genetics and Genomics*, 2020, 1: 1-13.
15. Du WJ, Wang M, Fu SX, Yu DY. Mapping WTLs for seed yield and drought susceptibility index in soybean (*Glycine max*) across different environments. *Journal of Genetics and Genomics*. 2009b, 36: 721-731.

16. Ebdon JS, Kopp KL. Relationships between water use efficiency, carbon isotope discrimination, and turf performance in genotypes of Kentucky bluegrass during drought. *Crop Science*, 2004, 44: 1754-1762.
17. Sloane RJ, Patterson RP, Carter JTE. Field drought tolerance of a soybean plant introduction. *Crop Science*, 1990, 30: 118-123.
18. Mian MAR, Bailey MA, Ashley DA, Wells R, Carter TE, Parrott WA, et al. Molecular markers associated with water use efficiency and leaf ash in soybean. *Crop Science*, 1996, 36: 1252-1257.
19. Hufstetler EV, Boerma HR, Carter TE, Earl HJ. Genotypic variation for three physiological traits affecting drought tolerance in soybean. *Crop Science*, 2007, 47: 25-35.
20. Abdel-Haleem H, Carter TE, Purcell LC, King CA, Ries LL, Chen P, et al. Mapping of quantitative trait loci for canopy-wilting trait in soybean (*Glycine max* Merr). *Theoretical and Applied Genetics*, 2012, 5: 837-846.
21. Oya T, Nepomuceno AL, Neumaier N, Farias JRB, Tobita S, Ito O. Drought tolerance characteristics of Brazilian soybean [*Glycine max*] cultivars: Evaluation and characterization of drought tolerance of various Brazilian soybean cultivars in the field. *Plant Production Science (Japan)*, 2004, 7: 129-137.
22. Dantas SAG, Silva FCS, Silva LJF, Silva L. Strategy for selection of soybean genotypes tolerant to drought during germination. *Genetics and Molecular Research*, 2017, 16: 4-8.
23. Dhanapal AP, Ray JD, Singh SK, Hoyos-Villegas V, Smith JR, Purcell LC, et al. Genome-wide association study (GWAS) of carbon isotope ratio ($\delta^{13}\text{C}$) in diverse soybean [*Glycine max* Merr] genotypes. *Theoretical and Applied Genetics*, 2015, 128: 73-91.
24. Kaler AS, Dhanapal AP, Ray JD, King CA, Fritschi FB, Purcell LC. Genome-wide association mapping of carbon isotope and oxygen isotope ratios in diverse soybean genotypes. *Crop Science*, 2017, 57: 3085-3100.
25. Specht JE, Chase K, Macrander M, Graef GL, Chung J, Markwell JP, et al. Soybean response to water. *Crop Science*, 2001, 41: 493-509.
26. Flintgarcia SA, Thornsberry JM, Bucklerive ES. Structure of linkage disequilibrium in plants. *Annual Review of Plant Biology*, 2003, 54: 357-374.
27. Ku YS, Au-Yeung WK, Yung YL, Li MW, Wen CQ, Liu X, et al. Drought stress and tolerance in soybean. A comprehensive survey of international soybean research-Genetics, Physiology, Agronomy and Nitrogen Relationships, 2013, pp 209-237.
28. Thabet SG, Moursi YS, Karam MA, Graner A, Alqudah AM. Genetic basis of drought tolerance during seed germination in barley. *PLoS ONE*, 2018, 13:
29. Liu C, Yang Z, Hu YG. Drought resistance of wheat alien chromosome addition lines evaluated by membership function value based on multiple traits and drought resistance index of grain yield. *Field Crops Research*, 2015, 179: 103-112.
30. Holland JB, Nyquist WE, Cervantesmartinez CT. Estimating and interpreting heritability for plant breeding: an update. *Plant Breeding Reviews*, 2010: 9-112.
31. Kisha TJ, Sneller CH, Diers BW. Relationship between genetic distance among parents and genetic variance in populations of soybean. *Crop Science*, 1997, 37: 1317-1325.
32. Mahajan A, Sim X, Ng HJ, Manning A, Rivas MA, Highland HM, et al. Identification and functional characterization of *G6PC2* coding variants influencing glycemic traits define an effector transcript at the *G6PC2-ABCB11* *PLoS Genetics*, 2015, 11: e1004876.
33. Zhao S, Jing W, Samuels DC, Sheng Q, Shyr Y, Guo Y. Strategies for processing and quality control of Illumina genotyping arrays. *Briefings in Bioinformatics*, 2018, 19: 765-775.
34. Liu KJ, Muse SV. PowerMarker: an integrated analysis environment for genetic marker analysis. *Bioinformatics*, 2005, 21: 2128-2129.
35. Li YH, Li DL, Jiao YQ, Schnable JC, Li YF, Li HH, et al. Identification of loci controlling adaptation in Chinese soya bean landraces via a combination of conventional and bioclimatic GWAS. *Plant biotechnology journal*, 2020, 18: 389-401.
36. Zeng A, Chen P, Korth K, Hancock F, Pereira A, Brye K, et al. Genome-wide association study (GWAS) of salt tolerance in worldwide soybean germplasm lines. *Molecular Breeding*, 2017, 37: 30.
37. Wen Z, Tan R, Yuan J, Bales C, Du W, Zhang S, et al. Genome-wide association mapping of quantitative resistance to sudden death syndrome in soybean. *BMC Genomics*, 2014, 15: 809.

