Indian J. Genet., 78(4): 460-469 (2018)
DOI: 10.31742/IJGPB.78.4.8

Genome-wide association studies (GWAS) reveal candidate genes for plant height and number of primary branches in soybean [*Glycine max* (L.) Merrill]

Jyotilekha Borah[2]*, Avinash Singode[1], Akshay Talukdar, Raju R. Yadav and Ramendra Nath Sarma[2]

Division of Genetics, ICAR-Indian Agricultural Research Institute, New Delhi 110 012; [1]ICAR-Indian Institute of Millets Research, Hyderabad; [2]Department of Plant Breeding and Genetics, Assam Agricultural University, Jorhat

(Received: July 2018; Revised: October 2018; Accepted: November 2018)

Abstract

Plant height and the number primary branches are important traits having bearing on yield. Understanding genetic control of these traits would help in the development of improved soybean genotypes suitable for water-stress and normal conditions. In this study, 63 diverse soybean accessions were genotyped with 2,84,923 single nucleotide polymorphism (SNP) markers, and phenotyping was done under optimal and moisture-stress conditions in two locations *viz*., New Delhi and Dharwad. Significant variation was observed among the genotypes for the traits in both the locations. Interaction of the genotypes with locations and the situations was also found to be significant. The population structure and genetic diversity analysis indicated existence of four sub-sets of population among the genotypes. GWAS conducted through Mixed Linear Model identified several Quantitative Trait Nucleotides(QTNs) associated with plant height and branches/plant. Genes associated with the identified QTNs and their functions have also been identified. Two novel QTNs for number of branches/plant were mapped on Chromosomes 3 and 16. The identified QTNs and the genes would facilitate designing of gene-specific markers which would pave the way for development of soybean genotypes with optimum height, branches and yield.

Key words: Water stress, QTL, single nucleotide polymorphism, association studies, soybean

Introduction

Soybean plant architecture is primarily determined by plant height and the primary branches (Huyghe, 1998). It also consists of metameric units *viz*., internode, trifoliate leaf and the associated reproductive branch born at the internodesthat grows continuously. Plant architecture and leaf arrangement influence light distribution on the plant. Biomass production and interception of solar radiation is highly correlated during both vegetative and reproductive phase in soybean (de Souza et al. 2009). Height of the plants and number of primary branches/plant haspositive correlation with grain yield (Aditya et al. 2013). Strong and positive correlation between plant height and seed yield was also reported by Mansur et al. (1996) and Cicek et al. (2006). Direct and positive association exists between number of primary branches and seed yield. However, for consistency of the results, trials on plant architecture need to be repeated across locations (Zhang et al. 2015). Advent of molecular markers has led to the identification of QTLs for various agronomic traits. Palomeque et al. (2010) noted that the marker Satt162 linked to seed yield QTL is associated with plant height and lodging. Similarly, association of markers and QTL has been reported for various traits including plant height (Orf et al. 1999; Hoeck et al. 2003; Zhang et al. 2004). Ning et al. (2016) suggested that the genomic loci should be studied in detail to elucidate underlying genetic mechanism in yield and yield related traits. For effective selection, the QTL needs to be consistent and strong in effect (Palomeque et al. 2009). Limitations of the bi-parental mapping of QTL (Li et al. 2008) were circumvented to a great extent by the genome-wide association studies (GWAS) (Gupta et al. 2005), a robust approach for fine mapping with great precision. It has been used in a number of crops including soybean (Hu et al. 2014; Li et al. 2016; Song et al. 2017). Precision of GWAS however,

*Corresponding author's e-mail: jyoti.lekha123@gmail.com
Published by the Indian Society of Genetics & Plant Breeding, A-Block, F2, First Floor, NASC Complex, IARI P.O., Pusa Campus, New Delhi 110 012; Online management by www.isgpb.org; indianjournals.com

depends on the availability and density of high-throughput makers, which can be achieved by using Genotyping-by-Sequencing (GBS) (Elshire et al. 2011). GBS platform is simple, cost effective and provides the desirable density expected for GWAS(He et al.2014;Kim et al. 2016; Shaheen et al. 2016). Kim et al.(2018) characterised 22 soybean mutants using GBS.The present study was undertaken to study genetic architecture of plant height and number of primary branches under water stressed condition through GWAS to identify candidate SNPs and genes for use in soybean improvement program.

Materials and methods

Plant materials

The experimental material for the present investigation comprised of 63 soybean genotypes (Table 1) obtained from the Indian Agricultural Research Institute, Pusa Campus, New Delhi. These genotypes include soybean accessions collected from USA, Taiwan and different states of India *viz*., Maharashtra, Punjab, Madhya Pradesh, Delhi, Gujrat, Himachal Pradesh, Jharkhand and Karnatak (Dharwad).

Soybean genomic DNA isolation, quantification and genotyping

Tender leaf samples from each genotype were harvested for DNA isolation. The genomic DNA was isolated using Qiagen® DNeasy Plant Mini Kit following their prescribed protocol. Quality of the DNA in the sample was assessed by digesting with *HindIII* restriction enzyme and 0.8% Agarose gel electrophoresis. DNA quantification was done using spectrophotometer. The DNAconcentration was maintained as 100ng/µl. A 50µl of the DNA samples were shipped to Institute of Genetic Diversity (IGD), Cornell University, Ithica NY, USA for sequencing. The DNA was digested with *ApeK1*. The adapters were ligated to the fragmented samples and barcoded. The sequencing libraries were made and sequencing was carried out on a Genome Analyzer II (Illumina, Inc.). The sequencing data was used for downstream analysis and SNP calling. The SNP calling was done using GBS pipeline employed in TASSEL software (Bradbury et al. 2007; Buckler et al. 2009). The raw data was filtered for quality and aligned to the reference genome. The data was separated chromosome wise into 20 workable files. All the sites with less than >10% missings were filtered out. Sites with <5% heterozygosity were retained. Similarly, the sites with lower proportion of minor alleles and were in bi-allelic state were retained. With this filtering parameter, a total of 284,923 SNPs were left for analysis.

