Population structure and association mapping for hundred seed weight in mungbean minicore

Muhammad Arslan Akhtar^{1*}, Muhammad Aslam¹, Roland Schafleitner², Muhammad Ahsan¹ and Ghulam Murtaza³

¹Department of Plant Breeding and Genetics, University of Agriculture, Faisalabad, Pakistan

²World Vegetable Center, Vegetable Diversity and Improvement, Tainan, Taiwan

³Institute of Soil and Environmental Sciences, University of Agriculture, Faisalabad, Pakistan

*Corresponding author: arslanpbg@gmail.com

Abstract

Mungbean is an important legume rich in protein and carbohydrates. It is cultivated mostly in south to south east Asia and east Africa. One of the important yield contributing traits in mungbean is 100 seed weight. Variation for this trait is available in the mungbean germplasm, and to facilitate improvement of this trait through breeding, genome-wide association mapping for 100 seed weight was carried out in the World Vegetable Center mungbean minicore collection. A total of 24,870 single nucleotide polymorphic markers were tested for association with 100 seed weight in plants grown in Pakistan, Bangladesh, Myanmar and Taiwan and candidate loci associated with seed weight on chromosomes 1, 4, 6, 8 and 9 were identified. None of the QTLs was stable across all environments, but loci on chromosomes 4 and 8 were significant in Pakistan and Bangladesh and other loci on chromosome 8 were significantly associated in plants grown in Myanmar and Taiwan.

Key words: Mungbean, seed weight, genome-wide association mapping, multilocation analysis

Introduction

Mungbean (*Vigna radiata* (L.) Wilczek), also known as green gram, is an important legume crop cultivated in south to south-east Asia and east Africa. Worldwide, it is currently cultivated on around 7.3 million hectares and produces about 5.3 million tons with an average yield of 0.73 tons per hectare. Major contributors towards mungbean production are India, China and Myanmar. The average yield of mungbean in Pakistan is comparable to the global average, but the overall production is less because of the small area under cultivation. The short vegetative period, the ability to withstand drought stress and fix atmospheric nitrogen along with high nutrient content makes mungbean a popular legume crop. 100 grams of boiled mungbean grains contain 7.02 g protein, 19.15 g carbohydrates, 7.60 g dietary fiber and 0.40 g fats, whereas 22% vitamin B1, 80% vitamin B9, 24% Mg, 16% Fe and 11% Zn on the basis of daily nutrient requirement [1]. Apart from grain, other plant parts are used as fodder and manure. The short cropping duration makes mungbean fit for many cropping patterns and combinations.

Large seeded mungbean achieves premium prices on markets, therefore large 100 seed weight is often preferred. Information on genetic loci contributing to this trait across environments would enhance breeding for larger grain. Molecular markers linked to various traits can help in the screening and selection of parents for the development of high yielding and disease resistant mungbean genotypes [2]. Loci associated with seed size in mungbean have been mapped previously in bi-parental populations [3,4]. Genome-wide association mapping or genome wide association studies (GWAS) is a complementary genetic approach used to associate variations in phenotype in germplasm panels, testing more alleles than in bi-parental

populations. In the present study, association between 100 seed weight and a genome-wide marker set was tested in a mungbean minicore collection at four locations in Pakistan (2016), Bangladesh (2017), Myanmar (2016) and Taiwan (2018).

Materials and methods

A total of 293 mungbean genotypes from the minicore set were cultivated in Pakistan (Islamabad, 33°40'42.6"N, 73°08'20.4" E, from March to May, 2016), Bangladesh (Ishurdi, 24°07'43"N, 89°03'57" E, from March to May 2017) Myanmar (Yezin, 19°50'02" N, 96°16'45" E, 133 m above sea level, from November 2016 to January 2017), and in Taiwan (World Vegetable Center, 23°01'30" N, 120°17'35" E, from September to November 2018). Data for 100 seed weight were determined on 3 replications of 5 plants per accession.

Genotyping data of the mungbean minicore accessions (Table 1) were obtained from Dart PRL as described by our fellows [5]. Imputation for the missing data was done using LD KNNi imputation protocol [6]. SNPs with more than 10% missing data after imputation and less than 5% minor allele frequency were filtered out.

