Identification of superior haplotypes in a diverse natural population for breeding desirable plant height in soybean

Javaid Akhter Bhat1,2 · Benjamin Karikari3 · Kehinde Adewole Adeboye4 · Showkat Ahmad Ganie5 · Rutwik Barmukh6 · Dezhou Hu1 · Rajeev K. Varshney6,7 · Deyue Yu1

Received: 13 January 2022 / Accepted: 28 April 2022 / Published online: 31 May 2022
© The Author(s) 2022

Abstract

Key message Plant height of soybean is associated with a haplotype block on chromosome 19, which classified 211 soybean accessions into five distinct groups showing significant differences for the target trait.

Abstract Genetic variation is pivotal for crop improvement. Natural populations are precious genetic resources. However, efficient strategies for the targeted utilization of these resources for quantitative traits, such as plant height (PH), are scarce. Being an important agronomic trait associated with soybean yield and quality, it is imperative to unravel the genetic mechanisms underlying PH in soybean. Here, a genome-wide association study (GWAS) was performed to identify single nucleotide polymorphisms (SNPs) significantly associated with PH in a natural population of 211 cultivated soybeans, which was genotyped with NJAU 355 K Soy SNP Array and evaluated across six environments. A total of 128 SNPs distributed across 17 chromosomes were found to be significantly associated with PH across six environments and a combined environment. Three significant SNPs were consistently identified in at least three environments on Chr.02 (AX-93958260), Chr.17 (AX-94154834), and Chr.19 (AX-93897200). Genomic regions of ~ 130 kb flanking these three consistent SNPs were considered as stable QTLs, which included 169 genes. Of these, 22 genes (including Dt1) were prioritized and defined as putative candidates controlling PH. The genomic region flanking 12 most significant SNPs was in strong linkage disequilibrium (LD). These SNPs formed a single haplotype block containing five haplotypes for PH, namely Hap-A, Hap-B, Hap-C, Hap-D, and Hap-E. Deployment of such superior haplotypes in breeding programs will enable development of improved soybean varieties with desirable plant height.
Introduction

Cultivated soybean (Glycine max L.) is an important oil-seed crop, grown globally as a major source of edible protein and oil (Kulkarni et al. 2016). However, in the past five decades, only a gradual increase in soybean yield has been observed in China. This can be witnessed by the fact that China imports > 80% of soybean and its products to meet its domestic requirements (Liu et al. 2018). Hence, development of soybean genotypes with desirable plant architecture and higher yield is a major objective of soybean breeders across the world, and especially from China. A significant positive correlation has been observed between PH and yield in soybean. However, taller soybean genotypes are highly prone to lodging, affecting the overall yield and seed quality (Liu et al. 2011; Lü et al. 2016). This makes PH a crucial trait in the breeding programs of soybean. Moreover, PH is a quantitative trait that is controlled by multiple loci/genes, besides being highly influenced by genotype × environment interactions (Wang et al. 2013; Yin et al. 2017). For instance, the past three decades of efforts in soybean breeding have witnessed identification of > 200 quantitative trait loci (QTLs), as documented in the SoyBase (https://www.soybase.org/), and these QTLs are located across 20 soybean chromosomes. Most of the previously reported studies used linkage mapping with low-resolution markers to identify such QTLs, which often possess low genomic resolution. This has hampered successful deployment of PH QTLs in breeding programs to develop improved soybean varieties with desired PH (Cao et al. 2019). An alternative approach of trait mapping such as GWAS detects alleles/QTLs linked to the trait of interest at a higher resolution using diverse natural populations (Schmutz et al. 2010). Technological advances in next-generation sequencing-based genotyping technologies have enabled GWAS to be used routinely in crop genetics research (Song et al. 2016; Cao et al. 2019). Several studies have demonstrated the efficiency of GWAS to identify marker-trait associations (MTAs) regulating important agronomic traits and to detect the underlying putative genes in different plant species such as Arabidopsis thaliana (Ren et al. 2019), rice (Xu and Crouch 2008; Li et al. 2014), maize (Yang and Zhu 2005), wheat (Wang et al. 1999), cotton (Lai et al. 2016), soybean (Hyten et al. 2008; Cao et al. 2017; Fang et al. 2017; Li et al. 2017; Chang et al. 2018), chickpea (Varshney et al. 2019a, b; Thudi et al. 2021), pigeonpea (Varshney et al. 2017a, b), and pearl millet (Varshney et al. 2017a, b).

With the availability of whole genome sequencing data (WGS), sequencing-based trait mapping strategies have been successful in identifying QTLs/MTAs in several crops (Varshney et al. 2019a, b). Notably, the use of WGS data has increased the mapping resolution of GWAS and enabled identification of haplotypes for the trait(s) of interest. In the post-sequencing era, ‘haplotype assembly’ has emerged as an efficient approach for the development of improved crop varieties (Bevan et al. 2017). In the past few years, haplo- pheno analysis has been successfully used for detecting superior haplotypes for agronomically important traits in some crop species. For example, Abbai et al. (2019) reported desirable haplotypes for 21 genes controlling quality and yield-related traits in rice, based on a panel of 3000 rice accessions. Similarly, superior haplotypes have been identified for drought and salinity stress adaptation in rice, soybean, and pigeonpea (Guan et al. 2014; Mishra et al. 2016; Kuroha et al. 2018; Chen et al. 2019; Sinha et al. 2020).

To understand the genetics of PH in soybean genotypes selected from major soybean growing areas of China, the present study used GWAS and haplotype analysis for identifying MTAs, candidate genes, and haplotypes. Superior haplotypes identified in this study hold potential for developing improved soybean varieties with desired plant architecture using haplotype-based breeding approach.

Material and methods

Plant material and field experiments

A natural population comprising 219 soybean germplasm accessions was evaluated for two consecutive years (2011 and 2012) at three different locations in China. This soybean germplasm represents the cultivated pool selected from different ecological regions of China, previously described by Wang et al. (2016) and Du et al. (2020) (Supplementary Table 1). The evaluated locations include the experimental fields of Nanjing Agricultural University (Nanjing); Jiangsu Yanjiang-Institute of Agricultural Sciences (Nantong); and Agricultural College of Yangzhou University (Yangzhou). Hence, the germplasm was evaluated at a total of six individual environments (location × year), namely E1–E6 (E1: Nanjing_2011, E2: Nantong_2011, E3: Yangzhou_2011, E4: Yangzhou_2012, E5: Nanjing_2012, E6: Nantong_2012). The germplasm represents the cultivated pool selected from major soybean growing areas of China, previously described by Wang et al. (2016) and Du et al. (2020) (Supplementary Table 1). The evaluated locations include the experimental fields of Nanjing Agricultural University (Nanjing); Jiangsu Yanjiang-Institute of Agricultural Sciences (Nantong); and Agricultural College of Yangzhou University (Yangzhou). Hence, the germplasm was evaluated at a total of six individual environments (location × year), namely E1–E6 (E1: Nanjing_2011, E2: Nantong_2011, E3: Yangzhou_2011, E4: Yangzhou_2012, E5: Nanjing_2012, E6: Nantong_2012).
Y_{ijk} = \mu + \text{Env}_i + \text{Rep}_j(\text{Env}_i) + \text{Gen}_k + \text{Env}_i \times \text{Gen}_k + \varepsilon_{ijk}

where Env_i and Env_i \times Gen_k are the ith environment and the \( G \times E \) interaction effects, respectively.