38. Farnir F, Coppieters W, Arranz JJ, Berzi P, Cambisano N, Grisart B, et al. Extensive genome-wide linkage disequilibrium in cattle. *Genome Research*, 2000, 10: 220-227.
39. Yu J, Holland JB, McMullen MD, Buckler ES. Genetic design and statistical power of nested association mapping in maize. *Genetics*, 2008, 178: 539-551.
40. Kosturkova G, Todorova R, Sakthivelu G, Akitha-Devi MK, Giridhar P, Rajasekaran T, et al. Response of Bulgarian and Indian soybean genotypes to drought and water deficiency in field and laboratory conditions. *General and Applied Plant Physiology*, 2008, 34: 239-250.
41. Hyten DL, Pantalone VR, Sams CE, Saxton AM, Landau-Ellis D, Stefaniak TR, et al. Seed quality QTL in a prominent soybean population. *Theoretical and Applied Genetics*, 2004, 109: 552-561.
42. Kabelka EA, Diers BW, Fehr WR, LeRoy AR, Baianu IC, You T, et al. Putative alleles for increased yield from soybean plant introductions. *Crop Science*, 2004, 44: 784-791.
43. Palomeque L, Liu L, Li W, Hedges B, Cober E, Smid M, et al. Validation of mega-environment universal and specific QTL associated with seed yield and agronomic traits in soybeans. *Theoretical and Applied Genetics*, 2010, 120: 997-1003.
44. Teng W, Han Y, Du Y, Sun D, Zhang Z, Qiu L, et al. QTL analyses of seed weight during the development of soybean (*Glycine max* Merr). *Heredity*, 2009, 102: 372-380.
45. Liang H, Yu Y, Yang H, Xu L, Dong W, Du H, et al. Inheritance and QTL mapping of related root traits in soybean at the seedling stage. *Theoretical and Applied Genetics*, 2014, 127: 2127-2137.
46. Yuan J, Njiti VN, Meksem K, Iqbal MJ, Triwitayakorn K, Kassem MA, et al. Quantitative trait loci in two soybean recombinant inbred line populations segregating for yield and disease resistance. *Crop Science*, 2002, 42: 271-277.
47. Stombaugh SK, Orf JH, Jung HG, Chase K, Lark KG, Somers DA. Quantitative trait loci associated with cell wall polysaccharides in soybean seed. *Crop Science*, 2004, 44: 2101-2106.
48. Panthee DR, Pantalone VR, West DR, Saxton AM, Sams CE. Quantitative trait loci for seed protein and oil concentration, and seed size in soybean. *Crop Science*, 2005, 45: 2015-2022.
49. Guzman PS, Diers BW, Neece DJ, Martin SK, LeRoy AR, Grau CR, et al. QTL associated with yield in three backcross-derived populations of soybean. *Crop Science*, 2007, 47: 111-122.
50. Li D, Sun M, Han Y, Teng W, Li W. Identification of QTL underlying soluble pigment content in soybean stems related to resistance to soybean white mold (*Sclerotinia sclerotiorum*). *Euphytica*, 2010, 172: 49-57.
51. Liu W, Kim MY, Van K, Lee YH, Li H, Liu X, et al. QTL identification of yield-related traits and their association with flowering and maturity in soybean. *Journal of Crop Science and Biotechnology*, 2011, 14: 65-70.
52. Moongkanna J, Nakasathien S, Novitzky WP, Kwanyuen P, Sinchaisri P, Srinives P. SSR markers linking to seed traits and total oil content in soybean. *Thai Journal of Agricultural Science*, 2011, 44: 233-241.
53. Han Y, Li D, Zhu D, Li H, Li X, Teng W, et al. QTL analysis of soybean seed weight across multi-genetic backgrounds and environments. *Theoretical and Applied Genetics*, 2012, 125: 671-683.
54. Wu JJ, Xu PF, Liu LJ, Zhang S, Wang JS, Lin WG, et al. Mapping QTLs for phosphorus-deficiency tolerance in soybean at seedling stage. *International Conference on Biomedical Engineering and Biotechnology*, 2012, pp 370-378.
55. Pan Analysis of fixed SNP reveals insight of morphology differences between wild and cultivated soybeans. Chinese University of Hong Kong, 2013.
56. Mirzaei S, Batley J, Ferguson BJ, Gresshoff PM. Transcriptome profiling of the shoot and root tips of *S562L*, a soybean *GmCLAVATA1A* Atlas *Journal of Biology*, 2014, 3: 183-205.
57. Wang X, Komatsu S. Proteomic approaches to uncover the flooding and drought stress response mechanisms in soybean. *Journal of Proteomics*, 2018, 172: 201-215.
58. Guo GY, Sun R, Hou M, Guo YX, Xin DW, Jiang HW, et al. Quantitative trait locus (QTL) analysis of pod related traits in different environments in soybean. *African Journal of Biotechnology*, 2011, 10: 11848-11854.
59. Vieira AJD, Oliveira DAD, Soares TCB, Schuster I, Piovesan ND, Martínez CA, et al. Use of the QTL approach to the study of soybean trait relationships in two populations of recombinant inbred lines at the F7 and F8 generations. *Brazilian Journal of Plant Physiology*, 2006, 18: 281-290.