Location and environment and phenotyping

Performance of 64 genotypes was evaluated under two water regimes (well-watered and water-limited drought block) with two replications at two locations *viz*., IARI, Pusa campus, New Delhi (Latitude and Longitude is 28.6353N and 77.225E, respectively) and Dharwad (Latitude and Longitude is 15.4923N and 74.9832E, respectively). In New Delhi, the trial was conducted during rainy season of 2016 in the research farm of Division of Genetics, IARI New Delhi. The average temperature at Delhi in the sowing time was 34^{o}C and humidity was 68% and the average rainfall in Delhi during the growing period was 65.5 mm with highest precipitation in the month of August. In the same year, trial was conducted at IARI-RRS Dharwad, Karnataka in the month of June. The average temperature and humidity at Dharwad during the sowing time were 25^{o}C and 82% respectively and the average rainfall during the cropping season was 72.5 mm with highest rainfall in the month of July (https://www.weather-ind.com/). The trials were conducted in Randomized Block Design with two replications (Singh and Choudhary, 1985). The inter-row distance was 75 cm and plant to plant distance was 10 cm. The plants were grown normally and stress was imposed by withholding irrigation 50 days after sowing till harvesting. Data for plant height and number of primary branches were recorded following standard procedure.

Statistical analyses

Analysis of variance (ANOVA) was performed to estimate the genetic variance at each location. Linear Mixed effect Model (LM) was performed to estimate the components of phenotypic variance. The genotypic effect, environmental effect and interaction effect was estimated using LM model as given below. The analysis was performed using the package '*lme4*' in R (Bates et al. 2015).

$$Y_{ijk} = \mu + g_i + l_j + gl_{ij} + b_{kj} + \varepsilon_{ijk}$$

where, μ is the total mean; g_i is the genetic effect of the *i*th genotype; l_j is the effect of the *j*th environment; $(gl)_{ij}$ is the interaction effect between the *i*th genotype and the *j*th environment (G × E); $b_{k(j)}$ is the random block effect within the *j*th environment; and e_{ijk} is a random error following N (0, σ_e^2). The Heritability and correlation coefficients of the drought related traits were calculated as per Singh and Choudhary, (1985).

Table 1. List of soybean genotypes used in the study

S.No.	Genotype	Pedigree	Place of collection	S.No.	Genotype	Pedigree	Place of collection
1	GP1	AMSS34	Maharashtra	35	GP71	G2133	Taiwan
2	GP2	AVRDC5	Taiwan	36	GP72	JS81-303	MP
4	GP3	AVRDC516	Taiwan	37	GP73	G2133	Taiwan
5	GP6	DS61	Delhi	38	GP78	G2344	Taiwan
6	GP7	-	Delhi	39	GP79	-	Delhi
7	GP8	DS178	Delhi	40	GP81	-	Delhi
8	GP10	DS241	Delhi	41	GP82	G2608	Taiwan
9	GP11	DS371	Delhi	42	GP83	-	Delhi
10	GP14	DS172	Delhi	43	GP85	G2650	Taiwan
11	GP15	DSb2	Dharwad	44	GP86	G2651	Pantnagar
12	GP16	EC112827	Taiwan	45	GP87	SL528	Punjab
13	GP18	DS-76-1-20-1	Delhi	46	GP88	G2658	Taiwan
14	GP20	DS-76-37-2	Delhi	47	GP91	G3023	Taiwan
15	GP26	DS-9-3	Delhi	48	GP95	G2132	Taiwan
16	GP32	EC112827	USA	49	GP99	GUJ-SOY-1	Gujarat
17	GP35	EC113398-B	Taiwan	50	GP100	HIMSO1574	Himachal Pradesh
18	GP38	-	Delhi	51	GP102	IC101449	Himachal Pradesh
19	GP39	-	Delhi	52	GP103	IC76151-W	Himachal Pradesh
20	GP39	-	Delhi	53	GP104	EC1021	USA
21	GP40	-	Delhi	54	GP105	IC76151-W	Himachal Pradesh
22	GP41	-	Punjab	55	GP106	IC10755	Himachal Pradesh
23	GP42	EC390981	Taiwan	56	GP107	IC141446	Himachal Pradesh
24	GP49	EC13969	USA	57	GP108	-	Delhi
25	GP55	EC1021	USA	58	GP110	J231	Jharkhand
26	GP56	-	Delhi	59	GP111	-	Punjab
27	GP57	EC14436	USA	60	GP112	G390	Taiwan
28	GP59	-	Delhi	61	GP113		Delhi
29	GP60	EC30221	USA	62	GP114	JS81-607	Madhya Pradesh
30	GP61	EC34141	Delhi	63	GP115	JS93-06	Madhya Pradesh
31	GP62	EC36961	USA				
32	GP64	EC97351	USA				
33	GP65	EC105790	USA				
34	GP67	-	Delhi				
35	GP71	G2133	Taiwan				

Population structure analysis

Population structure analysis was performed using the software STRUCTURE v.2.3.3 (Pritchard et al. 2000). The parameters of the software were set to default, specifying K value 1 to 10, with each K repeated for 5 times. The results obtained from the STRUCTURE analysis were subjected to Evanno "Delta-K" method (Evanno et al. 2005) in the STRUCTURE Harvester software (Earl and vonHoldt, 2012) to establish an optimum value of K. Once the optimum K valuewas

established, the final run was made with initial burning period of 100,000 and MCMC iteration to 1000,000 specifying K = 1 to 10 to generate a STRUCTURE 'Q' matrix. Software TASSEL 5.2.3 was used for analysing the Linkage Disequilibrium (LD) value (Bradbury et al. 2007; Buckler et al. 2009). Simple matching dissimilarity index was used to calculate the genetic distance and unweighted neighbour-joining method was used to construct the unrooted tree in the software DARWin v.6 (Perrier and Jaccqemoud-Collet, 2014).

