



Full Length Article

Intraspecific Variation in Soybean (*Glycine max*) Revealed by Multivariate Approaches to Morphometric Traits and Soybean Mosaic Virus Resistance

Haris Khurshid^{1,2*}, Doulat Baig¹, Nazakat Nawaz¹, Malik Ashiq Rabbani³ and Zabta Khan Shinwari²

¹*Oilseeds Research Program, National Agricultural Research Centre, Islamabad, Pakistan*

²*Department of Biotechnology, Quaid-i-Azam University Islamabad, Pakistan*

³*Bioresource Conservation Institute, National Agricultural Research Centre, Islamabad, Pakistan*

*For correspondence: hariskhurshid8@gmail.com

Received 09 December 2019; Accepted 29 January 2020; Published 02 April 2020

Abstract

Soybean crop improvement for grain yield and disease resistance has remained a primary target for plant breeders. This aim has been achieved by probing unique and distinct variation in diverse gene pools of the crop to enhance the expression level of economically important traits. In the present study, we employed diversity driven strategy using thirteen quantitative and qualitative descriptors to a set of 110 local and exotic soybean genotypes for elucidating its population structure. Experiment was planted for two years consecutively in Kharif 2016–2017 in augmented field design having 5m row of each genotype while plot was flanked by *begomovirus*-susceptible cotton and mungbean lines. Data were recorded on five representative plants from each row and soybean mosaic virus response was scored on a scale of 0–9, phenotypically. Multivariate procedure of principal component analysis revealed 51% variability in first two principal components. Higher eigenvectors were drawn on variable's plot for days to flower initiation, days to 50% flowering, number of pods plant⁻¹, plant height, seed yield and hundred seed weight showing variation. We observed strong association between disease resistance, oil content and higher seed yield. Heatmap based hierarchical cluster analysis was performed using Euclidean distances which classified all studied genotypes into five major clusters due to morpho-agronomic and geographic affinity. A total of 47 genotypes were identified as resistant followed by 22 highly resistant and 23 moderately resistant. We concluded that sufficient genetic variation exists for yield related traits in available soybean germplasm which can be utilized in breeding programs for developing ideotype with higher grain yield and resistance to soybean mosaic virus. © 2020 Friends Science Publishers

Keywords: Soybean; Population structure; Seed yield; SMV; Phenotyping

Introduction

Soybean (*Glycine max* L. Merrill) is most important oilseed crop widely grown for its high-quality protein and oil content. Its seed is major source of nearly half of the world's vegetable oil (Vollmann 2016) while its residual meal is utilized as animal feed (Khurshid *et al.* 2017). Globally, 337,452 MMT soybean was produced during 2017–2018 to meet the increasing demand of soybean grain for food and feed consumption (USDA 2018). Soybean was domesticated in China nearly 3000 years ago where it remained a local crop consumed for its grain until introduced to the USA in 1770 AD (Hart 2017). Since then, United States has been the dominant market leading as the largest soybean producer and consumption country. The crop is widely adapted to diverse agro-ecological conditions between ~35 S and ~54 N around the globe, however it grows relatively good in middle or high-latitude regions

under warm and humid conditions. The presence of genetic structure or broad genetic variability for agro-morphological traits and phenology is mainly due to earlier exploitation of the soybean crop in geographically diverse regions in China (Singh *et al.* 2007). This allows soybean breeders to design crop improvement programs for developing high yielding, adaptable cultivars which performs well under variable edaphoclimatic conditions and production systems. Grain yield in soybean is the most important trait from breeding standing point hence over the last few decades it has improved at the rate of 23 kg hectare⁻¹ year⁻¹ (Orf *et al.* 2004). However, the yield as a polygenic trait is expressed genetically via sophisticated mechanism of plethora of genes and their interaction while phenotypically as result of complex agronomic traits and physiological functions (Sleper and Poehlman 2006). Beside yield, protein and oil content concentration, plant height, seed weight and pest resistance has been some of the most pursued breeding

objectives in soybean crop (Bilyeu *et al.* 2016).

In Pakistan environmental conditions are well suited for soybean production however lack of new high yielding and disease resistant commercial varieties is major bottleneck for wide scale adaptation of the crop (Khurshid *et al.* 2017). The prevalence of pest mediated viral diseases of major crops in irrigated plains of the country also pose risk of infestation in soybean. Among them, soybean mosaic virus (SMV) is most prevalent and drastically reduces grain yield. Globally, the disease has been reported to cause yield losses up to 94% depending upon the cultivar, viral strain and climatic conditions (Ross 1977). In US alone, the disease has caused up to \$35 million losses to growers despite improved cultural practices and integrated pest management (Hill and Whitham 2014). The disease is transmitted through insect vectors such as white fly and aphids mostly in late growing Kharif crop season under hot humid condition (Khan *et al.* 2013). The vectors widespread presence in irrigated plains of Punjab and Sindh due to availability of hosts *i.e.*, cotton and vegetables cultivation poses sever threat to epidemic of the soybean mosaic virus (Arif *et al.* 2000). The SMV affected plants are characterized with stunted growth, curled leaves having yellow spots, low vigor and mottled seeds. However, the broad base genetic variation in soybean has augmented natural resistance against different diseases including soybean mosaic virus. Researchers have mapped three independent loci *i.e.*, *Rsv1*, *Rsv3*, *Rsv4* (Kiihl and Hartwig 1979; Buzzell and Tu 1989; Buss *et al.* 1997) responsible for conferring genetic resistance against SMV causing strains (G1 to G7) of *potyvirus* on chromosome 13, 14 and 2, respectively. Developing cultivars with natural resistance against SMV is the most economic, efficient and environment friendly method to mitigate the losses caused by the disease (Ahangaran *et al.* 2013). In addition to diseases, challenges like low yield potential, climate change and abiotic stresses are rendering old cultivars as obsolete (Khurshid *et al.* 2017). Continuous efforts to select agronomically superior soybean genotypes are crucial for development of new cultivars. Furthermore, the knowledge of plants collection, characterization and key information regarding quantitative and qualitative traits are important for efficient utilization of the germplasm's genetic potential. Therefore, it is imperative to screen available soybean germplasm against soybean mosaic virus resistance and other yield related traits under local agro-climatic conditions (Baig *et al.* 2018). Traditionally, this is achieved by selecting desirable genotypes based on morphometric, genealogical and phenological traits in soybean germplasm. There is dearth of new high yielding soybean cultivars in Pakistan which can cope with ever changing climate and biotic stress from viral diseases such as mosaic virus. In the present study we evaluated local and exotic soybean germplasm to select high yielding and adaptable genotypes having resistance against SMV. The efforts were made to identify elite lines for further utilization in soybean breeding program.

