Journal of Pure and Applied Agriculture (2022) 7(2): 30-37
ISSN (Print) 2617-8672, ISSN (Online) 2617-8680
http://jpaa.aiou.edu.pk/

Diversity of soybean (*Glycine max* L.) genotypes based on agro-morphological parameters

Diriba G. Beyene[1] and Zerihun Jalata[2]*

[1]Wollega University, Plant Breeding Program, P. O. Field 395 Nekemte, Ethiopia
[2]Wollega University, College of Agriculture, Department of Plant Sciences, P. O. Field 38, Shambu, Ethiopia

*Corresponding author: Zerihun Jalata (jaluu_z@yahoo.com)

Abstract

Assessment of genetic diversity existing in breeding material is a key step in planning effective breeding programs. This experiment was carried out to discover the prevailing genetic variation among soybean genotypes for diverse characters for breeding. The study contained hundred soybean genotypes that were planted in a lattice (10 x10) design. The result confirmed that cluster IX (23) contained the widest variety of genotypes observed by means of cluster IV (22), cluster V (19), cluster III (12), and cluster I (10), while the minimum number of genotypes were clustered in cluster II and VII (1). Besides this, specific cluster combinations between cluster VI and IX, between III and VI had comparably higher inter-cluster distance (397.8) indicating the prescience of highly divergent genotypes suitable for direct variety development and/or hybridization to produce a wide array of desirable segregants. And the first four principal components having Eigen values extra than one accounted for 61.96% of the total version in soybean genotypes tested. The cluster method analysis discovered that cluster II and cluster VII contained suitable yield characters which can be useful to expand a variety through selection and/or source of genes for hybridization. Generally, our findings showed the existence of particularly divergent genotypes which may be promising either for direct variety development and/or hybridization for advanced soybean yield.

Keywords: Cluster mean, Hybridization, Intra and inter-cluster, Principal components, Soybean

To cite this article: Beyene, D. G., & Jalata, Z. (2022). Diversity of soybean (*Glycine max* L.) genotypes based on agro-morphological parameters. *Journal of Pure and Applied Agriculture*, *7*(2), 30-37.

Introduction

Soybean (Glycine max [L.] Merr.) seed constitutes mainly protein (40%), oil (20%), and soluble carbohydrates (15%), making it one of the most economically important crops in the world (Hyun et al., 2021). Soybean contributes about 25% to global edible oil production, about two-thirds of the world's protein concentrate for livestock feeding, and is a valuable ingredient in formulated feeds for poultry and fish, besides its importance as raw material for food, pharma, and other industries (Maranna et al., 2021). Soybean has wide adaptation and higher productivity potential compared to other grain legumes (Asfaw et al., 2006). (In Ethiopia, currently, soybean productivity is about 2.3 t ha^{-1} which is produced on an area coverage of 36636 hectares (CSA, 2019) which is lower than its potential production and world productivity.

The limited availability of genetic variability has been a major constraint in soybean breeding in Ethiopia. To overcome the problem, introduction of germplasms from various sources has been done in the last four years and considerable effort has been made to enhance the genetic variability through hybridization (Mesfin and Abush, 2018). Therefore, better breeding strategies are required to efficiently utilize the available germplasm to enhance the productive potentials of soybean yield. Genetic diversity was reported by Bekele et al. (2012) of clustering of 49 soybean genotypes into 16 clusters. While Sareo et al. (2018) grouped 75 soybean genotypes into 9 clusters. In another study on 50 soybean genotypes, seeds per plant and seed yield per plant traits contributed maximum to the genetic diversity and the clustering pattern revealed no correlation existed between the geographical diversity and genetic diversity (Shadakshari et al., 2011).

The assessment of genetic variety in the present germplasm is important to devise successful breeding using various populations comprising desirable developments for hybridization and direct choice packages (Bhandari et al., 2017), and thus examining genetic diversity is useful in choosing the right type of parents for hybridization program to create new genetic stocks for plant improvement (Mahesh et al., 2017). In addition to this, evaluation of the genetic range of germplasm collections is critical for green conservation and utilization of germplasm sources (Davila et al., 1998; Manjunatha et al., 2006). Effective hybridization programs between genetically diverse parents will lead to a considerable amount of heterotic response in F1 hybrids and a broad spectrum of variability in segregating generations but the appropriate selection of the parents is essential to be used in crossing to enhance the genetic recombination for potential yield increase (Islam, 2004).