Genome-wide associations model

Genome wide SNP maker and trait association was performed using Mixed Linear Model including PCA and Kinship matrix (K) as covariates in the analysis. The analysis was performed using the standalone software TASSEL5. Statistical model as described by Henderson (1975) was used for the analysis. TASSEL provides functions to estimate PCA and K from a set of random markers covering the whole genome. To visualize the significant SNP marker loci, SNP marker sites were plotted against the respective *p* value using the Manhattan plot option in TASSEL. A logarithm of odds (LOD) value higher than 3 was used as threshold P-value for both marker-trait associations as per (Hwang et al. 2014) and only the peak SNP was considered and used to estimate the phenotypic variance. Functional annotation of the genes was performed using the genomic resources available in SoyBase (www.soybase.org).

Result and discussion

Variance, heritability, genetic advance and correlation analysis

The analysis of variance under non-stress and stress conditions (Table 2) inNew Delhi and Dharwad indicated that significant variation exists among the genotypes for plant height and number of primary branches. Location effect and Genotype x Location interaction was also found significant for both the traits. It was also observed that the mean plant height and number of primary branches/plant reduced under drought condition in both the locations (Fig. 1). Under water stress situation, the height of the plants in New Delhi and Dhardwad ranged between 33.83- 103.33cm and 20.25-106.1cm, respectively. Similarly, under well-watered condition, the plant height in New Delhi and Dhardwad ranged between 33.5-128.17cm and 20.25-111.1cm, respectively. Thus, height of the plants got reduced under water stress condition in both the locations. The high temperature prevailing during the period of growth also might have affected growth of the plants. Temperature beyond 35^0C reported to damage the photosystem-II in the thylakoid membranes of chloroplast, which reduces availability of the photo-assimilates needed for good growth (Hartman et al. 2012; Hemantaranjan et al. 2014).

The broad sense heritability estimates for plant height and number of primary branches were high in both the locations and conditions (Table 3). This suggest that additive gene actions are predominating expression of these traits in spite of the environmental effects. Thus, improvement of these traits would be possible through selection procedure aiming to exploit the additive gene effects, e.g. mass selection, progeny selection, etc. (Archana et al. 1999). Further, genotypic correlation coefficient between the two traits appeared to be positive. Thus, selection of one trait would facilitate improvement of the other.

Table 2. Pooled analysis of variance for yield and yield related traits

	Df	YPP	DF	DM	PPP	SPP	PH	HSW	PB
Rep	1	9.1	17.5	5	84	84	159	48.26**	207.21**
GP	62	48.2*	143.4**	54**	681**	681**	1340**	12.09**	6.21**
Env	1	25.8	1386.7**	284**	682 .	682**	41	133.16**	184.26**
LOC	1	2659.6*	142.5*	9395*	48129*	48129*	3346*	83.87**	207.64*
GP:LOC	62	40.6*	10.1	39*	875*	875*	643*	8.97*	1.95
GP:Env	62	22.5**	58.8**	38**	556**	556**	654**	9.01**	6.17**
Env:LOC	1	111.7**	340.1**	19	37	37	1047*	2.36	99.66**
Residuals	313	7.9	9.5	15	216	216	220	3.36	3.11

*, **Significant at P < 0.05 and P < 0.01, respectively
YPP: Yield per plant; DF: Days to 50% flowering; DM: Days to maturity; PPP: No. of pods per plant
SPP: No. of seeds per pod; PH: Plant Height; HSW: 100-seed weight; PB: No. of Primary Branches per plant;
GP: Genotype; Env: Environment; LOC: Location; Rep: Replication; Df: Degree of freedom