The broad-sense heritability for the combined and individual environment was estimated as \( \sigma_g^2/(\sigma_g^2 + \sigma_e^2/n + \sigma_r^2/r) \) and \( \sigma_g^2/(\sigma_g^2 + \sigma_e^2/n + \sigma_r^2/r) \), respectively, where \( \sigma_g^2 \) is the genetic variance, \( \sigma_{ge}^2 \) is the interaction of genotype and environment, \( \sigma_e^2 \) is the residual error, \( n \) is the number of environments, and \( r \) is the number of replication/block (Knapp et al. 1985).

To identify the relationship between PH evaluated across different environments, we performed a Pearson’s correlation analysis. In brief, Pearson correlation analysis for soybean PH measured across different environment was conducted based on the predicted means calculated across the environment and visualized using MVApp (Julkowska et al. 2019).

## Genome-wide association study

The NJAU 355 K SoySNP assay previously developed by our laboratory (Wang et al. 2016) was used in the present study. In brief, the sequencing data of 31 soybean genotypes (including 14 cultivated and 18 wild type) were used along with Williams 82 (Glyma.Wm82.a1.v1.1) as the reference genome (www.phytozome.net). The VIP variants include 60,800 SNPs from SoySNP50K15, 2103 SNPs from SoyBase, and 3671 SNPs from the sequencing data of 31 soybean genotypes. A total of 355,595 SNPs representing of 0.01 and missing genotype and taxa at 0.2 in TASSEL v5.2.73 (Bradbury et al. 2007).

Genome-wide association analysis was conducted with Genomic Association and Prediction Integrated Tool (GAPIT) package in R environment (Lipka et al. 2012) using a compressed mixed linear model (CMLM). This model was adopted to speed up computational time and optimize statistical performance (Lipka et al. 2012). UpSet plot was developed for significant SNPs by using the \texttt{UpSetR} package in R environment (Lax et al. 2014; Conway et al. 2017).

## Haplotype analysis

The LD level of pairwise SNPs was calculated using Haplovie 4.2 (Barnett et al. 2005). All markers associated with PH on chromosome 19 (Chr.19) within the LD ± 130 kb (Wang et al. 2016) were considered, and haplotype block was defined by “confidence intervals” algorithm (Gabriel
Genotypes were grouped into independent groups based on specific haplotype carried by them. To estimate the effect of haplotype on PH, groups were fitted into one-way ANOVA model in R environment as follows:

\[
\text{model} \leftarrow \text{aov(phenotype} \sim \text{group}, \ data = \text{data})
\]

where phenotype denotes PH values in the individual and combined environments. A pairwise comparison of means was conducted using Turkey’s HSD test and visualized in R environment.

### Identification of candidate genes

Within the LD decay distance of significant SNPs, all genes present within the ±130 kb interval flanking the stable SNP positions were retrieved from SoyBase website (https://www.soybase.org/), using Glyma2.0 gene model. The functional annotations of genes were downloaded from SoyBase, and were screened manually. The RNA-seq data available at SoyBase (https://www.soybase.org/) were extracted for the genes underlying the QTL intervals from different soybean tissues (leaf, flower, pod shell, seed, root, and nodule) at different growth and developmental stages, viz., young leaf, flower, one cm pod, pod shell 10 days to flowering (DAF), pod shell 14DAF, seed 10DAF, seed 14DAF, seed 21DAF, seed 25DAF, seed 28DAF, seed 35DAF, seed 42DAF, root, and nodule) (Supplementary Table 4). These RNA-seq data were previously generated and deposited in the SoyBase by Severin et al. (2010). The RNA-seq dataset was used to analyze the expression of putative genes in different soybean tissues and across multiple development stages. For developing a heat map, the fragments per kilobase of transcript per million fragment mapped (FPKM) values of the candidate genes were used. Hierarchical clustering was performed using Euclidean Distance Method with complete linkage approach (Tao et al. 2013). A heatmap of these candidate genes was constructed using TBtools_JRE1.6 software (Chen et al. 2020). Based on functional annotations of genes, available literature, and gene expression patterns, some genes were defined as putative candidates regulating PH.

### Results

**Phenotypic characterization and correlation analysis of PH in soybean accessions**

The descriptive statistics including mean, range (maximum and minimum values), standard deviation, skewness, kurtosis, broad-sense heritability, and coefficient of variation (CV) for PH in 219 soybean accessions evaluated across six different environments and the combined environment are presented in Table 1. In brief, the PH ranged from 8.24 cm in E6 to 311.67 cm in E4. The mean across individual environments ranged from 52.76 ± 0.83 cm in E3 to 105.30 ± 1.73 in E4. The CV in the combined environment was 49%, and ranged from 38% in E6 to 49.8% in E2. Furthermore, skewness and kurtosis were 1.34 and 2.84, respectively in the combined environment (Table 1). The broad-sense heritability ranged from 0.84 in E6 to 0.98 in E4, and 0.97 in the combined environment. The soybean accessions (G), environment (E), and the genotype×environment interaction (G×E) had significant effect (P < 0.0001) on PH (Table 2).

### Table 1 Descriptive statistics, broad-sense heritability, and coefficient of variation (CV) for plant height evaluated across 211 diverse soybean accessions

| Environment | Minimum | Maximum | Mean ± SE | SD | CV% | Skewness | Kurtosis | Broad-sense heritability |
|-------------|---------|---------|-----------|----|-----|----------|----------|-------------------------|
| CE          | 8.24    | 311.67  | 74.68 ± 0.60 | 36.69 | 49.13 | 1.34 | 2.84 | 0.97 |
| E1          | 35.00   | 243.00  | 85.83 ± 1.40 | 33.78 | 39.36 | 0.80 | 0.33 | 0.95 |
| E2          | 8.33    | 199.33  | 61.86 ± 1.26 | 30.82 | 49.83 | 1.34 | 2.24 | 0.94 |
| E3          | 21.00   | 195.00  | 52.76 ± 0.83 | 21.27 | 40.31 | 1.01 | 1.65 | 0.97 |
| E4          | 29.75   | 311.67  | 105.30 ± 1.73 | 44.25 | 42.02 | 1.20 | 2.15 | 0.98 |
| E5          | 18.67   | 191.67  | 69.61 ± 1.27 | 32.13 | 46.15 | 1.07 | 0.78 | 0.95 |
| E6          | 8.24    | 212.67  | 72.51 ± 1.10 | 27.72 | 38.23 | 1.01 | 1.95 | 0.84 |

**SE** Standard error
Moreover, Pearson correlation analysis of PH showed a significantly positive correlation ($P < 0.0001$) across different environments evaluated (Supplementary Table 5).

**Genome-wide distribution of SNPs in soybean natural population**

In the present study, a total of 211 soybean accessions and 291,962 SNP markers were retained after quality control analysis (Table 3). These markers covered 949,992,372 bp (949.99 Mb) of the soybean genome, representing ~85% of the genome. The number of SNPs in each chromosome varied from 12,035 to 19,880, with the lowest number of SNPs present on Chr.11, whereas Chr.18 contained the highest number of SNPs. The lowest and highest SNP density of 264.27 SNPs/Mb and 361.97 SNPs/Mb was found on Chr.01 and Chr.13, respectively (Supplementary Fig. 1; Table 3).

**GWAS revealed genetic architecture of PH in soybean**

A total of 128 SNPs were found to be significantly associated with PH across six environments and the combined environment, based on GWAS results (Fig. 1; Supplementary Table 6). The distribution of markers across soybean chromosomes and their intersections based on the individual and combined environments are presented in Fig. 2 as UpSet plot (Lex et al. 2014). The UpSet plots provide a convenient alternative to visualize the data sets by frequency. In Fig. 2a, the horizontal bar graph on the left-hand side of the UpSet plot showed the size of each set, i.e., total number of significant SNPs identified in each individual environment and the combined environment. For example, the horizontal bar linked to E5 environment indicated the presence of 58 significant SNPs identified in this environment. Similarly, the horizontal bars corresponding to other five individual environments (E1, E2, E3, E4, and E6) and the combined environment showed the number of significant SNPs identified in each of these environments (Fig. 2a). Furthermore, the vertical bars on the top indicate the individual set participation sizes (Fig. 2a). For instance, the vertical bar on the left most side shows that out of the 58 significant SNPs identified in E5 environment, 49 SNPs were specific for E5 environment, two SNPs were detected in E4 and E5; one SNP in E2, E3, and E5; four SNPs in E1, E3, E5, E6, and CE; and one SNP was detected in E1, E2, E3, E4, E5, E6, and CE (Fig. 2a).