Multivariate methods such as cluster analysis and principal components analysis (PCA) are important genetic diversity analysis and parental selection. Genotypes are clustered based on D^2 statistics (Mahalanobis, 1936) and Principal Component analysis (PCA) is another statistical tool that is beneficial for facts discount techniques relevant to quantitative facts. PCA transforms multi-correlated variables into every other set of uncorrelated variables

which may be further used for classifying genotypes (Bhandari et al., 2017). However, the potential genetic diversity available in the existing soybean genotypes was not studied. Thus, a better understanding of the extent of genetic variation is a basic foundation to exploit the genetic diversity through effective breeding programs in the future either to identify superior parents with desirable traits to achieve the best recombinants and/or varieties. Thus, this study aimed to explore the genetic diversity existing among soybean genotypes for future breeding programs and variety development.

Materials and Methods

Study site

The field experiment was conducted at Uke, East Wollega, Western Ethiopia in the 2018 *Meher* cropping season. Uke site is located at 365km from Addis Ababa with $8^{o}11^{'}52^{''}$ and 10^{o} $94^{'}44^{''}$ north latitude and 36^{o} 97'51'' and 37^{o} $11^{'}$ $52^{''}$ East longitude, and on the altitude of 1500-1700 masl above sea degree. The site has an average maximum and minimum temperature is 31^{O}c and 16^{O}c with 1400mm and 1200mm rainfall, respectively. The pH of the soil is acidic with the crimson colour of Nitosol a dominant soil kind in western Ethiopia.

Experimental materials and design

A total of 100 genotypes (97 genotypes and 3 varieties) of soybean were obtained from Jima Agricultural Research Centre, Ethiopia but the materials were introduced from abroad. The field experiment was planted in a 10 by 10 simple lattice design. Each genotype in the plot was grown in two rows of 4m lengths at 60cm spacing and 5cm between plants The crop was sown on 8th July 2018 and NPS fertilizer turned into implemented on the fee of 120 kg ha^{-1}at sowing time and all other recommended agronomic practices have been applied uniformly to all plots.

Data collection

Phonological data

Days to 50% emergence was recorded as the number of days from the date of sowing to the day on which 50% seedlings in a plot emerged. Days to 50% flowering was recorded as the number of days from the date of sowing to the day on which 50% of the plant flower in each plot. Days to 95% maturity was recorded as the number of days from the date of sowing to the day on which 95% of the plants in a plot mature. Grain filling period was computed as the number of days from 50% flowering to 95% maturity in each plot. The average height of the plant was measured from the bottom to the tip of five randomly selected plants in each plot at maturity time. Number of primary branches per plant was counted as the average number of main branches for five randomly selected plants in each plot at flowering time. Number of nodules per plant was counted as the average number of nodules per plant for five randomly selected plants at flowering time.

Yield and yield related data

Number of pods per plant was counted as average number of mature pods per plant for 10 randomly selected plants at maturity. Number of seeds per pod was counted as average number of seeds per pod, for five randomly selected plants in each plot at harvesting time. Hundred seed weight (g) was measured as the weight of 100 seeds for representative sample per plot at 12.5% moisture level. Biological yield (Kg ha^{-1}) was measured as the total weight of the above ground parts of the plants in each plot after harvested and sun dried converted to per hectare. Seed yield (Kg ha^{-1}) was measured as the weight of seeds for each plot at 12.5% moisture level converted to per hectare. The harvest index (%) was estimated by dividing total seed yield by biological yield.

Data analysis

All statistics measured on thirteen agro-morphological traits have been subjected to evaluation of variance using Proc GLM procedures of SAS version 9 (SAS, 2004**).** The agglomerative hierarchical clustering based on the Unweighted paired group method using the arithmetic mean (UPGMA) method was applied using the Proc CUSTER program of SAS (SAS, 2004). The data for all quantitative traits were standardized to mean zero and variance of one before clustering to avoid the difficulty of different scales that may have arisen due to differences in measurement scales. The genetic distance between clusters was calculated according to Mahalanobis (1936) formula as:

$$D^2_{(ij)} = (\bar{X}_i - \bar{X}j) S^{-1} (\bar{X}_i - \bar{X}j)$$

Where D^2ij = the square distance between any two genotypes i and j; Xi and Xj = the vectors for the values for genotypes i^{th} and j^{th} genotypes; and S^{-1}= the inverse of pooled variance-covariance matrix within groups. and the D^2 values obtained from pairs of clusters were considered as the calculated values of Chi-square ($\chi 2$) and were tested for significance at 1 and 5% probability levels against the tabulated values of $\chi 2$ at n-2 degrees of freedom, where, n is the number of characters considered (p = 13) (Urdan, 2005) while the first component extracted in a principal component analysis using SAS (2004).

