Genome-wide association studies to dissect the genetic loci underlying various agro-economical traits in mungbean (*Vigna radiata* L. Wilczek)

Comment [AA1]: The title needs to be modified. Dissect genetic refers to fine mapping and identification of causal genes/alleles.

Abstract

Mungbean (*Vigna radiata* L. Wilczek) is one of the food legumes grown mainly in Asia but consumed globally. It makes a substantial contribution to both nutritional security and environmental sustainability. The genetic basis of its agro-economic traits is still inadequately understood. To address this problem, 126 different mungbean genotypes were investigated for eleven agronomic traits under two environmental conditions in the current study. Significant phenotypic diversity was found among the analysed traits. In addition, we re-sequenced 126 diverse mungbean genotypes by genotyping-by-sequencing (GBS) method and produced 55,634 genomic variations, after annotation we retained 15926 SNPs and are used to analyse the genetic diversity,linkage disequilibrium (LD) of mungbean. Two distinct subgroups were identified within the genotypesand LD decayed in ~68 kilo base (kb). Genome-wide association studies (GWAS) with Bayesian-information and Linkage-disequilibrium Iteratively Nested Keyway (BLINK) model produced 50 signals (highest P-value > 0.0001 ($-\log_{10}(p)$) significantly associated with eleven agronomic traits depicted in Manhattan plot. Further, through in-silico analysis we recorded the genes with their protein present at 30 kb flanking region of each identified SNPs. Based on previous studies did in *Arabidopsis* on agronomic traits, we found and presumed that only eleven genes with their protein found in our GWAS are important for regulation of agro-economic traits (i.e., Days to 50% flowering (DF50) Plant Height (PH), Pod Length (PL), Pod Number (PN), Nitrogen status, Seeds Per Pods (SPP), 100-Seed Weight (100SW) and Yield Per Plant (YPP)) and are most likely to be the candidate genes. For instance, SNP loci S1_1401613 associated with candidate gene Vradi01g00800 encodes *histone-lysine N-methyltransferase* regulating Days to flowering and for yield, SNP S3_7458210 lies in proximity with Vradi03g06000 *(WD repeat-containing protein 89 homolog).* However, for traits like days to maturity (DM) and primary branch (PB) we did not find candidate genes, because proteins identified in our GWAS for these traits could not corresponds to proteins influencing agronomic traits in *Arabidopsis*. Finally, dissecting genetic control of agronomic traits has a pivotal contribution to fostering mungbean breeding and developing new varieties adaptable to changing climatic

conditions. Our GWAS results will assist in incorporating alleles into the elite mungbean germplasm through marker-assisted selection (MAS).

Keywords Mungbean·SNPs ·Diversity ·Association · Agro-economictraits.Candidate gene

Introduction

Mungbean [*Vigna radiata* (L.) R. Wilczek var. *radiata*] is an ancient grain legume originating from South Asia and popularly known as green gram. Due to its short life cycle (60 days from sowing to maturity), relative drought tolerance, and capacity to improve soil fertility through atmospheric nitrogen (N_2) fixation in symbiosis with *Rhizobium* and *Bradyrhizobium* bacteria in the soil, mungbean, a warm-season annual food legume, adapts well to different farming systems. When grown in rotation with cereals, this results in improved soil quality and a decrease in the amount of inorganic nitrogen fertiliser needed in the soil (Somta et al. 2022). As a result, mungbean extensively cultivated in South, East, and Southeast Asia, particularly India and China **(**Graham et al. 2003**)**.Apart from environmental benefits, mungbean known for excellent nutritional source includes proteins, amino acids, carbohydrates, vitamins, and minerals. The seeds contain about 20–30% protein and 60–70% carbohydrate. The seeds are also processed into sprouts, snacks, pastes, starches, noodles, protein isolates, and protein concentrates (Nair et al. 2020**)**. As an alternative to eggs and meat, mungbean seeds are a major plant-based protein source. As a consequence, the global consumption of mungbean has increased by 22–66% from 1984 to 2006 (Shanmugasundaram et al. 2009) the production area is about 7.5–8.0 million ha, about 80–90% of which is in Asia (Nair et al. 2020; Anonymous. 2021). Mungbean production areas outside Asia, including in America, Africa and Australia are increasing. This is driven by increasing consumer demand. The biggest producer and consumer of mungbean is India, with about 4.5 million ha cultivated and a total production of 2.5 million tons (Anonymous et al. 2021). Although mungbean production areas are increasing, the yield is low, at only about 115 kg/ha (Nair et al. 2020), and production is challenged by insect pests, diseases, and unsuitable environments (Pandey et al. 2018**;** Nair et al. 2019). Therefore, pilot research should be carried to address the concerns and to meet the needs of farmers, consumers, and processors.

Due to inadequate funding for breeding research at the national and international levels, particularly in the area of genomics, the mungbean still remains an orphan crop with little genetic information, despite its extensive environmental benefits, status as an important leguminous food source with a highly diverse landrace germplasm(Schafleitner et al. 2015),

and high socioeconomic importance. Mungbean is an ideal crop for genomics study because it is self-pollinating and diploid (2n = 2x = 22) and has a small genome size of 493.6 to 579.0 megabase (Mb) pairs (Arumuganathan et al. 1991**;** Kang et al. 2014**;** Liu et al.2016) and a short life cycle.

Understanding the genetic components of crucial agronomic features including seed coat colour, grain size, flowering time, and disease resistance is crucial to properly introgressing these traits to meet breeding objectives. Linkage mapping has historically been the main method for locating the genes responsible for a trait of interest. Merely a few genetic linkage maps in mungbean have been created.(Lambrides et al. 2000**;** Humphry et al. 2002**;** Isemura et al. 2012). The development of Next-generation sequencing (NGS) technology in the late 2000s and early 2010s transformed mungbean genome research. Another effective application of NGS is genotyping-by-sequencing (GBS), which allows for the discovery and genotyping of a large number of SNPs at a significantly lower cost (Elshire et al. 2011**)**. The greatest contribution of NGS to mungbean genomics is whole-genome sequencing (WGS). A draft reference genome (i.e. high density maps) of line "VC1973" from WorldVeg was constructed on the chromosome level using Illumina/Solexa and Roche 454 sequencing (Kang et al. 2014). More recently, the genome sequence of VC1973 was improved using third-generation sequencing, such as single-molecule real-time (SMRT) sequencing (Ha et al. 2021). Providing an opportunity to systemically identify and characterize the functions of genes and enable further advancement in alternative approaches to trait dissection, such as genome wide association mapping, also known as linkage disequilibrium (LD) mapping (Gupta et al. 2005**;** Abdurakhmonov et al. 2008). GWAS offers a better QTL resolution than biparental mapping. GWAS can therefore be used to pinpoint the genes that cause a particular trait. The resolution of a QTL mapped by GWAS depends on how quickly the LD decays over that distance. The power of GWAS depends on the strength of correlation (the degree of LD) between the genotypes of markers and those of relevant genes, which is a function of the distance between them (Myles et al. 2009). The LD extent is about 72–290 kb in cultivated mungbean(Noble et al. 2018**;** Sandhu et al. 2021**;** Ha et al. 2021) and 3–60 kb in wild mungbean(Noble et al. 2018).

In the past, GWAS was widely used in model and important crops where a large number of SNP markers were accessible. However, due to the completion of the mungbean reference genome sequence and the rapid advancements in high throughput sequencing technologies, it is also now possible to discover genomic variation in a significant number of mungbean

accessions. Numerous studies have examined the population structure and LD in mungbeanusing genotyping by sequencing (GBS) (Noble et al. 2018**;**Breria et al. 2020**;** Ha et al. 2021). Genetic loci associated with variation in mungbean seed coat colour (Noble et al. 2018) and seed coat luster (Breria et al. 2020) was identified through GWAS. Recently, 2,912 SNPs and 259 genes PAV (presence/absence variant) events associated with 33 agronomic traits were revealed by GWAS in mungbean (Liu et al. 2022). So far, there are only limited study has been reported focusing on agronomic traits.

In this study, we aimed to better understand the genetic diversity, population structure, LD and genetic basis of agronomic traits in diverse mungbean Association mapping (AM) panel from geographically diverse regions of India, Nepal, Bhutan, Myanmar and Thailand. We evaluated phenotypic variation across two environments and performed GWAS for eleven agronomic traits to identify genomic variation. Our results present a collection of genes that may be helpful for enhancing the genetic diversity of mungbean varieties, and provide valuable genomic information for future mungbean breeding programs.