Materials and Methods

Plant material

A total of one hundred and ten genotypes of soybean (*Glycine max* L.) acquired from gene bank of Bioresource Conservations Institute, National Agricultural Research Centre (NARC), Oilseeds Research Program, NARC, National Institute of Biotechnology and Genetic Engineering (NIBGE), Faisalabad, US Department of Agriculture (USDA) and Asian Vegetables Research and Development Center (AVRDC), Taiwan were used in the study.

Field experiment

The experiment was conducted consecutively for two years during Kharif 2016 and 2017 in experimental field at NARC Islamabad, located at 33° 40'24" N, 73° 7' 27" E and 502 m above sea. The climatic subtype of the site is categorized as humid subtropical "Cfa" according to Köppen-Geiger classification. The Kharif 2016 and 2017 at site was characterized as humid hot cropping season with limited rainfall (Fig. 1). A well-prepared moist clay loam soil was treated with 50 kg of DAP per acre before sowing. Augmented field design was used for the experiment where each accession was planted on a ridge using hand drill in a single row of 5 m length while plant to plant and row to row distance was maintained at 3–4 cm and 45 cm, respectively. The experiment plot was flanked by cotton (*Gossypium hirsutum*) and mungbean (*Vigna radiata*) lines susceptible to attack of whiteflies (*Bemisia tabaci*) and aphids (*Aphis glycines*) carrying *begomoviruses* to provide ideal conditions for Soybean Mosaic Virus (SMV) infestation. The germplasm was screened against soybean mosaic virus at BBCH stage 49 when harvestable main shoot was fully developed. The disease infection was scored on the lamina of infected leaves on the scale (0–9) as described by Khan *et al.* (2013). For agro-morphological characterization, data were recorded according to soybean descriptor of Bioversity international on five randomly selected plants of each accessions for days to flowering initiation (DFI), days to 50% flowering (DF50%), days to flower completion (DFC), days to maturity (DM), plant height (PH), primary branches plant⁻¹ (PBPP), number of pods plant⁻¹ (NPPP), seed yield (SY kg ha⁻¹), hundred seed weight (100SW g), flower color (FC) and plant habit (PHb) Table 1. Recommended agricultural practices were performed to obtain a good yield and individual accessions were harvested manually after reaching physiological maturity and 17% grain moisture content.

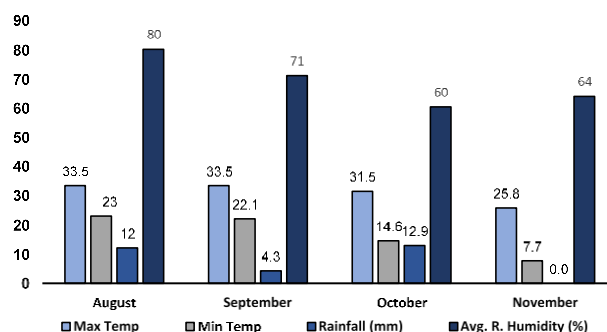
Data analysis

Mean data of all quantitative traits were used for descriptive statistics analysis in statistical package Statistica 7.0

Table 1: Details of phenotypic traits observed in the soybean population

Trait	Time	Method
Days to flower initiation	BBCH [*] stage 600	Number of days from sowing to first flower opening
Days to 50% flowering	BBCH stage 605	Number of days from sowing to 50% flowering
Days to flowering completion	BBCH stage 609	Number of days from sowing to last day of flowering
Days to Maturity	BBCH Stage 909	Number of days from sowing till browning and maturing of all pods with developed seeds
Plant height (cm)	BBCH stage 409	Length of plant main shoot from basal node to apical bud
Number of primary branches plant ⁻¹	BBCH stage 609	Number of pod-bearing branches arising from the main stem
Number of pods plant ⁻¹	BBCH stage 805	Number of fully developed grain filled pods on the plant
Seed yield (kg ha ⁻¹)	BBCH stage 909	Weight of seed obtained from a single row was converted to kg per hectare
100 Seed weight (g)	BBCH stage 909	Weight of hundred random seeds (g)
Flower color	BBCH stage 601	Color of corolla (white, purple etc.)
Plant Habit	BBCH stage 409	Plant stature (erect, prostrate or semi-erect)
Disease resistance (0-9)	BBCH stage 201-409	0= Immune, 1=Resistant, 3=Moderately resistant, 5=Moderately susceptible, 7= susceptible, 9= Highly susceptible

* Biologische Bundesanstalt, Bundessortenamt and Chemical industry manual of plant growth stages

**Fig. 1:** Mean climate data for *Kharif* 2016-17 recorded at NARC Islamabad

(Statsoft, Tulsa-USA) while frequency distribution of disease scoring, and quantitative traits was represented using Microsoft Excel 2016. The data was also subjected to multivariate and heatmap analysis using statistical packages factextra, Phangorn and ggplot2 in language R.

Results

The statistical analysis of agro-morphological data of 110 soybean genotypes dissected genetic structure of the population. Ten quantitative traits data were subjected to descriptive statistics (Table 2). Higher coefficient of variation (CV%) was observed for seed yield, number of pods plant⁻¹ and plant height. Among 110 genotypes white flower color was observed in 55 genotypes (50%) while the rest had purple flowers. The population showed diversity for plant growth habit as 73 genotypes (67%) were erect type followed by 20 semi-erect type (18%) while only 17 genotypes (15%) exhibited prostrate type growth. Pearson's correlation matrix was calculated for all the quantitative traits (Table 3). Highly significant correlation was observed between studied traits. Highly significant positive correlation was observed between flowering phenology parameters *i.e.*, Days to flowering initiation, days to 50% flowering and days to flowering completion. Days to flowering initiation and flowering completion also had positive association with plant height, number of pods plant⁻¹

and seed yield. Moreover, flowering initiation, 50% flowering and days to flowering completion were found to be negatively correlated to primary branches plant⁻¹ and hundred seeds weight. Seed yield showed positive correlation with pod plant⁻¹ and plant height.