Table 1. Mungbean minicore genotypes used under study

No.	VI no.	Origin	No.	VI no.	Origin
1	VI000020 AY	Thailand	150	VI003034 BG	India
2	VI000099 AG	India	151	VI003035 AG	India
3	VI000105 BG	India	152	VI003057 BG	India
4	VI000164 BG	Afghanistan	153	VI003062 BG	India
5	VI000170 B-BR	Afghanistan	154	VI003068 A-BR	India

6	VI000175 BY	India	155	VI003070 AG	India
7	VI000188 A-BLM	Pakistan	156	VI003083 BG	India
8	VI000203 B-BR	Afghanistan	157	VI003114 AG	India
9	VI000212 A-BLM	USA	158	VI003135 B-BL	India
10	VI000232 AG	Iran	159	VI003159 AG	India
11	VI000238 AG	Afghanistan	160	VI003172 BG	India
12	VI000253 AG	India	161	VI003181 B-GM	India
13	VI000316 AG	Pakistan	162	VI003183 AG	India
14	VI000317 BG	Pakistan	163	VI003187 BG	India
15	VI000319 AG	Pakistan	164	VI003212 B-BLM	India
16	VI000380 AG	Philippines	165	VI003220 AG	India
17	VI000461 BG	Thailand	166	VI003232 AG	India
18	VI000470 AG	Pakistan	167	VI003235 AG	India
19	VI000532 BG	India	168	VI003242 AG	India
20	VI000537 BG	India	169	VI003251 A-BL	India
21	VI000542 BY	India	170	VI003251 A-BLM	India
22	VI000551 AG	India	171	VI003252 BG	India
23	VI000554 AG	India	172	VI003255 AG	India
24	VI000559 AG	India	173	VI003276 BG	India
25	VI000578 AG	India	174	VI003329 AG	India
26	VI000589 B-BR	India	175	VI003332 AG	India
27	VI000616 BG	Brazil	176	VI003337 BG	India
28	VI000618 AG	India	177	VI003364 AG	India

29	VI000625 B-BR	India	178	VI003379 BG	India
30	VI000680 AG	USA	179	VI003382 BG	India
31	VI000723 AG	Iran	180	VI003407 AG	India
32	VI000732 AG	India	181	VI003413 BG	India
33	VI000735 BG	India	182	VI003440 AG	India
34	VI000736 AG	India	183	VI003455 AG	India
35	VI000749 AG	India	184	VI003456 AG	Unknown
36	VI000764 AG	India	185	VI003465 BG	India
37	VI000766 BG	India	186	VI003470 BG	India
38	VI000805 BG	India	187	VI003480 BG	India
39	VI000815 BG	India	188	VI003490 AG	India
40	VI000818 BG	India	189	VI003493 BG	India
41	VI000852 AG	India	190	VI003514 BG	India
42	VI000938 AG	India	191	VI003517 BG	India
43	VI000942 AG	India	192	VI003534 AG	India
44	VI000953 AG	India	193	VI003534 BG	India
45	VI000981 BG	Philippines	194	VI003548 AG	India
46	VI001023 BG	India	195	VI003554 AG	India
47	VI001066 BG	Australia	196	VI003560 BG	India
48	VI001096 AG	Australia	197	VI003563 A-BR	India
49	VI001124 AG	Australia	198	VI003577 AG	India
50	VI001126 BG	Australia	199	VI003602 AG	India
51	VI001162 AG	Australia	200	VI003642 AG	India