The distribution of markers across soybean chromosomes and their intersections based on the individual and combined environments are presented in Fig. 2 as UpSet plot (Lex et al. 2014). The UpSet plots provide a convenient alternative to visualize the data sets by frequency. In Fig. 2a, the horizontal bar graph on the left-hand side of the UpSet plot showed the size of each set, i.e., total number of significant SNPs identified in each individual environment and the combined environment. For example, the horizontal bar linked to E5 environment indicated the presence of 58 significant SNPs identified in this environment. Similarly, the horizontal bars corresponding to other five individual environments (E1, E2, E3, E4, and E6) and the combined environment showed the number of significant SNPs identified in each of these environments (Fig. 2a). Furthermore, the vertical bars on the top indicate the individual set participation sizes (Fig. 2a). For instance, the vertical bar on the left most side shows that out of the 58 significant SNPs identified in E5 environment, 49 SNPs were specific for E5 environment, two SNPs were detected in E4 and E5; one SNP in E2, E3, and E5; four SNPs in E1, E3, E5, E6, and CE; and one SNP was detected in E1, E2, E3, E4, E5, E6, and CE (Fig. 2a). In Fig. 2b, each vertical bar reveals the number of significant SNPs associated with PH identified on each chromosome. Notably, three significant SNPs were consistently identified in three or more environments (Fig. 2; Table 4). These included a SNP (AX-93897200) on Chr.19, which was detected in all six environments and the combined environment, a SNP

### Table 3

| Chromosome | Length (bp) | Length (Mb) | Number of SNPs | Inter-marker distance (kb) | Density (SNPs/Mb) |
|------------|-------------|-------------|----------------|----------------------------|-------------------|
| Chr.01     | 55,915,249  | 55.92       | 14,777         | 3.8                        | 264.27            |
| Chr.02     | 51,649,736  | 51.65       | 15,318         | 3.4                        | 296.57            |
| Chr.03     | 47,780,218  | 47.78       | 14,653         | 3.3                        | 306.68            |
| Chr.04     | 49,243,023  | 49.24       | 13,793         | 3.6                        | 280.10            |
| Chr.05     | 41,932,848  | 41.93       | 12,557         | 3.3                        | 299.45            |
| Chr.06     | 50,704,308  | 50.70       | 16,001         | 3.2                        | 315.57            |
| Chr.07     | 44,682,118  | 44.68       | 13,744         | 3.3                        | 307.60            |
| Chr.08     | 46,990,565  | 46.99       | 15,346         | 3.1                        | 326.58            |
| Chr.09     | 46,842,345  | 46.84       | 14,657         | 3.2                        | 312.90            |
| Chr.10     | 50,966,905  | 50.97       | 14,887         | 3.4                        | 292.09            |
| Chr.11     | 39,171,759  | 39.17       | 12,035         | 3.3                        | 307.24            |
| Chr.12     | 40,112,060  | 40.11       | 12,638         | 3.2                        | 315.07            |
| Chr.13     | 44,406,433  | 44.41       | 16,074         | 2.8                        | 361.97            |
| Chr.14     | 49,707,254  | 49.71       | 15,350         | 3.2                        | 308.81            |
| Chr.15     | 50,937,513  | 50.94       | 15,113         | 3.4                        | 296.70            |
| Chr.16     | 37,392,142  | 37.39       | 12,760         | 2.9                        | 341.25            |
| Chr.17     | 41,900,927  | 41.90       | 13,387         | 3.1                        | 319.49            |
| Chr.18     | 62,305,715  | 62.31       | 19,880         | 3.1                        | 319.07            |
| Chr.19     | 50,585,940  | 50.59       | 15,502         | 3.3                        | 306.45            |
| Chr.20     | 46,765,314  | 46.77       | 13,490         | 3.5                        | 288.46            |
| Total      | 949,992,372 | 949.99      | 291,962        | 65.15                      | 6166.32           |
Fig. 1 GWAS signals for plant height evaluated across six environments. Manhattan plot and quantile-quantile (Q-Q) plot for the GWAS for PH evaluated at a Nanjing_2011 (E1), b Nantong_2011 (E2), c Yangzhou_2011 (E3), d Nanjing_2012 (E4), e Nantong_2012 (E5), and f Yangzhou_2012 (E6). The red lines on the Y-axis designate the significance threshold (−log_{10} P<5.47). The numbers on the X-axis represent soybean chromosomes.
The theoretical and applied genetic (2022) 135:2407–2422

The regions within ~130 kb interval [based on the LD decay previously reported for the same population by using NJAU 355 K SoySNP assay (Wang et al. 2016)] flanking the significant SNPs (AX-93958260, AX-94154834, and AX-93897200) present on Chr.02, Chr.17, and Chr.19, were referred as qPH2, qPH17 and qPH19, respectively. These QTLs/genomic regions represented stable genetic elements regulating PH in soybean.

Table 4

| QTL  | Chromosome | Representative SNPa | Position (bp) | Number of significant SNPs | Environments Related QTL References |
|------|------------|---------------------|---------------|----------------------------|--------------------------------------|
| qPH2 | Chr.02     | AX-93958260         | 36,907,407    | 3                          | E2, E3, E5 Plant height 26–9         | Sun et al. (2006)                 |
| qPH17| Chr.17     | AX-94154834         | 2,929,975     | 1                          | E2, E4, E5 No related QTL Not available |                                      |
| qPH19| Chr.19     | AX-93897200         | 44,950,887    | 5                          | E1, E2, E3, E4, E5, E6, CE Plant height 6–1; Plant height 10–4 | Lark et al. (1995); Orf et al. (1999) |

aThe representative SNP with the minimum P-value; CE combined environment

Fig. 2 Intersection of significant QTLs and their frequency across soybean chromosomes. a UpSet plot showing the intersection of significant QTLs across six different environments (E1, E2, E3, E4, E5, and E6) and combined environment (CE). b Frequency of QTLs across soybean chromosomes.
Superior haplotypes for PH identified on soybean Chr. 19

The genomic region flanking three stable significant SNPs (AX-93958260, AX-94154834, and AX-93897200) within the LD decay of ±130 kb was subjected to haplotype analysis. The results showed that only significant markers identified within the LD decay flanking the SNP AX-93897200 (detected in all six environments and the combined environment) on Chr.19 formed a haplotype block. No haplotype block was detected for other two significant SNPs within the referenced LD decay (±130 kb) on Chr.02 and Chr.17. Hence, the SNP (AX-93897200) along with 11 other SNPs (AX-94195032, AX-93897192, AX-93659063, AX-94195035, AX-93897195, AX-93897196, AX-93955268, AX-93897197, AX-93897198, AX-93897199, and AX-93897201) were found to be in strong LD and formed a single haplotype block (Fig. 3). The length of this haplotype block was less than 0.15 Mb. These 12 SNPs within the haplotype block represented five haplotypes across 211 accessions of the soybean natural population. Notably, these haplotypes classified the genotypes of the studied soybean population into five groups. The PH of the genotypes varied significantly across the five haplotype groups. For instance, Haplotype-A (Hap-A) formed the largest group (n = 102), followed by Hap-B (n = 72), Hap-C (n = 16), Hap-D (n = 17), and Hap-E (n = 04) (Fig. 3c). The nucleotide variation detected for these haplotypes is represented in Supplementary Fig. 2.