Multivariate approach of principal component analysis was applied to ten economically important quantitative traits data to estimate genetic variation. Among 10 principal components (PCs), first three were selected with Eigen value equal or higher than 1. These accounted for 65% of the total variability within the studied germplasm. The first principal component (PC1) represented 34% of the variability which was mainly augmented by hundred seed weight, primary branches plant⁻¹, oil content (%) and days to maturity. Flowering initiation, 50% flowering and flowering completion contributed negatively to first PC. The second principal component showed 17% variation in population mainly due to positive effect of days to maturity, flowering initiation and days to 50% flowering. Primary branches plant⁻¹, hundred seed weight and seed yield were negative contributor to PC2. Similarly, PC3 depicted 13% variation which was attributed to positive effect of days to maturity, primary branches plant⁻¹ and seed yield. Moreover, oil content negatively affected the third principal component. Genotypes were scattered by PCA plot in all four quadrants showing presence of phenotypic diversity *viz-a-viz* studied traits (Fig. 2a). The variables interaction was observed with positive and negative effect (Fig. 2b). Vector of days to maturity moved in opposite direction of seed yield. However, days to flower completion, plant height, number of pods plant⁻¹ and seed yield were nearly parallel. The relatively higher loadings or vector's length of days to flowering initiation, days to 50% flowering, number of pods plant⁻¹ and primary branches plant⁻¹ showed that these traits were highly variable in the studied population.

Heatmap based hierarchical cluster analysis was performed using Euclidean distances between all the studied genotypes as shown in the dendrogram on left side (Fig. 3). All 110 genotypes were distributed in 5 major groups on the basis of agronomic traits. In the dendrogram (from top to bottom), group 1 included 15 genotypes. These genotypes showed trend of delayed flowering initiation and prolonged

Table 2: Descriptive statistics for 10 quantitative traits observed in 110 soybean genotypes

Traits	Mean	Minimum	Maximum	Range	Variance	Std. Dev.	St. Error	CV%
DFI	37.82	30	47	17	16.33	4.04	0.39	10.69
DF50%	47.13	36	53	17	13.30	3.65	0.35	7.74
DFC	52.76	43	57	14	9.58	3.09	0.30	5.87
DM	95.29	92	98	6	3.73	1.93	0.18	2.03
PH	43.88	14.30	92.70	78	231.37	15.21	1.45	34.66
PBPP	8.29	5	12	7	3.07	1.75	0.17	21.14
NPPP	75.92	12	271	259	2103.27	45.86	4.37	60.41
SY	660.89	18.89	3407.56	3388.67	4875.25	698.24	66.57	1.06
HSW	12.54	5.87	25.32	19.45	9.36	3.06	0.29	24.40
Oil %	19.68	14.41	22.89	8.48	1.82	1.35	0.13	6.85

* DFI= days to flowering initiation, DF50%= days to 50% flowering, DFC= days to flowering completion, DM= days to maturity, PH= plant height (cm), PBPP= primary branches plant-1, NPPP= number of pods plant-1, SY= seed yield (kg ha⁻¹), HSW= hundred seed weight (g), OC%= oil content %

Table 3: Correlation between quantitative traits in soybean germplasm

Trait	DFI	DF50%	DFC	DM	PH	PBPP	NPPP	SY	HSW
DFI	1								
DF50%	0.79**	1							
DFC	0.45*	0.63*	1						
DM	-0.04	0.03	-0.16	1					
PH	0.35	0.42*	0.21	-0.17	1				
PBPP	-0.55**	-0.44**	0.01	-0.15	-0.12	1			
NPPP	0.43*	0.51*	0.36*	-0.17	0.56*	-0.11	1		
SY	0.16	0.30	0.17	0.05	0.35	0.27	0.44	1	
HSW	-0.39	-0.42*	-0.13	-0.17	-0.14	0.39	-0.17	-0.08	1
OC%	-0.06	-0.14	-0.01	-0.12	-0.03	-0.05	0.02	-0.17	0.39

* = Significant at $\alpha = 0.005$

** = Highly Significant at $\alpha = 0.005$

DFI= days to flowering initiation, DF50%= days to 50% flowering, DFC= days to flowering completion, DM= days to maturity, PH= plant height (cm), PBPP= primary branches plant-1, NPPP= number of pods plant-1, SY= seed yield (kg ha⁻¹), HSW= hundred seed weight (g), OC%= oil content %

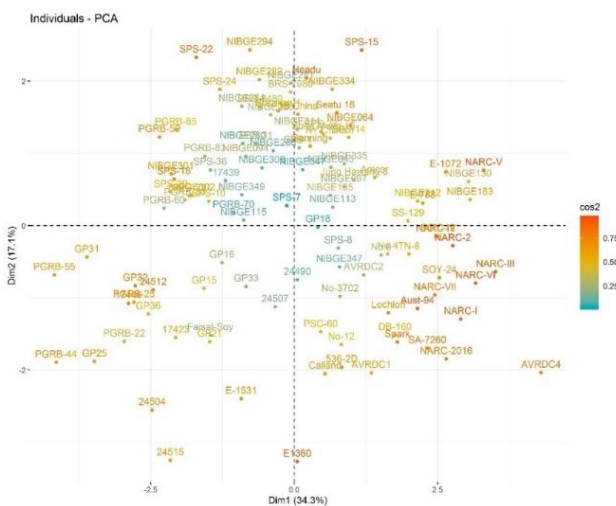


Fig. 2a: Distribution of 110 soybean genotypes in quadrants showing phenotypic variation

flowering. The plant height mean of these genotypes was highest (70 ± 14.3 cm) and early maturing (94 ± 1.6 days). These were also characterized with maximum pod bearing (145 ± 50 pods) and highest seed yielding (352.7 ± 35 g). The group 2 comprised of 43 medium statured lines (45.6 ± 10.6 cm) having lowest average number of primary branches plant⁻¹ (7.3 ± 1.08) but high oil content ($20 \pm$