60. Manavalan LP, Prince SJ, Musket TA, Chaky J, Deshmukh R, Vuong TD, et al. Identification of novel QTL governing root architectural traits in an interspecific soybean population. PLoS ONE, 2015, 10: e0120490.

Supplementary Tables

Table S1 Detailed information for the 410 soybean accessions

Test ID	Name	Geographic source
T001	BRS 132	Brazil
T002	BRS 155	Brazil
T003	Embrapa 58	Brazil
T004	CII 1271	Russia
T005	PSB471	Russia
T006	PSB543	Russia
T007	ДБ 2846	Russia
T008	ДБ 2849	Russia
T009	FUN-ZHUN/g	Russia
T010	ER-HUAN-YAN	Russia
T011	Jangbaeeg	Korea
T012	Namcheon	Korea
T013	Suwon123	Korea
T014	Bayfield	Canada
T015	CA26	Canada
T016	Rhodes	United States
T017	L60-246(Clark63)	United States
T018	L65-763	United States
T019	L67-3479	United States
T020	Harosoy	United States
T021	L65-756	United States
T022	T171	United States
T023	T219H	United States
T024	L83-544	United States
T025	L83-4387	United States
T026	L81-4590	United States
T027	Peking	United States
T028	Wilkin	United States
T029	Dunn	United States
T030	Vinton 81	United States
T031	Amsoy	United States
T032	Beeson	United States
T033	Century	United States
T034	Corsoy	United States
T035	Provar	United States

T036	Williams	United States
T037	Williams79	United States
T038	Williams82	United States
T039	Franklin	United States
T040	PI196160	United States
T041	PI157440	United States
T042	Bedford	United States
T043	DorchsoyB	United States
T044	AsgrowA1939	United States
T045	Sprite87	United States
T046	A2396	United States
T047	C1640	United States
T048	CX1038-14	United States
T049	L83-4744	United States
T050	L83-570	United States
T051	L82-1858	United States
T052	L81-4420	United States
T053	HP202	United States
T054	Amcor 89	United States
T055	Conrad	United States
T056	Newton	United States
T057	GR8836	United States
T058	Kunitz	United States
T059	Delsoy4900	United States
T060	Nile	United States
T061	Pharaoh	United States
T062	Cordell	United States
T063	Epps	United States
T064	Mack	United States
T065	Walters	United States
T066	Pickett	United States
T067	Sharkey	United States
T068	Twiggs	United States
T069	Gordon	United States
T070	Thomas	United States
T071	T308	United States

T072	T309	United States
T073	L85-1467	United States
T074	L79-842	United States
T075	L87-0482	United States
T076	Carlin	United States
T077	Accomac	United States
T078	BARC—7	United States
T079	Mercury	United States
T080	PN9394	United States
T081	Probst	United States
T082	TBD	United States
T083	9234	United States
T084	L89-2435	United States
T085	L84-2157	United States
T086	L88-8153	United States
T087	MN1301	United States
T088	Kovean	United States
T089	L72-920	United States
T090	PI 468903	United States
T091	M044	United States
T092	Tousan 83	United States
T093	Tousan kei NA5	United States
T094	Tousan kei NA75	United States
T095	Sargent	United States
T096	S99-3181	United States
T097	S01-9391	United States
T098	Osage	United States
T099	LS94-3207	United States
T100	□□□□	Japan
T101	□□□□	Japan
T102	Tokachi nagaha	Japan
T103	Tsurukogane	Japan
T104	danli	Japan
T105	guandong102	Japan
T106	zhongte1	Japan
T107	Yumeyutaka	Japan