Results and Discussion

The result indicated that one hundred genotypes were clustered into ten clusters based on D^2 analysis (Table 1). consequently, the result indicated that the largest quantity of genotypes were constituted in cluster IX (23) followed by cluster IV (22), cluster V (19), cluster III (12), and cluster I(10) whilst a minimum range of genotypes had been clustered in cluster II (1) and VII (1). In another finding ninety-three soybean accessions were classified into eight subgroups, indicating the genetic diversity

among the accessions, and similarities have been pronounced between 11 genotypes. The geographical starting place of the accessions turned into no longer continually associated with the clusters (Marconato et al., 2016), and cluster evaluation on sixteen soybean genotypes based on plant height, days to adulthood plant and hundred seeds weight and grain yield divided the genotypes into 4 clusters (Singh et al., 2020). Some other records confirmed seventy five soybean genotypes inclusive of seven assessments grouped into 5 clusters primarily based on D^2 values using Tocher's approach (Maranna et al., 2021). For ever-changing consumer preferences, desirable genes need to be reserved in cultivated and cultivable crops species in the form of germplasm resources. The existence of genetic diversity within and between crop plant species enables the breeders to select superior genotypes either to be directly developed a new variety or to be used as a parent in a hybridization program. Hence, the genetic diversity between two parents is essential to achieve heterosis and transgressive segregants (Bhandari et al., 2017).

Table 1 The distribution of a hundred soybean genotypes into clusters based totally on D^2 analysis

Cluster number	Number of genotypes	Genotypes
I	10	T29-15-T63-16-SA1(1),T52-15-T147-16-SA2 (28),T34-15-T70-16-SA1(84), T35-15-T79-16-SD1(13),T27-15-T57-16-SG2(43),,T29-15-T64-16-SB1(56),, T50-5-T139-16-SB2(100), T46-15-T119-16-SA1(36),T23-15-T45-16-SA1(2),T27-15-T51-16-SA2(71),
II	1	T53-15-T153-16-SC1(15)
III	12	T67-15-T203-16-SG2(68),T16-15-T-31-16-SK1(62),T35-15-T80-16-SE1(75),T73-15-T231-16-SF1(4),T54-15-T155-16-SA2(23),T55-15-T161-16-SC3(99),T73-15-T228-16-SC1(3),T52-15-T147-16-SA1(49),T54-15-T156-16-SB1(44),T45-15-T116-16-SA1(79),T44-15-T106-16-SD1(77),T47-15-T122-16-SA2(96)
IV	22	T49-15-T134-16-SC1(52),T27-15-T58-16-SH2(26),T16-15-T27-16-SG3(46),T52-15-T148-16-SB1(54),T28-15-T61-16-SC1(24),T71-15-T222-16 SA1(63),T44-15-T108-16-SF1(58), T47-15-T122-16-SA1(35),T44-15-T111-16-SI1(22),T15-15-T16-SD1(80), Afgat(82),T56-15-T164-16-SB1(19),T19-15-T39-16-SD1(85), T24-15-T46-16-SA1(89) ,T71-15-T224-16-SC1(31),T25-15-T47-16-SA1(51),T25-15-T47-16-SA2(78),T27-15-T52-16-SB1(70), T39-15-T96-16-SG1(93), T75-15-T240-16-SA1(32),T74-15-T240-16-SF2(18),T15-15-T20-16-SH1(73)
V	19	T19-15-T38-16-SC1 (27), T42-15-T97-16-SA1(21),T52-15-T149-16-SC1(16),T47-15-T122-16-SA1(35),T48-15-T128-16-SA1(76),T56-15-T163-16-SA1(73),T56-15-T165-16-SC1(90),T27-15-T51-16-SA3(59),T28-15-T62-16-SD1(83),T51-15-T142-16-SA1(95),Clarck 63 K (67) , T6-15-T2-16-SA2(69),T54-15-T155-16-SA1, (23) T74-15-T239-16-SE1(34), ,Nyala, (74), T45-15-T116-16-SC2(5),,T73-15-T230-16-SE1(29),,T72-15-T225-16-SA1(40),,T27-15-T57-16-SG1(38)
VI	4	T76-15-T241-16-SC1(92), T6-15-T223-16-SB(12), T74-15-T239-16-SE2(9), T27-15-T51-16-SA1(71)
VII	1	T15-15-T18-16-SF1(8)
VIII	3	T35-15-T77-16-SB1(8),T35-15-T78-16-SC1(8),T27-15-T58-16-SH1(8)
IX	23	T57-15-T167-16-SB1(98),T50-15-T139-16-SA1(100),T49-15-T132-16-SA1(45),T66-15-T193-16-SB1(14),T70-15-T219-16-SF2(50),T25-15-T48-16-SB1(53), T50-15-T141-16-SC1(41),T33-15-T69-16-SB1(55), T74-15-T236-16-SB1(66), T62-15-T181-16-SA1(65),T27-15-T53-16-SC1(64),T33-15-T68-16-SA1(81),T52-15-T149-16-SC1(16),T44-15-T107-16-SE1(47),T24-15-T46-16-SA2(39),T16-5-T23-16-SC(11)1,T39-15-T92-16-SC2(57),T16-15-T28-16-SH1(25),T37- 15-T81-16-SA1(60),T44-15-T111-16-SI2(22),T74-15-T240-16-SF1(18), T26-15-T50-16-SB1(87),T55-15-T159-16-SA1(72)
X	5	T47-15-T124-16-SC1(94),T51-15-T146-16-SE1(97),T63-15-T183-16-SB1(10), T55-15-T160-16-SB2(17),T53-15-T152-16-SB1(61)

Numbers in the parentheses following genotype name are their designated genotypic number.

