

Article

Not peer-reviewed version

Genome-Wide Association Analysis for Yield-Related Traits and Candidate Genes in Vegetable Soybean

[Hongtao Gao](#) , Guanji Wu , Felfei Wu , Yonggang Zhou , Xunjun Zhou , [Keheng Xu](#) , Yaxin Li , Wenping Zhang , Kuan Zhao , Jing Yan , Chen Feng , [Nan Wang](#) ^{*} , [Haiyan Li](#) ^{*}

Posted Date: 12 February 2024

doi: 10.20944/preprints202402.0682.v1

Keywords: Vegetable soybean; plant height; agronomic trait; fresh pods height; GWAS.



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Article

Genome-Wide Association Analysis for Yield-Related Traits and Candidate Genes in Vegetable Soybean

Hongtao Gao ¹, Guanji Wu ¹, Feifei Wu ¹, Xunjun Zhou ¹, Yonggang Zhou ¹, Keheng Xu ¹, Yaxin Li ¹, Wenping Zhang ¹, Kuan Zhao ², Yan Jing ¹, Chen Feng ¹, NanWang ^{2,*} and HaiyanLi ^{1,*}

¹ School of Breeding and Multiplication (Sanya Institute of Breeding and Multiplication), Hainan University, 572025, China.

² Changchun Academy of Agricultural Science, Changchun 130118, China

* Correspondence: hylh@hainanu.edu.cn (H.L.); 759774740@qq.com (N.W.)

Abstract: Owing to the rising demand for vegetable soybean products, there is an increasing need for high-yield soybean varieties to raise the output. However, complex correlation patterns among quantitative traits with genetic interactions bring a challenge for vegetable soybean breeding. Here, a genome-wide association study (GWAS) was applied for six yield-related traits with 188 vegetable soybean accessions. A total of 81 SNPs were identified for plant height, pod length, pods number, pod thickness, pod width, and fresh pod weight by Blink model. Among these significant SNPs, 79 were novel SNPs, while 2 overlapped with previously reported SNPs. A total of 220 genes were found in 100 kb upstream and downstream regions of significant SNPs, of which 17 genes were functional proteins. Among these function proteins, four candidate genes, *Glyma.13G109100*, *Glyma.03G183200*, *Glyma.09G102200*, and *Glyma.09G102300* were analyzed as significant haplotype variations, which encode MYB-related transcription factor, auxin-responsive family protein, F-box protein and CYP450, respectively. In addition, the relative expression with four candidate genes in the V030 and V071 (Among them, the plant height, pod number and fresh pod weight of V030 were lower than V071 strains) vegetable soybean were significantly different and these genes could be involved in plant growth and development via various pathways. Therefore, we proposed four genes from vegetable soybean germplasm that could be the key candidate genes for pod yield and plant height and these studies will accelerate the process of high-yield vegetable soybean breeding.

Keywords: Vegetable soybean; plant height; agronomic trait; fresh pods height; GWAS

1. Introduction

Soybean (*Glycine max*(L.) Merr.) is widely grown oil crop in the world. It is also an important source of edible oil and high-quality vegetable protein [1]. According to the different harvesting time and use of soybeans, it can be divided into two crops: food and vegetables. Vegetable soybeans are harvested at the R6 stage, when the pods are green and the seeds are full^[2]. Research has reported that the pods size, including large pods and grains, is an important visual quality unique to vegetable soybeans^[3,4]. Therefore, the size of vegetable soybean pods has been considered to be one of the most important traits in accelerating the breeding process of vegetable soybean. In addition, the yield of vegetable soybean was also affected by plant height, pod number, pod width, pod thickness, pod length and fresh pod weight per plant. With the development of economy and the growth of world population, the demand for vegetable soybean has increased dramatically [7], but there are fewer reports on the vegetable soybean yield-related trait at present. Therefore, it is necessary to analysis the genetic basis of soybean yield in R6 (maturity stage of vegetable soybean market) and improve the vegetable soybean yield.

Understanding the genetic variations related to yield traits was a necessary basis for breeding^[8]. Planting structure always been a key factor affecting the planting density and seed yield of soybeans in the field. The ideal soybean plant structure not only optimizes the structure of canopy, prevents

lodging, improves photosynthetic efficiency, also achieves the higher yield [7,8]. As a kind of special soybean, the yield trait of vegetable soybean was accordance as cultivated soybean, including the soybean yield per plant (YP), pod number (PN), pod length (PL), pod width (PW), pod thick (PT), plant height (PH) and branch number (BN). Therefore, dissecting the important SNP loci associated for yield traits has become an essential topic in the vegetable soybean breeding^[9-12].

Genome-wide association study (GWAS) is highly effective method to dissect the natural variation of quantitative traits based on linkage disequilibrium, which can provide a theoretical basis for analyzing the genetic structure and molecular improvement of soybean complex traits^[13-15]. For instance, there are 125 candidate selection regions and five potential candidate genes about agronomic traits were predicted using the GWAS [16]. Zhang et al. identified 22 loci of minor effect and predicted 3 important candidate genes on chromosome 19 using genome-wide association studies [17]. A total of 58 SNPs significantly correlated with seed weight (SW), internode number (IN) and plant height (PH) were identified by GWAS, which 28 candidate genes about yield trait were predicted [18]. In addition, 27 quantitative trait nucleotides (QTNs) associated with seed size correlations were identified using GWAS [19,20]. Although plant architecture and yield-traits about soybean have been verified, the mining of yield traits related genes in vegetable soybean (maturity stage of vegetable soybean market) was rarely reported.

Therefore, in this study, we collected and evaluated 188 diverse vegetable soybean genotypes for six yield-related traits, including number of pods, plant height, fresh pod weight, pod width, pod thick and pod length. Furthermore, GWAS was used to identify the genetic loci and candidate genes for yield traits, which provide a theoretical support for improving the yield of vegetable soybean.

Results

2.1. Phenotype Variations of Yield-Related Traits in Vegetable Soybean

To identify important yield-related genes from vegetable soybean germplasm resources, we conducted whole-genome re-sequencing and GWAS of a panel of 188 vegetable soybean accessions which include 115 and 58 genotypes originating from South and North in the China, and 14 genotypes for American, India, Brazil, Thailand, etc (Table S1). In this vegetable soybean, 44 accessions were landrace, and 144 were improved cultivars. Six important yield-related traits were investigated in R6 stage, including the plant height, fresh pod weight per plant, pod length, pod number per plant, pod width and pod thickness of vegetable soybean. These vegetable soybean germplasms were plant and conducted at Yazhou of Sanya, China, in 2023. Taking the average value of yield related traits of vegetable soybeans during the R6 period as phenotypic data, including the fresh pod weight, plant height, pod width, pod length and pod thickness, which present a normal distribution pattern(Fig.1). Besides, these yield-related traits were closely positive correlation to each other, except for the size of pod and pod number. The fresh pod weight was positively correlated with pod length, pod width, pod thick, plant height and pod number.