52	VI001191 BG	Philippines	201	VI003648 BG	India
53	VI001211 AG	Philippines	202	VI003658 BG	India
54	VI001221 AG	Philippines	203	VI003664 AG	India
55	VI001244 AG	Philippines	204	VI003678 BG	India
56	VI001268 BG	India	205	VI003685 AG	India
57	VI001282 AG	India	206	VI003699 B-BG	India
58	VI001284 AG	India	207	VI003720 BG	India
59	VI001339 AG	Philippines	208	VI003725 BG	India
60	VI001385 AG	India	209	VI003733 BG	India
61	VI001400 AG	India	210	VI003734 B-BR	India
62	VI001403 BR	India	211	VI003734 B-DG	India
63	VI001406 BG	Pakistan	212	VI003744 AG	India
64	VI001408 BG	India	213	VI003755 BG	India
65	VI001411 AG	India	214	VI003760 BG	India
66	VI001412 AG	India	215	VI003785 BG	India
67	VI001419 BG	India	216	VI003795 AG	India
68	VI001435 AG	USA	217	VI003801 BG	India
69	VI001448 A-BLM	India	218	VI003882 A-BLM	Afghanistan
70	VI001471 AG	India	219	VI003886 B-BR	India
71	VI001482 BG	India	220	VI003886 BY	India
72	VI001490 AG	Iran	221	VI003893 AG	India
73	VI001509 AG	Pakistan	222	VI003894 B-BLM	India
74	VI001514 AG	India	223	VI003907 AG	Iraq

75	VI001520 A-BLM	India	224	VI003914 AG	India
76	VI001533 BG	India	225	VI003925 B-BLM	India
77	VI001535 BG	India	226	VI003927 AG	India
78	VI001539 AG	India	227	VI003929 A-BL	India
79	VI001548 AG	India	228	VI003942 AG	Afghanistan
80	VI001556 BG	India	229	VI003944 B-BR	Afghanistan
81	VI001557 BG	USA	230	VI003947 B-BR	India
82	VI001562 AG	India	231	VI003948 B-BR	India
83	VI001576 BG	India	232	VI003951 AG	India
84	VI001579 BG	India	233	VI003954 BG	India
85	VI001605 BG	India	234	VI003957 AG	India
86	VI001612 AG	Unknown	235	VI003958 B-BLM	India
87	VI001628 AG	India	236	VI003959 BG	India
88	VI001651 BG	India	237	VI004006 A-GM	India
89	VI001652 BG	India	238	VI004010 AG	India
90	VI001654 BG	India	239	VI004024 AG	Australia
91	VI001678 BG	India	240	VI004044 BG	India
92	VI001692 AG	India	241	VI004045 A-DGM	India
93	VI001698 BG	India	242	VI004048 A-DGM	India
94	VI001728 AG	India	243	VI004069 BG	India
95	VI001733 BG	India	244	VI004096 AG	India
96	VI001743 BG	India	245	VI004096 BG	India
97	VI001756 BG	India	246	VI004129 A-BLM	Unknown

98	VI001762 A-GM	India	247	VI004138 BG	India
99	VI001806 AG	Pakistan	248	VI004145 B-BLM	Afghanistan
100	VI001806 BG	Pakistan	249	VI004184 AG	Netherlands
101	VI001820 BG	France	250	VI004243 B-BR	Turkey
102	VI001859 BG	Thailand	251	VI004244 B-BR	India
103	VI001974 BG	Korea	252	VI004297 AG	Afghanistan
104	VI001993 BG	Korea	253	VI004302 AG	Afghanistan
105	VI002009 BG	India	254	VI004307 AG	Afghanistan
106	VI002012 BG	India	255	VI004312 AG	India
107	VI002051 BG	India	256	VI004347 B-BLM	India
108	VI002063 BG	USA	257	VI004351 AG	India
109	VI002173 AG	India	258	VI004423 AG	Iran
110	VI002173 BG	India	259	VI004432 B-BR	Iran
111	VI002176 AG	India	260	VI004480 AG	Iran
112	VI002176 BG	India	261	VI004639 AG	Iran
113	VI002190 BG	India	262	VI004666 AG	Iran
114	VI002195 AG	Thailand	263	VI004691 AG	Iran
115	VI002197 BG	Korea	264	VI004694 BG	Iran
116	VI002206 AG	Philippines	265	VI004710 AG	Iran
117	VI002239 AG	Afghanistan	266	VI004734 AG	Iran
119	VI002284 BG	Afghanistan	267	VI004743 AG	India
120	VI002402 BG	Thailand	268	VI004789 BG	India
121	VI002432 AG	Thailand	269	VI004810 BG	India