The differences in PH based on different haplotype alleles identified within the haplotype block on Chr.19 in soybean GWAS population varied across environments. However, soybean accesses with Hap-D appeared to be the tallest across environments, those with Hap-E and Hap-B had a dwarf phenotype, and accesses with Hap-A and Hap-C were of intermediate height (Fig. 3c; Supplementary Fig. 3).

Identification of candidate genes within three QTL intervals

A total of 169 genes were identified within the physical genomic interval of all three major QTLs (qPH2, qPH17 and qPH19), which consisted of 18, 35, and 116 genes, respectively (Supplementary Table 7). Of the 169 genes detected, the RNA-seq data for only 124 genes were available online at SoyBase (https://www.soybase.org/) for 14 soybean tissues across different stages of growth and development (Supplementary Table 4). Expression pattern of the genes obtained from the RNA-seq data is represented in the form of a heatmap in Fig. 4. Furthermore, based on in silico analysis of gene expression data and gene annotations, we defined a total of 22 candidate genes underlying three stable genomic regions. This includes two genes underlying qPH2, five genes underlying qPH17, and 15 genes underlying qPH19 (Table 5). The selection of these candidate genes was based on gene function annotation (i.e., genes regulating cell division, meristem growth, plant growth hormone biosynthesis and signaling, cell elongation, and transition from vegetative to reproductive phase), literature studies, and gene expression profiles (Table 5; Supplementary Table 4). These 22 genes can be considered as putative candidates controlling PH. However, further functional validation of these genes is needed to determine their exact role in regulating PH in soybean.

Discussion

Plant height is an important determinant of soybean yield and quality (Lu et al. 2016; Yin et al. 2017). Taller soybean cultivars/varieties are often subjected to lodging, which leads to considerable reduction in soybean production as well as deterioration of its quality (Cao et al. 2019). Hence, tailoring PH in soybean to avoid qualitative and quantitative losses has been a long-term goal of soybean breeders. Determining the genetic basis of PH is a prerequisite to develop soybean varieties with desirable plant architecture. Studies have documented the complex inheritance of PH, besides the trait being strongly influenced by G×E interactions (Lee et al. 2015). Therefore, the present study was undertaken to identify the genetic loci/QTLs, and superior haplotypes regulating PH in soybean by using GWAS and haplotype analysis.

In the present study, ANOVA results revealed a significant difference among 211 accessions for PH (P<0.01). In accordance with previous studies (Zhang et al. 2004; Cao et al. 2019), our study confirmed that PH is significantly affected by G, E, and G×E interactions. This suggests the complex genetic nature of PH in soybean. Furthermore, the estimate of broad-sense heritability for PH was high (98.95%) in the soybean natural population across all six environments, which is consistent with previous studies (Cao et al. 2019). Higher broad-sense heritability reported for PH in this study suggests a high probability of achieving the same phenotypic performance for PH when grown in the same environment. In agreement with our findings, many authors have earlier reported similar results for PH in the soybean germplasm collection (Zhang et al. 2015; Chang et al. 2018; Xue et al. 2019).

To date, the genetic basis of PH in soybean has been documented by multiple research groups (Orf et al. 1999; Chapman et al. 2003; Liu et al. 2013; Yin et al. 2017; Cao et al. 2019). However, the majority of previously detected QTLs have not been validated, which limit their deployment in MAB efforts. To this end, a more powerful and high-resolution mapping approach called the LD-based GWAS
was used in the present study. Previous studies have demonstrated that GWAS is an effective strategy to determine the genetic basis of complex crop traits, such as PH, in a more precise and accurate manner (Kim et al. 2012; Cober and Morrison 2010). Hence, in the present study, we used GWAS for the identification of genomic regions associated with PH in soybean. We identified a total of 128 SNPs linked with PH across all the environments, which were found to be distributed on all chromosomes except Chr.06, Chr.18, and Chr.20. These results suggest the genome-wide distribution

Fig. 3 GWAS signal, haplotype block analysis of qPH19, and haplotype effect for plant height. a Manhattan plot and Q-Q plot for the GWAS for PH analyzed in the combined environment. Red line represents the Bonferroni correction threshold. b GWAS signal for PH obtained on Chr.19 and pairwise LD analysis. The pairwise LD diagram for significant variants ($P < 3.42 \times 10^{-6}$) is displayed. c The boxplot for the haplotypes based on the predicted values of PH in the combined environment (CE). The genotypes were grouped and pairwise comparisons were conducted using Turkey’s HSD Test at $P < 0.05$. The boxes with a common alphabet indicate no significant difference in PH. Number of accessions ($n$) in each sub-class is represented on top of each box.
of loci/genes regulating PH in soybean and reveal the complex genetic architecture of this trait. In agreement with our results, Zhang et al. (2015) reported the distribution of significant MTAs for PH on 15 different chromosomes of soybean. Similar results have also been reported by other researchers (Fang et al. 2017, 2020; Chang et al. 2018).
Further, the number of significant SNPs linked to PH was found to vary considerably across six different environments. For example, the highest number of SNPs were identified in E5, whereas the lowest number of SNPs were detected in E2. These findings suggest some effect of environment on the regulation of PH in soybean, which is in accordance with previous reports (Xue et al. 2019; Fang et al. 2020).

Importantly, three significant SNPs were consistently identified across multiple environments on Chr.02, Chr.17, and Chr.19. The genomic regions (~130 kb) flanking significant SNPs were referred as QTLs related to PH based on the LD decay. These QTLs/genomic regions represent stable genetic elements regulating PH in soybean. The QTL on Chr.02 (Plant height 26–9) associated with PH has been previously reported in the genomic region between 15,993,654 and 41,032,570 bp by Sun et al. (2006), and the genomic region underlying qPH2 was found to lie in the same physical interval. Therefore, qPH2 might be the same as Plant height 26–9, as earlier reported by Sun et al. (2006). However, when compared to Plant height 26–9, the physical interval of qPH2 was considerably narrowed down in the present study. Furthermore, qPH19 identified in the present study was found to co-locate with two previously identified QTLs viz., Plant height 6–1 (44,862,405–45,270,324 bp) and Plant height 10–4 (44,862,405–45,270,324 bp) on Chr.19 (Lark et al. 1995; Orf et al. 1999). However, there is no QTL identified to date that falls within the physical interval of qPH17. Hence, qPH17 can be regarded as a novel QTL identified in the current study. The physical intervals of qPH2 and qPH19 were considerably reduced in the present study. This may be because the earlier reported QTLs (Plant height 26–9, Plant height 6–1, and Plant height 10–4) were identified via linkage mapping by using low-density markers such as simple sequence repeat and restriction fragment length polymorphism, possessing a relatively low genomic resolution. These marker systems have low selection accuracy, which has been the biggest obstacle for their deployment in the MAB programs. QTL mapping involves the use of bi-parental mapping populations, which are associated with poor resolution in the detection of target loci regulating traits of interest (Kraakman et al. 2004). To this end, GWAS has emerged as a powerful approach using the ancestral recombination events and has high-resolution in the detection of MTAs. Therefore, high resolution of GWAS in the identification of the stable QTLs across multiple environments will facilitate their effective utilization in MAB programs for breeding soybean varieties with desirable PH.