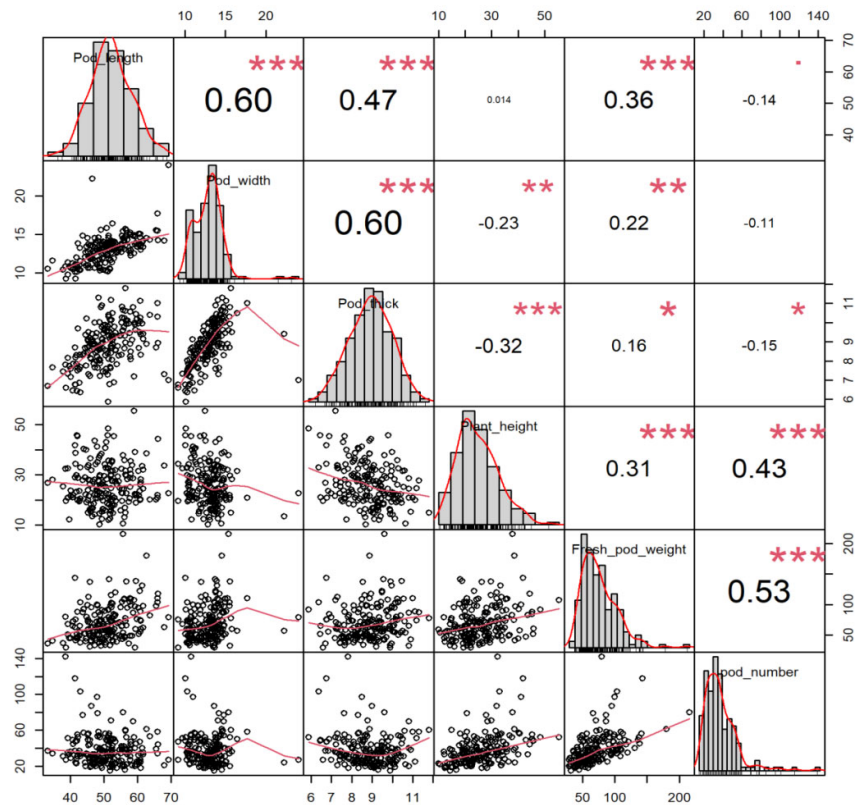


Figure 1. Correlation analysis of 188 vegetable soybean accessions between traits.

Plant height, pods number, fresh pods weight, pod width, pod thick, pod length were used for morphological analyses. One-way ANOVA was used to generate the *P* values. (*, **, ***:indicate *P* < 0.05,0.01, 0.001 respectively).

In addition, it can be seen that the variation coefficients of pod length, pod width, pod thickness and stem diameter were small (12.75%-14.65%), the variation coefficients of pod number, plant height, and fresh pod weight per plant were moderate (19.41%-40.91%), indicating that the variation potential of these yield related traits was large. It provides a rich genetic background for the selection of breeding materials (Table 1).

Table 1. Statistics and difference analysis of yield-related traits in the 188-vegetable soybean accessions.

Trait	Max	Min	Mean	SD	CV(%)
Pod length	69.33	33.30	50.90	6.73	13.22%
Pod width	24.04	9.19	12.97	1.90	14.65%
Pod thick	11.80	5.87	8.84	1.13	12.75%
Plant height	55.45	10.10	26.18	8.40	32.08%
Pods number	15.00	4.50	8.57	1.66	19.41%
Fresh pod weight	215.67	29.30	70.34	28.78	40.91%

2.2. Population Structure Analysis of Vegetable Soybean

Further genetic differentiation analysis was conducted for samples from North and South in China as well as other abroad countries. Ten subgroups were analyzed for the 188 genotypes vegetable soybean in the Admixture analysis (Fig.2A-B). The trait association mapping was performed at $k = 10$. According to the phylogenetic tree, there are ten main clusters; every group corresponded to a major subgroup of the admixture analysis, which divides the population into ten major groups (Fig. 2C). In addition, 188 vegetable soybean accessions were assessed for linkage disequilibrium (LD) using a subset of high-quality markers. At a threshold of $R^2 = 0.317$, the average decay distance of LD was 150.00kb for all 188 vegetable soybean accessions (Fig.2D). The above results indicate that GWAS can be used to identify significant marker trait associations in 188 vegetable soybeans.

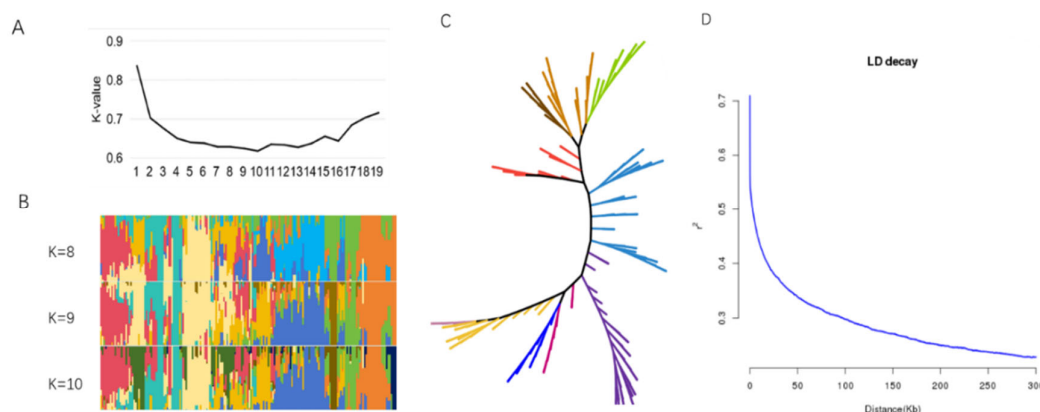


Figure 2. Population structure and linkage disequilibrium (LD) analysis of 188 vegetable soybeans. (A) Cross validation error rate for 188 samples based on clustering. (B) Bar plot divides the population into 10 cluster in this 188-vegetable soybean. (C) Phylogenetic tree of 188 soybean germplasm. (D) LD analysis for the 188 germplasm vegetable soybean.

