Ethiop. J. Agric. Sci. 34(4) 72-99 (2024)

Assessment of Genetic Variability in Soybean (*Glycine max* (L.) Merrill) Genotypes at Gondar, Ethiopia

Gobezie Chakelie[1], Mulugeta Atnaf[2] and Alemu Abate[3]

[1]Amhara Agricultural Research Institute, Gondar Agricultural Research Center, P.O.Box 1337, Gondar, Ethiopia; [2]Ethiopian Institute of Agricultural Research, Fogera National Rice Research and Training Center, P.O.Box 1937, Bahir Dar, Ethiopia; [3]Department of Plant Sciences, College of Agriculture and Environmental Sciences, Bahir Dar University, P.O.Box 79, Bahir Dar, Ethiopia

Corresponding author email: gobeziec88@gmail.com

Abstract

Identification and utilization of genetically diverse germplasm are the primary objectives of crop improvement. This study evaluated 81 soybean (Glycine max (L.) Merrill) genotypes in Metema and West Armachiho districts of Northwestern Ethiopia during 2019/2020 using a simple lattice design to assess phenotypic variability and propose effective selection strategies. Combined analysis of variance revealed significant differences ($P \leq 0.05$) among genotypes, locations, and genotype × location interactions for most traits. Estimations of genetic variability, heritability, and expected genetic advance indicate significant genetic variability among the tested genotypes. Key traits exhibited high broad-sense heritability (h^2b) and genetic advance as a percentage of the mean (GAM), including days to 50% flowering (94.84, 30.21), plant height (92.13, 38.63), branches per plant (80.89, 46.15), pods per plant (71.90, 46.22) and hundred seed weight (85.12, 21.93), suggesting significant potential for genetic improvement in these traits. Cluster analysis grouped the genotypes into six clusters, with Cluster I being the largest (63%), followed by Cluster III, which contained 11.11% of the genotypes. The maximum inter-cluster distance was between Clusters II and VI ($D^2 = 154.64$), indicating high genetic divergence suitable for hybridization. Principal component analysis attributed 77.98% of the total variation to the first four components, emphasizing traits critical for selection. In conclusion, the study demonstrated significant variability among the genotypes, which could be exploited in future soybean improvement programs.

Keywords: Soybean, Cluster, Genetic Advance, Heritability, Traits

Introduction

Soybean (*Glycine max* (L.) Merrill) is one of the most valuable crops globally, known for its high-quality protein content (40%) and vegetable oil (20%)(Naik *et al*. 2016; Khojely *et al*. 2018). It contributes 70% of the world's protein and 28% of its oil consumption, with an increasing role in biofuel production (Hartman *et al*. 2011; Jo *et al*. 2021). Beyond nutrition and economics, soybean enhances food security and agricultural sustainability through its nitrogen-fixing ability, which improves soil fertility and supports crop rotation systems (Mapope and Dakora, 2016; Ciampitti *et al*. 2021).

Over the past decades, global soybean production has steadily increased. It is cultivated worldwide on approximately 122 to 134 million hectares, with an annual production ranging from 335 to 350 million metric tons (FAOSTAT, 2022). Brazil, USA, Argentina, China, and India collectively account for over 80% of production. In Africa, South Africa, Nigeria, and Zambia are the top three producers, while Ethiopia ranks seventh on the continent. Soybean was introduced to Ethiopia in the 1960s (IAR, 1982). Although Ethiopia is not a major producer, it has experienced growth in soybean cultivation, producing 185,522.2 tons with an average yield of 2.4 tons per hectare in 2021/2022 (CSA, 2022). It is becoming an increasingly important commercial legume crop, serving as a significant source of protein, a raw material for edible oil production, animal feed, a means of earning foreign currency, and enhancing soil fertility through crop rotation. Given its nutritional benefits, soybean is expected to play a key role in Ethiopia's food security program. Additionally, soybean's wide agro-ecological adaptability and suitability for rain-fed agriculture align well with the country's agricultural practices.

Despite its growing importance, Ethiopia's soybean productivity (2.4 tons per hectare) remains below the global average (2.6 tons per hectare) (Sileshi, 2019a; FAOSTAT, 2022), while its potential yield can reach up to 12.7 tons per hectare (Winsor, 2021). Currently, over 38 soybean varieties have been released in Ethiopia (EAA, 2023), with breeding efforts primarily focused on developing new cultivars from available germplasm. However, a narrow genetic base has hindered progress, indicating the need to explore and characterize genetic resources. Additional constraints include low varietal stability, insect pests, diseases, limited access to improved seeds, and suboptimal agronomic practices (Tesfaye and Hailemariam, 2018). In the context of crop improvement, genetic diversity is fundamental to breeding programs, enabling the selection of superior traits and broadening the genetic base (Jing *et al*. 2010; Govindaraj *et al*. 2015). Understanding the genetic diversity of soybean genotypes can help interpret germplasm architecture, select parents with high diversity, and predict superior offspring combinations (Rahman *et al.* 2011; Bhatia *et al*. 2017).

Local breeding efforts have given limited attention to exploring the genetic diversity within the available germplasm for various reasons. However, the narrow genetic base of the crop has hindered its genetic improvement (Cornelious and Sneller, 2002). In Ethiopia, the low genetic diversity has slowed the development of new soybean varieties, highlighting the need to explore and characterize existing genetic resources. Previous studies by Mesfin (2018), Sileshi *et al.* (2019b), and Yirga *et al.* (2022) reported considerable genetic variability among the tested genotypes. However, as genetic materials are updated over time, previous genetic information may not fully represent new resources. Understanding genetic parameters such as genotypic and phenotypic variation,

heritability, and genetic advance is essential for crop improvement and selection (Aditya *et al.* 2011). Furthermore, characterizing the genetic background of soybean genotypes and estimating breeding values is crucial before initiating improvement programs (Arshad *et al.* 2009). Therefore, assessing genetic diversity using morphological trait markers is key to exploiting the existing variation in soybean genotypes for developing improved cultivars. This study aims to estimate the extent of genetic variability, heritability, and potential genetic advances in key traits and propose an effective selection strategy to harness promising genotypes for future soybean improvement programs.

Materials and Methods

Experimental sites

An experiment was conducted at Metema and West Armachiho, Gondar Agricultural Research Center sub-stations, Gondar, Ethiopia, during the main cropping season of 2019/2020 (Fig 1). Metema and West Armachiho, represent the lowland areas of the soybean growing areas of Gondar. Metema is located at 12° 47' 38"N, 36° 23' 41'¢ E, and at an altitude of 760 masl, while West Armachiho is found at 13° 28' 42" N, 36° 28' 23" E, with an altitude of 657 masl. The average annual rainfall for Metema is 1030 mm, while its average minimum and maximum temperature is 16 and 35 °C, respectively. The average annual rainfall and average minimum and maximum temperature of West Armachiho is 900 mm and 22.1–36.3 °C, respectively (Ayalew *et al*. 2012). The soil's textural class is predominantly clay loam. The rainfall in the two areas is monomodal, and the peak season is from June to September (the Ethiopian main *Meher season*) (Fig 2).