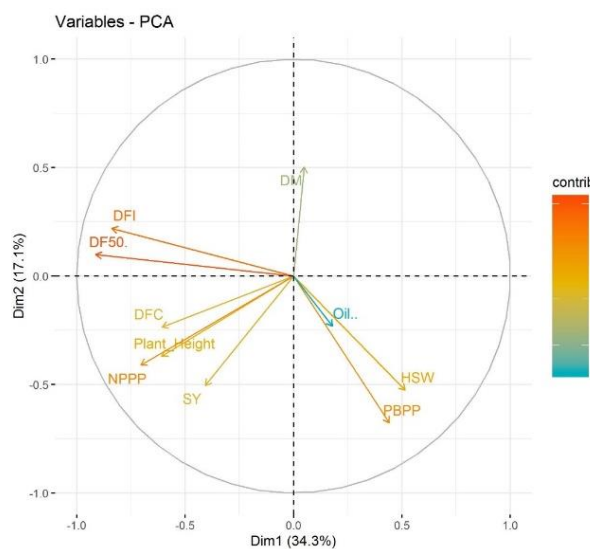


Fig. 2b: Contribution of agronomic traits in explaining phenotypic variation

1.05%). This group included accessions from Oilseeds Research Program NARC and material introduced from Korea and China. Likewise, 10 genotypes settled in group 3 based on their late maturity (97 ± 0.63 days), shortest height (32.81 ± 8.7 cm), lowest hundred seed weight (8.3 ± 1.32 g) and oil content ($17.5 \pm 1.66\%$). Most of these genotypes had

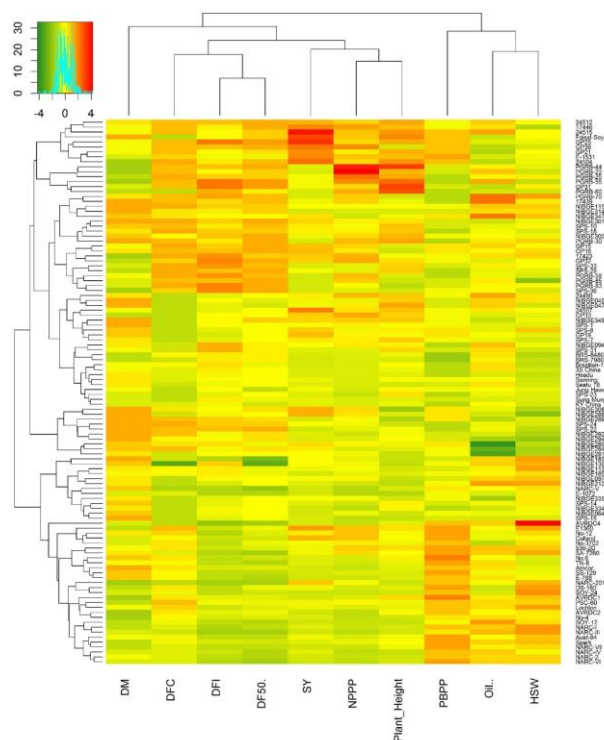


Fig. 3: Hierarchical clustering based classification of 110 soybean genotypes for quantitative traits in various groups

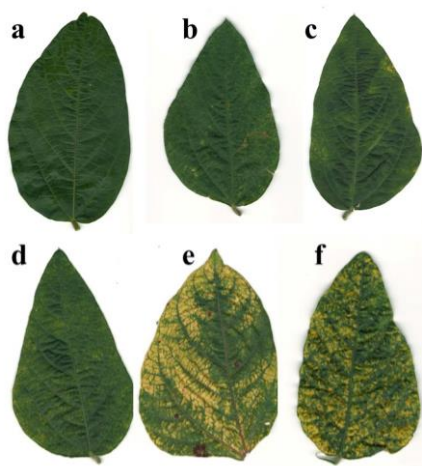


Fig. 4: Soybean mosaic virus disease severity categories; a: highly resistant (HR), b: resistant (R) , c: moderately resistant (MR), d: moderately susceptible (MS), e: susceptible (S) and f: highly susceptible (HS)

origin from gene banks of USDA. Group 4 had 13 genotypes with shortest flowering duration (49 ± 2.27 days), plant height (31.6 ± 6.75 cm) and minimum pods plant⁻¹ (46 ± 22.31). Also, these plants were late maturing (96 ± 1.03 days) but had bold seeds with mean HSW of 14.3 ± 2.06 g. The genotypes of this group were mostly USDA accessions. The fifth group comprised of 29 early flowering (33 ± 1.5

days) and maturing (94 ± 1.60 days) genotypes having highest number of branches plant⁻¹ (10 ± 0.8), HSW (15 ± 2.88 g) and oil content ($20.1 \pm 0.84\%$). Moreover, this group had below average performance for number of pods plant⁻¹ and seed yield. This group included most of the 1980s and 90s varieties introduced from USA or improved lines of Oilseeds Research Program. The upper dendrogram in Fig. 3. clubbed correlated traits together. From left to right, traits related to flowering phenology and days to maturity grouped together. Similarly, traits *i.e.*, seed yield, pods plant⁻¹, plant height and primary branches plant⁻¹, hundred seed weight and oil content (%) settled in same clusters, respectively.

Response to soybean mosaic virus disease

The higher average relative humidity and vicinity of the experiment to mungbean and cotton crop provided perfect hotbed for development of SMV vector *i.e.*, aphids and whitefly (Fig. 4). The disease scoring of all the 110 genotypes revealed highly variable response against soybean mosaic virus (Table 4). The leaf surface, lesions, and proportion of yellowed versus normal leaf were phenotypically observed (Fig. 4). On the scale from highly susceptible to highly resistant, 47 genotypes were found to be resistant followed by 23 moderately resistant and 22 highly resistant. Only six accessions *i.e.*, Sanning, 3S China, Headu, BRS-8480, BRS-7980 and Brazilian-1 showed symptoms of high susceptibility to the disease virulence. Likewise, 5 genotypes *i.e.*, TN8, Seatu 18, KY China, NARC-I and NARC-III were susceptible and while SS-129, Amcor, No.6, Jung Hawang 8, Sung Mung 15, No-4 and Spark were categorized as moderately susceptible.