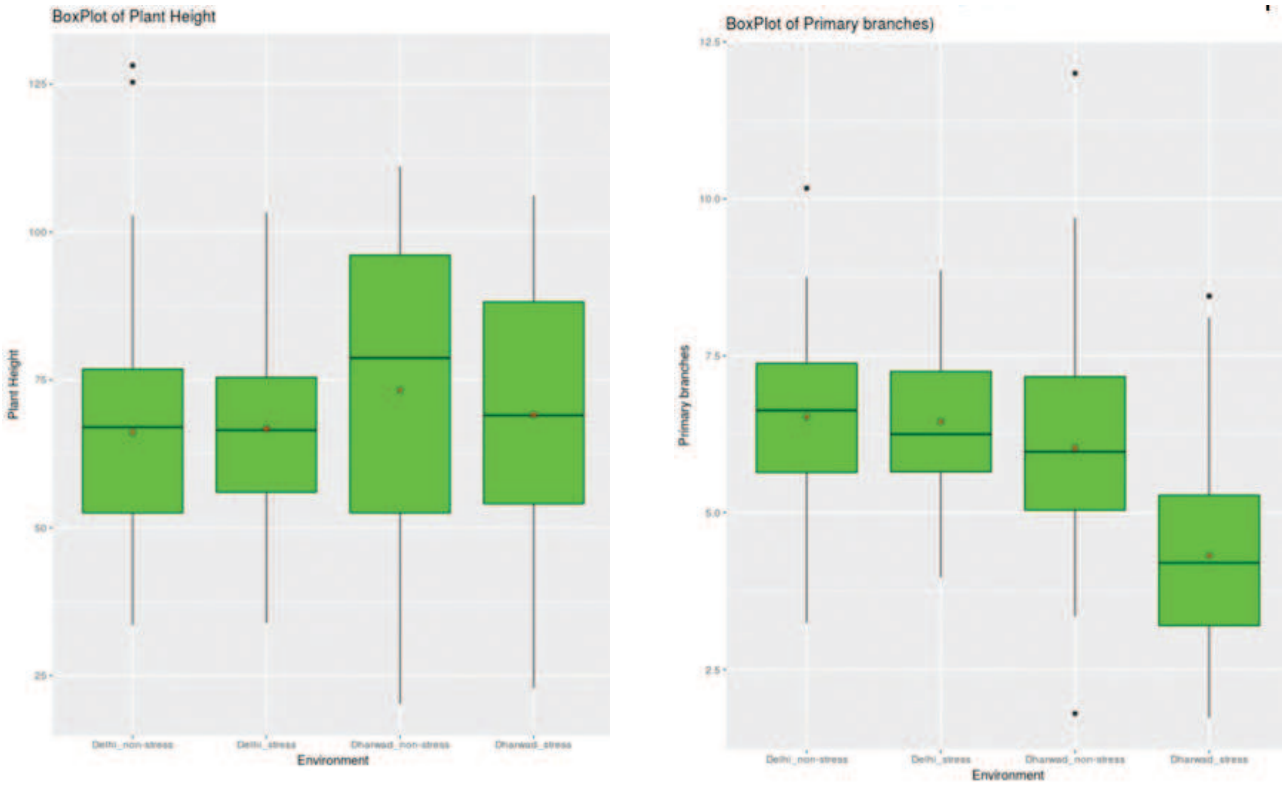

Fig. 1. Box Plot for the performances of plant height and number of primary branches

Population structure and linkage disequilibrium

The genotypes used in the present investigation of four sub-populations (Fig. 2), however, there were inter-mixing of genomes as indicated in the Fig. 3. This might have happened either through breeding involving diverse genotypes or through natural process of gene flow. The Q matrix indicated existence of high genetic diversity within sub-populations (Fig. 3). The distance based cluster analysis using Neighbour Joining method also grouped the genotypes in to four clusters (Fig. 4). Inclusion of genotype from diverse origin might have contributed towards enhanced diversity among the tested genotypes. Selection over time might have facilitated development of the sub-populations. Differences in allele frequency between different populations arising out of mutation, migration, genetic drift, local adaptation or geographical isolation, etc. leads to population structure (Knowler et al. 1988; Hwang et al. 2014). The diverse genotypes representing separate sub-population would be suitable for use in breeding program.

Pair wise LD between markers varied in the 20 chromosomes as revealed by r^2 value. LD value as high as 1 has been observed in some of the marker

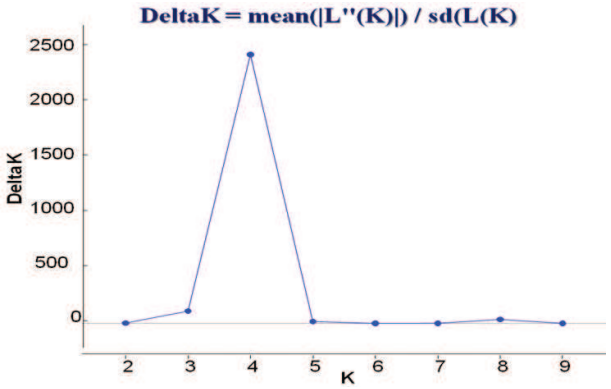

Fig. 2. Calculation of optimum (K) for 64 soybean genotypes based on magnitude of Delta K number

pairs. It indicated occurrence of historical recombination in the population used. The genotypes originated from USA and Taiwan are expected to differ in terms of alleles fixed. Selection pressure in the process of domestication and adaptation might have changedthe allele frequency. Change in allele frequencies or LD (or similar to population differentiation) can be attributed to mating system, mutation rate, founding effects, the magnitude of selection, admixture and genetic drift.