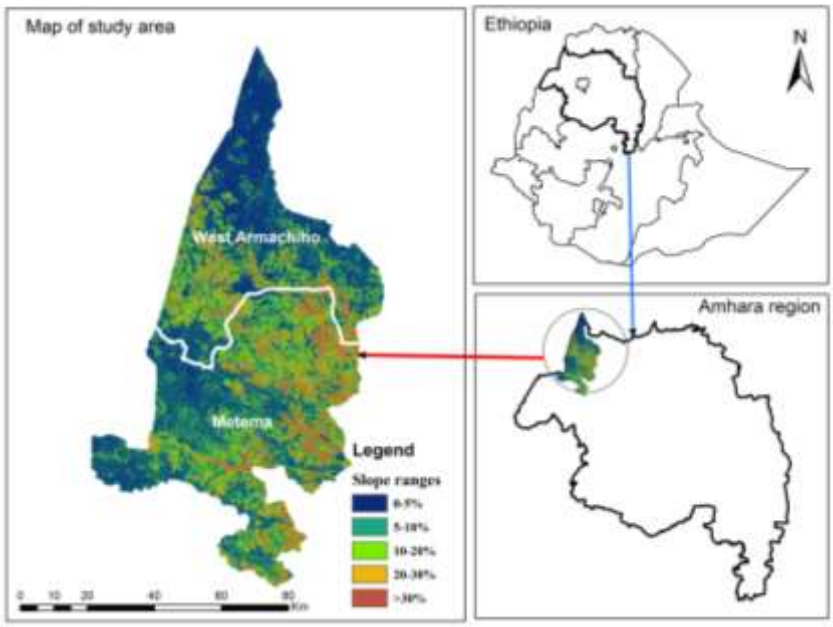

Fig 1. Geographical map of Metema and West Armachiho, Ethiopia

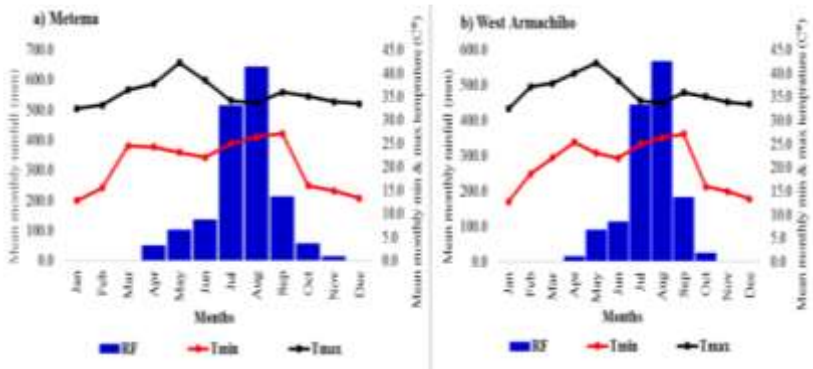

Fig 2. Meteorological data of Metema and West Armachiho for growing season (NMIE, 2020)

Experimental materials

The study was conducted using 81 soybean genotypes, which included nine released varieties (Andinet, AFGAT, Belessa-95, Gazale, Gishama, Gizo, Hawassa-04, Pawe-1, and Pawe-3). The test genotypes and varieties were obtained from the International Institute of Tropical Agriculture (IITA) (55) through Pawe Agricultural Research Center, Jimma (17), and Gondar (9) Agricultural Research Centers, Ethiopia (Table 1).

Table 1. List of 81 soybean genotypes used in the current study

Code	Genotype designation	Source/origin	Code	Genotype designation	Source/origin
G-1	TGX2009-14F	IITA	G-41	TGX-1889-62f	IITA
G-2	Gishama	PARC	G-42	Hawassa-04	HARC
G-3	TGX2025-9E	IITA	G-43	TGX2019-1E	IITA
G-4	TGX2016-2E	IITA	G-44	TGX2011-6F	IITA
G-5	TGX2025-19E	IITA	G-45	TGX2025-16E	IITA
G-6	Tgx-1990-40f	IITA	G-46	F6LG06-5920XU03-100612-01	USA
G-7	T34-15-T72-16-Sc1	IITA	G-47	TGX2016-3E	IITA
G-8	TGX2007-3F	IITA	G-48	TGX1987-14F	IITA
G-9	F6U03-300134XLG04-5187	USA	G-49	H3-15-SF-2	USA
G-10	TGX2010-14F	IITA	G-50	TGX2027-7E	IITA
G-11	Andinet	PARC	G-51	TGX-1987-28f	IITA
G-12	TGX2018-5E	IITA	G-52	TGX2017-6E	IITA
G-13	JM-ALM/H3-15-SC-1	JARC	G-53	TGX2016-4E	IITA
G-14	F6LG04-6000XLG04-5187-05	USA	G-54	T34-15-T73-16-SD1	IITA
G-15	TGX1834-10E	IITA	G-55	TGX2004-7F	IITA
G-16	TGX2013-2F	IITA	G-56	TGX2025-10E	IITA
G-17	TGX2009-1F	IITA	G-57	TGX-1835-10E	IITA
G-18	G7955-C3RPP (C1)	USA	G-58	F6LG04-6000XLG04-5187-06	USA
G-19	F6LG04-6000XLG04-5187-04	USA	G-59	TGX-1919-22F	IIAT
G-20	TGX2027-4E	IITA	G-60	TGX2015-1E	IITA
G-21	TGX2010-11F	IITA	G-61	F6LG04-6000XLG04-5187-01	USA
G-22	AFGAT	PARC	G-62	TGX-1987-11F	IITA
G-23	F6LG06-5920XU03-100612-03	USA	6-63	Belessa-95	PARC
G-24	TGX 2025-6E	IITA	G-64	TGX2025-14E	IITA
G-25	TGX1988-5F	IITA	G-65	TGX2023-1E	IITA
G-26	TGX2007-1F	IITA	G-66	TGX2027-1E	IITA
G-27	TGX1993-4FN	IITA	G-67	T47-15-T126-16-SF1	IITA
G-28	TGX 2009-16F	IITA	G-68	TGX2008-4F	IITA
G-29	TGX2020-1E	IITA	G-69	TGX-1989-40F	IITA
G-30	TGX2023-4E	IITA	G-70	Gizo	PARC
G-31	F6LG04-6000XLG04-5187-02	USA	G-71	TGX-1987-18f	IITA
G-32	TGX2017-5E	IITA	G-72	TGX1989-19F	IITA
G-33	Pawe-03	PARC	G-73	TGX-1988-5E	IITA
G-34	CRFRD-15-SB	USA	G-74	CRFRD-15-SE-2	USA
G-35	TGX1987-10F	IITA	G-75	TGX1485-1D	IITA
G-36	Pawe-1	PARC	G-76	TGX2022-4E	IITA
G-37	T44-15-T105-16Sc1	USA	G-77	TGX2023-3E	IITA
G-38	F6LG04-6000XLG04-5187-03	USA	G-78	TGX2004-13F	IITA
G-39	TGX-1989-65f	IITA	G-79	Gozela	PARC
G-40	TGX1951-4F	IITA	G-80	TGX2010-5F	IITA
			G-81	TGX-1990-5FP	IITA

IITA=International Institute of Tropical Agriculture, PARC, JARC and HARC=Pawe, Jimma and Hawassa Agricultural Research Center, USA=United States of America.