122	VI002437 BG	Korea	270	VI004811 BG	India
123	VI002456 AG	Korea	271	VI004822 BG	India
124	VI002469 AG	Philippines	272	VI004838 AG	India
125	VI002487 AG	Pakistan	273	VI004842 AG	India
126	VI002523 AG	Thailand	274	VI004853 BG	India
128	VI002532 AG	India	275	VI004871 BG	India
129	VI002537 AG	Turkey	276	VI004877 AG	India
130	VI002569 BG	Nigeria	277	VI004915 BG	India
131	VI002587 AG	Australia	278	VI004931 AG	Pakistan
132	VI002611 AG	Thailand	279	VI004933 AG	Pakistan
133	VI002646 AG	Thailand	280	VI004934 AG	Pakistan
134	VI002647 AG	Thailand	281	VI004937 AG	Pakistan
135	VI002672 AG	Thailand	282	VI004942 BG	Pakistan
136	VI002739 AG	Iran	283	VI004954 BG	Pakistan
137	VI002802 A-BR	Iran	284	VI004956 AG	Pakistan
138	VI002859 BG	Iran	285	VI004957 AG	Pakistan
139	VI002860 AG	Iran	286	VI004958 BG	Pakistan
140	VI002872 BG	Iran	287	VI004965 BG	Pakistan
141	VI002877 BG	Iran	288	VI004968 AG	Pakistan
142	VI002894 B-BR	Iran	289	VI004969 AG	Pakistan
143	VI002926 AG	India	290	VI004973 B-BLM	India
144	VI002934 AG	India	291	VI005022 BG	India
145	VI002986 AG	India	293	VI005030 BY	Mexico

146	VI002993 BG	India	294	VI005041 AG	Unknown
147	VI002999 AG	India	295	VI005066 A-GM	India
148	VI003019 A-BLM	Unknown	296	VI014178 BG	Kenya
149	VI003019 BG	Unknown			

Archaeopteryx tree was developed using complete SNP array of the minicore set following clustering method of neighbor joining [7] within TASSEL (Trait Analysis by aSSociation, Evolution and Linkage, version 5.0). Population structure analysis to be used as co-factor in GWAS was performed in STRUCTURE (version 2.3.4) to assess the population subgroups (K) using a Markov Chain Monte Carlo (MCMC) approach with a burn in period of 50,000 and 500,000 MCMC repeats for K values K = 2 to K = 15 [8,9]. The output was analyzed according to [10] using STRUCTURE HARVESTER [11], to identify the most likely population structure (K-value) for the germplasm panel under investigation. GWAS was performed in TASSEL. Candidate SNPs were first manually examined by comparing the phenotypic values with the marker alleles in Microsoft Excel.

Results and discussion

The 100 SW data obtained from 4 environments were compared. Overall, seed size among the locations was different, but correlated (Table 2, Fig 1). Seed size in average was largest in Pakistan and smallest in Bangladesh. Stability of 100 SW across environments was strongly different among the genotypes.

Table 2. Pearson correlation coefficients and T-test results of the comparison of 100 SW across 4 locations in Bangladesh, Pakistan, Taiwan and Myanmar

Pearson		Pakistan	Taiwan	Myanmar
	Bangladesh	0.48 (p<0.001)	0.6 (p<0.001)	0.56 (p<0.001)
	Pakistan		0.62 (p<0.001)	0.65 (p<0.001)
	Taiwan			0.81 (p<0.001)
T-test		Pakistan	Taiwan	Myanmar
(p-value)	Bangladesh	7.59 E-83	0.01	0.00003
	Pakistan		3.9 E-81	2.52 E-55

76

77 The minicore set was genotyped with 6,506 SNP markers distributed over all 11
78 chromosomes. Population structure analysis indicated that the minicore collection contained 3
79 sub-populations. L(K) graph (Figs 2a and 2b) shows sharp variation from cluster 2 to cluster 3,
80 and delta K (ΔK) showing peak at cluster 3 indicates that the genotypes under study belong to 3
81 main clusters contrary to the findings of [5] who found out number of possible sub-populations
82 to be 2 or 4. Seed weight was not evenly distributed over the subpopulations. In average, 100SW
83 was larger in subpopulation 2 (average 5.9 g) than in subpopulation 1 and 2 (average 3.8 g). The
84 size differences between subpopulation 2 and the other two subpopulations was significant (p<
85 0.001).