The SY (kg ha⁻¹) and oil content (%) of different genotypes with varying level of resistance to SMV was observed. The highly resistant (HR) or immune genotypes had average seed yield of 1777 kg ha⁻¹ to maximum of 3111 kg ha⁻¹ and 21% oil content. This was followed by resistant (R) lines with average yield of 1333 kg ha⁻¹ and 21% oil content. Moderately resistant genotypes also performed above average for yield and oil content. Moreover, all the three categories of susceptibility *i.e.*, susceptible, moderately susceptible and highly susceptible performed poor for these traits. Some susceptible lines showed consistent yield due to post pod formation development of SMV symptoms, but these were considered as outliers.

Discussion

Soybean landraces have been grown since long in certain areas of Northern Pakistan for domestic consumption. Moreover, as a modern day crop, it is relatively new entrant in the county's agricultural research and development whereas its varietal development started in 80s by national research institutes (Khurshid *et al.* 2017). The unavailability of indigenous genetic diversity in soybean necessitates its

Table 4: SMV Response and qualitative traits description of 110 soybean genotypes

S. No	Genotype	Origin	FC	PHb	Response	S. No	Genotype	Origin	FC	PHb	Response
1	NIBGE045	USDA	W	E	MR	56	SPS-7	NARC	W	E	MR
2	NIBGE047	USDA	W	E	MR	57	SPS-8	NARC	W	E	R
3	NIBGE064	USDA	P	E	MR	58	SPS-10	NARC	W	E	R
4	NIBGE094	USDA	P	Pr	MR	59	SPS-14	NARC	W	E	R
5	NIBGE097	USDA	P	E	R	60	SPS-15	NARC	W	E	MR
6	NIBGE113	USDA	P	Pr	HR	61	SPS-18	NARC	W	E	R
7	NIBGE115	USDA	P	Pr	MR	62	SPS-22	NARC	W	E	R
8	NIBGE130	USDA	P	Pr	MR	63	SPS-23	NARC	W	E	R
9	NIBGE183	USDA	P	Pr	MR	64	SPS-24	NARC	W	E	R
10	NIBGE185	USDA	W	Pr	R	65	SPS-29	NARC	W	E	R
11	NIBGE212	USDA	P	E	HR	66	SPS-31	NARC	W	E	MR
12	NIBGE280	USDA	P	Pr	HR	67	SPS-33	NARC	W	E	MR
13	NIBGE281	USDA	W	E	R	68	SPS-36	NARC	W	E	MR
14	NIBGE282	USDA	P	E	R	69	SS-129	NARC	P	SE	MS
15	NIBGE284	USDA	P	E	R	70	Amcor	USDA	P	SE	MS
16	NIBGE288	USDA	W	E	HR	71	No-3702	USDA	W	SE	HR
17	NIBGE289	USDA	W	Pr	R	72	Lochlon	USDA	P	SE	R
18	NIBGE294	USDA	W	Pr	R	73	TN-8	USDA	P	SE	S
19	NIBGE301	USDA	P	Pr	R	74	Aust-94	USDA	P	SE	MR
20	NIBGE302	USDA	W	Pr	R	75	E-1531	USDA	P	E	HR
21	NIBGE308	USDA	P	Pr	HR	76	Calland	USDA	W	E	HR
22	NIBGE314	USDA	W	Pr	MR	77	No.6	USDA	P	SE	MS
23	NIBGE334	USDA	P	Pr	R	78	E-1072	USDA	P	SE	R
24	NIBGE335	USDA	W	E	R	79	E-788	USDA	P	SE	R
25	NIBGE347	USDA	P	E	R	80	NARC-2	NARC	W	E	HR
26	NIBGE349	USDA	P	E	R	81	KYChina 1	China	P	SE	S
27	Faisal-Soy	Faisalabad	W	E	HR	82	Jung Hawang 8	S. Korea	W	SE	MS
28	GP15	NARC	W	E	R	83	Sung Mung 15	S. Korea	P	E	MS
29	GP16	NARC	P	E	R	84	Seatu 18	S. Korea	P	E	S
30	GP18	NARC	W	E	R	85	Sanning	S. Korea	W	E	HS
31	GP21	NARC	P	E	HR	86	3S China	China	W	SE	HS
32	GP25	NARC	W	E	HR	87	Headu	S. Korea	W	SE	HS
33	GP31	NARC	W	Pr	R	88	BRS-8480	Brazil	W	E	HS
34	GP32	NARC	W	E	R	89	BRS-7980	Brazil	W	E	HS
35	GP33	NARC	P	E	R	90	Brazilian-1	Brazil	W	SE	HS
36	GP36	NARC	P	E	HR	91	E1360	USDA	W	SE	HR
37	17423	NARC	P	E	R	92	No-4	USDA	P	E	MS
38	17439	NARC	W	E	R	93	Spark	USDA	P	E	MS
39	17446	NARC	P	E	HR	94	536-2D	USDA	P	E	HR
40	24490	NARC	P	E	R	95	No-12	USDA	P	SE	HR
41	24504	NARC	P	E	HR	96	DB-160	USDA	P	SE	MR
42	24507	NARC	W	E	R	97	SOY-12	S. Korea	P	E	MR
43	24512	NARC	W	E	MR	98	SOY-24	S. Korea	P	E	MR
44	24515	NARC	P	E	R	99	AVRDC1	Taiwan	P	E	HR
45	PGRB-22	NARC	P	E	R	100	AVRDC2	Taiwan	P	E	R
46	PGRB-25	NARC	W	E	R	101	AVRDC4	Taiwan	P	SE	R
47	PGRB-30	NARC	W	E	MR	102	PSC-60	NARC	W	SE	HR
48	PGRB-44	NARC	W	E	R	103	SA-7260	NARC	P	SE	HR
49	PGRB-55	NARC	W	E	R	104	NARC-I	NARC	P	E	S
50	PGRB-59	NARC	W	E	R	105	NARC-III	NARC	W	E	S
51	PGRB-60	NARC	W	E	R	106	NARC-IV	NARC	W	E	R
52	PGRB-70	NARC	W	E	MR	107	NARC-V	NARC	P	E	MR
53	PGRB-83	NARC	P	Pr	R	108	NARC-VI	NARC	P	E	MR
54	PGRB-85	NARC	W	Pr	R	109	NARC-VII	NARC	P	E	MR
55	SPS-1	NARC	W	E	R	110	NARC-2016	NARC	P	E	HR