Materials and Methods

Plant Materials

The mungbean Association mapping (AM) panel consists of 126 accessions, includes 37 released varieties (RV), 52 germplasm lines (GL) and 37 advanced breeding lines (ABL). The material originated from various sources Nepal, Thailand, Myanmar, China and India. Represents the widest range of phenotypic traits and characterized by the mungbean breeding team in Indian Agricultural Research Institute (IARI), New Delhi over the past years. Considering this previous information, we characterized for agronomic traits such as Days to 50% flowering (DF50), Days to 100% flowering (DF100), Days to maturity (DM), nitrogen status (using SPAD chlorophyll meter), Plant Height (PH), Primary Branch (PB), Pod Length (PL), Pod Number (PN), Seeds Per Pods (SPP), 100-Seed Weight (100SW) and Yield Per Plant (YPP).

Phenotyping

All 126 mungbean accessions were planted at IARI Research plot, New Delhi (28° 40' 44.6844" N, 77° 4' 10.9560" E) and Punjab Agricultural University (PAU), Ludhiana (30°54'3.47"N, 75°51'26.19"E), over kharif season of 2020. Delhi (DL) and Ludhiana (LUD) are located at Trans Ganga Plain of India and are situated at 218 m and 247m above sea level

and receive an average of 886 mm and 700 mm of rainfall per annum, respectively. The field trial design was made using Randomized Block Design (RBD) with two replications at each site. Accessions were planted in a single row of 4-meter length containing an average of 25 plants, spacing of 10 cm between plants within each row and 30 cm between rows maintained. Recommended agronomic practices were followed growing crop at both the locations. Days to flowering and Leaf nitrogen status (DF 50, DF 100, DM and SPAD) were measured in the full-bloom stage. Plant architecture related traits like PB and PH were measured manually at maturity. Yield-related traits PL, PN, SPP, 100SW and YPP were measured after harvesting manually.

Comment [AA2]: Describe the recommended practices (production technology).

Genotyping

Total genomic DNA from each accession was collected at early seedling stage using the Cetyltrimethyl Ammonium Bromide (CTAB) method (Chen et al. 1999). The samples were genotyped following an genotyping-by-sequencing (GBS) methodology involving complexity reduction of the genomic DNA to remove repetitive sequences using methylation sensitive restrictive enzymes prior to sequencing on next generation sequencing (NGS) platforms IlluminaHiSeq 4000 (Liu et al. 2012; Bastien et al. 2014). The sequence data generated were then aligned to the mungbean reference genome sequence using reference-based GBS pipeline approach of STACKS v1.01 (Kang et al. 2014) to identify single nucleotide polymorphisms (SNPs) markers. SNPs obtained from GBS were imputed for missing loci with LD KNNi imputation from TASSEL v.5.0 with default parameters. Further, SNPs were filtered to eliminate monomorphic markers, markers with a minor allele frequency (MAF) of less than 5%, missing data more than 10%, and heterozygote frequency greater than 50%, remaining 15926 SNPs were used in further analysis.

Estimation of Linkage Disequilibrium (LD)

The pairwise LD between SNPs genome-wide across 126 diverse mungbean genotypes was calculated based on the allele frequency correlations (r^2) using the TASSEL program (v5.0) **(**Tao et al. 2013**).** The LD decay graph was drawn by fitting a smooth spline of averaged r^2 over physical distance in R v3.3.1. The LD decay was calculated when the squared correlations of allele frequencies (r^2) decreased to half of its maximum value.

Analysis of Genetic Diversity

Classifying accessions into clusters we applied 1) Agglomerative hierarchical distance- based method, in which a pair-wise distance matrix is used as an input for analysis by a neighbor joining cluster algorithm, resulted output representation in Dendrogram/Tree depicting different clusters. 2) Innovative model-based clustering method based on Bayesian statistics in which clustering analysis was performed using STRUCTUREv2.3.4 programme (Pritchard 2000).

Genome-Wide Association Mapping of Agronomic traits

Association mapping was conducted using 'BLINK' model controlling for genetic background using PCA in GAPIT V.3.0 with default parameters (https://zzlab.net/GAPIT/gapit_help_document.pdf). P-value > 0.0001 was applied to set threshold P-value and significant SNPs were identified. The Manhattan plot and Q-Q plot were drawn. Identified SNP loci were compared with mungbean reference genome "*Vignaradiata* assembly v1.0" using BLAST search with 'J Browse' in "legume information system" platform (https://legacy.legumeinfo.org/genomes/jbrowse/?data=Vr1.0). Observed LD block size was 68 Kb, hence the annotated genes found in 30 Kb flanking the SNP loci were recorded, information provided in 'J-Browse' was used to identify the protein encoded by genes in SNP loci, that harbours many genes. However, one of which might contribute to variation of studied traits (i.e. causal gene). Therefore, follow up study was performed to compare the function of a protein encoded by genes in SNP locus to the model plant *Arabidopsis* for its homology and function. Further based on previous studies, we recorded the direct and indirect influence of those proteins on various agronomic traits in *Arabidopsis*. Finally, we presumed eleven genes with their associated SNPs are more likely to be the candidate genes in present GWAS for both environments.

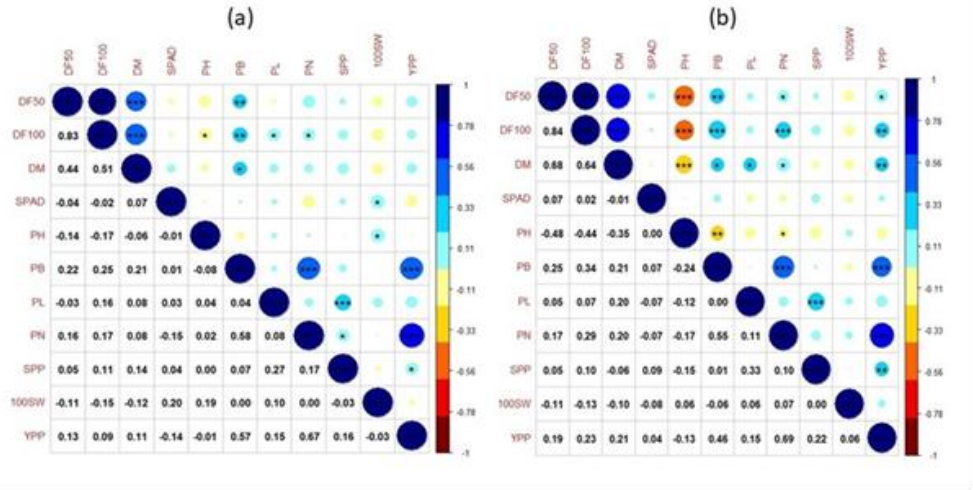
Results

Correlation Studies

Each quantitative traits evaluated at Delhi environment were positively correlated with the corresponding traits evaluated at Ludhiana environment with Pearson correlation coefficient ranging from 0.30 to 0.93 at P <0.001. The Pearson correlation coefficient was estimated to each component traits studied at two environments (Figure1). At Delhi environment, attribute YPP was positively correlated with traits such as PB, PN (at P <0.001) and SPP (at P <0.05). Traits DF50 and DF100 are positively correlated with traits DM (at P <0.001) as well as PB

(at P <0.01). Attribute DF100 is positively correlated with traits PL and PN. Trait 100SW is positively correlated with characters SPAD and PH (at P <0.05). Traits like DF50, DF100 and DM are positively correlated among themselves at P <0.001. However, negative correlation was observed between traits DF100 and PH (at P <0.05 (Figure1a). similarly, at Ludhiana environment, we witnessed positive correlation of yield with its component traits such as PB, PN (at P <0.001), DF100, DM, SPP (at P <0.01) and DF50 (at P <0.05). Attribute PN is significantly correlated with traits DF100 and PB (at P <0.01), DF50 and DM (at P <0.05). Among traits like Days to flowering and days to maturity are positively correlated among themselves (i.e. positive correlation among DF50, DF100 and DM). However, traits PH had shown negative correlation with traits DF50, DF100, DM (at P <0.001) and PB (at P <0.01) (Figure1b).