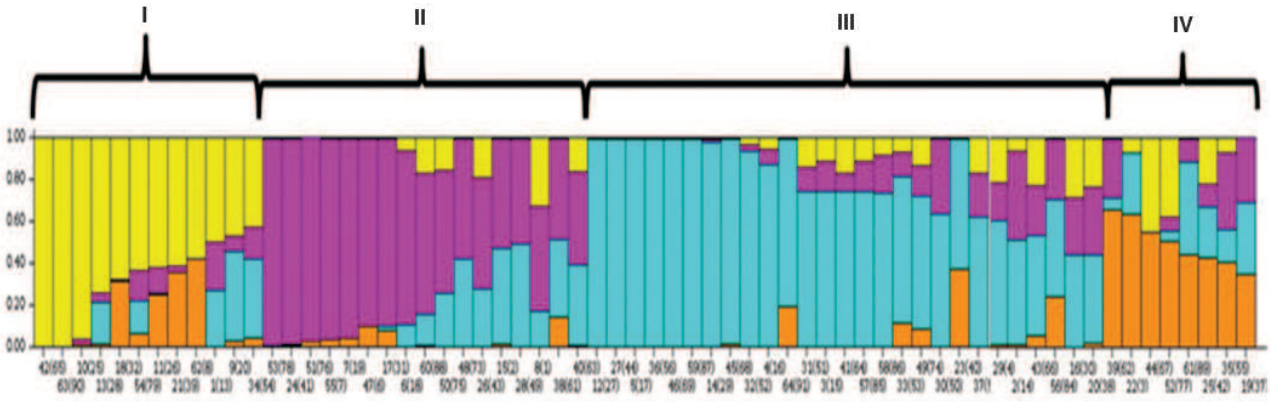

Fig. 3. Q Plot

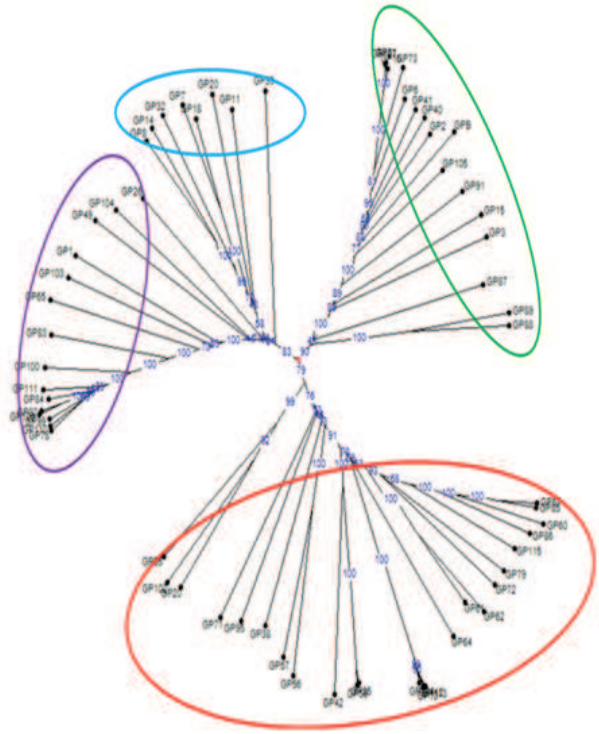

Fig. 4. Dendogram based on Neighbor Joining method

The LD decay in the 20 chromosomes varied from 138 kb to 1.76 Mb at $r^2 = 0.2$. Highest LD decay was in Chr. No. 20 and the lowest was in the Chr. No. 10 (1.5Mb). Some of the blocks spanned over a few mega-base pairs. Usually, the recombination rate varies greatly across the chromosome, higher in telomeric region and lower in centromeric region (Yu and Buckler, 2006). It also varies with crops; self-pollinated crops have long LD dacay and the cross-pollinated crops have short. In maize, a cross pollinated crops, LD decays at a much faster rate; it varies from 1 kb in landraces (Tenaillon et al. 2002) to 100 kb in commercial elite genotype (Ching et al. 2002).

GWAS for plant height and number of primary branches

Genome wide SNPs in 63 genotypes were deployed to identify markers associated with plant height and number of primary branches under normal and moisture stress condition at two locations, New Delhi and Dharwad. The genome wide scan for associated markers for the two traits was done using the MLM model with Kinship matrix and PCA. The significant SNPs were identified from the Manhattan plot.

Plant height

A highly significant SNP (LOD = 4.1) associated with plant height, under Delhi non-stress condition, was found in Chr. 11 position 1072146. The SNP explained 13.45% of the total phenotypic variation. The SNP is located in the gene *Glyma.11g015600* and encodes CCCH-type zinc finger family protein with RNA-binding domain. Zinc finger proteins are commonly associated with regulatory functions at transcription level (Peng et al. 2012). The reported gene is known to enhance tolerance to stress and resistance to fungal diseases (Guo et al. 2009). It is also involved in embryogenesis primarily at the apical domain of the embryo (Li and Thomas 1998). A major QTL, Ph24-4 reported by Chen et al. (2007) on Chromosome 11 (Satt426) was found in the vicinity of this putatively associated SNP. Sun et al. (2007) and Gai et al. (2007)also reported QTLs for plant height on Chr. 11.

Under moisture stressed condition in New Delhi, the SNP associated with plant height was located at the position 48603790 of Chr. 10. The LOD score of this SNP was 3.437 and the variation explained by it

was 11.066%. This SNP was located in the gene *Glyma10g40660* which encodes early growth response protein. The early growth response proteins are a family of zinc finger transcription factorswhich plays important role in regulating cell proliferation (Simmons et al. 1992). There exists a strong positive correlation between plant height and cell proliferation. Under water stressed condition, the *Glyma10g40660* gene might induce cell proliferation among the drought tolerant genotypes. Therefore, *Glyma10g40660* could be considered as the candidate gene for plant height. Reinprecht et al. (2006) also identified a QTL *Ph23-4* which is located close to the SNP identified in the present study.

Under non-stress situation in Dharwad, the SNP associated with the plant height was in the position 6079932 of chromosome 16. The LOD score of this SNP was 3.776 and was located in the gene *Glyma.16g062300* which encodes Transcription initiation factor TFIID, subunit BDF1 and related bromodomain proteins. The variation explained by this SNP was 27.47%. The bromodomain proteins plays critical role in transcription and are involved in a diverse range of functions, such as acetylating histones, remodeling chromatin, and recruiting other factors necessary for transcription (Josling et al. 2012). Specht et al. (2001) identified a QTL for plant height in the chromosome 16 i.e., *Ph13-5* and the SSR marker linked to this marker was Satt405 (Specht et al. 2001).

The SNP at position 23304601 on Chr. 1 showed significant association with plant height under stress condition at Dharwad. The LOD score of this SNP was 3.255 and is located in the *Glyma.01g081100* gene which codes for *APETELA2* domain. The phenotypic variation explained by this QTL was 29.45% of the total variation. The *APETELA2* are transcription factors which act primarily in the regulation of developmental programs and affects plant morphology, floral growth and ovule development. They are also found as the key regulators of internode elongation of deep-water rice (Licausi et al. 2013). Hu et al. (2013) reported a QTL for plant height i.e., Ph42-1 in the Chr. 1 with linked SSR marker Satt129.