Intra and inter-cluster distances

The result of intra-cluster distance ranged from 0.00 for cluster II and VII to 10.8 in cluster IX indicating intra-clusters varied from no intra-population distance to high variation within the population (Table 2). Whereas the inter-cluster distance, the maximum inter-cluster distance (397.8) was recorded between cluster VI and IX as well as between cluster III and VI (397.8) followed by cluster I and VI (360.4). Besides this, most of the clusters

combining with cluster VI and X had particularly higher inter-cluster distance V (Table 2). The low genetic range implies low genetic variability with low breeding value as genetically comparable genotypes share common alleles ensuing in little complementarity and show low heterosis because of low levels of heterozygosity in crosses (da Silva Rodrigue et al., 2016). Inter-cluster distance is a critical criterion for the choice of genotypes for breeding purposes. Consequently, genotype clusters that are with large inter-cluster distances are genetically more divergent and so that hybridization among such genotypes clusters is more likely to produce a wide array of desirable segregants (Sharma, 1998). It appears that an accelerated parental genetic distance implies the presence of a bigger wide variety of contrasting alleles on the preferred loci that would recombine at loci within the F2 and F3 generations at some stage in the crossing of distantly related mother and father creating more opportunity for the effective selection of yield and its additives (Ghaderi, 1984). Thus, the practical implication of grouping genotypes into different clusters and estimating the genetic distance among them is helpful to estimate genetic diversity (Bhatt, 1970).

Table 2 The intra (main diagonal) and inter-cluster distance (D^2) values along with their D value in soybean genotypes tested at Uke, Western Ethiopia

Cluster	I	II	III	IV	V	VI	VII	VIII	IX	X
I	**8.8**	99.5**	37.1**	72**	73.4**	360.4*	99.5**	119.1**	37.1**	276.6**
II		**0.0**	136.6**	27.5*	26.1*	261.2**	102.5**	19.6^{ns}	136.6**	177.1**
III			**9.2**	109**	110.5**	397.8**	186.6**	156.2**	121.8**	313.7**
IV				**6.0**	1.4^{ns}	288.7**	27.5*	47.1**	109.1**	313.7**
V					**2.7**	287.3**	26.1*	45.7**	110.5**	203.2**
VI						**3.6**	261.2**	241.6**	397.8**	84.1**
VII							**0.0**	19.6^{ns}	136.6**	177.1**
VIII								**1.6**	156.2**	157.5**
IX									**10.8**	313.7**
X										**8.3**

Chi-square ($\chi 2$) = 22.36, 27.69, *, ** indicate 5% and 1% probability levels, respectively.

Besides this, a comparison of cluster mean values of genotypes for various characters showed high variability (Table 3). Cluster II and VII had specifically the highest seed yield, biomass yield per ha, harvest index, number of pods per plant, and number of branches per plant. Cluster I and X had a fairly better variety of nodules per plant (Table 3). And, cluster III and V also showed a good harvest index. Moreover, clusters II and VIII had maximum height which may be useful in increasing the potential for more branches, pods per plant, and high biomass that may contribute to more seed yield. Thus, the cluster mean analysis evaluation discovered that cluster II and cluster VII contained vital yield characters which can be useful to gain improved seed yield via direct choice and supply of suitable genes for hybridization. This study partly conforms to Maranna et al. (2021) report of cluster mean of soybean genotypes on which days to flowering, days to maturity, number of pods/plant, 100-seed weight, and grain yield/plot caused highest in different clusters. Moreover, the report showed the symbiotic overall performance of 152 soybean genotypes was variable and the reaction of a few cultivars accumulating up to twice as much nodule dry weight than others (Hungria and Bohrer, 2000) which indicates that there was also genetic variability among soybean for their symbiotic association which is indispensable for nitrogen fixation.