Figure 1. Correlation coefficients and level of significance for agronomic traits of 126 mungbean accessions observed at (a) Delhi and (b) Ludhiana Environments.



Genomic Variants Discovery

The greatest contribution of Next-generation sequencing (NGS) to mungbean genomics is whole-genome sequencing (WGS). In present study, 126 diverse association mapping (AM) panel are sequenced using genotyping-by-sequencing (GBS) assay and generated high-quality sequence reads of 264.40 million, panel had an equal distribution of reads (mean, 1.83 million reads), 75% of these reads on an average were mapped to the *Vigna radiata* reference genome and identified a total of 76,160 high-quality SNPs (with read-depth 10, <5% missing

data, and 8% minor allelic frequency). Out of total, 55,634 chromosome-based SNPs with polymorphism were found in the genome of the AM panel. All SNPs were discovered to be distributed throughout the genome in a variety of places, including scaffold (19%) variants, intragenic (23%), intergenic (27%) and regulatory (31%). On chromosome 1, maximum of 2294 SNPs were mapped, whereas chromosome 3 had a minimum of 770 SNPs. On chromosome 10, there were only 3.5 SNPs per 0.1 Mb, which is a low SNP density (Table 1). A structural annotation of 55,634 SNPs identified 25,663 (46.12%) SNPs in 11,068 protein-coding genes (intragenic region) and 29,923 (53.78%) SNPs are in intergenic regions. The regulatory area included the greatest number of gene-based SNPs (33,724 SNPs, or 60.1%), followed by the CDS region (13,030 SNPs, or 23.4%), and the intron region (10,423 SNPs, or 18.73%). A total of 7,387 missense and synonymous SNPs, accounting for 56.6% of the coding SNPs, and 5,643 (43.4%) coding SNPs were identified. SNP density plot depicts the relative distribution of GBS-based SNPs on 11 mungbean chromosomes that demonstrate variation among AM panel 55,634 SNP-carrying genes were functionally annotated and it was discovered that 2,481 associated to growth, 7,741 related to development, 4,936 related to metabolism, and 764 signal transduction proteins. Furthermore, 55,634 SNPs were imputed with TASSEL software by LD KNNI method to remove ungenotyped markers from the haplotypes of other individuals. During imputation SNPs are filtered to remove monomorphic marker, marker containing minor allelic frequency <0.05and heterozygotefrequency more than 50%. Finally, we retained 15926 SNPs and utilised for genetic dissections of genomic variants underlying agronomic traits.

Table 1.SNP distribution on chromosomes.

Chromosome No	chromosome Average density (SNPs per 0.1Mb)	Size of the chromosome (bp)	No. of SNPs on chromosome
1	6.3	36501346.00	2294.00
2	5.5	25360630.00	1400.00
3	5.9	12950713.00	770.00
4	5.3	20812224.00	1098.00
5	4.0	37180910.00	1495.00
6	5.0	37436759.00	1857.00
7	4.1	55601358.00	2292.00
8	3.7	45727239.00	1695.00
9	6.0	21008463.00	1264.00
10	3.5	20996616.00	745.00
11	5.1	19732206.00	1015.00
Total	**4.8**	**333308464.00**	**15925.00**

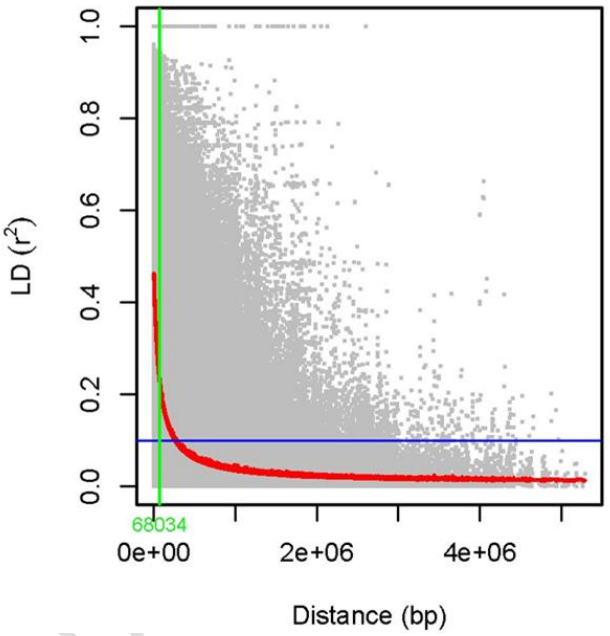
Comment [AA3]: It is mentioned din the abstract that ~50k SNPs were identified, then after trimming and cleaning ~15.9k were retained and used in GWAS. The original number should be added too in this table for each chromosome.

Estimation of Linkage Disequilibrium

The power of GWAS depends on the strength of correlation (i.e. degree of LD) between the genotypes of markers and those of causative genes, which is a function of the distance between them, and the resolution of a QTL mapped by GWAS depends on how rapidly the LD decays over that distance. In our study, Linkage disequilibrium was estimated between 15926 SNP markers over the 126 mungbean accessions. The squared correlations of allele frequencies r^2 of the mungbean population decreased to half of its maximum value at approximately 68 kb physical distance (Figure 2).

Comment [AA4]: It is mentioned in the introduction that LD in the wild V. radiate is 3-60 while 72-290 in the cultivated V. radiata. Because the association panel contain and significant portion of wild accession. I would suggest that carefully reanalyse the data and calculate the LD. In my opinion it might be greater than 68.

Figure 2. LD decay measured in association panel of 126 mungbean genotypes.



Analysis of Genetic Diversity and Population Structure

Comment [AA5]: Here it needs to be properly explained. How the population was structured (the basis). How many groups were formed. The number of accessions in each group, and what was the criteria for grouping i.e., geography etc. Provide the population structure figure.

Classifying AM panel consisting of 126 diverse accessions into clusters based on molecular marker (SNPs) dataset, we applied two types of clustering method; 1) Agglomerative hierarchical distance- based method, that results a Dendrogram depicting two diverse group of studied accession. However, we noticed some constraints of distance-based methods, to

address concern we used an 2) innovative model-based clustering method based on Bayesian statistics in which clustering analysis was performed using STRUCTUREv2.3.4 programme.

Genome-Wide Association Study of Agronomic Traits

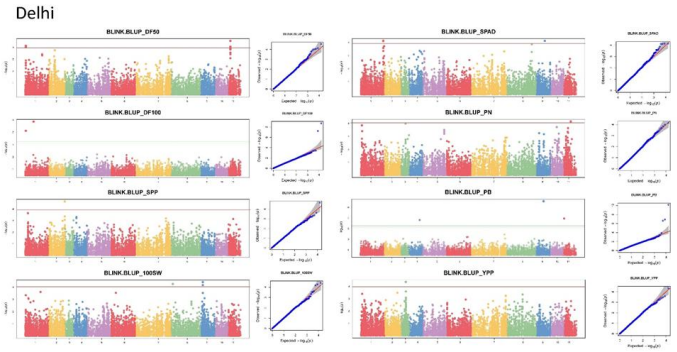
GAPIT V.3.0 is used to perform genome-wide association mapping for agronomic traits. These traits were chosen because they directly and indirectly influence mungbean yield. However, they are polygenic traits which vary based on the environment and showed less heritability. Using BLINK model with PCA as a covariate, a total of 50 significant SNPs spread across eleven different chromosomes associated with component traits of yield were identified under Delhi, Ludhiana and combined (C) BLUP condition based on P-value $P < 0.0001$ (Table 2). These results are depicted using Manhattan plot and Q-Q plot (Figure 3).Further, in-silico analysis was carried out to these significant SNPs by comparing the genomic position of them with *Vigna* reference genome; as a result, we detected several genes in 60kb window from each identified SNPs (i.e. several genes in the interval of each SNP locus). However, only one of which might contribute to variation of studied traits. Therefore, follow up study was performed to compare the homology and function of a protein encoded by genes in SNP locus to the model plant *Arabidopsis*. Based on these previous studies, we recorded the direct and indirect influence of those proteins on various agronomic traits in *Arabidopsis*. Finally, we presumed eleven genes with their associated SNPs are more likely to be the candidate genes in present GWAS for both environments (Table 3). However, we identified candidate gene for eight traits only, out of eleven traits studied. Identified SNPs were found across chromosomes 1,3,7,8 and 9 only. Chromosome 1 harboured four candidate genes governing different traits. For DF50, SNP S1_1401613 on chromosome 1 associated with candidate gene Vradi01g00800 encodes *histone-lysine N-methyltransferase*. For SPAD, two SNPs loci identified, S1_34950474 and S8_38348926 located on chromosome 1 and chromosome 8 respectively. SNP S1_34950474 linked with Vradi01g14220 encoding *Plant regulator RWP-RK family protein*, while, SNP S8_38348926 associated with Vradi08g17320 encoding *50S ribosomal protein L2*. For PH, two SNPs are reported SNP S1_33479087 and SNP S7_33956225 located on chromosome 1and chromosome 7 respectively. SNP S1_33479087 located in proximity with two candidate genes Vradi01g13770 (*polygalacturonase-like protein*) and Vradi01g13800 (*glutamine cyclotransferase protein*).While, SNP S7_33956225 associated with Vradi07g14210 encoding *DHHC-type zinc finger family protein*.For other agronomic traits like PN, PL, SPP, 100SW and YPP only one SNP loci on different chromosomes 3,8,7,9 and 3 respectively, was detected for each