Experimental design and management

A simple lattice design was used for evaluating the genotypes. Each genotype was sown on an experimental plot of 1.2 and 3 m in length and width, respectively, with a gross plot size of 3.6 m2. Each plot consisted of two rows with 60 and 5 cm between rows and plants spacing, respectively. Spacing between plots, blocks, and replications was 0.6, 1, and 2 m, respectively. Sowing was done manually, with

two seeds per hill to ensure a uniform stand. The recommended fertilizer rate, 121 kg NPS (19 N, 38 P_2O_5, and 7 S) per hectare, was used fully during sowing. Thinning was carried out two weeks after emergence, followed by three rounds of weeding.

Data collection

Data were recorded during the cropping season on the field and after harvesting. Five plants per plot were taken randomly for plant-based trait data and the net plot area for plot-based traits based on the descriptors of soybean (IBPGR, 1984).

Qualitative traits

Qualitative traits were recorded according to the IBPGR (1984) soybean descriptor:

Flower color was recorded at the vegetative stage (when the flower opens at one of the two uppermost nodes, 1 = white, 2 = purple).
Pubescence presence was recorded and scored at the beginning of maturity (1 = present, 2 = absent).
Pubescence color was recorded at the seed setting stage (full-size seed in the top four nodes, 1 = white, 2 = brown).
Seed color was scored after seed collection (1 = yellowish, 2 = yellow green).
Hilum color was recorded after seed collection (1 = yellow, 2 = black, 3 = grey).
Seed luster was also identified and recorded after seed collection as 1 = shiny, 2 = dull.

Quantitative traits

Days to 50% flowering, days to 95% maturity, grain filling period (the number of days from flowering to maturity), plant height (cm), number of branches per plant, number of pods per plant, number of seeds per pod, hundred seed weight (g), grain yield (kg ha^{-1}), and harvest index [(grain yield/biological yield)100] were recorded. In addition to the agronomic and qualitative data, oil and protein contents were also taken.

Protein and oil content: protein and oil analysis was done by taking 300 g of 162 soybean samples by near-infrared spectroscopy (NIRS) at the Amhara Regional Agricultural Research Institute (ARARI) laboratory in Bahir Dar. The NIRS spectral data were collected using the NIRS analyzer in the reflectance mode of the tool. Scanning of the sample was done twice in the 1100–2500 nm spectral range. The estimation accuracy of protein and oil content was considered as the reading indicating the same accuracy with a standard error of prediction of 0.22% compared to the earlier reading. The partial least squares calibrations were executed with un-scrambler 7.6 CAMO, and the Oslo regression method (Martens and Naes, 1989) was used to develop calibration models for determining the protein and oil content of the soybean samples based on the calibration sample set.

Statistical analysis

Shannon-Weaver diversity index

Morphological diversity for qualitative traits such as flower color, pubescence presence, pubescence color, pod pubescence color, seed color, hilum color and seed luster were estimated using the Shannon-Weaver diversity index (H′) (Dong *et al*. 2004), and the dominant and unique traits observed were recorded on a plot basis. The Shannon-Weaver diversity index (H′) used to characterize the phenotypic frequencies of the characters was defined as follows:

$$H' = \sum_{i=1}^{n} PilnPi$$

Where, n is the number of phenotypic classes for a character and Pi was the frequency of the i^{th} class of traits, and H' was estimated for each trait.

Analysis of Variance (ANOVA)

The simple lattice design was more efficient than the RCBD for most response variables. Thus, the data from all experimental units were analyzed using ANOVA and general linear models (Proc GLM) based on the lattice design. The analysis of variance (ANOVA) was done using SAS (PROC GLM, version 9.4) for both individual and combined locations, following a simple lattice design (SAS Institute Inc., 2013). Bartlett's test for homogeneity of variance was performed using SAS before the ANOVA to ensure the validity of each location's analysis. For the combined ANOVA, Hartley's F-max method (Hartley, 1950) was used to test the homogeneity of error variance, where the ratio of the larger mean square of error (MSE) from the separate analysis of variance to the smaller MSE. If the larger MSE is not three-fold larger than the smaller MSE, the error variance is considered homogeneous (Gomez and Gomez, 1984). Since all traits showed homogeneous error variance, ANOVA was performed separately and combined across locations using a linear random model to account for genotype and location effects. The mean squares of replication, interaction, and residual were combined to test the location effect. But the genotype random effect was tested against the interaction (genotype x location) mean square, while the interaction effect was tested against the residual mean square. Thus, the results of the two locations were interpreted and presented in combination. Tukey's honest significance test (Tukey's HSD) at 5% and 1% significance levels was used for genotype mean comparisons when differences were significant. The analysis of variance for individual locations and across locations was computed using the general linear model for a simple lattice design, as shown below:

The ANOVA model for individual location analysis

$P_{ijk} = \mu + g_i + b_{k(j)} + r_j + e_{ijk}$

Where, P_{ijk} = phenotypic value of i^{th} genotype under j^{th} replication and k^{th} incomplete block within replication j; μ = grand mean; g_i = the effect of i^{th} genotype; $B_{k(j)}$ = the effect of incomplete block k within replication j; R_j = the effect of replication j; and E_{ijk} = the residual or effect of random error.

The ANOVA model for over location analysis

$P_{ijkz} = \mu + g_i + B_{k(j)(z)} + R_{j(z)} + L_z + (gl)_{iz} + E_{ijkz}$

Where, P_{ijkz} = phenotypic value of i^{th} genotype under j^{th} replication at z^{th} location and k^{th} incomplete block within replication j and location z; μ = grand mean; g_i = the effect of i^{th} genotype; $B_{k(j)(z)}$ = the effect of incomplete block k within replication j and location z; $R_{j(z)}$ = the effect of replication j within location z; l_z = the effect of location z; $(gl)_{iz}$ = the interaction effects between genotype and location; and E_{ijkz} = the residual or effect of random error.

Table 2. The structure of ANOVA for individual location for simple lattice design

Source of variation	Degrees of freedom	Sum of squares (SS)	Mean square (MS)	F-value
Replication (r)	r-1	SSr	MSr	MSr/MSE
Genotypes (g unadj.)	g-1	SSg (unadj.)	MSg (unadj.)	MSg/MSE
Genotypes (adj.)	g-1	SSg(adj.)	MSg (adj.)	MSg/MSE
Block within replication (adj.)	r(b-1)	SSb (adj.)	MSb (adj.)	MSb/MSE
Intra–block error	(b-1) (rb-b-1)	SSe	MSe	
Total	rg-1	SST		

r = number of replications; g = No. of genotypes and b = number of plots in a block/block size

Table 3. Analysis of variances for combined over locations for simple lattice design

Source of variation	Degree of freedom	Mean square (MS)	Expected mean square (EMS)
Location (l)	l-1	MSl	$\sigma^2e + r\sigma^2gl + rg\sigma^2l$
Replication within location(r)	l (r-1)	MSr	$\sigma^2e + g\sigma^2rl$
Blocks within replication(b)	r(b-1)	MSb	$\sigma^2e + r\sigma^2gl + r\sigma^2 g$
Genotypes (g)	g-1	MSg	$\sigma^2e + r\sigma^2gl + rl\sigma^2g$
g x l interaction (i)	(g-1) (l-1)	MSgl	$\sigma^2e + r\sigma^2gl$
Error (e)	lg(r-1) - (rb-1) -(l -1)	MSe	σ^2e
Total	Lrb^2-1		

Where, b=intra blocks; σ^2g= genotypic variance, σ^2e = environmental variance, σ^2l=location variance, σ^2r = replication variance, and σ^2gl = genotype x location interaction variance, l = number of locations, g = number of genotypes and r = number of replications.