Table 3 Cluster mean values of ten clusters for eleven characters soybean genotypes

Traits	I	II	III	IV	V	VI	VII	VIII	IX	X
Days to 50% emergence (DE)	8.3	9.5	8.5	8.9	9.3	9.0	8.0	9.3	8.8	8.8
Days to 50% flowering (DF)	52.2	55.0	53.6	53.9	55.0	54.0	55.0	55.0	53.0	55.0
Days to 95% maturity (DM)	106.9	112.0	104.0	104.0	107.0	106.2	107.0	106.3	105.5	107.0
Grain filling period (GFP)	53.7	54.5	54.7	50.8	51.8	51.6	52.0	55.1	52.0	52.0
plant height, (cm)(PH)	60.6	70.0	66.0	60.0	61.8	61.8	66.0	70.0	62.5	64.0
Primary Branches/plant (NB)	4.4	6.8	4.8	4.8	5.2	5.1	6.1	4.0	4.8	3.6
Number of pod/plant (NP)	45.9	78.0	54.9	51.0	51.4	52.3	67.0	50.0	47.0	38.9
Number of seeds/pod (NS)	2.6	2.6	2.5	2.5	2.5	2.6	2.5	2.5	2.5	2.4
Number of nodules/plant (NN)	31.4	21.2	27.5	22.6	23.5	26.5	22.4	25.1	28.7	35.3
100 seed weight (g) (HSW)	14.8	16.6	14.49	14.9	15.8	15.5	15.8	14.3	15.1	15.7
Biological yield (kg ha^{-1})	3395	5743	3240	3682	3985	4619	5400	4330	3038	2444
Seed yield (kg ha^{-1})	1242	2897	1611	1581	1907	2274	2574	1559	1344	1083
Harvest index (HI)	0.37	0.49	0.46	0.43	0.47	0.48	0.48	0.41	0.42	0.41

Table 3 Principal component scores values of the first PCs of 100 soybean genotypes evaluated at Uke, East Wollega Zone

Traits	P1	P2	P3	P4
Eigenvalues	3.0762	2.4231	1.3794	1.1765
Differences	0.6531	1.0436	0.2029	0.1956
Proportion	0.2366	0.1864	0.1061	0.0905
Cumulative variance	0.2366	0.423	0.5291	0.6196
Days to 50 % emergence	0.0824	-0.0435	0.4915	0.2411
Days to 50 % flowering	0.2703	0.2063	0.3121	-0.5582
Days to 95 % maturity	0.3204	0.4991	-0.0321	0.0919
Grain filling period	0.1986	0.4539	-0.2185	0.4397
Plant height (cm)	0.2867	0.3701	0.0042	-0.1699
Number of primary branches per plant	0.3505	-0.0515	-0.082	-0.2637
Number of nodules per plant	0.1286	0.1758	0.1803	0.2762
Number of pods per plant	0.3451	-0.1357	0.0939	-0.1045
Number of seeds per pod	0.0467	-0.0568	0.5145	0.4125
hundred seed weight (g)	0.1378	-0.055	-0.5246	0.17
Biological yield (kg/ha)	0.3465	-0.3203	-0.1267	0.1667
Seed yield (kg/ha)	0.4167	-0.3547	-0.0679	0.1266
Harvest index (%)	0.3500	-0.2772	0.0606	0.0078

Principal component analysis

Large datasets are often difficult to interpret. Principal component analysis (PCA) is a method of reducing the dimensionality of large datasets, improving interpretability while minimizing the loss of information. It performs this task by creating new uncorrelated variables that successively maximize variance and this new variable, the principal components, reduces to solving an eigenvalue of eigenvector problem, and the new variables are defined by the dataset at hand (Jolliffe & Cadima, 2016). Accordingly, a look indicated there are 4 principal components with eigenvalues greater than one contributing 61.96% of the entire variation (Table 3). Thus, the study showed that PC1 contributed 23.66%, while PC2, PC3, and PC4 contributed 18.64%, 10.61%, and 9.05%, respectively to the total variation.