2.3. Yield-Related SNPs were Identified in Vegetable Soybean via GWAS

In this paper, Blink model was used to analyze yield traits of 188 vegetable soybean strains. A total of 81 significantly yield related SNPs were identified and p values were less than 0.00001 (Table 2). Of these 81 SNPs, 79 were novel, 2 SNPs overlapped with previously reported SNPs. Most of these SNPs have a positive effect on the vegetable soybean yield traits. Among these significantly SNPs, 11 SNPs of plant height were identified and located on chromosomes 1, 5, 7, 8, 10, 12, 13, 14, 17, 19 and 20; for the fresh weight of pods, 21 SNPs were detected and located on chromosomes 2, 5, 6, 7, 9, 10, 11, 14, 15, 16, 18, 19 and 20; for the pod of number, 19 SNPs were located on chromosomes 3, 7, 10, 12, 17 and 20; 9 SNPs of pod width were identified and located on chromosomes 4, 8, 9, 10, 11 and 20; 10 SNPs of pod thick were detected, which were located on chromosomes 1, 3, 6, 8, 11, 13, 18 and 20; 11 SNPs of pod length were detected and located on chromosomes 4, 14, 17 and 20.

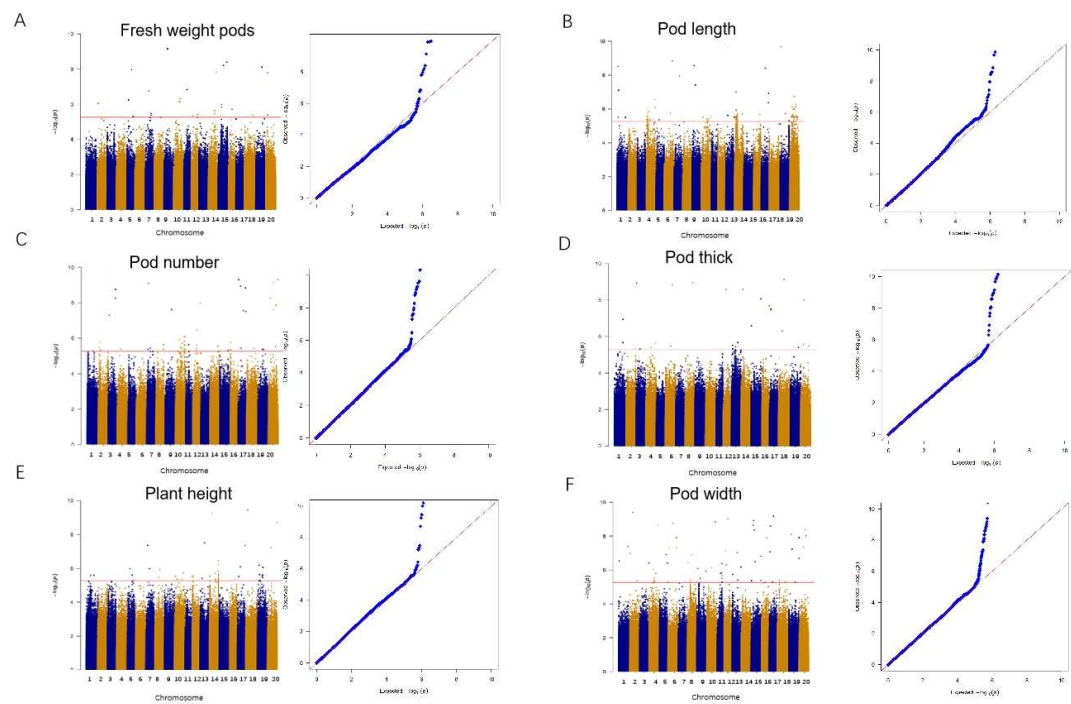


Figure 3. Genome-wide association analysis Manhattan plots and QQ plots of 188 vegetable soybean accession for fresh pod weight (A), pod length (B), pod number (C), pod thick (D), plant height (E), and pod width (F).

2.4. Candidate Genes Analysis Involved Ofyield Related Traits in Vegetable Soybean

In total, 81 significant SNPs were detected using Blink model and further used to analyzed candidate genes about vegetable soybean yield-trait. The candidate genes associated with vegetable soybean yield traits in the SNP regions was analyzed using the *Glycine max* reference genome database. Meanwhile, through the expression patterns and functional annotations of these genes by the website of phytozome and Dicots Plaza 5.0, a total of 17 genes were able to initially forecast and functional annotations are listed in Table 2.

Table 2. List of significant SNPs associated with vegetable soybean yield-related traits and candidate gene.

Trait	SNP	-log10(P)	Candidate	Gene annotation
Plant height	Chr7:8350061	4.49E-08	Glyma.07G089000	Vernalization-insensitive protein3
Plant height	Chr13:22150035	3.17E-08	Glyma.13G107400	Myosin-11-RELATED
Plant height	Chr13:22150035	3.17E-08	Glyma.13G109100	MYB-related transcription factors
Plant height	Chr14:44580744	1.29E-12	Glyma.14G182400	Hydroxyproline-rich glycoprotein family protein
Plant height	Chr14:14120295	5.18E-10	Glyma.14G116600	TPX2 protein family
Pod width	Chr03:24477441	1.8E-10	Glyma.03G211200	CW-type zinc-finger protein

Pod width	Chr04:45822312	2.28E-9	Glyma.04G187000	Histone deacetylase 2
Pod length	Chr14:8677626	9.71E-08	Glyma.14g093600	Myc down regulated-like protein
Pod number	Chr03:39469452	1.94E-15	Glyma.03G182100	Small auxin-up RNA
Pod number	Chr03:39469452	1.94E-15	Glyma.03G183200	Auxin-responsive family protein
Pod number	Chr07:17801936	3.83E-11	Glyma.07G148100	Terminal domain phosphatase like 2
Fresh pod weight	Chr09:18491673	3.00E-05	Glyma.09G102300	F-box protein
Fresh pod weight	Chr09:18491673	3.00E-05	Glyma.09G101100	WAT1-related protein
Fresh pod weight	Chr09:18491673	3.00E-05	Glyma.09G101200	Transcriptional regulator SNIP1
Fresh pod weight	Chr09:18491673	3.00E-05	Glyma.09G101300	Solute carrier family 35, member C2 (SLC35C2)
Fresh pod weight	Chr09:18491673	3.00E-05	Glyma.09G102200	CYP72A154