Comment [AA6]: Here important results from GWAS are missing.
1.How many SNPs/ loci were identified in each environment or BLUP dataset.
2.The identified significant SNPs/ loci were located on which chromosome.
This explanation is very important, and required as a base for your next discussion.
3. Also describe the P. value of each SNP, location on the chromosome, trait associated, environment in a table form.
4.Also mention the loci that affect more than one traits, identified I more than one environments etc.

trait. For PN, SNP loci S3_7575781 located on chromosome 3 linked to Vradi03g06110 encoding *exocyst complex component sec15B*. For PL, SNP S8_13451256 located on chromosome 8 associated with Vradi08g05940 encoding *receptor-like kinase protein*. For SPP, SNP loci S7_53198193 harboured on chromosome 7 linked with Vradi07g29450 encoding *ATP-dependent zinc metalloprotease FTSH protein*. For 100SW, SNP S9_2808918 located on chromosome 9 associated with Vradi09g02590 encoding *subtilisin-like serine protease 2*. For YPP, SNP loci S3_7458210 located on chromosome 3 associated with Vradi03g06000 encoding *WD repeat-containing protein 89 homolog*. However, for traits DF100, DM and PB, we did not find candidate genes because proteins identified in our GWAS for these traits could not corresponds to proteins influencing agronomic traits in *Arabidopsis*.

Figure 3. Manhattan (left) and quantile–quantile (Q–Q) (right) plots of various agronomic traits for (a) Delhi (b) Ludhiana.

Comment [AA7]: It was mentioned that GWAS was carried out for 11 traits. However you only provided the plots for 8 traits in DL environment while 6 plots for LUD environment.



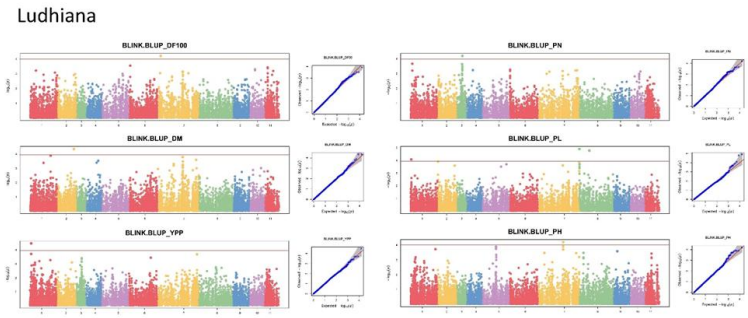


Table 2.List of 50 significant SNPs with their respective p value.

Sr. No.	Environment	Traits	SNP	Chromosome	Position	P. value	Effect	LOD
1	DL	DF50	S1_1401613	1	1401613	7.54E-05	1.820905	4.122823256
2	DL	DF50	S1_1401609	1	1401609	9.53E-05	1.798244	4.020976358
3	DL	DF50	S1_1401637	1	1401637	9.53E-05	-1.7982 4	4.020976358
4	DL	DF100	S1_13315197	1	13315197	1.88E-09	1.479525	8.72692225
5	DL	DF100	S1_1401613	1	1401613	5.73E-08	1.457305	7.241820165
6	DL	SPAD	S1_34950474	1	34950474	7.33E-05	-2.746	4.135070947
7	DL	SPAD	S1_34950467	1	34950467	7.70E-05	2.782531	4.113411529
8	DL	SPAD	S1_34950502	1	34950502	7.70E-05	2.782531	4.113411529
9	DL	SPP	S2_24801210	2	24801210	2.16E-05	0.206612	4.665144331
10	DL	YPP	S3_7458210	3	7458210	4.40E-05	0.881672	4.356821249
11	DL	PB	S4_16870061	4	16870061	2.37E-07	0.18313	6.62532092
12	DL	100SW	S8_2334614	8	2334614	5.36E-05	-0.17434	4.270981881
13	DL	100SW	S9_2808918	9	2808918	3.78E-05	-0.12378	4.422336513
14	DL	100SW	S9_2808960	9	2808960	6.73E-05	-0.123	4.171823384
15	DL	PB	S9_11221047	9	11221047	6.85E-11	0.221147	10.16447846
16	DL	SPAD	S9_12665401	9	12665401	7.43E-05	-3.20359	4.129288717
17	DL	DF50	S11_3265244	11	3265244	2.87E-05	-1.85768	4.54178281
18	DL	DF50	S11_3265377	11	3265377	8.68E-05	1.819192	4.061232812
19	DL	PB	S11_810817	11	810817	1.22E-07	-0.21666	6.914119182
20	DL	PN	S11_10602411	11	10602411	7.04E-05	6.291942	4.152301717
21	LUD	PL	S1_126006	1	126006	8.14E-	0.0883	4.089190

			4		4	05	36	473
22	LUD	SPP	S1_7517183	1	7517183	6.21E-05	0.314036	4.207110492
23	LUD	YPP	S1_2065382	1	2065382	3.35E-05	1.049634	4.474409919
24	LUD	DM	S2_21963318	2	21963318	4.52E-05	-0.74368	4.344521721
25	LUD	DM	S2_21963356	2	21963356	4.52E-05	-0.74368	4.344521721
26	LUD	PN	S3_7575781	3	7575781	6.33E-05	3.914049	4.198399 81
27	LUD	DF100	S7_4407415	7	4407415	6.12E-05	-1.3127	4.213363845
28	LUD	PH	S7_33956225	7	33956225	6.34E-05	0.545324	4.197810736
29	LUD	SPP	S7_53198193	7	53198193	8.07E-05	-0.2844	4.093325145
30	LUD	PL	S8_209897	8	209897	1.23E-05	0.118123	4.909680 26
31	LUD	PL	S8_13451256	8	13451256	1.64E-05	-0.12098	4.785668127
32	C BLUP	100SW	S1_23992177	1	23992177	8.09E-05	0.161901	4.091848474
33	C BLUP	PH	S1_33479087	1	33479087	8.47E-05	-1.6684 3	4.072281158
34	C BLUP	PH	S1_33479088	1	33479088	8.47E-05	-1.6684 3	4.072281158
35	C BLUP	YPP	S1_2065382	1	2065382	8.31E-05	0.978036	4.080232352
36	C BLUP	DM	S2_21963318	2	21963318	5.21E-05	-0.87261	4.282948308
37	C BLUP	DM	S2_21963356	2	21963356	5.21E-05	-0.87261	4.282948308
38	C BLUP	PL	S2_436374	2	436374	5.80E-05	-0.2611	4.236358074
39	C BLUP	SPP	S2_24810267	2	24810267	9.30E-05	0.297002	4.031569385
40	C BLUP	PN	S3_7458210	3	7458210	5.07E-05	4.190066	4.294618619
41	C BLUP	PL	S5_32684304	5	32684304	4.93E-06	-0.27801	5.306899136

42	C BLUP	PL	S7_55475454	7	55475454	2.64E-05	0.329244	4.577627255
43	C BLUP	PL	S7_33485549	7	33485549	5.10E-05	-0.27111	4.292489469
44	C BLUP	PL	S7_55475382	7	55475382	5.64E-05	-0.31253	4.249013883
45	C BLUP	100SW	S8_2334614	8	2334614	4.96E-05	-0.17977	4.304543915
46	C BLUP	PL	S8_13451256	8	13451256	9.48E-06	-0.34448	5.02333871
47	C BLUP	PL	S8_209897	8	209897	2.05E-05	0.320249	4.688467318
48	C BLUP	SPAD	S8_38348926	8	38348926	2.82E-05	0.953522	4.550264815
49	C BLUP	100SW	S9_2808918	9	2808918	3.88E-05	-0.12691	4.410679916
50	C BLUP	100SW	S9_2808960	9	2808960	7.95E-05	-0.12514	4.099381041

DL: Delhi, LUD: Ludhiana, C BLUP: Combined Best Linear Unbiased Predictors. DF50: Days to 50% flowering, DF100: Days to 100% flowering, DM: Days to maturity, SPAD: nitrogen status, PH: Plant Height, PB: Primary Branch, PL: Pod Length, PN: Pod Number, SPP: Seeds per Pods, 100SW: 100-Seed Weight and YPP: Yield Per Plant.