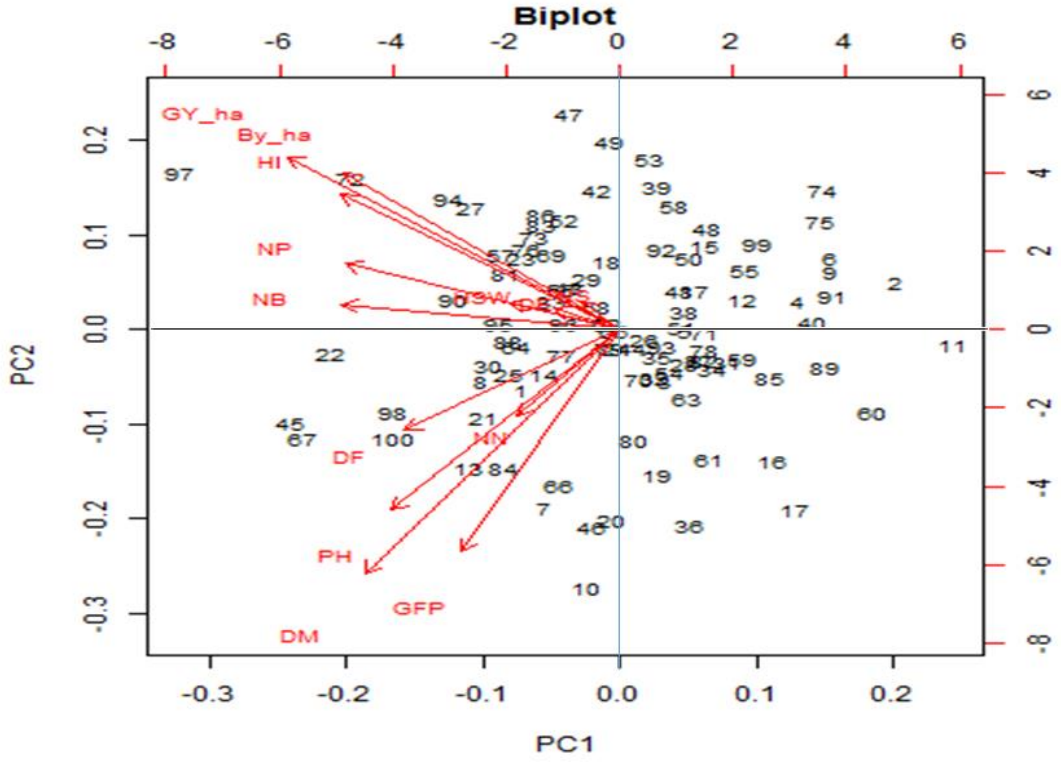
The various soybean characteristics which contributed more to PC1 include many primary branches/plant, variety of pods/plant, harvest index, biological yield, and seed yield implying that these characteristics are the major contributors for the total variation among genotypes resulting in a higher percentage of variation contributed by the first PCs to the total variation. While in PC2, the characters which contributed more protected: days to 50% flowering, days to 95% adulthood, grain filling period, plant top (cm), harvest index, biological yield, and seed yield The third PC that accounted for 10.61% of the total variation was influenced by characters such as days to flowering, grain filling period, the number of seeds per pod, and hundred seeds per pod. The fourth PC was influenced by characters' grain filling period and the number of seeds per pod.

This study partly agrees with Singh et al. (2020) who stated PC1 and PC2 had eigenvalues higher than unity explaining 76.6% of total variability among soybean genotypes attributable to plant height, days to maturity, number of pods/plants, 100 seed weight, and grain yield. While in this study four four PCA had eigenvalues higher than unity. In another study, genetic diversity in black gram germplasm accessions and the first four PCA with eigenvalues greater than one contributed 79.5% of total variability among accessions (Ghafoor et al., 2001). Iqbal et al. (2010), said the primary 3 computers with eigen values extra than one contributed 77% of the total variability in soybean genotypes and PC1 become implicated by seed yield/plant, biological yield/plant, and harvest index (%). The present result is also in agreement with the report of Hashash (2016) and Mohammadi and Prasanna (2003). In keeping with Chahal and Gosal (2002)

within the first primary component, characters with the largest absolute value toward cohesion have more influence on clustering than people with a lower absolute fee in the direction of zero. Consequently, within the gift look at, seed yield, number of branches in keeping with plant, wide variety of pods in line with plant, organic yield, and harvest index had a distinctly excessive contribution to the total version inside the clustering of soybean genotypes.

The primary main component had high nice factor loading from seed yield (0.4167), quantity of primary branches in keeping with plant (0.3505), harvest index (0.3500), biological yield (0.3465), and number of pods in keeping with plant (0.3451). Similarly, the major contributing characters to the diversity in the PC2 have high positive component loading from days to 95% maturity (0.4991), grain filling period (0.4539), and plant height (03701). Primary contributing characters for the range inside the 1/3 main aspect (PC3) had high nice factor loading from numerous seeds per pod (0.5145), days to 50% emergence (0.4915), days to 50% flowering(0.3121), and excessive poor loading from hundred seed weight (-zero.5246) and grain filling period (-0.2185). In principal component four (PC4) high positive component loading from grain filling period (0.4397), many seeds per pods (0.4125), and many nodules per plant (0.2762) but high negative loading from days to 50 % flowering (-0.5582). The positive and negative loading shows the trends of associations between the components and the variables. Hence, the characters which load high positively or negatively contributed more to the diversity and these characters were the ones that most differentiated the clusters. In another study, clustering and principal coordinate analysis done based on molecular analysis showed similarity in explaining the extent of genetic diversity within the soybean accessions tested. And, PCA obtained from phenotypic traits resulted two clusters whereas booth clustering-based SSR data and PCA on phenotypic data showed similar results showing the assembled germplasm was diverse genetically especially with high variation in flowering, maturity period, and main yield components (Denwar et al., 2019). Besides this, days to flowering (52.25%) exhibited greater variation and contribution to diversity among genotypes followed by days to maturity (10.38%) and plot yield (9.23%). While yield/plant (1.15%), harvest index (0.22%), and branches/plant (1.73%) contributed comparatively less to the total diversity (Maranna et al., 2021).