FC= Flower colour, W= white, P= purple, PHb= plant growth habit, E= erect, Pr= prostrate, SE= semi-erect, HR= highly resistant, R= resistant, MR= moderately resistant, MS= moderately susceptible, S= susceptible, HS= highly susceptible

introduction from diverse regions for detailed analyses of genetic structure for breeding new varieties. The present study focused on the objective to elucidate genetic variation among 110 soybean accessions and varieties for economically important descriptors as well as to identify potential high yielding and SMV resistant promising lines. Recently, similar studies have been conducted by curators to

assess genetic variation in soybean in different countries (Malek *et al.* 2014; Andayanie *et al.* 2017; Oliveira *et al.* 2017). Soybean yield is dependent on number of pods plant¹, number of seeds pod¹ 100 seed weight, number of branches plant¹ and plant height (Carpenter and Board 1997; Liu *et al.* 2010). In the present study soybean population exhibited significantly higher variance and coefficient of variation for

these traits. This implies presence of genetic divergence for these agronomic parameters which can be used as selection criterion while improving grain yield. Previously, Arshad *et al.* (2006) and Malik *et al.* (2011) also reported higher variation for these quantitative traits while studying phenotypic variation in soybean germplasm. As polygenic traits these parameters are highly prone to environmental variation. However, as important components of the seed yield, variation in these traits can be useful for developing high yielding cultivars (Baig *et al.* 2018).

In our study, correlation between different quantitative traits was determined to analyze association for efficient ideotype selection based on traits inter-relatedness. We observed strong positive correlation between phenological descriptors *i.e.*, flowering initiation, 50% flowering, flowering completion and days to maturity. These results were in conformity with Malek *et al.* (2014) while studying genetic divergence and character association in soybean mutant lines. Arshad *et al.* (2006) reported strong highly significant association between days to maturity, branches plant⁻¹ and pods plant⁻¹ while a negative association between days to maturity, days to flowering completion, 100 seed weight and seed yield was found. Likewise, they reported highly significant positive correlation between number of pods plant⁻¹, number of branches plant⁻¹ and 100 seed weight. These results were in congruence with findings of our study. Kumar *et al.* (2019) confirmed our results while investigating association between soybean yield and related traits under western Himalayan conditions. It is imperative for soybean breeders to opt for negative selection of traits having adverse relationship with seed yield and to devise strategies for breaking undesirable linkages. Researchers have extensively used multivariate analyses of plant morphometric data for estimation of genetic variation in different crops (Arif *et al.* 2015; Saleem *et al.* 2017; Shah *et al.* 2018; Khurshid *et al.* 2019). In the present study we utilized principal component analysis (PCA) for quantitative traits to show variation in soybean genotypes and explain interactivity of these traits or variables. First two principal components (PCs) revealed a total of 51% variation mainly in highly diverse polygenic traits such as plant height, number of pods plant⁻¹, number of branches plant⁻¹, seed yield and hundred seed weight. Hence these traits can be perfectly characterized as ideal candidates for selection criteria in a breeding program. Our results were in line with those of Khan *et al.* (2014) as they employed PCA on 11 quantitative traits data of 115 soybean accessions whereupon first two PCs dissected 55% of the total phenotypic variation. Malek *et al.* (2014) also reported similar pattern of variability among quantitative traits of a soybean population. Variables' vectors in biplot indicated strong relationships among important traits such as oil content, primary branches plant⁻¹ and hundred seed weight. The vector for seed yield was drawn somewhat parallel to plant height and number pods plant⁻¹. This implies that new selections in soybean germplasm should be directed at higher number of branches

and pods plant⁻¹ to breed high yielding varieties with better oil content. Zhao *et al.* (2007) employed PCA to investigate phenotypic diversity of soybean varieties and their results supported our findings. In individuals' plot all 110 genotypes were dispersed across four quadrants showing diversity. The distribution pattern of soybean genotypes in plot was found to be driven by their respective origins *i.e.*, accessions or varieties from NIBGE and NARC settled closely due to possible genetic inter-relatedness. Shah *et al.* (2018) reported similar trend of germplasm accessions or varieties to end up closely in PCA cluster based on common origin or institute of development. The quantitative traits data were used for genotypes classification using hierarchical cluster analysis. Broadly the population studied was distributed into five major clusters on the basis of morphometric similarities. Variation in traits *i.e.*, primary branches plant⁻¹, plant height, days to maturity, seed yield, number of pods plant⁻¹ and 100 seed weight was instrumental to classify genotypes. Malek *et al.* (2014) found similar pattern of Euclidean distances based grouping among 27 soybean mutants for phenotypic and phenological variation. Ojo *et al.* (2012) findings supported our results as they observed diversity among 40 soybean accessions for morphological traits and obtained seven clusters. Previous studies on soybean population structure (Harer and Deshmukh 1992; Cui *et al.* 2001; Aditya *et al.* 2011) also reported high level of morphological variation and corroborated with our results. Moreover, we observed that the grouping pattern in present study correlated with the origin of genotype or variety *i.e.*, accession from USDA or varieties of Oilseeds Research Program, NARC Islamabad usually settled in a common clad. These findings were earlier also reported by Perry and McIntosh (1991). This affinity can be attributed to similar genetic background as usually varieties of various crops developed at same research institutes have earlier been found to share common pedigree (Westman and Kresovich 1999; Khurshid *et al.* 2019).