T108	AGS162	Thailand
T109	Syella	Italy
T110	Dekabig	Italy
T111	Shunyiheidou	Beijing, China
T112	Miyunlaoyelian	Beijing, China
T113	Zhongpin661	Beijing, China
T114	YZY20041515W83	Beijing, China
T115	YZY200415W90	Beijing, China
T116	AXN155	Beijing, China
T117	Zhonghuang68	Beijing, China
T118	Youhuangdou	Gansu, China
T119	Lvhuangdou	Gansu, China
T120	Tueryan	Hebei, China
T121	Baiqidawandou	Hebei, China
T122	Yangtianxiaohuangdou	Hebei, China
T123	Nanguanxiaopiqing	Hebei, China
T124	Doushidaqingdou	Hebei, China
T125	Chichenglvhuangdou	Hebei, China
T126	Datunxiaohaidou	Hebei, China
T127	Xiataizimoshidou	Hebei, China
T128	Maoyandou	Hebei, China
T129	Dongnong42A	Heilongjiang, China
T130	Dongnong42C	Heilongjiang, China
T131	Keqixiaoliheidou	Neimenggu, China
T132	Xiheheidou	Ningxia, China
T133	Loutianwaziheidou	Ningxia, China
T134	Nidinghuameidou	Ningxia, China
T135	Baipihuangdou	Shanxi, China
T136	Tianedan	Shanxi, China
T137	Tianedan	Shanxi, China
T138	Huanggandou	Shanxi, China
T139	Daheidou	Shanxi, China
T140	Chuanmanheidou	Shanxi, China
T141	Xiaheidou	Shanxi, China
T142	Hongdadou	Shanxi, China
T143	Xiaohuangdou	Shanxi, China

T144	Gingkeyuandou	Shanxi, China
T145	Huangdou2	Shanxi, China
T146	Xiaohuangdou	Shanxi, China
T147	Yuxuan13hao	Shanxi, China
T148	Bailudou	Shanxi, China
T149	Liushiribaidou	Shanxi, China
T150	Xiaobaidou2	Shanxi, China
T151	Xiaoheidou	Shanxi, China
T152	Dongshan69	Shanxi, China
T153	Lvpihuangdou	Shanxi, China
T154	Tiefeng31	Shanxi, China
T155	Zaoshuhuangdou	Shanxi, China
T156	Xiaoheidou	Shanxi, China
T157	Xiaoheidou	Shanxi, China
T158	Laoheidou	Shanxi, China
T159	Yanqihuangdou	Xinjiang, China
T160	Changjihuangdou1	Xinjiang, China
T161	Dongnong4	Heilongjiang, China
T162	Fengshou1	Heilongjiang, China
T163	Heihe1hao	Heilongjiang, China
T164	Mufeng1	Heilongjiang, China
T165	Suinong1hao	Heilongjiang, China
T166	Jingshanpu	Heilongjiang, China
T167	Tujiazi	Heilongjiang, China
T168	Baimaoshuang	Heilongjiang, China
T169	Liushitianhuanjia	Heilongjiang, China
T170	Huananxiaojindou	Heilongjiang, China
T171	Qingdou	Heilongjiang, China
T172	Lvrangheidou	Heilongjiang, China
T173	Qinganheidou	Heilongjiang, China
T174	Fangzhengmoshidou	Heilongjiang, China
T175	Nenfeng11hao	Heilongjiang, China
T176	Hefeng24hao	Heilongjiang, China
T177	Hefeng25hao	Heilongjiang, China
T178	Dongnong36hao	Heilongjiang, China
T179	Heihexiaohuangdou	Heilongjiang, China