The SNP for plant height-22150035, pod number-39469452, fresh pod weight-18491673 was significant loci with close genetic relationship. During haplotype analysis, four candidate genes were identified with two or three different haplotypes(Fig.4). The SNP for plant height (22150035) on the chromosome 13, the candidate gene *Glyma.13g109100* has three haplotypes, 22.34% of landraces and 34.5% of improved cultivars possessed Hap1(TA), which had greater plant height than Hap3(AA). The SNP for pod number (39469452) on the chromosome 3, the candidate gene *Glyma.03g138200* has three haplotypes, 8% of landraces and 26% of improved cultivars possessed Hap1(GTTCAG), which had greater pod numbers than Hap3(CATCCA). At the same time, the SNP for fresh pod weight (18491673) on the chromosome 9, the two candidate genes were identified with two haplotypes. Among these two genes, *Glyma.09g102200* was identified three haplotypes, 13% of landraces and 34% of improved cultivars possessed Hap1(TCC), which had greater fresh pod weight than Hap3(TCT); *Glyma.09g102300* was identified two haplotypes, 22.8% of landraces and 50% of improved cultivars possessed Hap1(CT), which had greater fresh pod weight than Hap3(AA).

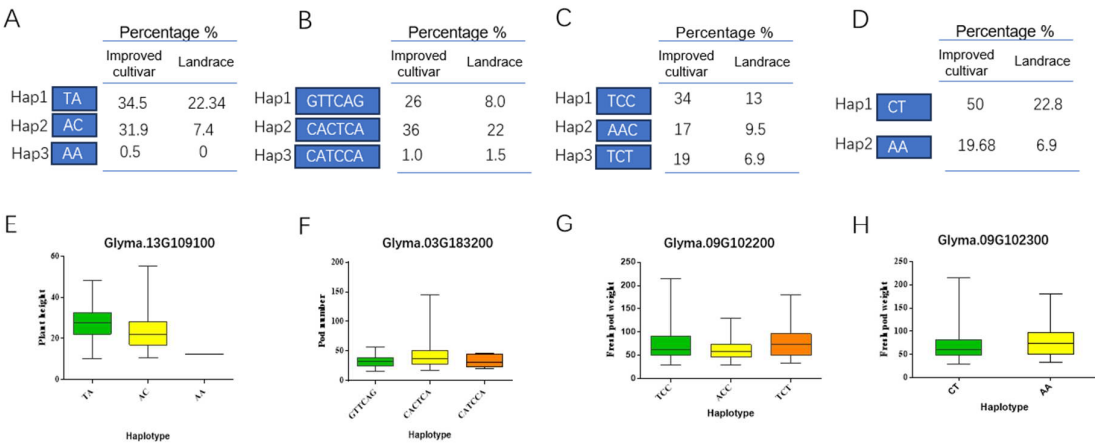


Figure 4. Haplotype analysis of candidate genes for yield-related traits in 188 vegetable soybean. A-E: Analysis for the haplotype analysis of plant height candidate genes *Glyma.13g109100* in the 188 vegetable soybean ($p<0.05$); B-F: Analysis for the haplotype of pod number candidate genes *Glyma.03g183200* in the 188 vegetable soybean ($p<0.05$); C-G: Analysis for the haplotype of fresh pod weight candidate genes *Glyma.09g102200* in the 188 vegetable soybean ($p<0.05$); D-H: Analysis for the haplotype of fresh pod weight candidate genes *Glyma.09g102300* in the 188 vegetable soybean ($p<0.05$).

2.5. Different Expression Pattern of Candidate Genes were Observed in Pods and Stem

The agronomic characters of 188 vegetable soybeans were analyzed statistically, we found that the plant height, pod number and fresh pod weight of V030 were weaker than those of V071. Therefore, to further confirm that the candidate genes regulate the plant height, fresh pods weight and pods numbers of vegetable soybean, the expression patterns of the candidate genes *Glyma.09g102300*, *Glyma.09g102200*, *Glyma.13g109100* and *Glyma.03g138200* were tested via qRT-PCR in the pods and stem between V30 and V71 at R5, R6 and R7stage. Relative expression of the potential candidate genes associated with fresh pod weight, *Glyma.09g102200* and *Glyma.09g102300* showed significant differences between V071 and V030 at the R5, R6 and R7 stages ($P \leq 0.05$) (Fig. 5). The plant height candidate genes *Glyma.13g109100* showed significant different expression pattern between V071 and V030 in stem at the R6 stages ($P \leq 0.05$). The relative expression for candidate genes of *Glyma.03g138200* showed significant differences between V071 and V030 at the R5, R6 and R7 stages ($P \leq 0.05$). Therefore, we assumed that *Glyma.13g109100*, *Glyma.03g183200*, *Glyma.09g102300* and *Glyma.09g10220000* may be the regulatory genes for vegetable soybean yield traits.

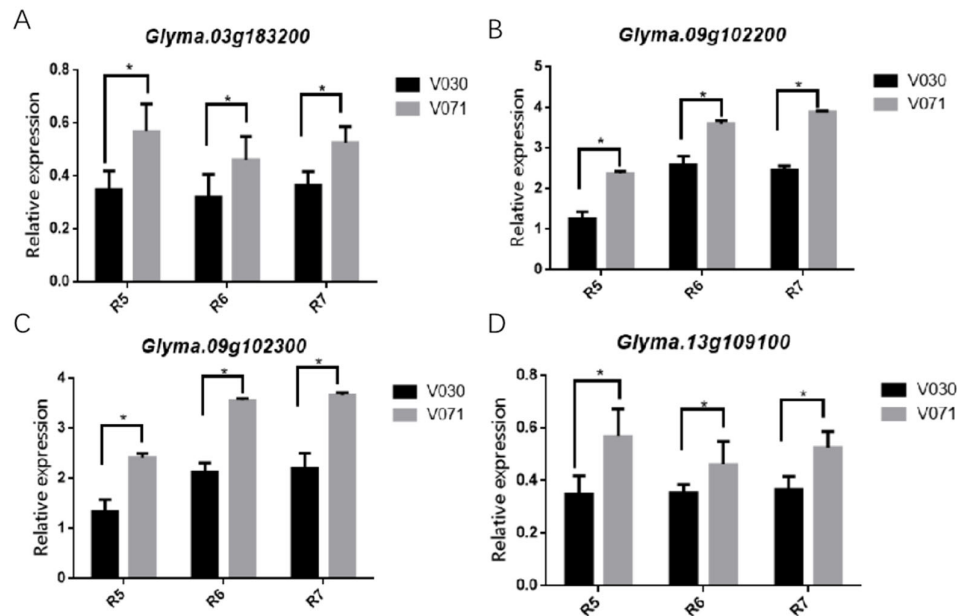


Figure 5. Expression analysis of potential candidate genes in V030 and V071 at three growth developmental stages. A-C: Analysis for the expression levels of vegetable soybean pod number and fresh pod weight candidate gene *Glyma.03g183200*, *Glyma.09g102200* and *Glyma.09g102300* in the pods between V030 and V071 strains at the stage of R5, R6, R7.; D: Analysis for the expression levels of plant height candidate genes *Glyma.13g109100* in the stem between V030 and V071 strains at the stage of R5, R6, R7.