Estimation of phenotypic and genetic parameters

The phenotypic and genotypic variance components, along with the coefficients of phenotypic and genotypic variability, were estimated using the respective mean square values using the method suggested by Burton and De Vane (1953). The total variance was partitioned into components due to genotype (σ^2g), genotype-by-location interaction (σ^2gl), and environment (σ^2e) variances, assuming the

observed mean squares were equal to their expected mean squares, as suggested by Singh and Choudhary (1985).

Environmental variance (σ^2e) = error mean square = MSe (individual location)
Environmental variance (σ^2e) = error mean square = MSe (combined over locations)
Genotypic variance (σ^2g) = (MSg - MSgl)/rl (combined over locations)
Genotypic variance (σ^2g) = (MSg - MSe)/r (individual location)
Phenotypic variance (σ^2p) = (σ^2g) + (σ^2e) (individual location)
Phenotypic variance (σ^2p) = $\sigma^2g + \sigma^2gl/l + \sigma^2e/rl$ (combined over locations)
Genotype x location interaction variance (σ^2gl) = (MSgl-MSe/r)
Where: MSgl = mean square due to genotypes by location interaction, MSg = mean square due to genotypes, r = number of replications, l = number of locations.

The coefficient of variation at phenotypic and genotypic levels was estimated using the methods suggested by Singh and Choudhary (1985) and Deshmukh *et al*. (1986).

Phenotypic Coefficient of Variation (PCV) = $\frac{\sqrt{\sigma^2 p}}{\bar{\chi}}\ x\ 100$

Genotypic Coefficient of Variation (GCV) = $\frac{\sqrt{\sigma^2 g}}{\bar{\chi}}$ x 100

Where: $\bar{\chi}$ = grand mean of the character under study. The classification for GCV and PCV given by Deshmukh *et al.* (1986) as low (<10%), moderate (10-20%), and high (>20%).

Estimation of heritability in broad sense and genetic advance

Heritability in broad sense (H^2b) is expressed as a percentage of the ratio of the genotypic variance (σ^2g) to the phenotypic variance (σ^2p) estimated by using a method proposed by Hanson *et al*. (1956) and Allard (1960).

Heritability (H^2b) = $\frac{\sigma^2 g}{\sigma^2 p}$ x 100

Where, H^2b = heritability in broad sense, σ^2p = phenotypic components of variance, σ^2g = genotypic components of variance. As demonstrated by Robinson *et al*. (1949), heritability can be categorized as low (0-30%), moderate (30 -60%), and high (60% and above).
Expected genetic advance (GA) for desirable traits under selection was computed by the formulae described by Johnson *et al*. (1955).

Expected genetic advance$(GA) = H^2b * k * \sigma\rho$
Genetic advance as percent of mean (GAM) was computed to compare the extent of the predicted advance of different traits under selection using the formula suggested by Johnson *et al*. (1955) and classified as low (<10%), moderate (10-20%) and high (>20%).

Expected genetic advance (GAM) $= \frac{GA}{\mu} \times 100$

Where, σρ = phenotypic standard deviation on mean basis, H^2b = heritability in broad sense, k = selection differential (where k = 2.06 at 5% selection intensity) and μ = grand mean of the trait under consideration.

Cluster analysis

Cluster analysis was used to group breeding materials based on genotype performance. Genotypes were clustered using the PROC CLUSTER procedure in SAS software version 9.4 (SAS Institute Inc., 2013). The number of clusters was determined by examining cut-off points where local peaks of the pseudo-F statistic aligned with small values of the pseudo t^2 statistic, followed by a larger pseudo t^2 for the next cluster fusion. An agglomerative hierarchical approach was employed, with trait means standardized to zero mean and unit variance to avoid measurement scale biases. Clustering was performed using the average linkage and squared Euclidean distance method with Minitab software version 19.0 (Minitab, 2019), and a dendrogram was created as a measure of dissimilarity using JMP software version 14 Pro (JMP, 2018).

The Mahalanobis's D^2 statistic (Mahalanobis, 1936) was employed to assess the genetic distance between populations. The generalized distance, or squared distances (D^2), between pairs of genotype combinations were calculated using the following formula:

$D^2ij = (xi - xj)\ cov^{-1}\ (xi - xj)$

where, D^2ij = the distance between cases i and j; xi and xj = vectors of the values of the variables for cases i and j; and cov^{-1} = the pooled within groups variance-covariance matrix.

The D^2 values obtained for pairs of clusters were considered as calculated Chi-square (χ^2) values and tested for significance at 1% and 5% probability levels against the tabulated χ^2 values for 'p' degrees of freedom, where 'p' represents the number of traits considered. Average intra- and inter-cluster D^2 values were calculated using the formula provided by Singh and Choudhary (1985).

Average intra-cluster $D^2 = \frac{\sum D_i{}^2}{n}$; where, $\sum D^2_i$ is the sum of distance between all possible combinations, (n) is the population/genotypes included in a cluster.

Average inter cluster $D^2 = \sum D^2_i$ / ninj; Where; $\sum D^2_i$ = sum of distance between all possible combinations, ni and nj = number of genotypes in cluster i and j, respectively.

Principal Component Analysis

Principal component analysis (PCA) was computed to find out the traits that contributed most to the total variation (Jeffers, 1967). Prior to conducting PCA,

the data were standardized to a mean of zero and a variance of one. The PCA was based on the correlation matrix was calculated using Past software version 4.03 (Hammer *et al*. 2020). Eigenvalues greater than or equal to one were considered significant in explaining the observed variability (Jeffers, 1967). Additionally, the correlations between the original traits and their respective principal components (PCs) were also estimated. The PCA was computed using the following equation:

PC1= b11(x1) + b12 + b1p = xp
Where, PC1 = the subjects score on PC1 (the first component extracted), b1p = the regression coefficient (weight) for observed variable p, as used in creating principal component 1 and xp = the subjects score on observed variable p.

Results and Discussion

Variability of qualitative traits

The genetic similarity of 81 soybean genotypes was assessed using the Shannon diversity index (H') to measure phenotypic diversity across qualitative traits, as presented in Table 4. The analysis revealed an average Shannon diversity index of 0.496. Among the seven qualitative traits evaluated, hilum color demonstrated the highest variation with an index of 0.865, followed by seed luster at 0.680 and seed color at 0.530. In contrast, pubescence exhibited no variation, as all genotypes displayed pubescence in their morphology. In the analysis of qualitative traits, 82.72% of genotypes (67 out of 81) had purple flowers, while 17.28% had white. For pubescence color, 80.25% showed brown pubescence, and 19.75% had white, with a diversity index of 0.497. These high diversity indices for these traits indicate significant variation among the genotypes. Similar results were observed by Kumar *et al*. (2020) and Dong *et al.* (2001, 2004). Qualitative traits like flower and seed color are stable across environments and are useful markers for identifying soybean varieties, as noted by Gupta *et al.* (2010).