Primary branches

SNP at position 6236769 of Chr. 16 showed significant association with number of primary branches under optimal condition in New Delhi. The SNP position is within the gene *Glyma.16g063400*. This SNP had LOD score 3.71 and had 12.52% effect on the trait. The gene *Glyma.16g063400* encodes zinc finger family protein. It is involved in a number of pathways that effect wide range of developmental processes including embryogenesis, hormone signalling and senescence (Moon et al. 2004). The SNP located in the position 49423269 of chr. 15 was significantly associated with number of primary branches/plant under water stress condition in New Delhi. The LOD score of the SNP was 4.172 and the variation explained by the QTL was 17.26%. This SNP is in the gene *Glyma.15g261900* which is known to be involved in synthesis of Phosphoglycerate kinase family protein. These proteins are located in the thylakoid, mitochondrion, chloroplast, membrane and are responsible for glycolysis. Chen et al. (2007) mapped a QTL with linked SSR marker Satt263 for branching on Chr. 15, however, functional relationship of the QTL could not be revealed. SNP 45156351 in Chr. 3 is significantly associated with number of primary branches under nornal condition in Dharwad. The LOD score of this SNP was 3.778 and was located in the gene *Glyma.03g256700* which encodes WRKY DNA-binding protein. The variation explained by the QTL harbouring this SNP was 28.82%, which can be considered as a major one.The WRKY families of transcription factors are present in plant species only. They are characterized by diverse biological function related to resistance to diseases, tolerance to abiotic stress and in embryogenesis, seed and trichome development, senescence, as well as additional developmental and hormone-controlled processes (Bakshi and Oelmüller, 2014). The SNP at position 41845126 of Chr. 10 with LOD score of 4.033 was significantly associated with number of primary branches under stressed condition in Dharwad. This SNP was located within the gene *Glyma.10g185400* that codes for CRM (chloroplast RNA splicing and ribosome maturation) family member 3A. The QTL for this trait explained 29.46% of the total variations. The CRM 3A is a RNA-binding domain involved in RNA folding and splicing (Keren et al. 2008; Asakura and Barkan, 2007). Li et al. (2008a) also mapped QTL *Branching 2-1* on Chr. 10 with the marker Satt581.

In this study, two novel SNPs for branching were mapped on Chr. 3 and 16. Unlike biparental QTL mapping approach, GWAS uses whole genome scanning that enhances chance of detecting QTL which otherwise may escape. New genotypes may throw novel QTL (Raychaudhuri 2011; Ladouceur et al. 2012). In this study, genotypes of diverse origin including breeding lines were used which might have

contributed the novel QTL for branching.

Plant height and number of primary branches are important architectural traits for yield. GWAS could identify a number of putative candidate genes for these traits which will be useful in breeding program to develop genotypes with suitable plant height with optimum number of primary branches. Validation of the SNPs identified in this study and their utilization would pave the way for development of soybean genotypes with proper architecture and higher yield.

Author's contribution

Conceptualization of research (JLB, AS); Designing of the experiments (JLB, AT); Contribution of experimental materials (AT); Execution of field/lab experiments and data collection (JLB, RY); Analysis of data and interpretation (JLB, AS, RS); Preparation of manuscript (JLB).

Declaration

The authors declare no conflict of interest.

Acknowledgement

Authors thank DST for providing funds under WOS-A scheme and ICAR-IARI for providing necessary facility to carry out the research.

References

Aditya J. P. P. Bhartya and Anuradha B. 2013. Genetic variability, heritability and character association for yield and component character in soybean. J. Cent. Eur. Agric., **12**(1): 27-34.

Archana T. P. W., Khorgade R. B., Ghorade G., Manjusha A., Thorat and Ghodke M. 1999. Variability, heritability and genetic advance in soybean [*Glycine max* (L.) Merrill]. J. Soils Crop, **9**: 198-200.

Asakura Y. and Barkan A. 2007. A CRM Domain Protein Functions Dually in Group I and Group II Intron Splicing in Land Plant Chloroplasts. Plant Cell Online, **19**(12): 3864-3875.

Bakshi M. and Oelmüller R. 2014. Wrky transcription factors jack of many trades in plants. Plant Signal. Behav., **9**(FEB): 1-18.

Bates D. M., Maechler B., Bolker and Walker S. 2015. Package lme4. J. Stat. Softw., **67**(1): 1-91.

Bradbury P. J. Z., Zhang D. E., Kroon T. M., Casstevens Y., Ramdoss and Buckler E. S. 2007. TASSEL: Software for association mapping of complex traits in diverse samples. Bioinformatics, **23**(19): 2633-2635.

Buckler E., Casstevens T., Bradbury P. and Zhang Z. 2009. Analysis by aSSociation, evolution and linkage (TASSEL) version 2.1. user manual. Cornell University, Ithaca.

Chen P. C. H., Sneller L. C., Purcell T. R., Sinclair C. A., King and Ishibashi T. 2007. Registration of Soybean Germplasm Lines R01-416F and R01-581F for Improved Yield and Nitrogen Fixation under Drought Stress. J. Plant Regist., **1**(2): 166.

Ching A. K., Caldwell S., Jung M., Dolan M., Smith O. S., Tingey S., Morgante M. and Rafalski A. J. 2002. SNP frequency, haplotype structure and linkage disequilibrium in elite maize inbred lines.

Cicek M. S., Chen P., Maroof M. A. S. and Buss G. R. 2006. Interrelationships among agronomic and seed quality traits in an interspecific soybean recombinant inbred population. Crop Sci., **46**(3): 1253-1259.

Earl D. A. and vonHoldt B. M. 2012. STRUCTURE HARVESTER: A website and program for visualizing STRUCTURE output and implementing the Evanno method. Conserv. Genet. Resour. **4**(2): 359-361.

Elshire R. J., Glaubitz J. C., Sun Q., Poland J. A., Kawamoto K., Buckler E. S. and Mitchell S. E. 2011. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS One, **6**(5): 1-10.

Evanno G. S., Regnaut and Goudet J. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. Mol. Ecol., **14**(8): 2611-2620.

Gai J. Y., Wang X., Wu and Chen S. 2007. A comparative study on segregation analysis and QTL mapping of quantitative traits in plants-with a case in soybean. Front. Agric. China, **1**(1): 1-7.

Guo Y. H., Yu Y. P., Wang D., Wu C. A., Yang G. D., Huang J. G. and C. C. Zheng. 2009. GhZFP1, a novel CCCH-type zinc finger protein from cotton, enhances salt stress tolerance and fungal disease resistance in transgenic tobacco by interacting with GZIRD21A and GZIPR5. New Phytol. **183**(1): 62-75.