3. Discussion

Vegetable soybeans are a very important type of legume vegetable, especially in Asian countries such as Japan and China. Due to its superior nutrition, appearance, and taste, the global demand for plant-based soybeans continues to grow. Since the 1990s, the demand for plant-based soybeans in the United States has been increasing, reaching 10000 tons in 2000 [22,23]. However, due to the lack of good varieties, the demand for vegetable soybeans cannot be met. China is the country with origin of soybeans and has the largest soybean genetic resources in the world. Based on the abundant soybean resources in China, the yield genetic structure of vegetable soybean was analyzed by GWAS to provide beneficial genes, functional markers and specific materials for molecular breeding.

At present, GWAS method has been used to analyze and calculate the association between genotype and phenotype variation, and dissecting the genetic basis of important traits [24–26]. In the study conducted by Zeng et al[27], phenotype variation of the association panel of grain yield in plant (GYP) and tassel branch number (TBN) was 42.37% and 49.79%; respectively; the CV of grain width (GW) and 100 kernel weight (HKW) were 11%, 19%. Meanwhile, some study shown that the phenotype variation of GYP is 40%; the CV of HKW is 17%; the CV of kernel number per row (KNR) was 28%, thus, significant phenotype variation in the population will be beneficial for analyzing the genetic structure of yield traits[27]. Therefore, the plant structure, especially the number of nodes and branches, largely determines the pods number and yield of soybeans. Other plant structural, including the pods of number and the size of pods is another major component that affects the overall structure of the plant and grain yield. It is great significance to explore the yield related trait candidate genes of vegetable soybean for the evaluation of vegetable soybean yield and genetic breeding.

In this paper, 188 vegetable soybean accessions were divided into 9 categories by population structure analysis, which indicating that there was some variation within the population. In addition, these results were similar to phylogeny analyses and it could help prevent false positives from the result of GWAS[28]. Besides, in the process of population analysis, the acceptable distance between candidate genes and markers were determined by LD, the variation of LD was varies with different populations[29]. And the LD decay distance of 188 vegetable soybean varieties in this study was

150.00 kb ($r^2 = 0.375$), within the previously reported range (90 kb to 574 kb)^[30–32]. Therefore, it is sufficient for a valid GWAS analysis.

In the previously reported, the QTL region related to plant height was located in Chr13:21,937,082–23,937,081 [31,32]. In this paper, a SNP (Chr13:22150035) identified from plant height trait, is also located in this interval, and *Glyma.13G109100* is located at 154.23kb away from chr13:22150035. Besides, *Glyma.13G109100* showed a higher expression level in long-stalked variety compared to short-stalked variety in our study (Fig. 5), indicates that it may differentially regulate. This gene encodes MYB-related transcription factors, a very conserved transcription factor family in eukaryotes, which involved in many developmental processes, such as root hair development, pollen formation, seed germination, flower stem strength, yield, etc. It also plays a role in abiotic stresses such as drought, ultraviolet light, cold stress, high temperature stress and salt stress [33,34]. MYB-related transcription factors also could reduction the cell size regulate plant structure such as internode length, petiole length, leaf area, and plant height via brassin steroid (BR) pathway^[33]. The studies also shown that *OsMYB110* not only changes the plant height, but also endows rice yield by increasing the number of grains per panic and grain size^[35]. Therefore, *Glyma.13G109100* was predicted to be the candidate gene to regulate the vegetable soybean height.

In the recently report, there are 294 seed-weight related SNPs were identified on the soybean chromosomes (<http://www.soybase.org/>) [36]. Besides, some SNPs have been identified that were correlated with yield traits, such as seed size, maturity, flowering time and plant height. For the seed weight, Chr09:18491673 overlaps with a known reported SNP locus that controls seed weight [38]. Combine with annotation information analysis, *Glyma.09G102300* and *Glyma.09G102200* were identified as a candidate gene for controlling seed yield in vegetable soybean. Among them, *Glyma.09G102300* mainly encodes F-box related protein and F-box proteins are a family of proteins that exist widely in eukaryotes and contain the F-box domain. It plays an important role in cell cycle regulation, transcription regulation, apoptosis, cell signal transduction, growth and development [39,40]. In addition, studies have shown that F-box proteins OsFBX206, OsFBK12 can regulate the synthesis of ethylene and boisterousness for improving the grain size and yield in rice^[41–43]. Besides, we found that NADH dehydrogenase1 alpha sub-complex was predicted to be an interacting protein of the fresh pod weight candidate gene *Glyma.09G102300*. The stability of NADH dehydrogenase1 has been demonstrated that play an important role to regulate seed germination, yield and growth retardation in Arabidopsis and maize [44,45]. In addition, As a candidate gene related to seed yield of vegetable soybean, *Glyma.09G102200* encodes cytochrome P450, which is a large family of self-oxidizing heme proteins. As a member of this family, CYP78A has been shown to regulate grain morphology and yield per plant in Arabidopsis [46], soybean [47], rice [48], and wheat [49]. Therefore, *Glyma.09g102300* and *Glyma.09g102200* were predicted to be the candidate gene to regulate the vegetable soybean pod weight.

4. Method

4.1. Plant Material and Field Experiments

The vegetable soybean in this study consisted of 188 different genotypes, 173 of them from 25 provinces in China and 15 genotypes from the Japan, Brazil, United States, etc and these vegetable soybean were preserved in our laboratory. The seeds were sown in 3.0 m row and 0.5 m row spacing. As previously described that soybean germplasm planting follows normal agronomic practices [50].

4.2. Vegetable Soybean Yield-Related Trait Data Analysis

Vegetable soybean yield traits measured in this study included plant height, pod number, fresh pod weight, pod width, pod thick and pod length at the R6 stage. The standard deviation and coefficient of variation of vegetable soybean yield trait were calculated by SPSS16 software. Among them, the coefficient of variation = SD (Standard Deviation)/Mean.