Table 4. Qualitative trait diversity in soybeans with an estimated phenotypic diversity index (H')

Morphological traits	Category	Genotypes (No.)	Frequency (%)	Diversity Index (H′)
Flower color	White	14	17.28	0.460
	purple	67	82.72	
Pubescence presence	present	81	100.00	0.000
	absent	0	0.00	
Pubescence color	White	16	19.75	0.497
	brown	65	80.25	
Seed color	Yellow	18	22.22	0.530
	Yellow green	63	77.78	
Hilum color	yellow	15	18.52	0.865
	black	12	14.82	
	grey	54	66.66	
Seed luster	shiny	34	41.98	0.680
	dull	47	58.02	

Analysis of variance, range and mean performances

The combined analyses of variance for quantitative traits across two locations for 81 soybean genotypes showed highly significant mean squares for genotypes ($P \leq 0.01$) for all traits (Table 5). This indicates substantial genetic variation among the genotypes, providing valuable opportunities to enhance selection in soybean breeding. The mean squares for the genotype (G) x location (L) interaction showed highly significant differences ($p \leq 0.01$) for most traits, except for hundred seed weight ($p \leq 0.05$), seeds per pod, and oil content ($p \geq 0.05$) (Table 5). This significant interaction (MSgl) indicates that the genotypes exhibited different responses across the two locations for the traits evaluated. In addition, the mean squares for location (MSl) were highly significant ($p \leq 0.01$) for 12 traits and significant ($p \leq 0.05$) only for oil content (Table 5). This suggests that the phenotypic expression of these traits varied between the two locations, highlighting the substantial impact of environmental factors on soybean genotype performance. Differences in environmental conditions, such as soil and climate, likely contributed to these variations. Similar findings were also reported by Guleria *et al*. (2019), Sileshi (2019a), and Yirga *et al*. (2022), who observed variability among soybean genotypes for various traits.

Table 5. Combined analysis of variance (ANOVA) for 13 traits in 81 soybean genotypes tested at two locations, 2019/2020

Source of Var.	Mean Squares						
	MSl (1)	MSg (80)	MSgl (80)	Rep within Location (2)	Block within Rep (16)	MSe (144)	CV (%)
DF	1995.11**	264.40**	13.64**	17.83**	2.52^{ns}	1.79	2.55
DM	474.27**	185.83**	21.69**	8.34^{ns}	8.47**	3.27	1.61
GFP	523.90**	132.07**	38.17**	1.78^{ns}	7.49*	4.52	3.57
PH	50545.03**	1057.78**	83.27**	100.89^{ns}	100.7**	34.71	7.37
BPP	143.87**	6.20**	1.18**	11.29**	1.46**	0.36	13.45
PPP	73441.0**	1386.59**	389.66**	258.49^{ns}	231.75**	89.57	15.86
SPP	0.96**	0.11**	0.08^{ns}	0.047^{ns}	0.058**	0.06	9.46
PL	11.07**	0.19**	0.18**	0.23^{ns}	0.13^{ns}	0.12	9.62
GY	1633797.5**	2193634**	661101.3**	118301.6^{ns}	89986.8^{ns}	62893.7	9.36
HSW	1134.48**	14.45**	2.15*	0.51^{ns}	5.51**	1.51	8.08
HI	0.026**	0.018**	0.0096**	0.0041*	0.0012^{ns}	0.00099	8.15
OC	4.044*	8.77**	1.09^{ns}	1.00^{ns}	0.81^{ns}	0.93	4.58
PC	124.94**	15.54**	2.38**	10.31**	1.27^{ns}	1.39	2.77

Note *, ** significant at 0.05 and 0.01 probability levels, respectively; NS = Non-Significant, figures in parenthesis indicate degrees of freedom; MSl = Mean Square of Location, MSg = Mean Squares of Genotypes; MSgl = Mean Square Due to Genotype by Location, MSe = Mean Squares of Error; CV = coefficient of variation (%), DF = Days to 50% Flowering, DM = Days to Maturity, GFP = Grain Filling Period, PH = Plant height(cm), BPP = Branches per Plant, PPP = Pods per Plant, SPP = Seed per Pod, PL = Pod Length(cm), GY = Grain Yield (kg/ha), HSW = Hundred Seed Weight (g), HI = Harvest Index, OC = Oil content, and PC = Protein Content, Number in parenthesis show respective degrees of freedom.

The range and mean values of the 13 traits, along with their respective coefficients of variation combined across the two locations, are presented in Table 6. The combined mean for days to flowering among the genotypes ranged from 34.75 to 65.25 days, with an overall mean of 52.57 days, while days to maturity ranged from 97.25 to 125.0 days. The latest maturity was observed in genotype

TGX2004-13F (125 days), while the earliest maturity was recorded in genotype F6LG04-6000XLG04-5187-03 (97.25 days). The grain filling period also showed significant variation, ranging from 41.25 to 71.75 days. The wide variation observed in days to maturity among the genotypes offers an opportunity to develop soybean varieties suited to diverse agro-ecological zones. Hence, this genetic diversity facilitates the breeding of both early- and late-maturing varieties tailored to different conditions influenced by rainfall patterns. Sileshi (2019a) and Liu *et al.* (2020) also noted similar variability, identifying different maturity groups in soybean genotypes.

The plant height trait displayed a wide range of mean values, from 40.70 cm to 116.37 cm, with genotypes exceeding the mean of 79.9 cm being potential candidates for breeding programs aimed at increasing height. This is consistent with Viotto *et al.* (2020) and Shilpashree *et al.* (2021). Significant differences were also observed in the number of branches per plant (ranging from 1.10 to 7.20), pods per plant (22.40 to 104.75), seeds per pod (2.25 to 3.05), and pod length (3.05 to 4.13 cm) (Table 6). Notably, 50.61% of the genotypes exceeded the grand mean for pods per plant, highlighting genetic variability and selection potential. Similar variations in branches, pods, and seeds per pod have been reported by Kumar *et al.* (2020).

Grain yields among the genotypes ranged from 341.80 to 4,499.00 kg ha^{-1}, with approximately 61.73% exceeding the grand mean of 2,679.13 kg ha^{-1}. The highest yield was recorded in genotype TGX1951-4F at 4,499.00 kg ha^{-1}, followed by TGX2010-5F at 4,267.50 kg ha^{-1}, with yield advantages of 40.5% and 33.2%, respectively, over the variety AFGAT. Conversely, the lowest yield was observed in genotype H3-15-SF-2 (341.80 kg ha^{-1}), followed by F6LG04-6000XLG04-5187-05 (477.80 kg ha^{-1}). This substantial variation indicates significant genetic diversity among the genotypes, highlighting the potential for soybean yield improvement through selection. These findings align with previous research by Getnet (2018) and Sileshi (2019a), which also reported significant variations in grain yield.

Among agronomic traits, harvest index and hundred seed weight showed considerable variability, while oil and protein content exhibited significant mean ranges. The mean ranges were 0.2312–0.5002 for harvest index, 10.96–19.12 g for hundred seed weight, 17.97–24.45% for oil content, and 37.20–46.75% for protein content. The highest hundred seed weight was observed in CRFRD-15-SB (19.12 g), and the lowest was in Pawe-03 (10.96 g). Comparable variations in these traits were documented by Aditya *et al.* (2011) and Rasyad *et al.* (2017). For the quality traits of oil and protein content, the highest values were observed in genotypes F6LG06-5920XU03-100612-03 (24.45%) and TGX1987-10F (46.75%), respectively. Such significant variation among genotypes is crucial for plant

breeding, enhancing the potential for effective crossing and selection. High-performing genotypes should be prioritized for improving traits. Rasyad *et al.* (2017) observed similar trends in oil and protein content, while Ramteke *et al.* (2010) noted significant variations in hundred seed weight, oil, and protein content. Overall, the assessed traits show substantial genetic variability among the tested soybean genotypes.