In the practical breeding programs, the identification of potential markers/candidate genes for the trait of interest underlying a particular genomic region is the ultimate objective (Ganie and Ahammer 2021; Ganie et al. 2021). To date, minimal efforts have been made to identify candidate genes regulating PH in soybean (Liu et al. 2013; Lee et al. 2015; Cao et al. 2017), and only few genes have been characterized (Liu et al. 2010; Ping et al. 2014; Zhang et al. 2018). Hence, based on literature research and gene annotations, we defined 22 candidate genes underlying qPH2, qPH17, and qPH19 that may possibly regulate PH. For example, Dt1 (Glyma.19G194300) gene lying within the physical interval of qPH19 was considered as the candidate gene. The same locus on Chr.19 underlying the Dt1 locus was also identified previously (Sonah et al. 2015; Kato et al. 2019). Liu et al. (2010) observed regulation of the growth habits and PH in soybean by Dt1. In addition to the growth habit genes, maturity and flowering time genes also have a major impact on PH (Cober and Morrison 2010; Xia et al. 2012; Zhang et al. 2015; Cao et al. 2017). Gene function related to growth hormones, mitosis, cell division, cell elongation or functions that are directly or indirectly related to the vegetative growth are involved in regulating PH. To this end, a total of 169 gene models were found underlying the physical intervals of three stable QTLs, and 22 genes (including Dt1) were defined as putative candidates controlling PH by taking into consideration their annotations and gene expression patterns. However, further verification and functional validation of these genes are required to precisely pinpoint their role in the regulation of PH in soybean.

Recently, haplotype-based breeding has emerged as a promising approach to develop custom-designed crop varieties (Varshney et al. 2021). However, this breeding approach needs detection of superior haplotypes for their ultimate use in breeding programs. Haplotype analysis allows plant breeders to utilize the genetic variation underlying key genes/loci to the greatest extent. For example, in soybean, a candidate gene regulating salinity tolerance (GmCHX1) was subjected to haplotype analysis. Among various haplotypes identified for GmCHX1, SV-2 was found to provide maximum salinity tolerance (Patil et al. 2016). Moreover, Wang et al. (2017) identified superior haplotypes for grain quality traits (cooking and eating quality traits) in rice. Sinha et al. (2020) performed haplotype analysis of five genes controlling drought tolerance in pigeonpea. In a given genomic region, many factors regulate haplotype variation, such as mutation, recombination rates, and selection (Zaitlen et al. 2017).
### Table 5  Potential candidate genes underlying the three QTLs (qPH2, qPH17, and qPH19) and their gene annotation

| Serial no. | Gene ID (Wm82.a4.v1) | Gene ID (Wm82.a2.v1) | Biological functions |
|------------|----------------------|----------------------|---------------------|
| 1          | Glyma02g35210        | Glyma.02g195300      | Regulation of stomatal movement; response to auxin stimulus; response to ethylene stimulus |
| 2          | Glyma02g35350        | Glyma.02g195900      | MAPK cascade; intracellular signal transduction; protein ubiquitination |
| 3          | Glyma17g04210        | Glyma.17g037900      | Aerobic respiration; cell redox homeostasis; response to fructose stimulus; response to glucose stimulus; response to sucrose stimulus |
| 4          | Glyma17g04300        | Glyma.17g038700      | Cytokinesis; cytokinesis by cell plate formation; microtubule-based movement; nuclear division; nucleolus organization; regulation of DNA replication; regulation of G2/M transition of mitotic cell cycle |
| 5          | Glyma17g04340        | Glyma.17g039100      | S-adenosylmethionine biosynthetic process; lignin biosynthetic process; methionine metabolic process |
| 6          | Glyma17g04350        | Glyma.17g039200      | Auxin polar transport; regulation of anion channel activity; root development |
| 7          | Glyma17g04360        | Glyma.17g039300      | Auxin polar transport; regulation of anion channel activity; root development |
| 8          | Glyma19g37380        | Glyma.19g189700      | Bract formation; floral meristem determinacy; leaf morphogenesis; positive regulation of transcription, regulation of timing of transition from vegetative to reproductive phase; regulation of transcription |
| 9          | Glyma19g37410        | Glyma.19g190000      | Determination of bilateral symmetry; meristem initiation; polarity specification of adaxial/abaxial axis; regulation of meristem growth; regulation of transcription |
| 10         | Glyma19g37530        | Glyma.19g191100      | Embryo development ending in seed dormancy; lipid storage; meristem structural organization; photomorphogenesis; photoperiodism, flowering; regulation of flower development; seed germination; vegetative to reproductive phase transition of meristem |
| 11         | Glyma19g37570        | Glyma.19g191600      | Gibberellin biosynthetic process; regulation of post-embryonic root development; regulation of stem cell division; regulation of timing of transition from vegetative to reproductive phase; response to fructose stimulus; response to sucrose stimulus |
| 12         | Glyma19g37610        | Glyma.19g191900      | DNA recombination; chromatin silencing by small RNA; meiosis; meiotic chromosome segregation; regulation of telomere maintenance; sister chromatid cohesion; telomere maintenance in response to DNA damage |
| 13         | Glyma19g37660        | Glyma.19g192300      | Determination of bilateral symmetry; meristem initiation; negative regulation of biological process; negative regulation of organ growth; organ morphogenesis; polarity specification of adaxial/abaxial axis; regulation of flower development; regulation of meristem growth; regulation of transcription, DNA-dependent; xylem and phloem pattern formation |
| 14         | Glyma19g37680        | Glyma.19g192500      | Biological process; cell differentiation; tissue development |
| 15         | Glyma19g377750       | Glyma.19g192800      | Amylopectin biosynthetic process; carbohydrate metabolic process; plant-type cell wall organization; polysaccharide catabolic process; starch biosynthetic process; starch metabolic process |
| 16         | Glyma19g38730        | Glyma.19g200100      | Anther development; double-strand break repair via homologous recombination; embryo development; embryo sac development; establishment of meiotic sister chromatid cohesion; leaf phyllotactic patterning; meiotic chromosome segregation; mitotic sister chromatid segregation; root development; sister chromatid cohesion |
| 17         | Glyma19g38760        | Glyma.19g200400      | Biological process; vegetative to reproductive phase transition of meristem |
| 18         | Glyma19g38790        | Glyma.19g200700      | Embryo development ending in seed dormancy; mevalonate-independent pathway; ovule development; thylakoid membrane organization; vegetative to reproductive phase transition of meristem |
| 19         | Glyma19g37770        | Glyma.19g193100      | RNA processing; mitotic cell cycle |
| 20         | Glyma19g37870        | Glyma.19g194100      | Axial/abaxial axis specification; autophagy; leaf shaping; leaf vascular tissue pattern formation; meristem maintenance; regulation of meristem growth |
| 21         | Glyma19g37890        | Glyma.19g194300      | Negative regulation of cell aging; negative regulation of flower development; photoperiodism, flowering; protein targeting to vacuole; regulation of flower development; response to sucrose stimulus |
2005). Hence, it is a prerequisite to include heterozygous haplotypes to capture maximum haplotype variation for the analysis of haplotype diversity (Sinha et al. 2020). In our study, the genomic region flanking the most stable significant SNP (AX-93897200) that was in strong LD with 11 significant SNPs formed a single haplotype block. Five haplotypes (Hap-A, Hap-B, Hap-C, Hap-D, and Hap-E) were detected in this haplotype block. The results showed that five haplotypes identified in the block on Chr.19 regulate PH in soybean from dwarf → intermediate → tall. However, our results also reported that these haplotype alleles showed significant environment interaction across six different environments in the regulation of PH. This suggests the presence of haplotype × environment interaction, which needs to be fixed before the utilization of superior haplotype(s) in crop improvement.