Table 3. Putative candidate genes at 30 Kb flanking region of SNPs and corresponding proteins produced by them

Traits	SNP	Gene	Position	Definition	Function	Reference
DF50	S1_140161613	Vradi01g00800	Vr01:1373450..1375142 (- strand)	histone-lysine N-methyltransferase	Early flowering in short days (EFS) regulates the seed size in *Arabidopsis*	Cheng et al., 2018
SPAD	S1_34950474	Vradi01g14220	Vr01:34942979..34954116 (- strand)	Plant regulator RWP-RK family protein	RWP-RK proteins have a key role in regulating responses to nitrogen availability	Chardin et al., 2014
SPAD	S8_38348926	Vradi08g17320	Vr08:38346422..38349356 (- strand)	50S ribosomal protein L21	Our studies suggest that the chloroplast ribosomal protein L21 gene is required for chloroplast development and embryogenesis in *Arabidopsis*.	Yin et al., 2014
PH	S7_33956225	Vradi07g14210	Vr07:33933301..33946641 (+ strand)	DHHC-type zinc finger family protein	A DHHC-type zinc finger protein gene regulates shoot branching in *Arabidopsis*	Xiang et al., 2010
PH	S1_334	Vradi01g	Vr01:33445581..334492	polygalacturonase like protein	Plant Polygalacturonases Involved in Cell Elongation	Babu

	79087	13770	64 (+ strand)			et al., 2014
PH	S1_33479087	Vradi01g13800	Vr01:33471587..33479361 (- strand)	glutamine cyclotransferase	cytosolic Gln production and plant development, ROS production and stress tolerance	Ji et al., 2019
PN	S3_75757781	Vradi03g06110	Vr03:75518220..7554216 (- strand)	exocyst complex component sec15B	An Exocyst Complex Functions in Plant Cell Growth in*Arabidopsis*and Tobacco	Hál aet al., 2008
PL	S8_13451256	Vradi08g05940	Vr08:13447049..13449653 (+ strand)	receptor-like protein kinase 1	Growth, development, stress responses, and disease resistance.	Jose et al., 2020
SPP	S7_53198193	Vradi07g29450	Vr07:53174689..53182782 (- strand)	ATP-dependent zinc metalloprotease FTSH protein	chloroplast development; photosynthesis	Kato et al., 2018
100SW	S9_28089188	Vradi09g02590	Vr09:27807888..2789841 (- strand)	subtilisin-like serine protease 2	regulation of stomatal density and distribution in *Arabidopsis* thaliana	Berger et al., 2000
Y	S3_	Vrad	Vr03:74271	WD repeat-	WD40 domain proteins have varied interacting partners and are involved in as diverse	Vil

P P	745 821 0	i03g 0600 0	44..7431208 (- strand)	containing protein 89 homolog	functions as cell motility, division and cytokinesis, apoptosis, light signaling and vision, environmental stress, flowering and floral development, and meristem organization	lan uev a et al., 20 16

DF50: Days to 50% flowering, SPAD: nitrogen status, PH: Plant Height, PL: Pod Length, PN: Pod Number, SPP: Seeds per Pods, 100SW: 100-Seed Weight and YPP: Yield Per Plant.

Discussion

Correlation Study among Agronomic Traits

Before choosing cultivars with the best combination of attributes and including them in crop improvement programmes, correlation studies among significant traits contributing to yield must be conducted. Recognizing the importance of this step in crop breeding, many researchers have studied the correlation using a variety of methods, primarily the path analysis method among various yield component traits in mungbean, including the days to flowering, plant height, and number of seeds per plant.[Manivannan et al. 2002**;** Patil et al. 2001**;** Bisht et al. 2005**;** Makeen et al. 2007**;** Das et al. 2008). Positive correlation of seed yield with number of pods per plant and plant height has been reported by (Upadhaya et al. 1980). (Khan et al. 2001) shown that several yield contributing traits are correlated to yield. There is also a positive correlation between grain yield and the number of branches, according to (Reddy et al. 1991 and Khan et al. 2001). (Malik et al. 1983) found that the number of primary branches per plant and the number of pods per plant were positively correlated with seed yield. (Rubio et al. 2004) observed a positive association between flowering time and seed yield. In our present study at Delhi environment, traits DF50 and DF100 are positively correlated with attribute such as DM and PB. Similarly, at Ludhiana environment, significant positive correlation was observed among traits like DF50, DF100 and DM. Dates to flowering and maturity dates are correlated may be due to late and early type of growth habit of the crop plant. Trait DF100 is positively correlated with both characters PL and PN, it may due to late maturing plants accumulate higher photosynthates, and it leads to higher yield. However, PH had shown negative correlation with traits DF50, DF100 and DM.

LD in Mungbean

The power of GWAS approach depends on degree of LD (i.e. strength of correlation) between the genotypes of markers and those of causative loci, which is determined by distance between them. In addition, the resolution of a QTL mapped by GWAS and density of marker coverage needed for GWAS depends on how rapidly the LD decays over that

distance (Myles et al.2009). If LD decays faster than expected, a higher marker density is required to capture markers associated to causal loci (Lin et al. 2019). In the current study, the genome-wide LD decayed at genomic distances of about 68 kb (figure 2). Previous studies revealed that The LD extent is about 72–290 kb in cultivated mungbean (Noble et al. 2018**;** Ha et al. 2021; Sandhu et al. 2021) and 3–60 kb in wild mungbean (Noble et al. 2018). Mungbean LD pattern was determined to be distinct from chickpea (Saxena et al. 2014**;** Kujur et al. 2016) but presumably similar to other self-pollinated crop species, such as soybean (Patil et al. 2016).

High-Resolution Association Mapping Study

GWAS take full advantage of ancient recombination events happening in a group of germplasms, mainly in landraces to identify either causative/predictive gene for the trait of interest, or to unravel the genetic architecture of complex traits by finding DNA markers, usually SNPs, underlying particular trait at relatively high resolution. GWAS well-known as linkage disequilibrium (LD) mapping and is done by scanning genotype–phenotype associations along the chromosomes of all given germplasms. This study needs a huge number of germplasms with high genetic diversity and a large number of SNPs. GWAS provides a better QTL resolution, often to the gene level than biparental mapping. Therefore, it can be used to pinpoint the genes for particular trait. This shows that GWAS method is a useful and robust approach corresponding to classical biparental mapping and has the power to genetically map multiple traits concurrently. Earlier, GWAS carried out successfully in major crops including rice (Huang,et al. 2010**;** Li et al. 2020**),** wheat (Gupta et al. 2020), maize (Wang et al. 2020), cotton **(**Ma et al. 2018), soybean (Fang et al. 2017), and food legumes **(**Varshney et al. 2017**;** Varshney et al. 2019**;** Wu et al. 2020). And models crops*Arabidopsis thaliana* (Koornneef et al. 2014) where a large number of SNPs were available. Rapid development in high throughput sequencing technologies, computational method and the completion of the mungbean reference genome sequence (Kang et al. 2014) allowing the possibility of using GWAS in orphan crops like mungbean accessions. Genotyping by sequencing (GBS) has been employed in several studies) to examine population structure in mungbean (Noble et al. 2018**;** Breria et al. 2020**;** Ha et al. 2021). Genomic loci associated with variation in mungbean seed coat color (Noble et al. 2018) and seed coat lusters (Breria et al. 2020) were discovered through GWAS. Very recently, GWAS in mungbean identified 2,912 SNPs and 259 gene PAV events underlying 33 agronomic characteristics (Liu et al. 2022).