Table 6. Descriptive statistics for 13 traits in 81 soybean genotypes tested over locations, 2019/2020

	Statistics			
Traits	Mean ± SE mean	Range	CV (%)	R^2
DF	52.57±0.07	34.75 - 65.25	2.55	99.02
DM	112.13±0.10	97.25 – 125.00	1.61	97.45
GFP	59.56±0.12	41.25 - 71.75	3.57	95.90
PH	79.90±0.33	40.70 - 116.37	7.37	96.88
BPP	4.49±0.03	1.10 - 7.20	13.45	94.27
PPP	59.66±0.52	22.40 - 104.75	15.86	94.63
SPP	2.62±0.01	2.25 - 3.05	9.46	66.41
PL	3.57±0.02	3.05 - 4.13	9.62	72.92
GY	2679.13±13.93	341.80 - 4499.00	9.36	98.57
HSW	15.20±0.07	10.96 - 19.12	8.08	92.66
HI	0.3866±0.002	0.2312 - 0.5002	8.15	94.41
OC	21.00±0.05	17.97 - 24.45	4.58	86.71
PC	42.59±0.06	37.20 - 46.75	2.77	89.81

SE = Standard Error, CV = Coefficient of Variation (%), R^2 = Coefficient of Determination, DF = Days to 50% Flowering, DM = Days to Maturity, GFP = Grain Filling Period, PH = Plant height(cm), BPP = Branches per Plant, PPP = Pods per Plant, SPP=Seed per Pod, PL= Pod Length(cm), GY = Grain Yield (kg ha^{-1}), HSW = Hundred Seed Weight(g), HI = Harvest Index, OC = Oil content, and PC = Protein Content

Estimates of Genetic Parameters

Estimates of variance components

The analysis of various traits, including estimates of phenotypic variance (σ^2_p), genotypic variance (σ^2g), environmental variance (σ^2e), genotypic coefficient of variation (GCV), phenotypic coefficient of variation (PCV), broad-sense heritability (H^2), genetic advance (GA), and genetic advance as a percentage of the mean (GAM), is presented in Table 7. The results indicated that the magnitude of the phenotypic coefficient of variation (PCV) was slightly higher than the genotypic coefficient of variation (GCV). The minimal difference between GCV and PCV suggests a stronger genetic contribution to trait variation among the evaluated genotypes. Bisht *et al.* (2018) and Guleria *et al.* (2019), who reported similar patterns of minimal variation between PCV and GCV for traits such as days to flowering, plant height, and grain yield in soybean genotypes.

In this study, PCV values ranged from 4.63% for protein content to 37.76% for grain yield, while GCV values ranged from 1.24% for pod length to 34.58% for grain yield (Table 7). Traits such as plant height, branches per plant, pods per plant, and grain yield exhibited high PCV values, while intermediate PCV values were observed for days to flowering, hundred seed weight, and harvest index. In

contrast, low PCV values were recorded for days to maturity (6.08%), grain filling period (9.65%), seeds per pod (6.34%), pod length (6.16%), oil content (7.05%), and protein content (4.63%). This observation aligns with Guleria *et al.* (2019), who found high PCV values for branches per plant, pods per plant, grain yield, and hundred seed weight. Similar results were reported by Bisht *et al.* (2018) for plant height, branches per plant, seeds per pod, and hundred seed weight; Guleria *et al.* (2019) for pod length and seeds per pod; and Kumar *et al.* (2020) for days to 50% flowering and hundred seed weight in soybean genotypes. Low PCV values for oil and protein content indicate minimal phenotypic variation, making improvement through phenotypic selection may be challenging. Hence, enhancing these traits may require alternative methods, such as crossing or mutagenesis, followed by selection. This aligns with Baraskar *et al*. (2014), who reported low PCV values for days to maturity, oil, and protein content, and Kumar *et al.* (2020) for days to maturity.

Traits with high GCV values include branches per plant, pods per plant, and grain yield, while days to 50% flowering, plant height, hundred seed weight, and harvest index were categorized as having medium GCV. Traits exhibiting low GCV values were days to maturity, grain filling period, seeds per pod, pod length, oil content, and protein content (Table 7). The higher GCV estimates for these traits indicate significant genetic variability among the tested genotypes, suggesting that selection could effectively enhance these traits. High GCV was found for pods per plant and grain yield by Getnet (2018), and for branches per plant, pods per plant, and grain yield by Guleria *et al*. (2019). Reni and Rao (2013) observed moderate GCV for days to 50% flowering, plant height, and hundred seed weight, and low GCV for oil and protein content, consistent with this study. In general, the observed wider variability on phenotypic and genotypic coefficients of variation generally indicates the existence of substantial variability in the studied traits. However, traits with low PCV and GCV values (Table 7) demonstrate limited variability, indicating a narrower scope for improvement through selection. This underscores the necessity of employing alternative methods to generate variability. Moreover, traits exhibiting high PCV, GCV, broad-sense heritability, and genetic advance indicate a greater potential for effective selection based on genotypic variation (Baraskar *et al*. 2014).

Estimates of broad sense heritability and genetic advance

According to Robinson *et al*. (1949) heritability estimates were classified as low (<30%), medium (30-60%), and high (≥60%). Heritability estimates in this study ranged from 4.03% for pod length to 94.84% for days to flowering. High heritability was observed for most traits, including days to flowering (94.84%), plant height (92.13%), days to maturity (88.32%), oil content (87.49%), hundred seed weight (85.12%), protein content (84.68%), grain yield (83.85%), branches per plant (80.89%), pods per plant (71.90%), and grain filling period (71.09%),

while harvest index exhibited moderate heritability (48.08%). These high and moderate heritability values indicate that genetic factors predominantly influence these traits, with minimal environmental impact. As a result, a significant proportion of the variation is heritable, making it feasible to enhance these traits through direct selection of superior genotypes based on their phenotypic traits. High heritability estimates for such traits were also reported by Neelima *et al.* (2018) and Adetiloye *et al*. (2020). Similarly, Chandrawat *et al*. (2017) and Shilpashree *et al.* (2021) found high heritability for protein and oil content.

According to Johnson *et al.* (1955), genetic advance as a percent of mean (GAM) is classified as low (<10%), moderate (10-20%), or high (>20%). The genetic estimation showed that genetic advance as a percentage of the mean ranged from 0.51% (pod length) to 65.22% (grain yield). High genetic advance was found for traits like days to 50% flowering, plant height, branches per plant, pods per plant, hundred seed weight, and grain yield. Moderate genetic advance was observed for days to maturity, grain filling period, harvest index, and oil content, while seeds per pod, pod length, and protein content showed low genetic advance. The high estimates of genetic advance for these traits suggest a significant potential for improvement through selection. High genetic advance values indicate additive gene action, while low values suggest non-additive gene action (Singh and Narayanan, 1993). In line with this, Neelima *et al.* (2018) reported high genetic advance for plant height, moderate genetic advance for days to maturity, and low genetic advance for branches per plant, hundred seed weight, oil content, and protein content. Mesfin (2018) found high genetic advance for traits such as plant height, branches per plant, pod length, harvest index, grain yield, and hundred seed weight, but low genetic advance for days to maturity and pods per plant. Similarly, Malek *et al.* (2014) found comparable results, while Chandrawat *et al.* (2017) observed contrasting genetic advance for days to 50% flowering, days to maturity, plant height, pods per plant, and harvest index.