A previous study has documented only two haplotypes underlying the Dt1 locus viz., Gm19_Hap42b and Gm19_Hap42a regulating two different phenotypes of PH (Contreras-Soto et al. 2017). This haplotype block showed only two haplotypes (biallelic), and the significant SNP within this region was also biallelic. Therefore, this study showed no difference between the uses of the haplotype or SNP markers within the target region for MAB. The main objective behind the deployment of haplotype markers is that they are multi-allelic and thus provide better opportunity to modify the plant trait. For example, in the present study we identified five different haplotypes for PH underlying the Dt1 locus viz., Hap-A, Hap-B, Hap-C, Hap-D, and Hap-E which regulated five different phenotype classes of PH in soybean. This provides a greater possibility to modify PH in soybean. Therefore, the incorporation of these haplotypes via haplotype-based breeding in soybean breeding programs will be very useful to tailor plant architecture and improve yield and quality.

**Conclusion**

The present study identified 128 significant SNPs and three stable QTLs/genomic regions (qPH2, qPH17, and qPH19) associated with PH, which were consistently detected in at least three environments. Out of these three QTLs, the QTLs co-localizing with qPH2 and qPH19 have been previously detected by other researchers. However, no QTL was found to be associated with qPH17, implying this locus to be a novel genomic region regulating PH. A total of 22 candidate genes underlying these three QTL regions were also prioritized. The putative candidates defined within the intervals of qPH2, qPH17, and qPH19 need further functional validation before they can be used in soybean breeding programs. The haplotype analysis detected five haplotypes (Hap-A, Hap-B, Hap-C, Hap-D, and Hap-E) within the stable genomic region of Chr.19, which regulated PH ranging from dwarf → intermediate → tall. The superior haplotypes identified in the Chinese soybean germplasm will serve as a potential resource for haplotype-based breeding of PH in soybean.

**Supplementary Information** The online version contains supplementary material available at https://doi.org/10.1007/s00122-022-04120-0.

**Author contribution statement** JAB, DY, and RKV conceived and planned the experiments. JAB and DH performed the experiments. KAA, JAB, BK, SAG, RB, and DH analyzed the data. KAA, JAB, BK, and SAG wrote the initial draft of the manuscript. JAB, DY, and RKV supervised the project. All the authors read and revised the final manuscript.

**Funding** Open Access funding enabled and organized by CAUL and its Member Institutions. This work was supported in part by National Natural Science Foundation of China (No. 32090065), Ministry of Science and Technology (No. 2017YFE0111000), and Horizon 2020 of European Union (EUCLEG/ No. 727312).

**Data availability** All relevant data can be found within the manuscript and in Supplementary Material.

**Code availability** The scripts used for performing different analysis in the present study are provided in Supplementary File 1.

**Declarations**

**Conflict of interest** The authors declare no competing interests.

**Ethics approval** The authors declare that no ethical standards have been violated during the course of the study.

**Consent to participate** Not applicable.

**Consent for publication** Not applicable.
Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

Abbai R, Singh VK, Nachimuthu VV, Sinha P, Selvaraj R, Vippalra AK, Singh AK, Singh MS, Varshney RK, Kumar A (2019) Haplotype analysis of key genes governing grain yield and quality traits across 3K RG panel reveals scope for the development of tailor-made rice with enhanced genetic gains. Plant Biotechnol J 17:1612–1622

Allen FL (1994) Usefulness of plant genome mapping to plant breeding. In: Gresshoff PM (ed) Plant genome analysis. CRC Press, Boca Raton, pp 11–18

Bevan MW, Uary C, Wulff BB, Zhou J, Krasileva K, Clark MD (2017) Genomic innovation for crop improvement. Nature 543:346

Bhat JA, Deshmukh R, Zhao T, Patil G, Deokar A, Shinde S, Chaudhary J (2020) Harnessing high-throughput phenotyping and genotyping for enhanced drought tolerance in crop plants. J Biotechnol 324:248–260

Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckstaff R, Reynolds MD, Jung HJ, Zhang Y, lidski H, Markow TA, Miklas PN,例如 J, McImity E, Ornish D, Benner A, Sillers K, Cannon H, Rek KA, Zhang X, Teskey QK, Deloulme J, Shindo S, Perrier X, Kim TW, Schaffner SF, Lee SH, Yakup JS, Yandell M, Painter NW, Beavis WD, Fung A, Ramsay SC, Conyer S, Horvath O, Dernoncourt F, Weigel D, Hansen LC, Bell S, Ecker JR (2009) TASSEL: software for association mapping of complex traits in diverse samples. Theor Appl Genet 23:2633–2635

Cao Y, Li S, He X, Chang F, Kong J, Gai J, Zhao T (2017) Mapping QTLs for plant height and flowering time in a Chinese summer planting soybean RIL population. Euphytica 213:39

Cao Y, Li S, Chen G, Wang Y, Bhat JA, Karikari B, Kong J, Gai J, Zhao T (2019) Deciphering the genetic architecture of plant height in soybean using two RIL populations sharing a common M8206 parent. Plants 8:373

Chang F, Guo C, Sun F, Zhang J, Wang Z, Kong J, He Q, Sharmin RA, Zhao T (2018) Genome-wide association studies for dynamic plant height and number of nodes on the main stem in summer sowing soybeans. Front Plant Sci 9:1184

Chapman A, Pantalone V, Ustun A, Allen F, Landau-Ellis D, Trigiano R, Gresshoff P (2003) Quantitative trait loci for agronomic and seed quality traits in an F2 and F4: 6 soybean population. Euphytica 129:387–393

Chen K, Zhang Q, Wang CC, Liu ZX, Jiang YJ, Zhai LY, Zheng TQ, Xu JL, Li ZK (2019) Genetic dissection of seedling vigour in a diverse panel from the 3000 Rice (Oryza sativa L.) genome project. Sci Rep 9:4804

Chen C, Chen H, Zhang Y, Thomas HR, Frank MH, He Y, Xia R (2020) TBiools: An integrative toolkit developed for interactive analyses of big biological data. Mol Plant 13:1194–1202

Cober ER, Morrison MJ (2010) Regulation of seed yield and agronomic characters by photoperiod sensitivity and growth habit genes in soybean. Theor Appl Genet 120:1005–1012

Contreras-Soto RI, Mora F, de Oliveira MAR, Higashi W, Scapim CA, Schuster I (2017) A genome-wide association study for agronomic traits in soybean using SNP markers and SNP-based haplotype analysis. PLoS ONE 12:e0171105

Du W, Ning L, Liu Y, Zhang S, Yang Y, Wang Q, Chao S, Yang H, Huang F, Cheng H, Yu D (2020) Identification of loci and candidate gene GmSPX-RING1 responsible for phosphorus efficiency in soybean via genome-wide association analysis. BMC Genom 21:1–16

Fang C, Ma Y, Wu S, Liu Z, Wang Z, Yang R, Hu G, Zhou Z, Yu H, Zhang M, Pan Y, Zhou G, Ren H, Du W, Yan H, Wang Y, Han D, Shen Y, Liu S, Liu T, Zhang Q, Qin H, Yuan J, Yuan X, Kong F, Liu B, Li J, Zhang Z, Wang G, Zhu B, Tian Z (2017) Genomewide association studies dissect the genetic networks underlying agronomical traits in soybean. Genome Biol 18:161

Fang Y, Liu S, Dong Q, Zhang K, Tian Z, Li X, Li W, Qi Z, Wang Y, Tian X, Song J, Wang J, Yang C, Jiang S, Li W-X, Ning H (2020) Linkage analysis and multi-locus genome-wide association studies identify QTNs controlling soybean plant height. Front Plant Sci 11:9

Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, Blumenstiel B, Higgins J, DeFelice M, Lochner A, Faggart M, Liu-Cordero SN, Rotimi C, Adyemto A, Cooper R, Ward R, Lander ES, Daly MJ, Altshuler D (2002) The structure of haplotype blocks in the human genome. Science 296:2225–2229