86

Population structure graph developed by STRUCTURE (Fig 3) showed the distribution of the genotypes under study in their respective sub-populations. Most of the accessions showed at least some degree of admixture.

The population structure was in line with a phylogenetic tree That contained 3 major clusters, divided into sub-clusters (Fig 4). Quantile-quantile (QQ) graphs plot significance of associated SNP markers to the studied trait. The inclined ascending line shows the expected probability of the association of SNPs assuming there is no association of the trait of interest with SNPs studied, while the dots around the line show actual calculated probabilities. Deviation from the expected probability indicates association between SNPs and the trait of interest. The greater the difference between the expected line and calculated probability, the stronger will be the association [12]. QQ plots from all four locations showed different levels of association. The largest number of associated SNPs was obtained for the trial in case Myanmar, followed by Bangladesh, Pakistan and Taiwan.

Manhattan plots were developed to depict the association of SNP positions with 100 seed weight (Fig 6). Random dispersion of the SNPs indicates that there is no significant association, while dots above a certain threshold, here assumed as 4.0 ($-\log_{10}$ of p-value) indicate significant association. Vertical alignment of associated SNPs is formed if linked SNPs at a locus are significantly associated with the trait. This facilitates the identification of nearby genes as candidates for being involved with the trait variation. The analysis of the marker-trait association from Pakistan showed vertically aligned markers in chromosomes 2, 7 and 8 (Fig 6a) pinpointing

loci associated with 100 SW variation. In case of Bangladesh and Taiwan, aligned markers were present at chromosome 6, 7 and 8 (Figs 6b and 6d), and for Myanmar candidate markers were found on chromosomes 7 and 8 (Fig 6c).

The candidate SNPs were aligned along with genotypes and 100 seed weight arranged in ascending order to depict the association between SNP genotype and increased 100 seed weight. This helped in the screening of the false positive markers and selection of markers that were showing some degree of relatedness. It can be seen from table 2 that most of the associated sites are located on chromosome 8. Additionally, some sites are also commonly found associated with 100 seed weight across multiple locations. It means that the gene(s) for 100 seed weight is possibly locate around said positions of the markers on chromosome 8 (Table 3).

Genetic analysis of domestication related traits in a mungbean population derived from the wild accession JP211874 collected in Myanmar and the mungbean cultivar Sukhothai pointed towards loci on linkage group 8 (corresponding to chromosome 8 of the reference sequence [13] between markers CEDG059 (position ~ 15,462,360 bp) and VM37 (position ~29,308.079) [3] overlapping with markers found associated with seed size in Pakistan and Bangladesh. The seed size QTLs mapped by [4] could not be attributed to specific chromosomes as for most associated markers nor sequence was disclosed that could be attributed to a physical position.

Association of markers to 100 SW was mostly specific for a specific trial. Only a few loci on chromosome 8 were significant in trials in Myanmar and Taiwan, or between Pakistan and Bangladesh. The trial in Myanmar was performed in the monsoon season, while the trial in Taiwan was planted in the pre-monsoon season.

131 **Table 3. SNPs showing significant association with 100SW across locations**

Chromosome	Pakistan SNP position	Bangladesh SNP position	Myanmar SNP position	Taiwan SNP position
1		32,026,195	31,425,058	25,049,038
4	7,032,154	7,032,154		
	11,559,392			
6		5,433,578 34,701,957 34,712,555 34,744,525 34,913,611 35,030,031		
	36,662,459			
8			9,321,413 9,379,508 9,427,417 9,427,833	
			9,427,846	9,427,846
			9,432,667	9,432,667
			9,588,677	9,588,677

	27,458,129			
	27,246,365	27,246,365		
		28,480,139		
		28,480,205		
			29,255,058	
			35,992,362	
			35,978,574	
9	9,495,582			
	9,488,553		9,488,553	

Conclusion

Mungbean minicore set used in present study splits into 3 sub-populations. Most of the markers putatively associated to 100 seed weight in mungbean are located on chromosome 8. The association between SW and the marker at position 8:27,246,365 bp was common for the trials in Pakistan and Bangladesh data, whereas markers 8:9,588,667, 8:9,427,846 and 9,432,667 were common in trials in Myanmar and Taiwan. These markers are likely associated with 100 seed weight in mungbean and further investigations and validations are needed to confirm these results.