4.3. Population Analysis

The 188 accessions of soybean were re-sequenced approximately 10Gb raw data were collected each genome. TRIMMOMATIC (v.0.36) were used to filter the adaptors and low-quality bases of raw data [51], and then BOWTIE 2 (v.2.4.1) were used to map clean data to Williams82.v2.1 genome with the default parameters [52]. The sentieon software (<https://www.sentieon.com/>) were used for global realignment of reads and produce VCF files. VCF tools were further employ to filter the raw VCF file with parameters (minDP 4, min-alleles 2, max-alleles 2, minQ 30, maxDP 60, max missing 0.9, maf 0.05), nucleotide diversity (p) (with 500 kb window no overlap), and Tajima's D(with 500 kb window no overlap) [53]. Population structure analysis was performed using the admixture [54,55]. A phylogenetic tree was constructed using the neighbor-joining method implemented in iTOL [56]. PopLDdecay was used to analysis population [57].

4.4. Candidate Gene Analysis and Gene-Based Association Mapping

All of the potential candidate genes were extracted within 100 kb with the detected SNP sites. Gene expression and annotation analysis was obtained from the soybean database (<https://www.soybase.org/>), which related to the vegetable soybean yield-related trait. At the same time, the haplotypes analysis was also conducted to identify the important genes that regulate vegetable soybean yield-related traits.

4.5. Real-Time Quantitative Reverse Transcription PCR Analysis for the Yield-Related Candidate Genes of Vegetable Soybean

According to the yield data of vegetable soybean, we selected vegetable soybean V071 and V031 with large differences in plant height, pod number and fresh pod weight. To analyze the expression patterns of candidate genes in V070 and V031 materials, the pod and main stem at R5, R6 and R7 periods was extracted and expression levels of yield related candidate genes in R5, R6 and R7 growth stages of vegetable soybean were analyzed. The quantitative primers are shown in Table 3.

5 Conclusions

In this paper, we genotyped 188 vegetable soybean and identified 11, 19, 21 SNP associated with plant height, pod number and fresh pod weight by GWAS, respectively. Based on gene annotation and expression analysis, we proposed *Glyma.09G102300*, *Glyma.09G102200*, *Glyma.03g183200* and *Glyma.13G109100* as candidate genes for plant height, pod number and fresh pod weight. These results provide new supply for marker-assisted breeding to improve the yield of vegetable soybean.

Author Contributions: H.G. drafted the manuscript; G.W., X.Z. and F.W. performed the computational analysis; G.Z. K.Z. and K.X collected the soybean materials; Y.J. and C.F attributed the method; W.Z. draw the picture. Y.L. revised this manuscript. N.W. and H.L. participated in the design and supervised the study; All of the authors have agreed to published this manuscript.

Funding: This work was supported by the Hainan Province Science and Technology Special Fund (ZDYF2022XDNY142), Project of Sanya Yazhou Bay Science and Technology City (SCKJ-JYRC-2022-53), Hainan Province Science and Technology Special Fund (ZDYF2023GXJS153, ZDYF2023XDNY180), Postdoctoral Station Program of Hainan Province, China (2021-BH-01).

Conflicts of Interest: The author of the article declares that they have no conflict of interest.

References

1. Liu N., Niu Y., Zhang G., Feng Z., Bo Y., Lian J., Wang B., Gong Y. Genome sequencing and population resequencing provide insights into the genetic basis of domestication and diversity of vegetable soybean. *Horticulture Research*.2022;9.
2. Kao C., He S., Wang C., Lai Z., Lin D., Chen S. A modified roger's distance algorithm for mixed quantitative-qualitative phenotypes to establish a core collection for Taiwanese vegetable soybeans. *Frontiers in Plant Science*.2021, 11:612106.