In the present study, high heritability combined with high genetic advance was found for plant height (92.13, 38.63%), and pods per plant (71.9, 46.22), respectively indicating the potentials for effective improvement through selection methods. Days to 50% flowering and days to maturity exhibited high heritability with moderate genetic advance, indicating they are mainly influenced by both additive and non-additive gene action and are less affected by environmental factors. Conversely, grain filling period, branches per plant, hundred seed weight, oil content, and protein content exhibited high heritability with low genetic advance, indicating non-additive gene action and suggesting that recombination breeding and recurrent selection may be more effective for these traits, as indicated by Hakim and Suyamto (2017). In conformity, Shilpashree *et al.* (2021) reported high heritability and high genetic advance for plant height and pods per plant, but high heritability with low genetic advance for pod length. A

contradicting report on days to 50% flowering and maturity (high H^2 and low GA), and protein content (high H^2 with high GA) were also demonstrated by these authors. Jain *et al.* (2018), indicated dissimilar report where moderate heritability together with a high genetic advance on days to maturity, plant height, pods per plant, hundred seed weight, harvest index, and grain yield in soybean genotype.

Table 7. Variance, heritability, genotypic and phenotypic coefficient of variation, and genetic advance for the 13 traits of 81 soybean genotypes combined over location, 2019/2020

Traits	σ^2g	σ^2gl	σ^2p	σ^2e	GCV (%)	PCV (%)	H^2 (%)	GA	GAM (%)
DF	62.69	5.92	66.10	1.79	15.06	15.46	94.84	15.88	30.21
DM	41.03	9.21	46.46	3.27	5.71	6.08	88.32	12.40	11.06
GFP	23.47	16.83	33.02	4.52	8.13	9.65	71.09	8.41	14.13
PH	243.63	24.28	264.44	34.71	19.54	20.35	92.13	30.86	38.63
BPP	1.25	0.41	1.55	0.36	24.91	27.69	80.89	2.07	46.15
PPP	249.23	150.04	346.65	89.57	26.46	31.21	71.90	27.57	46.22
SPP	0.007	0.01	0.03	0.06	3.11	6.34	24.03	0.08	3.14
PL	0.002	0.03	0.05	0.12	1.24	6.16	4.03	0.02	0.51
GY	383133.2	629654.5	713683.8	62893.7	23.10	31.53	53.68	934.25	34.87
HSW	3.08	0.32	3.61	1.51	11.54	12.51	85.12	3.33	21.93
HI	0.002	0.004	0.005	0.001	12.20	17.59	48.08	0.07	17.42
OC	1.92	0.08	2.19	0.93	6.60	7.05	87.49	2.67	12.71
PC	3.29	0.49	3.88	1.39	4.26	4.63	84.68	3.44	8.07

σ^2p = Phenotypic variation, σ^2g = Genotypic variation, σ^2gl = Variance for genotype x location interaction, σ^2e = Environmental variance, GCV (%) = Genotypic coefficient of variation, PCV (%) = Phenotypic coefficient of variation, H^2(%) =Broad sense heritability, GA (5%) = Genetic advance at 5% selection intensity, GAM =Genetic advance as percent of mean, DF = Days to 50% Flowering (days), DM = Days to Maturity (days), GFP = Grain Filling Period (days), PH= Plant Height (cm), BPP = Number of Branches per Plant (No.), PPP = Number of Pods per Plant (No.), PL = Pod Length (cm), HSW(g) = Hundred Seed Weight, HI = Harvest Index (%), PC = protein Content (%)

Multivariate Analysis

Clustering of genotypes

The 81 soybean genotypes were grouped into six distinct clusters based on quantitative traits (Table 8). Cluster I was the largest, comprising 51 genotypes (62.96%), followed by Cluster III with 9 genotypes (11.11%), Cluster II with 8 genotypes (9.88%), Cluster IV with 7 genotypes (8.64%), and Cluster VI with 5 genotypes (6.17%). Cluster V had the lowest number of genotypes, containing only one genotype (a singleton). These results indicate a high degree of diversity among the tested soybean genotypes

Previous studies have similarly documented the diversity among soybean genotypes grouped into distinct clusters. Iqbal *et al.* (2008) and Oliveira *et al.* (2017) employed multivariate analysis to assess genetic diversity and found significant variation among the genotypes. Similarly, Marconato *et al.* (2016) grouped genotypes into eight clusters, and Singh and Shrestha (2019) identified five clusters, all indicating significant genetic variation. Assessing the genetic diversity and relationships among genotypes is essential for selecting parents with distinct genetic backgrounds, a key factor in the success of any breeding program. Genotypes clustered together exhibit a closer genetic affinity compared to those in other clusters. Consequently, genotypes within the same cluster tend to be genetically more homogeneous, while displaying significant divergence from those in different clusters. The genotype distribution across clusters and their relationships in the dendrogram are shown in Fig 3.

Table 8. Distribution of genotypes into clusters based on D^2 analysis for 81 soybean genotypes combined over location, 2019/2020

Cluster	No. of genotypes	Proportion (%)	Genotypes/cluster membership
I	51	62.96	TGX2009-14F, Gishama, TGX2025-9E, TGX2016-2E, TGX2025-19E, TGX-1990-40f, TGX2018-5E, TGX2013-2F, TGX2009-1F, TGX2027-4E, TGX2010-11F, TGX 2025-6E, TGX1988-5F, TGX2007-1F, TGX1993-4FN, TGX 2009-16F, TGX2020-1E, TGX2023-4E,TGX2017-5E, TGX-1989-65f, TGX1951-4F, TGX2019-1E, TGX2011-6F, TGX2025-16E, TGX2016-3E, TGX1987-14F, TGX-1987-28f, TGX2017-6E, TGX2016-4E, T34-15-T73-16-SD1, TGX2004-7F, TGX2025-10E, TGX-1835-10E, TGX-1919-22F, TGX2015-1E, TGX-1987-11F, Belessa-95, TGX2025-14E,TGX2023-1E, TGX2027-1E, TGX2008-4F, TGX-1989-40F, Gizo,TGX-1987-18F, TGX-1988-5E,TGX1485-1D, TGX2022-4E, TGX2023-3E, TGX2004-13F, TGX2010-5F, TGX-1990-5FP
II	8	9.88	TGX2007-3F, TGX1835-10E, TGX1989-19F, TGX-1889-62F, Pawe-03, TGX2010-14F, TGX1987-10F, TGX2027-7E
III	9	11.11	T34-15-T72-16-Sc1, JM-ALM/H3-15-SC-1, Hawassa-04, Pawe-1, T44-15-T105-16Sc1, T47-15-T126-16-SF1, AFGAT, Gozela, G7955-C3RPP(C1)
IV	7	8.64	F6U03-300134XLG04-5187, F6LG06-5920XU03-100612-03, F6LG04-6000XLG04-5187-04, F6LG04-6000XLG04-5187-03, F6LG06-5920XU03-100612-01, F6LG046000XLG04-5187-02, F6LG04-6000XLG04-5187-01
V	1	1.24	Andinet
VI	5	6.17	F6LG04-6000XLG04-5187-05, F6LG04-6000XLG04-5187-06, CRFRD-15-SE-2, CRFRD-15-SB, H3-15-SF-2