Acknowledgement

The above research was conducted at the World Vegetable Center, Taiwan, and was funded by Higher Education Commission (HEC) of Pakistan under International Research

Support Initiative Program (IRSIP) and by the long-term strategic donors to the World Vegetable Center, Taiwan, UK aid from the UK government, United States Agency for International Development (USAID), Australian Centre for International Agricultural Research (ACIAR), Germany, Thailand, Philippines, Korea, and Japan.

References

1. USDA, NRCS. 2020. The PLANTS Database (<http://plants.usda.gov>, April 2020). National Plant Data Team, Greensboro, NC 27401-4901 USA.
2. Collard BC, Mackill DJ. Marker-assisted selection: an approach for precision plant breeding in the twenty-first century. *Phil Trans. R Soc B.* 2008; 363: 557–572.
3. Isemura T, Kaga A, Tabata S, Somta P, Srinives P, Shimizu T, et al. Construction of a genetic linkage map and genetic analysis of domestication related traits in mungbean (*Vigna radiata*). *PLoS one.* 2012; 7(8): 1-14.
4. Alam AKMM, Somta P, Mukhtadir MA, Srinives P, et al. Quantitative Trait Loci Associated with Seed Weight in Mungbean (*Vigna radiata* (L.) Wilczek). *Kasetsart J - Natural Sci.* 2014; 48(2): 197-204.
5. Breria CM, Hsieh CH, Yen JY, Nair R, Lin CY, Huang SM, et al. Population Structure of the World Vegetable Center Mungbean Mini Core Collection and Genome-Wide Association Mapping of Loci Associated with Variation of Seed Coat Luster. *Trop Plant Biol.* 2020; 13: 1-12.
6. Money D, Gardner K, Migicovsky Z, Schwaninger H, Zhongand GY, Myles S, et al. LinkImpute: Fast and Accurate Genotype Imputation for Nonmodel Organisms. *G3: Genes, Genomes, Genetics.* 2015; 5: 2383-2390.

7. Saitou N, Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol.* 1987; 4(4): 406-425.
8. Pritchard JK, Stephens M, Donnelly P, et al. Inference of Population Structure Using Multilocus Genotype Data. *Genetics.* 2000; 155: 945–959.
9. Pritchard JK, Wen X, Falush D, et al. Documentation for structure software Version 2.3. 2010. https://web.stanford.edu/group/Pritchardlab/software/readme_structure2.pdf.
10. Evanno G, Regnaut S, Goudet J, et al. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol Ecol.* 2005; 14: 2611–2620.
11. Earl DA, Vonholdt BM. STRUCTURE Harvester: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv Genet Resour.* 2012; 4: 359–61.
12. Kang Y, Sakiroglu M, Krom N, Geddes JS, Wang M, Lee YC, et al. Genome-wide association of drought-related and biomass traits with HapMap SNPs in *Medicago truncatula*. *Plant Cell Environ.* 2015; 38: 1997–2011.
13. Kang YJ, Kim SK, Kim MY, Lestari P, Kim KH, Ha BK, et al. Genome sequence of mungbean and insights into evolution within *Vigna* species. *Nat Commun.* 2014. 5(5443): 1-9.

Figure 1

Figure 2a

Figure 2b

Figure 3

bioRxiv preprint doi: <https://doi.org/10.1101/2021.05.04.442577>; this version posted May 4, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Figure 4

Expected -Log₁₀(P-Value) vs. -Log₁₀(P-Value)

Figure 5a

Expected -Log10(P-Value) vs. -Log10(P-Value)

Figure 5b

Expected -Log10(P-Value) vs. -Log10(P-Value)

Figure 5c

Expected -Log₁₀(P-Value) vs. -Log₁₀(P-Value)

Figure 5d

P-Values by Chromosome for Pak

Figure 6a

P-Values by Chromosome for Ban

Figure 6b

P-Values by Chromosome for Taiw

Figure 6d