T180	Longquandadou(heqi)	Heilongjiang, China
T181	Xiaolimoshidou	Heilongjiang, China
T182	Suinong14hao	Heilongjiang, China
T183	Hedou2(MN413)	Heilongjiang, China
T184	Heihe38	Heilongjiang, China
T185	Hefeng52	Heilongjiang, China
T186	Heinong47	Heilongjiang, China
T187	HLT2	Heilongjiang, China
T188	Ha123510	Heilongjiang, China
T189	Jilin3	Jilin, China
T190	Xiaojinhuang1	Jilin, China
T191	Fengdihuang	Jilin, China
T192	Jinyuan1	Jilin, China
T193	Ha1	Jilin, China
T194	Huichundou	Jilin, China
T195	Jiaohezihua1	Jilin, China
T196	Lanqi	Jilin, China
T197	Changchunmancangjin	Jilin, China
T198	Niumaohuang	Jilin, China
T199	Baodigao	Jilin, China
T200	Chasedou	Jilin, China
T201	Heimoshidou	Jilin, China
T202	Heimodou	Jilin, China
T203	Zihua2hao	Jilin, China
T204	Fuyuduludou	Jilin, China
T205	Jiutaibaodigao	Jilin, China
T206	Huaidebaihuadali	Jilin, China
T207	Helongyoutai	Jilin, China
T208	Tonghuapingdingxiang	Jilin, China
T209	Baichengmoshidou	Jilin, China
T210	Jinshanchamoshidou	Jilin, China
T211	Jilinchalihua	Jilin, China
T212	Huangdali	Jilin, China
T213	Hefeng37hao	Jilin, China
T214	Dongsheng1	Jilin, China
T215	Jilin30	Jilin, China

T216	Jiyu67	Jilin, China
T217	Jilinxiaolidou	Jilin, China
T218	Jiyu86	Jilin, China
T219	Jiyu109	Jilin, China
T220	Tiefeng18	Liaoning, China
T221	Jindou33	Liaoning, China
T222	Jinzhou41	Liaoning, China
T223	Dabaimei	Liaoning, China
T224	Tianedan	Liaoning, China
T225	Daheiqi	Liaoning, China
T226	Heiqi	Liaoning, China
T227	Dadou2	Liaoning, China
T228	Tiejiajinpings	Liaoning, China
T229	Huangqi	Liaoning, China
T230	Xiaobaiqi	Liaoning, China
T231	Xiaohuangdou	Liaoning, China
T232	Niumaohuang	Liaoning, China
T233	Qingpipingdingxiang	Liaoning, China
T234	Baitiejia	Liaoning, China
T235	Baiheidou	Liaoning, China
T236	Daliheidou	Liaoning, China
T237	Liushitianhuancang	Liaoning, China
T238	Yushidou	Liaoning, China
T239	Jiyu72	Liaoning, China
T240	Liaodou11	Liaoning, China
T241	Liaodou16	Liaoning, China
T242	Dongnong50	Liaoning, China
T243	Tiefeng29	Liaoning, China
T244	Liaodou32	Liaoning, China
T245	Liao08012	Liaoning, China
T246	Liao08Q104	Liaoning, China
T247	Liao08024	Liaoning, China
T248	Liao10Q015	Liaoning, China
T249	Chi382	Neimenggu, China
T250	Jindou36	Neimenggu, China
T251	Suiningpingdinghuang	Jiangsu, China

T252	Pixianhongmaoyou	Jiangsu, China
T253	Pixiandazihuacao	Jiangsu, China
T254	Pixiansilicao	Jiangsu, China
T255	Huaiyangchundou	Jiangsu, China
T256	Muyangchunheidoubing	Jiangsu, China
T257	Pudou206	Jiangsu, China
T258	Hualvhuangdou	Gansu, China
T259	Diliuhuangdou2	Hebei, China
T260	Sijiaoqihuangdou	Hebei, China
T261	Bendidahuangdou	Hebei, China
T262	Heidou	Hebei, China
T263	Huaheihu	Hebei, China
T264	Jidou7hao	Hebei, China
T265	Qingdou	Hebei, China
T266	Miyangxiaozihuang	Henan, China
T267	Xichuanjiwohuang	Henan, China
T268	Miyangniumaohuang	Henan, China
T269	Zhechengxiaohongdou	Henan, China
T270	Boaihongpizaojiaozi	Henan, China
T271	Xinyangyangyandou	Henan, China
T272	Zheng8516	Henan, China
T273	Zheng84240B1	Henan, China
T274	Shanning7	Henan, China
T275	Pixianlayanghuang	Jiangsu, China
T276	Tongshanqingdadou	Jiangsu, China
T277	Guanyunhaibaihua	Jiangsu, China
T278	Sidou2hao	Jiangsu, China
T279	Shengli3hao	Shandong, China
T280	Siliyuan	Shandong, China
T281	Pingdinghuangdou	Shandong, China
T282	Dabaipi	Shandong, China
T283	Dahuangdou	Shandong, China
T284	Datianedan	Shandong, China
T285	Xiaomidou	Shandong, China
T286	Lvcaodou	Shandong, China
T287	Douliheidou	Shandong, China