In addition, the biplot graph was constructed from PC1 and PC2 to display the association of the different agronomic traits and genotypes (Fig. 1). Genotypes that have PC1 scores greater than 0 are in a positive direction and high yielding potential while PC1 scores less than 0 are low yielders. Thus genotypes 11, 2, and 60 are among the top-performing genotypes. On other hand, grain yield, biomass yield, and harvest index THere was also close relation among grain filling period, days to maturity, and plant height (Fig. 1). And, genotypes that are found closer to the nearby agronomic characters are said to be the best performer in that agronomic traits.had closely associated.



NB: The red colored designated parameters are characters (Table 3) while numbers are genotypes (Table 1).
Fig. 1 Principal component analysis biplot of soybean genotypes and their characters

Conclusion

The result showed that cluster IX (23) contained the largest number of genotypes followed by cluster IV (22), cluster V (19), cluster III (12), and cluster I (10). Besides this, comparably the greatest inter-cluster distance (397.8) was observed between cluster combinations of VI and IX,, between III and VI implying highest divergent genotypes suitable for direct variety development and/or hybridization to produce a wide array of desirable segregants for high yielding potential cultivars The cluster method analysis discovered that cluster II and cluster VII contained suitable yield characters which can be useful to expand a variety through selection and/or source of genes for hybridization. The principal components analysis showed that the first four PCs contributed 61.96% of the total variation of soybean genotypes. The traits such as number of primary branches per plant, number of pods per plant, harvest index, biological yield, and seed yield contributed more to PC1... Consequently, it may be concluded that there exists sufficient genetic variability among soybean genotypes tested for desirable traits in future breeding programs.

Acknowledgments: The authors would like to thank Wollega University for their financial support in completing this study.

Conflicts of interest: The authors declare that there is no competing interest.

References

Asfaw, A., Abush, T., & Sentayehu, A. (2006). Soybean genetic improvement in Ethiopia. In: K. Ali, et al. (Eds. *Food and forage legumes of Ethiopia: progress and prospects. Proc. of the workshop on food and forage legume*, Ghion Hotel, Ethiopia. pp. 22-26.

Bekele, A., Getint, A., & Habtamu, Z. (2012). Genetic divergence among soybean (*Glycine max* (L) Merrill) introductions in Ethiopia based on agronomic traits *Journal of Biology, Agriculture and Healthcare, 2*(6), 1-12.

Bhandari, H. R., Bhanu, A. N., Srivastava, K., & Hemantaranjan, A. (2017). Assessment of genetic diversity in crop plants: An overview. *Advances in Plants and Agricultural Research*, *7*(3), 279–286.

Bhatt, G. M. (1970). Multivariate analysis approach to the selection of parents for hybridization aiming at yield improvement in self-pollinated crops. *Australian Journal of Agricultural Research, 20*, 1-7.

Chahal, G. S., & Gosal, S. S. (2002). *Principles and, procedures of plant breeding: Biotechnology and conventional approaches*. Narosa Publishing House, New Delhi.

CSA (Central Statistical Agency). (2019). Agricultural sample survey 2018/2019: Report on area and production of major crops (private peasant holdings, *Meher* Season). Volume I, Statistical Bulletin 584, 2019, Addis Ababa, Ethiopia. https://www.statsethiopia.gov.et/work-with-us/

da Silva Rodrigues, J. I., Antunes Arruda, K. M., Deniz Piovesan, N., de Barros, E. G., & Moreira, M. (2016). Selection of progenitors for increase in oil content in soybean. *Revista. Ceres, 63*(5), 661-667.

Davila, J. A., Hoz, M. P. S., Loarce, Y., & Ferrer, E. (1998). The use of random amplified microsatellite polymorphic DNA and coefficients of parentage to determine genetic relationships in barley. *Genome*, *41*, 477-486.

Denwar, N. N., Awuku, F. J., Diers, B., Addae-Frimpomaah, F., Chigeza, G., Oteng-Frimpong, R., Puozaa, S. K., & Barnor, M. T. (2019). Genetic diversity, population structure, and key phenotypic traits driving variation within soyabean (*Glycine max*) collection in Ghana. *Plant Breeding*, *138*(5), 577-587.

Ghaderi, A., Adams, M. W., & Nassib, A. M. (1984). Relationship between genetic distance and heterosis for yield and morphological traits in dry edible bean and faba bean. *Crop Science*, *24*, 37-42.