Gupta P. K., Rustgi S. and Kulwal P. L. 2005. Linkage disequilibrium and association studies in higher plants: Present status and future prospects. Plant Mol. Biol., **57**(4): 461-485.

Hartman Y., Hooftman D. A. P., Uwimana B., van de Wiel C. C. M., Smulders M. J. M., Visser R. G. F. and van Tienderen P. H. 2012. Genomic regions in crop-wild hybrids of lettuce are affected differently in different environments: Implications for crop breeding. Evol. Appl., **5**(6): 629-640.

He J., Zhao X., Laroche A., Lu Z.-X., Liu H. and Li Z. 2014. Genotyping-by-sequencing (GBS), an ultimate marker-assisted selection (MAS) tool to accelerate plant breeding. Front. Plant Sci., **5**(September): 484.

Hemantaranjan A., Bhanu A. N., Singh M., Yadav D., Patel P., Singh R. and Katiyar D. 2014. Heat stress responses and thermotolerance. Adv. Plants Agric. Res., **1**(3): 1-9.

Henderson C. R. 1975. Best Linear Unbiased Estimation and Prediction under a Selection Model. Source: Biometrics, **31**(2): 423-447.

Hoeck J. A., Fehr W. R., Shoemaker R. C., Welke G. A., Johnson S. L. and Cianzio S. R. 2003. Molecular marker analysis of seed size in soybean. Crop Sci., **43**(May): 68-74.

Hu Z., Zhang H., Kan G., Ma D., Zhang D., Shi G., Hong D., Zhang G. and Yu D. 2013. Determination of the genetic architecture of seed size and shape via linkage and association analysis in soybean (*Glycine max* L. Merr.). Genetica, **141**(4-6): 247-254.

Hu Z., Zhang D., Zhang G., Kan G., Hong D. and Yu D. 2014. Association mapping of yield-related traits and SSR markers in wild soybean (*Glycine soja* Sieb. and Zucc.). Breed. Sci., **63**(5): 441-9.

Huyghe C. 1998. Genetics and genetic modifications of plant architecture in grain legumes: a review. Agronomie, **18**: 383-411.

Hwang E.-Y., Song Q., Jia G., Specht J. E., Hyten D. L., Costa J. and Cregan P. B. 2014. A genome-wide association study of seed protein and oil content in soybean. BMC Genomics, **15**: 1.

Joshi O. P. and Bhatia V. S. 2003. Stress management in soybean. In: Singh H and Hegde D M (Eds.), Souvenir. National Seminar on Stress Management in Oilseeds for Attaining Self-reliance in Oilseeds for Attaining Self Reliance in Vegetable Oils, Indian Society of Oilseeds Research, Hyderabad, pp. 13-25.

Josling G. A., Selvarajah S. A., Petter M. and Duffy M. F. 2012. The role of bromodomain proteins in regulating gene expression. Genes (Basel), **3**(2): 320-343.

Keren I., Klipcan L., Bezawork-Geleta A., Kolton M., Shaya F. and Ostersetzer-Biran O. 2008. Characterization of the molecular basis of group II intron RNA recognition by CRS1-CRM domains. J. Biol. Chem., **283**(34): 23333-23342.

Kim W. J., Ryu J., Im J., Kim S. H., Kang S. Y., Lee J. H., Jo S. H. and Ha B. K. 2018. Molecular characterization of proton beam-induced mutations in soybean using genotyping-by-sequencing. Mol. Genet. Genomics, **0**(0): 1-12.

Kim K. S., Vuong T. D., Qiu D., Robbins R. T., Shannon J. G., Li Z. and Nguyen H. T. 2016. Advancements in breeding , genetics , and genomics for resistance to three nematode species in soybean. Theor. Appl. Genet., **129**(12): 2295-2311.

Knowler W. C., Williams R. C., Pettitt D. J. and Steinberg A. G. 1988. Gm3;5,13,14 and type 2 diabetes mellitus: an association in American Indians with genetic admixture. Am. J. Hum. Genet., **43**(4): 520-6.

Ladouceur M., Dastani Z., Aulchenko Y. S., Greenwood C. M. T. and Richards J. B. 2012. The empirical power of rare variant association methods: Results from sanger sequencing in 1,998 individuals. PLoS Genet., **8**(2).

Li L., Lin F., Wang W., Ping J., Fitzgerald J. C., Zhao M., Li S., Sun L., Cai C. and Ma J. 2016. Fine mapping and candidate gene analysis of two loci conferring resistance to Phytophthora sojae in soybean. Theor. Appl. Genet., **129**(12): 2379-2386.

Li D., Pfeiffer T. W. and Cornelius P. L. 2008a. Soybean QTL for Yield and Yield Components Associated with Alleles. Crop Sci., **48**(2): 571.

Li Z. and Thomas T. L. 1998. PEI1, an embryo-specific zinc finger protein gene required for heart-stage embryo formation in Arabidopsis. Plant Cell, **10**(3): 383-398.

Li W., Zheng D., Van K. and Lee S. 2008. QTL mapping for major agronomic traits across two years in soybean (*Glycine max* L. Merr.). J. Crop Sci. Biotech., **11**(3): 171-190.

Licausi F., Ohme-Takagi M. and Perata P. 2013. APETALA2/Ethylene Responsive Factor (AP2/ERF) transcription factors: Mediators of stress responses and developmental programs. New Phytol., 639-649.

Mansur L. M., Orf J. H., Chase K., Jarvik T., Cregan P. B. and Lark K. G. 1996. Genetic mapping of agronomic traits using recombinant inbred lines of soybean. Crop Sci., **36**(5): 1327-1336.

Moon J., Parry G. and Estelle M. 2004. The ubiquitin-proteasome pathway and plant development. Plant Cell Online, **16**(December): 3181-3195.