T288	Pingdinghei	Shandong, China
T289	Chadou	Shandong, China
T290	Maodou	Shandong, China
T291	Qisiwa	Shandong, China
T292	Gaozuoxuan1hao	Shandong, China
T293	Jilin36	Shandong, China
T294	Mengdou14	Shandong, China
T295	Niumaohuang	Shanxi, China
T296	Huichaxiaohuangdou	Shanxi, China
T297	Niupihuangdou	Shanxi, China
T298	Laoshupi	Shanxi, China
T299	Jianghuangdou	Shanxi, China
T300	Baomuji	Shanxi, China
T301	ZDD04918	Anhui, China
T302	ZDD04959	Anhui, China
T303	WeiJ127	Anhui, China
T304	Jindou21	Anhui, China
T305	Huaidou4	Anhui, China
T306	Doushanbaimadou	Fujian, China
T307	Dalihuang	Fujian, China
T308	Daqingren	Fujian, China
T309	Xiamentengzidou	Fujian, China
T310	Tonganzihongdou	Fujian, China
T311	Pudou451	Fujian, China
T312	Quanbian11	Fujian, China
T313	Zhaoanqiudadou	Fujian, China
T314	Shaxianqingdou	Fujian, China
T315	Shaxianwudou	Fujian, China
T316	Baiqiu1hao	Fujian, China
T317	Dabaimaodou	Guangdong, China
T318	Longchuanhuangniumao	Guangdong, China
T319	Lianjiangpohuangdou	Guangdong, China
T320	Qingyuandaqingdou	Guangdong, China
T321	Yingdehedou	Guangdong, China
T322	Dahuangdou2	Guangdong, China
T323	Madaiqingdou2	Guangdong, China

T324	Doupingqingdou	Guangdong, China
T325	Madaiheidou3	Guangdong, China
T326	Baizhidou	Guangxi, China
T327	Dawudou	Guangxi, China
T328	Mashanrenfenghuangdou	Guangxi, China
T329	Daimaodou	Guizhou, China
T330	Xihuangdou8	Guizhou, China
T331	Xihuangdou9	Guizhou, China
T332	Doujizaodou2	Guizhou, China
T333	Zaohuangdou	Guizhou, China
T334	Dahuangdou1	Guizhou, China
T335	Zaojiaodou	Guizhou, China
T336	Zadou6	Guizhou, China
T337	Qiyuehuang1	Guizhou, China
T338	Heikewudou	Hainan, China
T339	Jinghuang35yi	Hubei, China
T340	Daimidou	Hubei, China
T341	Zhongdou24	Hubei, China
T342	8216	Hubei, China
T343	Chihuangdou2	Hubei, China
T344	Shuguanghuangdou	Hubei, China
T345	Chahuangdaidou1	Hubei, China
T346	Shanzibaihuangdou	Hubei, China
T347	Chihuangdou1	Hubei, China
T348	Huameidou	Hubei, China
T349	Xiaokehuangdou	Hubei, China
T350	Honghuliuyuebao	Hubei, China
T351	Nidou	Hubei, China
T352	8470	Hubei, China
T353	Huangmaodou	Hunan, China
T354	Hongzhudou	Hunan, China
T355	Changshanidou	Hunan, China
T356	Aishengnidou1	Hunan, China
T357	Yizhangliuyuehuang	Hunan, China
T358	Wujiangwuyueniumaohuang	Jiangsu, China
T359	Yizhengdalihuangdou	Jiangsu, China

T360	Taixingheidou	Jiangsu, China
T361	Taixingaijiaohong	Jiangsu, China
T362	77-391-1	Jiangsu, China
T363	Shaxindou	Jiangxi, China
T364	Ruijinqingpidou	Jiangxi, China
T365	Dahuangzhu	Jiangxi, China
T366	Xinyudaliqing	Jiangxi, China
T367	Shangraobayuebai	Jiangxi, China
T368	Yantianqingpidou	Jiangxi, China
T369	Hengfengwudou	Jiangxi, China
T370	Wuyuehuang	Jiangxi, China
T371	Duchangwudou	Jiangxi, China
T372	Fengchengzaowudou	Jiangxi, China
T373	Dahuadou	Sichuan, China
T374	Wuyanwo	Sichuan, China
T375	Shiyuehuang	Sichuan, China
T376	Zengjialvhuangdou	Sichuan, China
T377	Jiangehualinjiwodou	Sichuan, China
T378	Qionglaihuangmaozi	Sichuan, China
T379	Qionglaiyoujiangheidou	Sichuan, China
T380	Hanyuanbalixiaoheidou	Sichuan, China
T381	Douhuangdou1	Sichuan, China
T382	Liuyuebao2	Sichuan, China
T383	Zaohuangdou4	Sichuan, China
T384	Baimaozaodouzi	Sichuan, China
T385	Touxinlv	Sichuan, China
T386	Lvdouzi	Sichuan, China
T387	Lvlanzi	Sichuan, China
T388	Xiaobaimao	Sichuan, China
T389	Bazhongtiankandou2	Sichuan, China
T390	Quxianbayuehuang	Sichuan, China
T391	Pixianxiaohuangdou	Sichuan, China
T392	Zizhongliuyuezao	Sichuan, China
T393	Jianweiquanshuidou	Sichuan, China
T394	Changshoushiyuehuang	Sichuan, China
T395	Suiningfengtaijiangsedou	Sichuan, China