3. Zhang B., Lord N., Kuhar T., Duncan S., Huang H., Ross J., Rideout S., Arancibia R., Reiter M., Li S., et al. 'VT Sweet': A vegetable soybean cultivar for commercial edamame production in the mid-Atlantic USA. *Journal of Plant Register*.2022; 16:29–33.
4. Chen Z., Zhong W., Zhou Y., Ji P., Wan Y., Shi S., Yang Z., Gong Y., Mu F., Chen S. Integrative analysis of metabolome and transcriptome reveals the improvements of seed quality in vegetable soybean (*Glycine max*(L.) Merr.) *Phytochemistry*.2022, 200: 113216.
5. Xu W., Liu H., Li S., Zhang W., Wang Q., Zhang H., Liu X., Cui X., Chen X., Tang W., et al. GWAS and identification of candidate genes associated with seed soluble sugar content in vegetable soybean. *Agronomy Journal*.2022; 12:1470.
6. Nair RM, Boddepalli VN, Yan MR, Kumar V, Gill B, Pan RS, Wang C, Hartman GL, Silva E Souza R, Somta P. Global Status of Vegetable Soybean. *Plants (Basel)*. 2023, 12(3):609.
7. Zhang H, Hao D, Siteo HM, Yin Z, Hu Z, Zhang G, et al. Genetic dissection of the relationship between plant architecture and yield component traits in soybean (*Glycine max*) by association analysis across multiple environments. *Plant Breed*.2015; 134(5):564–572.
8. Zhao X, Dong H, Chang H, Zhao J, Teng W, Qiu L, Li W, Han Y. Genome wide association mapping and candidate gene analysis for hundred seed weight in soybean [*Glycine max* (L.) Merrill]. *BMC Genomics*. 2019 Aug 14;20(1):648.
9. Li X, Zhang X, Zhu L, Bu Y, Wang X, Zhang X, Zhou Y, Wang X, Guo N, Qiu L, Zhao J, Xing H. Genome-wide association study of four yield-related traits at the R6 stage in soybean. *BMC Genetics*. 2019, 29;20(1):39.
10. Cao Y, Jia S, Chen L, Zeng S, Zhao T, Karikari B. Identification of major genomic regions for soybean seed weight by genome-wide association study. *Molecular Breed*. 2022, 42(7):38.
11. Ayalew H, Schapaugh W, Vuong T, Nguyen HT. Genome-wide association analysis identified consistent QTL for seed yield in a soybean diversity panel tested across multiple environments. *Plant Genome*. 2022, 15(4): e20268.
12. Wang J, Hu B, Jing Y, Hu X, Guo Y, Chen J, Liu Y, Hao J, Li WX, Ning H. Detecting QTL and Candidate Genes for Plant Height in Soybean via Linkage Analysis and GWAS. *Frontiers in Plant Science*. 2022, 21; 12:803820.
13. Chang F, Guo C, Sun F, Zhang J, Wang Z, Kong J, He Q, Sharmin RA, Zhao T. Genome-Wide Association Studies for Dynamic Plant Height and Number of Nodes on the Main Stem in Summer Sowing Soybeans. *Front Plant Sci*. 2018;20:9:1184.
14. Liu Z, Li H, Fan X, Huang W, Yang J, Wen Z, Li Y, Guan R, Guo Y, Chang R, Wang D, Chen P, Wang S, Qiu LJ. Phenotypic characterization and genetic dissection of nine agronomic traits in Tokachi nagaha and its derived cultivars in soybean (*Glycine max* (L.) Merr.). *Plant Science*. 2017, 256:72-86.
15. Ayalew H, Schapaugh W, Vuong T, Nguyen HT. Genome-wide association analysis identified consistent QTL for seed yield in a soybean diversity panel tested across multiple environments. *Plant Genome*. 2022,15(4): e20268.
16. Wen Z, Boyse JF, Song Q, Cregan PB, Wang D. Genomic consequences of selection and genome-wide association mapping in soybean. *BMC Genomics*. 2015;16(1):671.
17. Zhang J, Song Q, Cregan PB, Jiang GL. Genome-wide association study, genomic prediction and marker-assisted selection for seed weight in soybean (*Glycine max*) *Theor Appl Genet*.2016;129(1):117–130.
18. Assefa T, Otyama PI, Brown AV, Kalberer SR, Kulkarni RS, Cannon SB. Genome-wide associations and epistatic interactions for internode number, plant height, seed weight and seed yield in soybean. *BMC Genomics*.2019; 20(1):527.
19. Zhao X, Li W, Zhao X, Wang J, Liu Z, Han Y, et al. Genome-wide association mapping and candidate gene analysis for seed shape in soybean (*Glycine max*). *Crop Pasture Science*.2019;70(8):684–693.
20. Zhang X, Ding W, Xue D, Li X, Zhou Y, Shen J, Feng J, Guo N, Qiu L, Xing H, Zhao J. Genome-wide association studies of plant architecture-related traits and 100-seed weight in soybean landraces. *BMC Genom Data*. 2021, 6;22(1):10.
21. Zhan X, Wang B, Li H, Liu R, Kalia RK, Zhu JK, Chinnusamy V. Arabidopsis proline-rich protein important for development and abiotic stress tolerance is involved in microRNA biogenesis. *Proc Natl Acad Sci U S A*. 2012;109(44):18198-18203.
22. Li YX, Li CH, Bradbury PJ, Liu XL, Lu F, Romay CM, et al. Identification of genetic variants associated with maize flowering time using an extremely large multi-genetic background population. *Plant Journal*. 2016; 86:391–402.
23. Lin C. Frozen edamame: global market conditions. USA: Second International Vegetable Soybean conference; 2001. pp. 93–97.
24. Nguyen VQ. Edamame (vegetable green soybean) Australia: Rural Industries Research & Development. The new rural industries: a handbook for farmers and investors; 2001. pp. 49–56.

25. Yano, K.; Yamamoto, E.; Aya, K.; Takeuchi, H.; Lo, P.C.; Hu, L.; Yamasaki, M.; Yoshida, S.; Kitano, H.; Hirano, K.; et al. Genome-wide association study using whole-genome sequencing rapidly identifies new genes influencing agronomic traits in rice. *Nature Genetics*. 2016;48, 927.
26. Hao, H.; Li, Z.; Leng, C.; Lu, C.; Luo, H.; Liu, Y.; Wu, X.; Liu, Z.; Shang, L.; Jing, H.C. Sorghum breeding in the genomic era: Opportunities and challenges. *TAG. Theor. Appl. Genetics. Theor. Und Angew. Genet*. 2021;134, 1899–1924.
27. Zeng T, Meng Z, Yue R, Lu S, Li W, Li W, Meng H, Sun Q. Genome wide association analysis for yield related traits in maize. *BMC Plant Biology*. 2022, 21;22(1):449.
28. Jiao, X.; Lyu, Y.; Wu, X.; Li, H.; Cheng, L.; Zhang, C.; Yuan, L.; Jiang, R.; Jiang, B.; Rengel, Z.; et al. Grain production versus resource and environmental costs: Towards increasing sustainability of nutrient use in China. *Journal of Experiment Botany*. 2016;67, 4935–4949.
29. Eltaher S, Sallam A, Belamkar V, Emara HA, Nower AA, Salem KFM, Poland J, Baenziger PS. Genetic Diversity and Population Structure of F3:6 Nebraska Winter Wheat Genotypes Using Genotyping-By-Sequencing. *Frontiers Genetic*. 2018; 9:76.
30. Li Y, Reif JC, Hong H, Li H, Liu Z, Ma Y, et al. Genome-wide association mapping of QTL underlying seed oil and protein contents of a diverse panel of soybean accessions. *Plant Science*. 2018; 266:95–101.
31. Li S, Cao Y, Wang C, Yan C, Sun X, Zhang L, Wang W, Song S. Genome-wide association mapping for yield-related traits in soybean (*Glycine max*) under well-watered and drought-stressed conditions. *Frontiers in Plant Science*. 2023;14:1265574.
32. Paterne A. A., Norman P. E., Asiedu R., Asfaw A. Identification of quantitative trait nucleotides and candidate genes for tuber yield and mosaic virus tolerance in an elite population of white Guinea yam (*Dioscorea rotundata*) using genome-wide association scan. *BMC Plant Biology*. 2021;21, 552.
33. Zhang H , Hao D , Siteo, Hélder Manuel, et al. Genetic dissection of the relationship between plant architecture and yield component traits in soybean (*Glycine max*) by association analysis across multiple environments[J]. *Plant Breeding*, 2015, 134(5):564-572
34. Zhang J , Song Q , Cregan P B , et al. Genome-wide association study for flowering time, maturity dates and plant height in early maturing soybean (*Glycine max*) germplasm. *BMC Genomics*, 2015, 16(1):217
35. Chen L, Yang H, Fang Y, Guo W, Chen H, Zhang X, Dai W, Chen S, Hao Q, Yuan S, Zhang C, Huang Y, Shan Z, Yang Z, Qiu D, Liu X, Tran LP, Zhou X, Cao D. Overexpression of GmMYB14 improves high-density yield and drought tolerance of soybean through regulating plant architecture mediated by the brassinosteroid pathway. *Plant Biotechnology Journal*. 2021;19(4):702-716.
36. Qi X, Tang W, Li W, He Z, Xu W, Fan Z, Zhou Y, Wang C, Xu Z, Chen J, Gao S, Ma Y, Chen M. Arabidopsis G-Protein β Subunit AGB1 Negatively Regulates DNA Binding of MYB62, a Suppressor in the Gibberellin Pathway. *International Journal Molecular Science*. 2021;22(15):8270.
37. Wang T, Jin Y, Deng L, Li F, Wang Z, Zhu Y, Wu Y, Qu H, Zhang S, Liu Y, Mei H, Luo L, Yan M, Gu M, Xu G. The transcription factor MYB110 regulates plant height, lodging resistance, and grain yield in rice. *Plant Cell*. 2023, 268.
38. Zhang J, Song Q, Cregan PB, Jiang GL. Genome-wide association study, genomic prediction and marker-assisted selection for seed weight in soybean (*Glycine max*). *Theor Appl Genet*. 2016, 129(1):117-30.
39. Li S, Cao Y, Wang C, Yan C, Sun X, Zhang L, Wang W, Song S. Genome-wide association mapping for yield-related traits in soybean (*Glycine max*) under well-watered and drought-stressed conditions. *Frontiers in Plant Science*. 2023;14:1265574.
40. Rani R, Raza G, Ashfaq H, Rizwan M, Razzaq MK, Waheed MQ, Shimelis H, Babar AD, Arif M. Genome-wide association study of soybean (*Glycine max* [L.] Merr.) germplasm for dissecting the quantitative trait nucleotides and candidate genes underlying yield-related traits. *Frontiers in Plant Science*. 2023;14:1229495
41. Xu K, Zhao Y, Zhao Y, Feng C, Zhang Y, Wang F, Li X, Gao H, Liu W, Jing Y, Saxena RK, Feng X, Zhou Y, Li H. Soybean F-Box-Like Protein GmFBL144 Interacts With Small Heat Shock Protein and Negatively Regulates Plant Drought Stress Tolerance. *Frontiers in Plant Science*. 2022;13:823529.
42. Bu Q, Lv T, Shen H, Luong P, Wang J, Wang Z, Huang Z, Xiao L, Engineer C, Kim TH, Schroeder JI, Huq E. Regulation of drought tolerance by the F-box protein MAX2 in Arabidopsis. *Plant Physiology*. 2014;164(1):424-439.
43. Chen Y, Xu Y, Luo W, Li W, Chen N, Zhang D, Chong K. The F-box protein OsFBK12 targets OsSAMS1 for degradation and affects pleiotropic phenotypes, including leaf senescence, in rice. *Plant Physiology*. 2013;163(4):1673-85.
44. Zhou S, Yang T, Mao Y, Liu Y, Guo S, Wang R, Fangyue G, He L, Zhao B, Bai Q, Li Y, Zhang X, Wang D, Wang C, Wu Q, Yang Y, Liu Y, Tadege M, Chen J. The F-box protein MIO1/SLB1 regulates organ size and leaf movement in *Medicago truncatula*. *Journal of Experiment Botany*. 2021;72(8):2995-3011.
45. Sun X, Xie Y, Xu K, Li J. Regulatory networks of the F-box protein FBX206 and OVATE family proteins modulate brassinosteroid pathway to regulate grain size and yield in rice. *Journal of Experiment Botany*. 2023, erad397.