Soybean mosaic virus is one of the deadliest pathogen which can cause up to 94% production losses to soybean seed. As a major challenge to seed production plant breeders have traditionally focused on harnessing genetic resistance in soybean against the disease (Kang *et al.* 2005; Klepadlo *et al.* 2016). We evaluated 110 accessions for genetic resistance against prevalence of SMV in the field. The pathogenicity of virus varied against studied accessions from highly susceptible to highly resistant. Similar screening strategies against mosaic virus virulence have been applied by researchers to identify resistance in large soybean populations (Asad *et al.* 2006; Gadde 2006; Khan *et al.* 2013; Baruah *et al.* 2014). We reported a total of 42% genotypes as resistant followed by 20% Highly resistant and 21% moderately resistant. The proportion of highly susceptible was 5.5% while moderately susceptible and susceptible symptoms were observed in 6.4 and 4.5% of the genotypes, respectively. NARC varieties released in 1990s were found to be moderately resistant except for NARC-II and Ajmeri which were found highly resistant. NARC-I and

NARC-III were susceptible to disease which can be due to loss of function of avirulence in these genotypes or adaptation and evolution of the new pathotype (Khan *et al.* 2016). Korean (Sanning, Headu), Chinese (3S China) and Brazilian (BRS8480, BRS7980) accessions were highly susceptible to soybean mosaic virus resistance and also showed symptoms of soybean dwarf virus (SbDV). Almost 13% of NIBGE germplasm having USDA gene banks background were characterized as resistant against smv followed by 5% as highly resistant. Hence this material can be utilized for gene pyramiding against mosaic virus in soybean cultivars. Arif *et al.* (2000) screened soybean varieties for resistance against soybean mosaic virus and reported nine varieties and accessions to be resistant. We found that SMV resistant lines had high grain yield and oil content. These results were in conformity with Andayanie *et al.* (2017) as they screened soybean mosaic virus resistant breeding population and reported higher grain yield to be strongly associated with disease resistance. Moreover, to validate selection, disease resistance should be thoroughly investigated by detailed phenotyping and genomic analyses.

Conclusion

The present study reported significant genetic variation in economically important agronomic traits *i.e.*, pods plant⁻¹, seed yield kg ha⁻¹, 100 seed weight and plant height. Furthermore, phenotypic evaluation for soybean mosaic virus resistance identified exotic and local genotypes having resistance or tolerance against soybean mosaic virus. Mostly Chinese, Korean and Brazilian origin genotypes *i.e.*, Sanning, 3S China, Headu, BRS-8480 and BRS-7980 were highly susceptible to SMV. Moreover, accessions *i.e.*, 24515, GP-25, Faisal-Soy and NARC-2016 were found as high yielding and disease resistant. These genotypes can be utilized by researchers and breeders for soybean crop improvement. We recommend that further studies should be conducted to investigate the molecular basis of resistance against ever evolving mosaic viruses.

Acknowledgment

The authors acknowledge the contribution of Bio-resource Conservation Institute, NARC Islamabad, Asian Vegetables Research and Development Center (AVRDC), Taiwan, United States Department of Agriculture (USDA) and National Institute for Biotechnology and Genetic Engineering (NIBGE), Faisalabad for providing soybean germplasm.

References

Aditya J, P Bhartiya, A Bhartiya (2011). Genetic variability, heritability and character association for yield and component characters in soybean (*G. max* L. Merrill). *J Centr Euro Agric* 12:27–34

- Ahangaran A, MK Habibi, GHM Mohammadi, S Winter, F García-Arenal (2013). Analysis of soybean mosaic virus genetic diversity in Iran allows the characterization of a new mutation resulting in overcoming *Rsv4*-resistance. *J Gen Virol* 94:2557–2568
- Andayanie WR, V Santosa, M Rahayu (2017). Resistance to soybean mosaic virus with high yield on F7 soybean lines. *Intl J Agric Biol* 19:226–232
- Arif M, H Khurshid, SU Siddiqui, SA Jatoi, SA Jan, M Ilyas, SA Khan, A Khan, MI Ibrahim, N Saleem (2015). Estimating spatial population structure through quantification of oil content and phenotypic diversity in Pakistani castor bean (*Ricinus communis* L.) germplasm. *Sci Technol Dev* 34:147–154
- Arif M, A Ali, M Stephen, S Hassan (2000). Evaluation of resistance in soybean germplasm against soybean mosaic potyvirus. *Pak J Biol Sci* 3:1921–1925
- Arshad M, N Ali, A Ghafoor (2006). Character correlation and path coefficient in soybean *Glycine max* (L.) Merrill. *Pak J Bot* 38:121–130
- Asad S, S Fauzia, A Mansab, S Hussain, A Nazir (2006). Screening of soybean entries against various diseases under natural conditions during autumn 2001 at NARC. *Pak J Agric Res* 19:78–80
- Baig D, H Khurshid, M Arshad, SA Jan, MA Khan, N Nawaz (2018). Evaluation of soybean genotypes for yield and other economically important traits under rainfed condition. *Pure Appl Biol* 7:1–7
- Baruah S, M Sarma, D Baishya, A Sharma, R Borah, J Bhuyan (2014). Genetic variation for seed yield and yellow mosaic virus resistance in Soybean [*Glycine max* (L.) Merrill]. *Intl J Sci Res Publ* 4:1–10
- Bilyeu K, MB Ratnaparkhe, C Kole (2016). *Genetics, genomics, and breeding of soybean*. CRC Press, Boca Raton, Florida, USA
- Buss G, G Ma, P Chen, S Tolin (1997). Registration of V94-5152 soybean germplasm resistant to soybean mosaic potyvirus. *Crop Sci* 37:1987–1988
- Buzzell R, J Tu (1989). Inheritance of a soybean stem-tip necrosis reaction to soybean mosaic virus. *J Hered* 80:400–401
- Carpenter A, J Board (1997). Branch yield components controlling soybean yield stability across plant populations. *Crop Sci* 37:885–891
- Cui Z, TE Carter, JW Burton, R Wells (2001). Phenotypic diversity of modern Chinese and North American soybean cultivars. *Crop Sci* 41:1954–1967
- Gadde PM (2006). Genetic investigations in soybean (*Glycine max* (L.) Merrill). *Thesis*. University of Agricultural Sciences, Dharwad, India
- Harer PN, RB Deshmukh (1992). Genetic variability, correlation and path coefficient analysis in soybean (*Glycine max* (L.) Merrill). *J Oilseeds Res* 9:65–65
- Hart C (2017). The economic evolution of the soybean industry. In: *The Soybean Genome. Compendium of Plant Genomes*. Nguyen H, M Bhattacharyya (eds.). Springer, Cham, Switzerland
- Hill JH, SA Whitham (2014). Control of virus diseases in soybeans. In *Advances in virus research*, Vol., 90, pp:355–390. Academic Press, London
- Kang BC, I Yeam, MM Jahn (2005). Genetics of plant virus resistance. *Annu Rev Phytopathol* 43:581–621
- Khan M, M Karim, M Haque, A Karim, M Mian (2014). Variations in agronomic traits of Soybean genotypes. *SAARC J Agric* 12:90–100
- Khan M, S Tyagi, Z Dar (2013). Screening of soybean (*Glycine max* (L.) Merrill) genotypes for resistance to rust, yellow mosaic and pod shattering. In: *Soybean- Pest Resistance*, pp:173-184. El-Shemy HA (ed.). IntechOpen, London
- Khan Q, AS Mumtaz, H Khurshid, SA Jan, N Ahmad, SA Khan, N Saleem, SH Shah, MI Ibrahim, M Ilyas (2016). Exploring durable genetic resistance against leaf rust through phenotypic characterization and *Lr34*-linked STS marker in wheat germplasm. *Biosci J* 32:986–998
- Khurshid H, M Arshad, MA Khan, N Ali, ZK Shinwari, MA Rabbani (2019). Genetic structure of Pakistani Oilseed *Brassica* cultivars revealed by morphometric and microsatellite markers. *Pak J Bot* 51:1331–1340
- Khurshid H, D Baig, SA Jan, M Arshad, MA Khan (2017). Miracle crop: the present and future of soybean production in Pakistan. *MOJ Biol Med* 2:189–191