T396	Shifangluosidou	Sichuan, China
T397	8307-8-1	Sichuan, China
T398	Gongdou7hao	Sichuan, China
T399	Liuyuehuang	Sichuan, China
T400	Pengshanhuangkezi3	Sichuan, China
T401	Xicangdadou12	Xicang, China
T402	Xuanza	Yunnan, China
T403	Huangdou	Yunnan, China
T404	Yangyandou	Yunnan, China
T405	Songzidou	Yunnan, China
T406	Malanzaochadou	Yunnan, China
T407	Zaoshumaopengqing	Zhejiang, China
T408	Cudou	Zhejiang, China
T409	Fudou9765	Zhejiang, China
T410	Quxian3	Zhejiang, China

Table S2 Detailed information for the six soybean accessions used for PEG screening

Test ID	Name	Geographic source
D001	Muyangchunheidoubing	Jiangsu, China
D002	Xiaomidou	Shandong, China
D003	Douliheidou	Shandong, China
D004	Jindou21	Anhui, China
D005	Qisiwa	Shandong, China
D006	Sidou2hao	Jiangsu, China

Table S3 Descriptive statistics of four germination-related traits under 0% PEG (C) and 20% PEG (D) conditions for the 410 soybean accessions

Traits	Treat	Range	Mean	SD	CV
GR	D	0.00~100.00	15.07	18.69	124.01
	C	60.00~100.00	96.92	5.95	6.14
GE	D	0.00~96.67	13.58	17.37	127.92
	C	55.00~100.00	96.33	6.78	7.04
GDI	D	0.00~135.83	17.49	23.40	133.80
	C	82.50~250.00	189.20	35.12	18.56
GI	D	0.00~4.69	0.68	0.86	127.79
	C	2.61~14.92	7.50	1.90	25.36

GR, germination rate; GE, germination energy; GDI, germination drought index; GI, germination index; *SD* standard deviation; *CV* coefficient of variation.

Table S4 Genetic parameters revealed by the analysis of 117,811 polymorphic SNP markers in the 410 soybean accessions

	Minimum	Maximum	Mean
Minor allele frequency	0	0.5030	0.2228
Gene diversity	0	0.5061	0.3043
Heterozygosity	0	0.4070	0.0237
PIC	0	0.3843	0.2458

Table S5 SNPs positioned near genes and functional annotated information for these genes