Ning L., Kan G., Du W., Guo S., Wang Q., Zhang G., Cheng H. and Yu D. 2016. Association analysis for detecting significant single nucleotide polymorphisms for phosphorus-deficiency tolerance at the seedling stage in soybean. Breed. Sci., **66**(2): 191-203.

Orf J. H., Chase K., Jarvik T., Mansur L. M., Cregan P. B., Adler F. R. and Lark K. G. 1999. Genetics of soybean agronomic traits: I. Comparison of three related recombinant inbred populations. Crop Sci., **39**(6): 1642-1651.

Palomeque L., Li-Jun L., Li W., Hedges B., Cober E. R. and Rajcan I. 2009. QTL in mega-environments: II. Agronomic trait QTL co-localized with seed yield QTL detected in a population derived from a cross of high-yielding adapted/high-yielding exotic soybean lines. Theor. Appl. Genet., **119**(3): 429-436.

Palomeque L., Liu L. J., Li W., Hedges B. R., Cober E. R., Smid M. P., Lukens L. and Rajcan I. 2010. Validation of mega-environment universal and specific QTL associated with seed yield and agronomic traits in soybeans. Theor. Appl. Genet., **120**(5): 997-1003.

Patterson N., Price A. L. and Reich D. 2006. Population structure and eigenanalysis. PLoS Genet., **2**(12): 2074-2093.

Peleman J. D. and Van Der Voort J. R. 2003. Breeding by

design. Trends Plant Sci., **8**(7): 330-334.

Peng X., Zhao Y., Cao J., Zhang W., Jiang H., Li X., Ma Q., Zhu S. and Cheng B. 2012. CCCH-type zinc finger family in maize: Genome-wide identification, classification and expression profiling under abscisic acid and drought treatments. PLoS One, **7**(7).

Perrier X. and Jaccqemoud-Collet J. 2014. Dissimilarity Analysis and Representation for Windows DARWIN. 115.

Pritchard J. K., Stephens M., Rosenberg N. A. and Donnelly P. 2000. Association mapping in structured populations. Am. J. Hum. Genet., **67**(1): 170-81.

Raychaudhuri S. 2011. Mapping rare and common causal alleles for complex human diseases. Cell, **147**(1): 57-69.

Reinprecht Y., Poysa V. W., Rajcan I., Ablett G. R. and Pauls K. P. 2006. Agronomic performance of soybean with seed lipoxygenase nulls and low linolenic acid content. Can. J. Plant Sci.

Shaheen T., Rahman M.-, Shahid Riaz M., Zafar Y. and Rahman M.- 2016. Soybean production and drought stress. p. 177-196. *In* Abiotic and Biotic Stresses in Soybean Production.

Simmons D. L., Neel B. G., Stevens R., Evett G. and Erikson R. L. 1992. Identification of an early-growth-response gene encoding a novel putative protein kinase. Mol. Cell. Biol., **12**(9): 4164-9.

Singh R. K. and Chaudhary B. D. 1985. Biometrical methods in quantitative genetic analysis. Kalyani Publishers, New Delhi, Ludhiana, India. 39-78.

Song Q., Yan L., Quigley C., Jordan B. D., Fickus E., Schroeder S., Song B.-H., Charles An Y.-Q., Hyten D., Nelson R., Rainey K., Beavis W. D., Specht J., Diers B. and Cregan P. 2017. Genetic Characterization of the Soybean Nested Association Mapping Population. Plant Genome, **10**(2).

de Souza P. J. D. P., Ribeiro A., da Rocha E. J. P., Farias J. R. B., Loureiro R. S., Bispo C. C. and Sampaio L. 2009. Solar radiation use efficiency by soybean under field conditions in the Amazon region. Pesqui. Agropecu. Bras., **44**(10): 1211-1218.

Specht J. E., Chase K., Macrander M., Graef G. L., Chung J., Markwell J. P., Germann M., Orf J. H. and Lark K. G. 2001. Soybean response to water: A QTL analysis of drought tolerance. Crop Sci., **41**(2): 493-509.

Sun J., Jiang H., Xu Y., Li H., Wu X., Xie Q. and Li C. 2007. The CCCH-Type Zinc Finger Proteins AtSZF1 and AtSZF2 Regulate Salt Stress Responses in Arabidopsis. Plant Cell Physiol., **48**(8): 1148-1158.

Tenaillon M. I., Sawkins M. C., Anderson L. K., Stack S. M., Doebley J. and Gaut B. S. 2002. Patterns of Diversity and Recombination Along Chromosome 1. Genetics, **162**(3): 1401-1413.

Yu J. and Buckler E. S. 2006. Genetic association mapping and genome organization of maize. Curr. Opin. Biotechnol., **17**(2): 155-160.

Zhang J., Song Q., Cregan P. B., Nelson R. L., Wang X., Wu J. and Jiang G.-L. 2015. Genome-wide association study for flowering time, maturity dates and plant height in early maturing soybean (*Glycine max*) germplasm. BMC Genomics, **16**(1): 217.

Zhang W. K., Wang Y. J., Luo G. Z., Zhang J. S., He C. Y., Wu X. L., Gai J. Y. and Chen S. Y. 2004. QTL mapping of ten agronomic traits on the soybean (*Glycine max* L. Merr.) genetic map and their association with EST markers. Theor. Appl. Genet., **108**(6): 1131-1139.

Zhou Z., Jiang Y., Wang Z., Gou Z., Lyu J., Li W., Yu Y., Shu L., Zhao Y., Ma Y., Fang C., Shen Y., Liu T., Li C., Li Q., Wu M., Wang M., Wu Y., Dong Y., Wan W., Wang X., Ding Z., Gao Y., Xiang H., Zhu B., Lee S.-H., Wang W. and Tian Z. 2015. Resequencing 302 wild and cultivated accessions identifies genes related to domestication and improvement in soybean. Nat. Biotechnol., **33**(4): 408-14.