In present study, GWAS for agronomic traits such as DF50, DF100, DM, SPAD, PH, PB, PL, PN, SPP, 100SW and YPP in a mungbean collection of 126 accessions was performed using 15926 SNPs across two different environments and identified 50 significant SNPs spread across eleven different chromosomes. These results are depicted in pictorial form using Manhattan plot and Q-Q plot (Figure 3). Further, in-silico analysis was done by comparing the genomic position of identified SNPs with *Vigna* reference genome; as a result we detected several genes across 60kb interval from each reported SNPs (i.e. several genes in the interval of one SNP locus). However, only one of which associated with studied traits. Therefore, follow up study required to pinpoint the causal gene. In line with this, we compare the function of a protein encoded by genes in SNP locus to the model plant *Arabidopsis* for its homology and function. Further based on previous studies, we recorded the direct and indirect influence of those proteins on agronomic traits in *Arabidopsis*. Finally, proteins identified in our GWAS could correspond to proteins influencing agronomic traits in *Arabidopsis*were recorded based on this; we presumed that eleven genes are more likely to be the candidate genes in present GWAS for both environments (Table 3). Similarly (Sandhu and Singh, 2021) was carried out GWAS for agronomic traits (plant height and days to flowering) and seed size (100 Seed weight) in a USDA mungbean collection of 482 accessions using 264,550 SNPs and discovered Three SNP loci on different chromosomes were detected for each trait. So far, only few studies have been reported for GWAS on agronomic traits in mungbean.

Days to flowering is pivotal trait responsible for adaptation and it showed maximum sensitivity to environmental photoperiod and temperature in various crops [Hung et al. 2012**;** Wei et al. 2020**;** Lu et al. 2020). Present study reported a SNP S1_1401613 located near to candidate gene Vradi01g00800 encoding *histone-lysine N-methyltransferase*, whose function in *Arabidopsis* is regulation of early flowering in short days as well as controlling the variation of seed size (Liu et al. 2016; Cheng et al. 2018). Also, (Sandhu and Singh, 2021) for Days to flowering, detected SNP loci namely SNP 1_11367629 on chromosome 1, SNP 5_4604047 on chromosome 5 showed R^2-values higher than 25%, while the other SNPs showed R^2-values of about only 1%. SNP 1_11367629, located within LOC106774729 and producing *receptor like protein kinase FERONIA (FER)*. FER has diverse functions in plant growth and development, including hypocotyl and root elongation, root hair development, and flowering time (Deslauriers et al. 2010; Duan et al. 2010; Wang et al. 2020;Zhu et al. 2020). Further, they looked into an 83 kb region (Vr01:11309527...11393240) covering

LOC106774729 and found five other FER genes located next to LOC106774729. It will therefore be difficult to determine the causative FER genes for these traits. Similarly, they also surveyed the region around SNP 5_4604047 and found that this marker is about 25 kb away from a Phytochrome gene. However, none of the days to flowering SNPs detected in this study were in the same region as VrPHYA, the candidate gene for days to flowering reported byXiong et al., 2015 (Hwang et al. 2017**).**

Leaf chlorophyll concentration measured using a portable and handy SPAD meter. Leaf N content per leaf area and SPAD readings is highly affected by environmental factors (Xiong et al. 2015). We identified two SNPs loci S1_34950474 and S8_38348926 present on chromosome 1 and 8 respectively, they present proximity to two candidate genes Vradi01g14220 and Vradi08g17320 encoding Plant regulator *RWP-RK family protein* and *50S ribosomal protein L21* respectively. Previous work did in *Arabidopsis* show that, they regulates development of chloroplasts and embryogenesis (Yin et al. 2012) and the way *Arabidopsis* responds to nitrogen availability (Chardin et al. 2014).

Plant height and primary branch are two major traits that affect the plant architecture. Agronomic performance of crop species depends on its plant architecture (Huyghe et al. 1998; Liet al. 2016**;** Song et al. 2009). We identified two SNP loci, S1_33479087and S7_33956225 for plant height, harboured on chromosomes 1 and 7 respectively. Single SNP S1_33479087 associated with two candidate genes Vradi01g13770 and Vradi01g13800. Vradi01g13770 encodes *polygalacturonase* [in *Arabidopsis* it controls cell elongation] (Babu et al. 2014) and Vradi01g13800 encodes *glutamine cyclotransferase* [it regulates various function in *Arabidopsis* includes cytosolic Gln production, plant development and stress tolerance**(**Ji et al. 2019)]. SNP loci S7_33956225 located proximity to Vradi07g14210, encodes *DHHC-type zinc finger family protein* [it regulates shoot branching in *Arabidopsis* (Xiang et al. 2010)]. Similarly,(Sandhu and Singh, 2021) identified Three SNP loci on different chromosomes for plant height, among the three SNP loci, only SNP 1_11367629, with R^2 of about 30%, appeared to be correctly identified, while the others, with R^2 -values of 0%, were likely false positives. However, we could find candidate gene for trait primary branch.

For other agronomic traits like pod number, pod length, seeds per plant, 100 seed weight and yield per plant, only one SNP loci on different chromosomes were detected for each trait. For pod number, SNP loci S3_7575781 harboured on chromosome 3 associated to candidate gene

Vradi03g06110 encodes *exocyst complex component sec15B* [in *Arabidopsis*it controls plant cell growth] (Hála et al. 2008). For pod length, SNP S8_13451256 on chromosome 8 linked to Vradi08g05940 coding *receptor-like kinase 1 protein* [it regulates various function like growth, development, stress responses, and disease resistance in *Arabidopsis* (Jose et al. 2020)]. For seeds per plant, SNP S7_53198193 on chromosome 7 found proximity to Vradi07g29450 encodes *ATP-dependent zinc metalloprotease FTSH protein* [it controls chloroplast development and photosynthesis in *Arabidopsis*(Kato et al. 2018)]. For 100 seed weight, SNP S9_2808918 on chromosome 9 associated with Vradi09g02590 encoding *subtilisin-like serine protease 2* [it regulates stomatal density and distribution in *Arabidopsis thaliana* (Berger et al. 2000)]. Also, (Sandhu and Singh, 2021) reported major loci for 100 seed weight on chromosomes 1 and 7, and each QTL contributed 10–13% of the seed weight variation. For yield per plant, SNP S3_7458210 on chromosome3 located proximity to Vradi03g06000 encodes *WD repeat-containing protein* [previous study in *Arabidopsis* showed that same protein regulates diverse functions such as cell motility, division and cytokinesis, apoptosis and light signalling(Villanueva et al. 2016)].

SNPs and candidate genes underlying the agronomic traits identified in this GWAS could not corresponds to any those SNPs which have been identified in previous GWAS for similar traits. So far, only few studies have been carried out for GWAS on agronomic traits in mungbean. Hence, it difficult to pinpoint precisely which candidate gene in the interval of SNP locus significantly regulates agronomic traits in mungbean, more research need to be done. Therefore, SNPs information linked with distinct candidate genes found in this GWAS analysis needs further validation either in different diverse populations or by using laboratory tests such as overexpression and knockout of candidate genes. Further, true causals genes can be effectively deployed for developing new superior cultivars in mungbean through MAS.

References

Abdurakhmonov I. Y. and Abdukarimov A. 2008. Application of association mapping to understanding the genetic diversity of plant germplasm resources. Int. J. Plant Genomics, 2008: 574927.

Anonymous (2021). Greengram Outlook Report. Available online at: https://angrau.ac.in/downloads/AMIC/GREENGRAM%20OUTLOOK%20REPORT%20 January%20to%20May%202021.pdf (accessed April 7, 2022).

Arumuganathan K. and Earle E. D. 1991. Nuclear DNA content of some important plant species. Plant mol. Biol. Rep., **9**: 208-218.

Babu Y. and Bayer M. 2014. Plant Polygalacturonases involved in cell elongation and separation—the same but different?. Plants, **3**(4): 613-623.