Ghafoor, A. A., Ahmad, M. A., Rabbani, M. A., Sharif, Z., & Zahid, M. A. (2001). Genetic diversity in black gram (*Vigna mungo* (L.) Hepper). *Field Crops Research*, *69*, 183-190.

Hashash, E. F. (2016). Genetic diversity of soybean yield based on cluster and principal component analyses. *Journal of Advanced Biology*, *10*(3), 1-4.

Hungria, M., & Bohrer, T. R. J. (2000). Variability of nodulation and dinitrogen fixation capacity among soybean cultivars. *Biology and Fertility of Soils, 31*, 45–52.

Hyun, J., Lee, J. Y., Cho, H., Choi, H. J., Son, C. K., Bae, J. S., Bilyeu, K., Song, J. T., & Lee, J. (2021). Genetic diversity of soybeans (*Glycine max* (L.) Merr.) with black seed coats and green cotyledons in Korean Germplasm. *Agronomy*, *11(3)*, 581; https://doi.org/10.3390/agronomy11030581

Iqbal, Z., Arshad, M., Ashraf, M., Naeem, R., Malik, M. F., & Waheed, A. (2010). Genetic divergence and correlation studies of soybean [*Glycine max* (L.) Merrill.] genotypes. *Pakistan Journal of Botany*, *42*(2), 971-976.

Islam, M. R. (2004). Genetic diversity in irrigated rice. *Pakistan Journal of Biological Sciences*, *2*, 226-229.

Jolliffe, I. T., & Cadima, J. (2016). Principal component analysis: A review and recent developments. *Philosophical Transactions of the Royal Society A, 374*, 20150202. http://dx.doi.org/10.1098/rsta.2015.0202

Mahalanobis, C. (1936). On the generalized distance in statistics. *National Institute of Sciences of India*, *2*(1), 49-55.

Mahesh, M., Sudhanshu, J., & Bhadoria, V. S. (2017). Genetic divergence studies in soybean (*Glycine max* L. Merrill). *International Journal of Agriculture Sciences*, *9*(19), 4196-4197.

Manjunatha, T., Bisht, I. S., Bhat, K. V., & Singh, B. P. (2006). Genetic diversity in barley (*Hordeum vulgare* L.) landraces from Uttaranchal. *Genetic Resources and Crop Evolution*, *54*(1), 55-65.

Maranna, S., Nataraj, V., Kumawat, G., Chandra, S., Rajesh, V., Ramteke, R., Manohar Pate, R., Ratnaparkhe, M. B., Husain, S. M., Gupta, S., & Khandekar, N. (2021). Breeding for higher yield,

early maturity, wider adaptability and waterlogging tolerance in soybean (*Glycine max* L.): A case study. *Scientific Reports*, *11*, 22853; https://doi.org/10.1038/s41598-021-02064-x

Marconato, M. B., Pereira, E. M., Silva, F. M., Bizari, E. H., Pinheiro, J. B., Mauro, A. O., & Unêda-Trevisoli, S. H. (2016). Genetic divergence in soybean (*Glycine max*) diversity panel based on agro-morphological traits. G*enetics and Molecular Research*, *15*(4), doi: 10.4238/gmr15048980

Mesfin, H., & Abush, T. (2018). Progress of soybean [*Glycine max* (L.) Merrill] breeding and genetics research in Ethiopia: A review. *Journal of Natural Sciences Research*, *8*(13), 67-77.

Mohammadi, S. A., & Prasanna, B. M. (2003). Analysis of genetic diversity in crop plants-salient statistical tools and considerations. *Crop Science*, *43*, 1235-1248.

Sareo, H., Devi, H. N., Devi, T. H. R, Devi, T. S., Karam, N., & Devi, L. S. (2018). Genetic diversity analysis among soybean (*Glycine max*) genotypes based on agro morphological characters. *Indian Journal of Agricultural Sciences*, *88*(12), 1839-1842.

SAS Institute Inc. (2004). SAS/STAT user's guide. Version 9.2. Fourth Edition. Statistical Analysis Institute Inc., Cary North Carolina.

Shadakshari, T. V., Kalaimagal, T., Senthil, N., Boranayaka, M. B., Gowda, R. K., & Rajesha, G. (2011). Genetic diversity studies in soybean [*Glycine max* (L.) Merrill] based on morphological characters. *Asian Journal of Bio Science*, *6*(1), 7-11.

Sharma, J. R. (1998). Statistical and biometrical techniques in plant breeding. New Age International, New Delhi, pp.432.

Singh, P. K., Shrestha, J., & Kushwaha, U. K. S. (2020). Multivariate analysis of soybean genotypes. *Journal of Agriculture and Natural Resources*, *3*(1), 69-76.

Urdan, T. C., (2005). *Statistics in Plain English*, 2nd Ed., Lawrence Erlbaum Associates, Mahwah, New Jersey, pp. 211.