- Kiihl RA, E Hartwig (1979). Inheritance of Reaction to Soybean Mosaic Virus in Soybeans. *Crop Sci* 19:372–375
- Klepadlo M, P Chen, C Wu (2016). Genetic Analysis of Resistance to Soybean Mosaic Virus in PI 438307 Soybean Accession. *Crop Sci* 56:3016–3023
- Kumar S, V Kumari, V Kumar (2019). Genetic variability and character association studies for seed yield and component characters in soybean [*Glycine max* (L.) Merrill] under North-western Himalayas. *Legum Res* DOI: 10.18805/LR-4006
- Liu B, X Liu, C Wang, Y Li, J Jin, S Herbert (2010). Soybean yield and yield component distribution across the main axis in response to light enrichment and shading under different densities. *Plant Soil Environ* 56:384–392
- Malek M, MY Rafii, SS Afroz, UK Nath, M Mondal, (2014). Morphological characterization and assessment of genetic variability, character association, and divergence in soybean mutants. *Sci World J* 2014; Article 968796
- Malik MFA, M Ashraf, AS Qureshi, MR Khan (2011). Investigation and comparison of some morphological traits of the soybean populations using cluster analysis. *Pak J Bot* 43:1249–1255
- Ojo D, A Ajayi, O Oduwaye (2012). Genetic relationships among soybean accessions based on morphological and RAPDs techniques. *Pertan J Trop Agric Sci* 35:237–248
- Oliveira M, L Sousa, M Reis, ES Junior, D Cardoso, O Hamawaki, A Nogueira (2017). Evaluation of genetic diversity among soybean (*Glycine max*) genotypes using univariate and multivariate analysis. *Genet Mol Res* 16:1–10
- Orf JH, BW Diers, HR Boerma (2004). Genetic improvement: conventional and molecular-based strategies. *Soybeans Improv Prod Uses* 16:417–450
- Perry M, M McIntosh (1991). Geographical patterns of variation in the USDA soybean germplasm collection: I. Morphological traits. *Crop Sci* 31:1350–1355
- Ross JP (1977). Effect of aphid-transmitted soybean mosaic virus on yields of closely related resistant and susceptible soybean lines 1. *Crop Sci* 17:869–872
- Saleem N, SA Jan, MJ Atif, H Khurshid, SA Khan, M Abdullah, M Jahanzaib, H Ahmed, SF Ullah, A Iqbal, S Naqi (2017). Multivariate based variability within diverse Indian Mustard (*Brassica juncea* L.) genotypes. *Open J Genet* 7:69–83
- ShahJA, M Mujtaba, SA Jan, JS Akbar, RM Ashiq, A Ghafoor, M Jahanzaib, H Khurshid (2018). Estimation of spatial genetic structure in inter-regional populations of *Trigonella foenum-graceum* L. species through phenotypic variation and seed protein profiling. *Genetika* 50:171–185
- Singh RJ, RL Nelson, G Chung (2007). Soybean (*Glycine max* (L.) Merr.). Genetic resources, chromosome engineering, and crop improvement. *Oilseed Crops* 4:13–50
- Sleper DA, JM Poehlman (2006). *Breeding Field Crops*. Blackwell Publishing, Oxford, UK
- USDA (2018). *Oilseeds: World markets and trade*. Foreign Agricultural Service, Washington DC, USA
- Vollmann J (2016). Soybean versus other food grain legumes: a critical appraisal of the United Nations International Year of Pulses 2016. *J Land Manage Food Environ* 67:17–24
- Westman AL, S Kresovich (1999). Simple sequence repeat (SSR)-based marker variation in *Brassica nigra* gene bank accessions and weed populations. *Euphytica* 109:85–92
- Zhao Y, ZD Geng, LP Bao, TJ Wang (2007). Principal component analysis and cluster analysis of local soybean varieties of Yunnan. *J Hunan Agric Univ* 33:120–122