Ganie SA, Ahammed GJ (2021) Dynamics of cell wall structure and related genomic resources for drought tolerance in rice. Plant Cell Rep 40:437–459

Ganie SA, Wani SH, Henry R, Hensel G (2021) Improving rice salt tolerance by precision breeding in a new era. Curr Opin Plant Biol 60:101996

Guan R, Qu Y, Guo Y, Yu L, Liu Y, Jiang J, Chen J, Ren Y, Liu G, Tian L, Jin L, Liu Z, Hong H, Chang R, Gillihan M, Qiu L (2014) Salinity tolerance in soybean is modulated by natural variation in GmSALT3. Plant J 80:937–950

Hyten DL, Song Q, Choi YI, Yoon MS, Specht JE, Matukumalli LK, Nelson RL, Shoemaker RC, Young ND, Cregan PB (2008) High-throughput genotyping with the golden gate assay in the complex genome of soybean. Theor Appl Genet 116:945–952

Kato S, Sayama T, Taguchi-Shiobara F, Kikuchi A, Ishimoto M, Cober E (2019) Effect of change from a deterministic to a semi-deterministic growth habit on the yield and lodging resistance of soybeans in the northeast region of Japan. Breed Sci 69:151–159

Kim KS, Diers BW, Hyten DL, Mian MAR, Shannon JG, Nelson RL (2012) Identification of positive yield QTL alleles from exotic soybean germplasm in two backcross populations. Theor Appl Genet 125:1353–1369

Knapp SJ, Stroup WW, Ross WM (1985) Exact confidence intervals for heritability on a progeny mean basis. Crop Sci 25:192–194

Kraakman ATW, Niks RE, Van den Berg PMMM, Stam P, Van Eeuwijk FA (2004) Linkage disequilibrium mapping of yield and yield stability in modern spring barley cultivars. Genetics 168:435–446

Kulkarni KP, Kim M, Shannon JG, Lee JD (2016) Identification of quantitative trait loci controlling soybean seed weight in recombinant inbred lines derived from PI 483463 (Glycine soja) × Hutcheson (G. max). Plant Breed 135:614–620

Kuroha T, Nagai K, Gamuyao R, Wang DR, Furuta T, Nakamori M, Kitakoa T, Adachi K, Minami A, Mori Y, Mashiguchi K, Seto Y, Yamaguchi S, Kojima M, Skakibara H, Wu J, Ebana K, Mitsuda N, Ohme-Takagi M, Yanagisawa S, Yamasaki M, Yokoyama R, Nishitani K, Moschizuki T, Tamiya G, McCouch SR, Ashikari M (2018) Ethylene-gibberellin signaling underlies adaptation of rice to periodic flooding. Science 361:181–186

Lai Y, Cheng J, He Y, Yang B, Wang Z, Zhan H (2016) Identification of QTLs with additive, epistatic, and QTLx environment interaction effects for seed vigor in rice. Plant Mol Biol Rep 34:160–171

Lark KG, Chase K, Adler F, Mansur LM, Orf JH (1995) Interactions between quantitative trait loci in soybean in which trait variation

 Springer
at one locus is conditional upon a specific allele at another. Proc Natl Acad Sci 92:4656–4660

Lee S, Jun TH, Michel AP, Mian MAR (2015) SNP markers linked to QTL conditioning plant height, lodging, and maturity in soybean. Euphytica 203:521–532

Li JY, Wang J, Zeigler RS (2014) The 3000 rice genomes project: new opportunities and challenges for future rice research. Gigascience 3:8

Li B, Fan S, Yu F, Chen Y, Zhang S, Han F, Yan S, Wang L, Sun J (2017) High-resolution mapping of QTL for fatty acid composition in soybean using specific-locus amplified fragment sequencing. Theor Appl Genet 130:1467–1479

Lipka AE, Tian F, Wang Q, Peiffer J, Li M, Bradbury PJ, Gore MA, Buckler ES, Zhang Z (2012) GAPT: genome association and prediction integrated tool. Bioinformatics 28:2397–2399

Liu BH, Watanabe S, Uchiyama T, Kong FJ, Kanazawa A, Xia ZJ, Nagamatsu A, Ariai M, Yamada T, Kitamura K, Masuta C, Harada K, Abe J (2010) The soybean stem growth habit gene Dt1 is an ortholog of Arabidopsis TERMINAL FLOWER1. Plant Physiol 153:198–210

Liu W, Kim MY, Van K, Lee YH, Li H, Liu X, Lee SK (2011) QTL identification of yield-related traits and their association with flowering and maturity in soybean. J Crop Sci Biotechnol 14:65–70

Liu YL, Li YH, Reif JC, Mette MF, Liu ZX, Liu B, Zhang SS, Yan L, Chang RZ, Qiu LJ (2013) Identification of quantitative trait loci underlying plant height and seed weight in soybean. Plant Genome 6:841–856

Liu J, Dou Y, Batistella M, Challies E, Connor T, Friis C, Milling-Stratton of three related recombinant inbred populations. Crop Sci 153:198–210

Liu J, Guo D, Barry K, Futrell-Griggs M, Abernathy B, Du J, Tian Z, Qiu L, Nelson RL, Clemente TE, Specht JE, Mao D (2014) Genetic mapping of agronomic traits using recombinant inbred lines of soybean. Crop Sci 36:1327–1336

Mishra S, Singh B, Panda K, Singh BP, Singh N, Misra P, Rai V (2016) Association of SNP haplotypes of HKT family genes with salt tolerance in Indian wild rice germplasm. Rice 9:15

Orf JH, Chase K, Jarvik T, Mansur LM, Cregan PB, Adler FR, Lark KG (1999) Genetics of soybean agronomic traits: I. comparison of three related recombinant inbred populations. Crop Sci 39:1642–1651

Patil G, Do T, Vuong TD, Villiyodan B, Lee J-D, Chaudhary J, Shannong JG, Nguyen HT (2016) Genomic-assisted haplotype analysis and the development of high-throughput SNP markers for salinity tolerance in soybean. Sci Rep 6:19199

Ping J, Liu Y, Sun L, Zhao M, Li Y, She M, Sui Y, Lin F, Liu X, Tang Z, Nguyen H, Tian Z, Qiu L, Nelson RL, Clemente TE, Specht JE, Ma J (2014) D12 is a gain-of-function MADS-domain factor gene that specifies semi-determinacy in soybean. Plant Cell 26:2831–2842

Ren D, Wang X, Yang M, Yang L, He G, Deng XW (2019) A new regulator of seed size control in Arabidopsis identified by a genome-wide association study. New Phytol 222:895–906

Schmutz J, Cannon SB, Schlueter J, Ma JX, Mitros T, Nelson W, Hyten DL, Song QI, Thelen JJ, Cheng J, Xu D, Hellsten U, May GD, Yu Y, Sakurai T, Umezawa T, Bhattacharryya MK, Sandhu B, Villiyodan B, Lindquist E, Peto M, Grant D, Shu S, Goodstein D, Barry K, Futrell-Griggs M, Abernathy B, Du J, Tian Z, Zhu L, Gill N, Joshii T, Libault M, Sethuraman A, Zhang XC, Shinozaki K, Nguyen HT, Wing RA, Cregan P, Specht J, Grimwood J, Rokhsar D, Stacey G, Shoemaker RC, Jackson SA (2010) Genome sequence of the palaeopolyploid soybean. Nature 465:120

Severin AJ, Woody JL, Bolon YT, Joseph B, Diers BW, Farmer AD, Muehlbauer GJ, Nelson RT, Grant D, Specht JE, Graham MA (2010) RNA-seq atlas of Glycine max: a guide to the soybean transcriptome. BMC Plant Biol 10:1–16