Bastien M., Sonah H. and Belzile F. 2014. Genome wide association mapping of Sclerotinia sclerotiorum resistance in soybean with a genotyping- by- sequencing approach. The Plant Genome, **7**(1): plantgenome2013-10.

Berger D. and Altmann T. 2000. A subtilisin-like serine protease involved in the regulation of stomatal density and distribution in Arabidopsis thaliana. Genes and dev., **14**(9): 1119-1131.

Bisht I. S., Bhat K. V., Lakhanpaul S., Latha M., Jayan P. K., Biswas B. K. and Singh, A. K. 2005. Diversity and genetic resources of wild Vigna species in India. Genet. Resour. Crop Evol., **52**(1): 53-68

Breria C. M., Hsieh C. H., Yen J. Y., Nair R., Lin C. Y., Huang S. M. and Schafleitner R. 2020. Population structure of the world vegetable centermungbean mini core collection and genome-wide association mapping of loci associated with variation of seed coat luster. Trop. Plant Biol., **13**: 1-12.

Chardin C., Girin T., Roudier F., Meyer C. and Krapp A. 2014. The plant RWP-RK transcription factors: key regulators of nitrogen responses and of gametophyte development. J. Exp. Bot., **65**(19): 5577-5587.

Chen D. H. and Ronald P. C. (1999). A rapid DNA minipreparation method suitable for AFLP and other PCR applications. Plant Mol. Bio. Rep., **17**(1): 53-57.

Cheng L., Shafiq S., Xu W. and Sun Q. 2018. Early flowering in short days (EFS) regulates the seed size in Arabidopsis. Sci. China Life Sci., **61**(2): 214-224.

Das A., Biswas M. and Dastidar K. G. 2010. Genetic Divergence in Green Gram (Vignaradiata L. Wilczek). J. Agro., **9**: 126-130.

Deslauriers S. D. and Larsen P. B. 2010. FERONIA is a key modulator of brassinosteroid and ethylene responsiveness in Arabidopsis hypocotyls. Mol.plant, **3**(3): 626-640.

Duan Q., Kita D., Li C., Cheung A. Y. and Wu H. M. 2010. FERONIA receptor-like kinase regulates RHO GTPase signaling of root hair development. Proc. Nat. Acad. Sci., **107**(41): 17821-17826.

Elshire R. J., Glaubitz J. C., Sun Q., Poland J. A., Kawamoto K., Buckler E. S. and Mitchell S. E. 2011. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PloS one, **6**(5): e19379.

Fang C., Ma Y., Wu S., Liu Z., Wang Z., Yang R. and Tian Z. 2017. Genome-wide association studies dissect the genetic networks underlying agronomical traits in soybean. Genome biol., **18**(1): 1-14.

Graham P. H. and Vance C. P. 2003. Legumes: importance and constraints to greater use. Plant physiol., **131**(3): 872-877.

Gupta O. P., Singh A. K., Singh A., Singh G. P., Bansal K. C. and Datta S. K. (2022). Wheat Biofortification: Utilizing natural genetic diversity, genome-wide association mapping, genomic selection, and genome editing technologies. Front. Nutr., **9**: 826131

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

Ha J., Satyawan D., Jeong H., Lee E., Cho K. H., Kim M. Y. and Lee S. H. 2021. A near- complete genome sequence of mungbean (Vignaradiata L.) provides key insights into the modern breeding program. The plant genome, **14**(3), e20121.

Hála M., Cole R., Synek L., Drdová E., Pečenková T., Nordheim A. and Žárský V. 2008. An exocyst complex functions in plant cell growth in Arabidopsis and tobacco. The Plant Cell, **20**(5): 1330-1345.

Huang X., Wei X., Sang T., Zhao Q., Feng Q., Zhao Y. and Han B. 2010. Genome-wide association studies of 14 agronomic traits in rice landraces. Nat. gen., **42**(11): 961-967.

Humphry M., Konduri V., Lambrides C., Magner T., McIntyre C., Aitken, E. and Liu C. 2002. Development of a mungbean (Vignaradiata) RFLP linkage map and its comparison with lablab (Lablab purpureus) reveals a high level of colinearity between the two genomes. Theor. Appl. Genet., **105**: 160-166.

Hung H. Y., Shannon L. M., Tian F., Bradbury P. J., Chen C., Flint-Garcia S. A. and Holland J. B. 2012. ZmCCT and the genetic basis of day-length adaptation underlying the post domestication spread of maize. Proc. Nat. Acad. Sci., **109**(28): E1913-E1921.

Huyghe C. 1998. Genetics and genetic modifications of plant architecture in grain legumes: a review. Agronomie, **18**(5-6): 383-411.

Hwang W. J., Ha J., Lee T., Jeong H., Kim M. Y., Kim S. K. and Lee S. H. 2017. A candidate flowering gene in mungbean is homologous to a soybean Phytochrome A gene. Euphytica, **213**: 1-9.

Isemura T., Kaga A., Tabata S., Somta P., Srinives P., Shimizu T. and Tomooka N. 2012. Construction of a genetic linkage map and genetic analysis of domestication related traits in mungbean (Vignaradiata). PLoS ONE, **7**(8).

Ji Y., Li Q., Liu G., Selvaraj G., Zheng Z., Zou J. and Wei Y. 2019. Roles of cytosolic glutamine synthetases in Arabidopsis development and stress responses. Plant Cell Phys., **60**(3): 657-671.

Jose J., Ghantasala S. and Roy Choudhury S. 2020. Arabidopsis transmembrane receptor-like kinases (RLKs): a bridge between extracellular signal and intracellular regulatory machinery. Int. J. Mol. Sci., **21**(11): 4000.

Kang Y. J., Kim S. K., Kim M. Y., Lestari P., Kim K. H., Ha B. K. and Lee S. H. 2014. Genome sequence of mungbean and insights into evolution within Vigna species. Nat. com., **5**(1): 5443.

Kato Y. and Sakamoto W. 2018. FtsH protease in the thylakoid membrane: physiological functions and the regulation of protease activity. Front. Plant Sci., **9**: 855.

Khan S., Shah S. and Akbar H. 2001. Effect of planting geometry on yield and yield components in mungbean. Sarhad J. Agri., **17**(4): 519-524.

Koornneef M., Alonso-Blanco C. and Vreugdenhil D. 2004. Naturally occurring genetic variation in Arabidopsis thaliana. Annu. Rev. Plant Biol., **55**: 141-172.

Kujur A., Bajaj D., Upadhyaya H. D., Das S., Ranjan R., Shree T. and Parida S. K. 2015. Employing genome-wide SNP discovery and genotyping strategy to extrapolate the natural allelic diversity and domestication patterns in chickpea. Front. Pl. sci., **6**: 162.

Lambrides C. J., Lawn R. J., Godwin I. D., Manners J. and Imrie B. C. 2000. Two genetic linkage maps of mungbean using RFLP and RAPD markers. Aust. J. Agri. Res., **51**(4):415-425.

Li F., Chen B., Xu K., Gao G., Yan G., Qiao J. and Wu X. 2016. A genome-wide association study of plant height and primary branch number in rapeseed (Brassica napus). Plant Sci., **242**: 169-177.

Li X., Chen Z., Zhang G., Lu H., Qin P., Qi M. and Liang, C. 2020. Analysis of genetic architecture and favourable allele usage of agronomic traits in a large collection of Chinese rice accessions. Sci. China Life Sci., **63**: 1688-1702.

Lin Y., Yi X., Tang S., Chen W., Wu F., Yang X. and Liu Y. 2019. Dissection of phenotypic and genetic variation of drought- related traits in diverse Chinese wheat landraces. The Plant Genome, **12**(3), 190025.

Lipka A. E., Tian F., Wang Q., Peiffer J., Li, M., Bradbury P. J. and Zhang Z. 2012. GAPIT: genome association and prediction integrated tool. Bioinformatics, **28**(18): 2397-2399.

Liu B., Berr A., Chang C., Liu C., Shen W. H. and Ruan Y. 2016. Interplay of the histone methyltransferases SDG8 and SDG26 in the regulation of transcription and plant flowering and development. BBA-Gene Reg. Mech., **1859**(4): 581-590.