46. Lee K, Han JH, Park YI, Colas des Francs-Small C, Small I, Kang H. The mitochondrial pentatricopeptide repeat protein PPR19 is involved in the stabilization of NADH dehydrogenase 1 transcripts and is crucial for mitochondrial function and Arabidopsis thaliana development. *New Phytologist*. 2017, 215(1):202-216.
47. Cai M, Li S, Sun F, Sun Q, Zhao H, Ren X, Zhao Y, Tan BC, Zhang Z, Qiu F. Emp10 encodes a mitochondrial PPR protein that affects the cis-splicing of nad2 intron 1 and seed development in maize. *Plant Journal*. 2017, 91(1):132-144.
48. Zhao B, Dai A, Wei H, Yang S, Wang B, Jiang N, Feng X. Arabidopsis KLU homologue GmCYP78A72 regulates seed size in soybean. *Plant Molecular Biology*. 2016, 90(1-2):33-47.
49. Wang X, Li Y, Zhang H, Sun G, Zhang W, Qiu L. Evolution and association analysis of GmCYP78A10 gene with seed size/weight and pod number in soybean. *Molecular Biology Reports*. 2015, 42(2):489-496.
50. Zhou C, Lin Q, Ren Y, Lan J, Miao R, Feng M, Wang X, Liu X, Zhang S, Pan T, Wang J, Luo S, Qian J, Luo W, Mou C, Nguyen T, Cheng Z, Zhang X, Lei C, Zhu S, Guo X, Wang J, Zhao Z, Liu S, Jiang L, Wan J. A CYP78As-small grain4-coat protein complex II pathway promotes grain size in rice. *Plant Cell*. 2023 Nov 30;35(12):4325-4346.
51. Guo L, Ma M, Wu L, Zhou M, Li M, Wu B, Li L, Liu X, Jing R, Chen W, Zhao H. Modified expression of TaCYP78A5 enhances grain weight with yield potential by accumulating auxin in wheat (*Triticum aestivum* L.). *Plant Biotechnology Journal*. 2022, 20(1):168-182.
52. Zhang S., Hao D., Zhang S., Zhang D., Wang H., Du H., et al. Genome-wide association mapping for protein, oil and water-soluble protein contents in soybean. *Mol. Genet. Genomics*, 2021, 296, 91–102.
53. Bolger, A.M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, 2014, 30, 2114–2120.
54. Langmead, B. & Salzberg, S.L. Fast gapped-read alignment with Bowtie 2. *Nature Methods*, 2012, 9, 357–359.
55. Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A. et al. The variant call format and VCFtools. *Bioinformatics*, 2011, 27, 2156–2158.
56. Earl, D.A. & vonHoldt, B.M. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genet Resour*, 2012, 4, 359–361.
57. Felsenstein, J. PHYLIP (Phylogeny Inference Package) version 3.6. Distributed by the author. Department of Genome Sciences, University of Washington, Seattle. 2005.
58. Ivica Letunic, Peer Bork. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation, *Nucleic Acids Research*, 2021, 49(1):293–296.
59. Sabeti, P.C., et al. Detecting recent positive selection in the human genome from haplotype structure. *Nature*, 2002;419(6909):832-837.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.