Marker	Chr.	Position	Site	Gene	Homologous gene in Arabidopsis	Functional annotation
Gm01_35877607	1	35877607	Intergenic region	Glyma.01g106000; Glyma.01g106100	AT3G09270	Glutathione S-transferase U1-related
Gm01_38948188	1	38948188	Intergenic region	Glyma.01g113500; Glyma.01g113600	AT2G24670	Domain of unknown function
Gm01_47042336	1	47042336	Intergenic region	Glyma.01g141000; Glyma.01g141100	AT5G12060	Genomic DNA, chromosome 3, P1 clone: MDJ14-related
Gm01_48619013	1	48619013	nonsynonymous	Glyma.01g149300	AT1G31850	Methyltransferase PMT21-related
Gm02_6357585	2	6357585	synonymous	Glyma.02g072600	AT3G03860	5'-adenylylsulfate reductase-like 5-related
Gm03_39037	3	39037	Intergenic region	Glyma.03g000200; Glyma.03G000300	AT2G31820	Ankyrin repeats-containing protein
Gm04_4484515	4	4484515	Intergenic region	Glyma.04g055500; Glyma.04g055600	AT1G76880	NA
Gm04_50945875	4	50945875	nonsynonymous	Glyma.04g241400; Glyma.03G000300	AT1G21460	Bidirectional sugar transporter sweet1
Gm05_38540838	5	38540838	Intergenic region	Glyma.05g201700; Glyma.05G201800	AT5G50915	Transcription factor BLLH137
Gm06_9791913	6	9791913	nonsynonymous	Glyma.06g120400	AT1G55200	Interleukin-1 receptor-associated kinase 1 (IRAK1)
Gm07_24735482	7	24735482	Intergenic region	Glyma.07G165100; Glyma.07g165200	AT2G43630	Glycine-rich protein
Gm08_1438457	8	1438457	intronic	Glyma.08g017800	AT1G63940	Monodehydroascorbate reductase, chloroplastic
Gm08_4052111	8	4052111	nonsynonymous	Glyma.08g052100	AT3G18050	Genomic DNA, Chromosome 3, P1 Clone: MRC8
Gm08_7972856	8	7972856	synonymous	Glyma.08g103900	AT1G67980	Flavonoid 3',5'-methyltransferase
Gm09_11414508	9	11414508	Intergenic region	Glyma.09g087500; Glyma.09g087600	AT1G09040	Atrophin-related // subfamily not named
Gm09_18023730	9	18023730	Intergenic region	Glyma.09g099300; Glyma.09g099400	AT3G42170	Finger-related // subfamily not named
Gm11_30280479	11	30280479	intronic	Glyma.11g210400	AT2G18950	homogentisate phytyltransferase / homogentisate geranylgeranyltransferase
Gm13_35517964	13	35517964	synonymous	Glyma.13g246400	NA	NA
Gm14_46603856	14	46603856	Intergenic region	Glyma.14g200900; Glyma.14G201100	AT4G35160	O-methyltransferase// subfamily not named
Gm15_11950665	15	11950665	UTR3	Glyma.15g145200	AT4G16110	Response regulator of two-component system // subfamily not named
Gm15_47429024	15	47429024	Intergenic region	Glyma.15g248700; Glyma.15G248800	AT2G01050	Domain of unknown function (DUF4283)
Gm19_49449499	19	49449499	UTR3	Glyma.19g248400	AT3G04490	Exportin-4

Gm20_4618170	20	4618170	Intergenic region	Glyma.20g033800; Glyma.20G033900	AT5G15290	Casparian strip membrane protein 5
Gm20_13921498	20	13921498	Intergenic region	Glyma.20g056300; Glyma.20G056400	NA	gag-polypeptide of LTR copa-type (UBN2)
Gm20_34956219	20	34956219	Intergenic region	Glyma.20G106800; Glyma.20g106900	AT1G34360	Translation initiation factor IF-3 // subfamily not named
Gm20_36902659	20	36902659	UTR3	Glyma.20g126800	AT4G35220	Arylformamidase / Kynurenine formamidase

Figures

Figure 1

Frequency distribution of four germination-related traits GR (A), GE (B), GDI (C) and GI (D) for the six soybean accessions. (GR, germination rate; GE, germination energy; GDI, germination drought index; GI, germination index; 0%, 10%, 15%, 20%, 25% represent different PEG treatment concentration; *, significantly different from the control. D001-006 represent six soybean accessions (Table S2))

Figure 2

Distribution of five drought tolerance indices (RGR, relative germination rate; RGE, relative germination energy; GDTI, germination drought tolerant index; GSI, germination stress index; MFV, Membership function value)

Figure 3

Distribution of the genetic diversity of 136,337 SNPs across 410 accessions. A is for minor allele frequency; B is for gene diversity; C is for heterozygosity; and D is for polymorphic information content (PIC).

Figure 4

Linkage disequilibrium (LD) decay across soybean genome.

Figure 5

Genetic structure and relatedness of the 410 soybean accessions. (A) Neighbor-joining tree constructed using SNP data, Foreign soybean germplasm expressed as green; soybean accessions from north are shown in red; from the Huanghuaihai valley region are shown in blue; and from the south valley region are shown in orange. (B) Principal component analysis for the entire panel of soybean accessions; (C) PCA eigenvalue performed by GAPIT using the pruned set of 200K SNP. As presented the total variance explained by each principal component (PC) decreased from PC1 to PC4 and, after PC4, the variance explained by each further PC remained low and stable; (D) Clustering for PCA =4 for the entire panel of soybean accessions. Each individual is represented by a vertical bar, partitioned into colored segments with the length of each segment representing the proportion of the individual's genome from groups when PCA=4.

Figure 6

Manhattan and Quantile-quantile plot from the PCA+K model across the five drought tolerance indices RGR (A), RGE (B), GDTI (C), GSI (D), and MFV (E).

Supplementary Files

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

- [FigS1.png](#)
- [FigS2.png](#)
- [Supplementarydata.xlsx](#)