Liu C., Fan B., Cao Z., Su Q., Wang Y. A. N., Zhang Z. and Tian J. 2016. A deep sequencing analysis of transcriptomes and the development of EST-SSR markers in mungbean (Vignaradiata). J. Gen., **95**: 527-535.

Liu C., Wang Y., Peng J., Fan B., Xu D., Wu J. and Tian J. 2022. High-quality genome assembly and pan-genome studies facilitate genetic discovery in mungbean and its improvement. Plant Comm., **3**(6): 100352.

Liu L., Li Y., Li S., Hu N., He Y., Pong R. and Law M. 2012. Comparison of next-generation sequencing systems. J. Biomed. Biotechnol., 2012(251364): 251364.

Lu S., Dong L., Fang C., Liu S., Kong L., Cheng Q. and Kong F. 2020. Stepwise selection on homeologous PRR genes controlling flowering and maturity during soybean domestication. Nat. Gen., **52**(4): 428-436.

Ma Z., He S., Wang X., Sun J., Zhang Y., Zhang G. and Du X. 2018. Resequencing a core collection of upland cotton identifies genomic variation and loci influencing fiber quality and yield. Nat. gen., **50**(6): 803-813.

Makeen K., Abrahim G., Jan, A. and Singh A. K. 2007. Genetic variability and correlations studies on yield and its components in mungbean (Vignaradiata (L.) wilezek). J. Agro., **6**: 216-218.

Malik B. P. S. 1983. Genetics of some metric traits in greengram. Indian J. Agric. Sci., **53**: 1002-1005.

Manivannan N. 2002. Genetic diversity in cross derivatives of greengram [Vignaradiata (L.) Wilczek]. Leg. Res. Int. J., **25**(1): 50-52.

Myles S., Peiffer J., Brown P. J., Ersoz E. S., Zhang Z., Costich D. E. and Buckler E. S. 2009. Association mapping: critical considerations shift from genotyping to experimental design. The Plant Cell, **21**(8): 2194-2202.

Nair R. and Schreinemachers P. 2020. "Global status and economic importance of mungbean", in The Mungbean Genome, ed R. M. Nair, R. Schafleitner, and S. H. Lee (Cham: Springer International Publishing), 1–8.

Nair R. M., Pandey A. K., War A. R., Hanumantharao B., Shwe T., Alam A. K. M. M. and Schafleitner R. 2019. Biotic and abiotic constraints in mungbean production—progress in genetic improvement. Front. Plant Sci., **10**: 1340.

Noble T. J., Tao Y., Mace E. S., Williams B., Jordan D. R., Douglas C. A. and Mundree S. G. 2018. Characterization of linkage disequilibrium and population structure in a mungbean diversity panel. Front. Plant Sci., **8**: 2102.

Pandey A. K., Burlakoti R. R., Kenyon, L. and Nair R. M. 2018. Perspectives and challenges for sustainable management of fungal diseases of mungbean [Vignaradiata (L.) R. Wilczek var. radiata]: a review. Front. Environ. Sci., **6**: 53.

Patil B. L., Hegde V. S. and Salimath P. M. 2001. Studies on genetic divergence over stress and non-stress environment in mungbean [Vignaradiata (L.)Wilczek]. Indian J. Gen. Plant Breed., **63**(01): 77-78.

Patil G., Do T., Vuong T. D., Valliyodan B., Lee J. D., Chaudhary J. and Nguyen H. T. 2016. Genomic-assisted haplotype analysis and the development of high-throughput SNP markers for salinity tolerance in soybean. Sci. Rep., **6**(1): 1-13.

Pritchard J. K., Stephens M., Rosenberg N. A. and Donnelly P. 2000. Association mapping in structured populations. The American J. Human Gen., **67**(1):170-181.

Reddy TD, Mishra R K, Yadav RK. 1991. Genetic variability and correlation coefficient related to yield and other quantitatives and use of path coefficient in mungbean. Indian J. Pulses Res. **4**:100-104

Rubio J., J. I. Cubero L. M. Suso and F. Flores. 2004. Biplot analysis of trait relation of white lupin in Spain. Euphytica. **135**: 217-224.

Sandhu K. and Singh A. 2021. Strategies for the utilization of the USDA mungbean germplasm collection for breeding outcomes. Crop Sci., **61**(1): 422-442.

Saxena M. S., Bajaj D., Kujur A., Das S., Badoni S., Kumar V. and Parida, S. K. 2014. Natural allelic diversity, genetic structure and linkage disequilibrium pattern in wild chickpea. PloS one, **9**(9): e107484.

Schafleitner R., Nair R. M., Rathore A., Wang Y. W., Lin C. Y., Chu S. H. and Ebert A. W. 2015. The AVRDC–The World Vegetable Centermungbean (Vignaradiata) core and mini core collections. BMC genom., **16**(1): 1-11.

Shanmugasundaram S., Keatinge J. D. H. and Hughes J. D. A. 2009. Counting on beans: mungbean improvements in Asia. Millions fed: proven successes in agricultural development, 103-108.

Somta P., Laosatit K., Yuan X. and Chen X. 2022. Thirty Years of Mungbean Genome Research: Where Do We Stand and What Have We Learned?. Front. Plant Sci., **13**: 944721.

Song X. and Zhang T. 2009. Quantitative trait loci controlling plant architectural traits in cotton. Plant Sci., **177**(4): 317-323.

Tao Y., Jiang L., Liu Q., Zhang Y., Zhang R., Ingvardsen C. R. and Xu M. 2013. Combined linkage and association mapping reveals candidates for Scmv1, a major locus involved in resistance to sugarcane mosaic virus (SCMV) in maize. BMC Plant Biol., **13**:1-13.

Upadhaya LP, Singh RB, Agerwal RK 1980. Characters association in green gram population of different maturity group. Ind. J. Agric. Sci. 6: 473- 476.

Varshney R. K., Saxena R. K., Upadhyaya H. D., Khan A. W., Yu Y., Kim C. and Datta S. K. 2017. Whole-genome resequencing of 292 pigeonpea accessions identifies genomic regions associated with domestication and agronomic traits. Nat. Gen., **49**(7): 1082-1088.

Varshney R. K., Thudi M., Roorkiwal M., He W., Upadhyaya H. D., Yang W. and Liu X. 2019. Resequencing of 429 chickpea accessions from 45 countries provides insights into genome diversity, domestication and agronomic traits. Nat. gen., **51**(5): 857-864.

Villanueva M. A., Islas-Flores T. and Ullah H. 2016. Signalling through WD-repeat proteins in plants. Front. plant sci., **7**: 1157.

Wang B., Lin Z., Li X., Zhao Y., Zhao B., Wu G. and Wang H. 2020. Genome-wide selection and genetic improvement during modern maize breeding. Nat. Gen., **52**(6): 565-571.

Wang L., Yang T., Lin Q., Wang B., Li X., Luan S. and Yu F. 2020. Receptor kinase FERONIA regulates flowering time in Arabidopsis. BMC plant biol., **20**: 1-16.

Wei H., Wang X., Xu H. and Wang L. 2020. Molecular basis of heading date control in rice. Abiotech, **1**(4): 219-232.

Wu J., Wang L., Fu J., Chen J., Wei S., Zhang S. and Wang S. 2020. Resequencing of 683 common bean genotypes identifies yield component trait associations across a north–south cline. Nat. gen., **52**(1):118-125.

Xiang J., Lin J., Tang D., Zhou B., Guo M., He R. and Liu X. 2010. A DHHC-type zinc finger protein gene regulates shoot branching in Arabidopsis. African J. Biotech., **9**(45): 7759-7766.

Xiong D., Chen J., Yu T., Gao W., Ling X., Li Y. and Huang J. 2015. SPAD-based leaf nitrogen estimation is impacted by environmental factors and crop leaf characteristics. Sci. rep., **5**(1): 1-12.

Yin T., Pan G., Liu H., Wu J., Li Y., Zhao Z. and Zhou Y. 2012. The chloroplast ribosomal protein L21 gene is essential for plastid development and embryogenesis in Arabidopsis. Planta, **235**(5): 907-921.

Zhu S., Estévez J. M., Liao H., Zhu Y., Yang T., Li C. and Yu F. 2020. The RALF1–FERONIA complex phosphorylates eIF4E1 to promote protein synthesis and polar root hair growth. Mol. Plant, **13**(5): 698-716.