Sinha P, Singh VK, Saxena RK, Khan AW, Abbari R, Chitikineni A, Desai A, Molla J, Upadhyaya HD, Kumar A, Varshney RK (2020) Superior haplotypes for haplotype-based breeding for drought tolerance in pigeonpea (Cajanus cajan L.). Plant Biotechnol J 18:2482–2490

Sonah H, O’Donoghue L, Cober E, Rajcan I, Belzile F (2015) Identification of loci governing eight agronomic traits using a GBS-GWAS approach and validation by QTL mapping in soya bean. Plant Biotechnol J 13:211–221

Song Q, Jenkins J, Jia G, Hyten DL, Pantalone V, Jackson SA, Schmutz J, Cregan PB (2016) Construction of high-resolution genetic linkage maps to improve the soybean genome sequence assembly Glyma.01. BMC Genomics 17:1–11

Sun D, Li W, Zhang Z, Chen Q, Ning H, Qiu L, Sun G (2006) Quantitative trait loci analysis for the developmental behavior of soybean (Glycine max L. Merr.). Theor Appl Genet 112:665–673

Tao T, Zhao L, Lv Y, Chen J, Hu Y, Zhang T, Zhou B (2013) Transcriptome sequencing and differential gene expression analysis of delayed gland morphogenesis in Gossypium australis during seed germination. PLoS ONE 8:e75323

Thudi M, Chen Y, Pang J, Kalavikatte D, Bajaj P, Roorkiwal M, Chitikineni A, Ryan MH, Lambers H, Siddique KHM, Varshney RK (2021) Novel genes and genetic loci associated with root morphological traits, phosphorus-acquisition efficiency and phosphorus-use efficiency in chickpea. Front Plant Sci 12:636973

Varshney RK, Saxena RK, Upadhyaya HD, Khan AW, Yu Y, Kim C, Rathore A, Kim D, Kim J, An S, Kumar V, Anuradha G, Yamini KN, Zhang W, Muniswamy S, Kim J-S, Pennetsa RV, Wettberg E, Datta SK (2017a) Whole-genome resequencing of 292 pigeonpea genotypes identifies genomic regions associated with domestication and agronomic traits. Nat Genet 49:1082

Varshney RK, Shri C, Thudi M, Mariac C, Wallace J, Qi P, Zhang H, Zhao Y, Wang X, Rathore A, Srivastava RK, Chitikineni A, Fan G, Bajaj P, Punnuri S, Gupta SK, Wang H, Jiang Y, Couederc M, Katta MAVSK, Fadel DR, Mangra KD, Chen W, Harris-Shultz RK, Garg V, Desai N, Doddhamani D, Kane NA, Conner JA, Ghatke A, Chaturvedi P, Subramaniam S, Yadav OP, Berthouly-Salazar C, Hamidou F, Wang J, Liang X, Cloutaup J, Upadhyaya HD, Cubry P, Rhoné B, Gueye MC, Sunkar R, Dupuy C, Sparvoni F, Cheng S, Mahala RS, Singh B, Yadav BS, Lyons E, Datta SK, Hash CT, Devos KM, Buckler E, Benmetenet JL, Paterson AH, Ozias-Akins P, Grando S, Wang J, Mohapatra T, Weckwerth W, Reif JC, Liu X, Viguouroux Y, Xu X (2017b) Pearl millet genome sequence provides a resource to improve agronomic traits in arid environments. Nat Biotechnol 35:969–976

Varshney RK, Thudi M, Roorkiwal M, He W, Upadhyaya HD, Wang B, Bajaj P, Cubry P, Rathore A, Jian J, Doddhamani D, Khan AW, Garg V, Chitikineni A, Xu D, Gaur PM, Singh NP, Chaturvedi SK, Nadigatla GVPR, Krishnamurthy L, Dixit GP, Fikre A, Kimurto PK, Sreeman SM, Bhardwaj C, Tripathi S, Wang J, Lee SH, Edwards D, Polavarapu KKB, Pennetsa RV, Crossa J, Nguyen HT, Siddique KHM, Colmer TD, Sutton T, von Wettberg E, Viguouroux Y, Xu X, Liu X (2019a) Resequencing of 429 chickpea accessions from 45 countries provides insights into genome diversity, domestication and agronomic traits. Nat Genet 51:857–864

Varshney RK, Pandey MK, Bohra A, Singh VK, Thudi M, Saxena RK (2019b) Toward the sequence-based breeding in legumes in the post-genome sequencing era. Theor Appl Genet 132:797–816
Varshney RK, Bohra A, Yu J, Graner A, Zhang Q, Sorrells ME (2021) Designing future crops: genomics-assisted breeding comes of age. Trends Plant Sci 26:631–649

Wang D, Zhu J, Li Z, Paterson AH (1999) Mapping QTLs with epistatic effects and QTL × environment interactions by mixed linear model approaches. Theor Appl Genet 99:1255–1264

Wang D, Graef GL, Procopiuk AM, Diers BW (2004) Identification of putative QTL that underlie yield in interspecific soybean backcross populations. Theor Appl Genet 108:458–467

Wang J, Chu S, Zhang H, Zhu Y, Cheng H, Yu D (2016) Development and application of a novel genome-wide SNP array reveals domestication history in soybean. Sci Rep 6:1–10

Wang X, Pang Y, Zhang J, Wu Z, Chen K, Ali J, Ye G (2017) Genome-wide and gene-based association mapping for rice eating and cooking characteristics and protein content. Sci Rep 7:17203

Xia Z, Watanabe S, Yamada T, Tsubokura Y, Nakashima H, Zhai H, Anai T, Sato S, Yamazaki T, Lu S, Wu H, Tabata S, Harada K (2012) Positional cloning and characterization reveal the molecular basis for soybean maturity locus E1 that regulates photoperiodic flowering. Proc Natl Acad Sci USA 109: E2155–E2164

Xu Y, Crouch JH (2008) Marker-assisted selection in plant breeding: from publications to practice. Crop Sci 48:391–407

Xue H, Tian X, Zhang K, Li W, Qi Z, Fang Y, Li X, Wang Y, Song J, Li WX, Ning H (2019) Mapping developmental QTL for plant height in soybean [Glycine max (L.) Merr.] using a four-way recombinant inbred line population. PLoS ONE 14:e0224897

Yang J, Zhu J (2005) Methods for predicting superior genotypes under multiple environments based on QTL effects. Theor Appl Genet 110:1268–1274

Yin Z, Qi H, Chen Q, Zhang Z, Jiang H, Zhu R, Hu Z, Wu X, Li C, Zhang Y, Liu C, Hu G, Xin D, Qi Z (2017) Soybean plant height QTL mapping and meta-analysis for mining candidate genes. Plant Breed 136:688–698

Zaitlen NA, Kang HM, Feolo ML, Sherry ST, Halperin E, Eskin E (2005) Inference and analysis of haplotypes from combined genotyping studies deposited in dbSNP. Genome Res 15:1594–1600

Zhang WK, Wang YJ, Luo GZ, Zhang JS, He CY, Wu XL, Gai JY, Chen SY (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 108:1131–1139

Zhang JP, Song QJ, Cregan PB, Nelson RL, Wang XZ, Wu JX, Jiang GL (2015) Genome-wide association study for flowering time, maturity dates and plant height in early maturing soybean (Glycine max) germplasm. BMC Genomics 16:1

Zhang X, Wang W, Guo N, Zhang Y, Bu Y, Zhao J, Xing H (2018) Combining QTL-seq and linkage mapping to fine map a wild soybean allele characteristic of greater plant height. BMC Genomics 19:226

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.