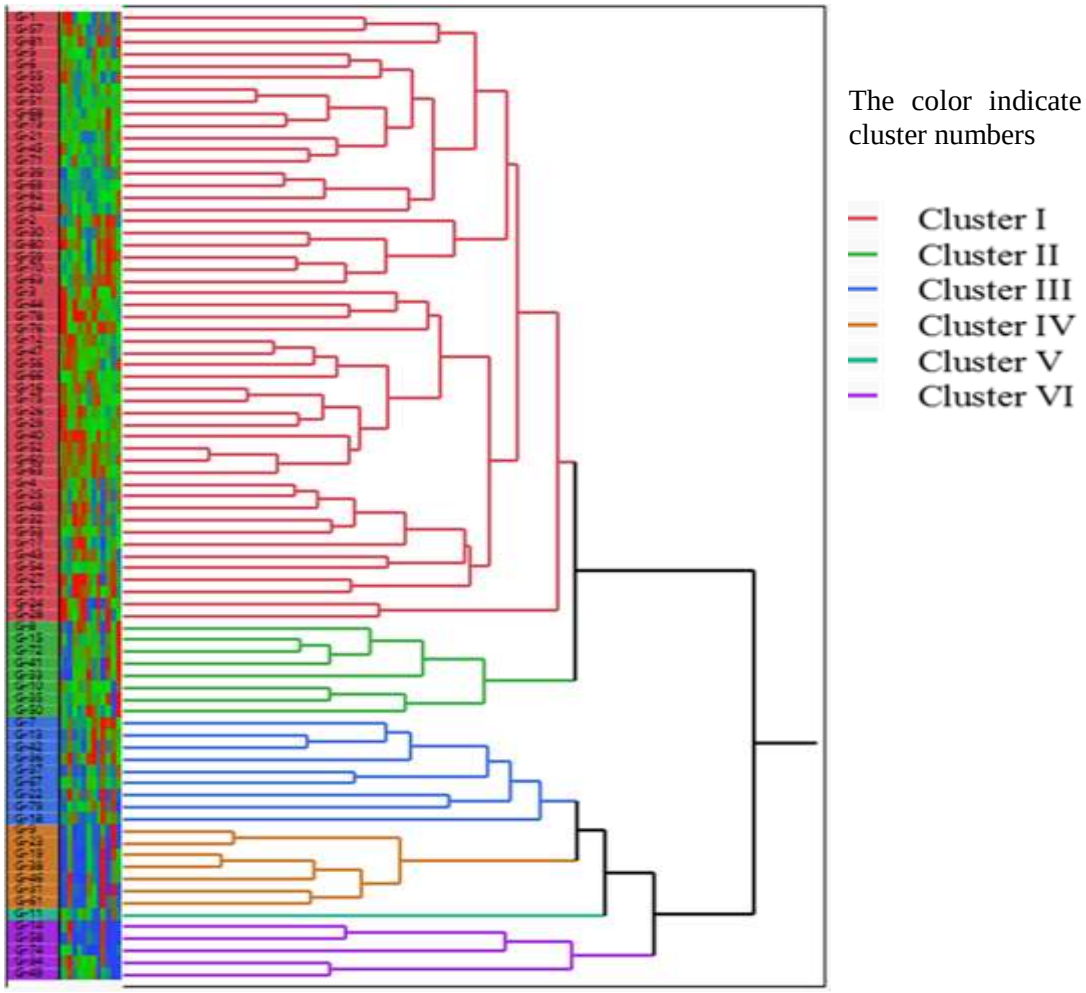

Fig 3. Dendrogram showing relationships among 81 soybean genotypes

Cluster means and distance analysis

The standardized Mahalanobis D^2 statistics revealed significant genetic differences between cluster pairs, with all inter-cluster divergences being significant ($p \leq 0.05$ and $p \leq 0.01$). However, intra-cluster divergences were non-significant. The average intra- and inter-cluster D^2 values and their corresponding distances are presented in Table 9. Intra-cluster distances ranged from 1.91 (cluster IV) to 3.94 (cluster VI), indicating low D^2 values, which suggests more similarity within clusters. Thus, genotypes in the same cluster were less divergent compared to those in different clusters. Sharma *et al. (*2005) reported similar findings, with intra-cluster D^2 values ranging from 0.0 to 4.43 in 62 soybean varieties.

The inter-cluster analysis showed the greatest distance between clusters II and VI (D^2 = 154.64**), followed by clusters II and IV (D^2 = 132.39**) (Table 9),

indicating significant genetic diversity among the soybean genotypes. This broad divergence is valuable for breeders, as genotypes from highly divergent clusters can be used to develop lines with diverse genetic backgrounds, aiding future soybean improvement through existing variability. The minimum inter-cluster distance (D^2 = 22.53*) was observed between clusters I and III, with nearly a similar distance between clusters III and IV (D^2 = 23.23*), indicating closer proximity (Table 9). Adie and Krisnawati (2017) categorized genotypes into ten clusters, and Sharma *et al.* (2005) into fifteen clusters. In conclusion, genotypes from clusters I, II, and IV have the potential to serve as gene sources for developing new soybean varieties through hybridization.

Table 9. Pair-wise generalized intra- (**bolded diagonal**) and- inter (**off-diagonal**)-cluster distances (D^2) between cluster values of 81 soybean genotypes, 2019/2020

Cluster	I	II	III	IV	V	VI
I	**2.53**					
II	23.70*	**2.53**				
III	22.53*	59.78**	**3.94**			
IV	77.35**	132.39**	23.23*	**1.91**		
V	64.04**	111.52**	35.18**	27.55**	**0**	
VI	93.32**	154.64**	63.21**	62.77**	56.84**	**3.08**

** = Significant at $p < 0.05$ for $x^2 = 21.03$; ** = significant at $p < 0.01$ for $x^2 = 26.22$ and ns = non-significant*

The mean values of all 13 traits for each cluster group are presented in Table 10, indicating variations among the six clusters for different traits. Cluster I had moderate trait values but the highest for days to maturity, plant height, branches per plant, and grain yield. Cluster II, the third-largest, exhibited high values for days to 50% flowering, seeds per pod, harvest index, and protein content but lower grain-filling period and hundred seed weight. Cluster III, with 9 genotypes, showed the highest pod length. Cluster IV, containing 7 genotypes, had the highest hundred seed weight and oil content but the lowest values for most traits. The genotype in Cluster IV is notably early-maturing, making it valuable for breeding programs targeting early soybean maturity. Cluster V, with one genotype, was notable for high pods per plant. Cluster VI, with 5 genotypes, had the longest grain-filling period but the lowest values for seeds per pod, grain yield, oil content, and protein content. Generally, the cluster analysis grouped genotypes by morphological similarities, enabling the selection of representative genotypes from each cluster for hybridization breeding. Correspondingly, Khan *et al.* (2014) identified six clusters among 115 soybean genotypes, revealing significant differences in mean values for nearly all traits.

Table 10. Mean values of 13 quantitative traits of the six clusters of 81 soybean genotypes for combined data, 2019/2020

Traits	Clusters					
	I	II	III	IV	V	VI
DF	56.16	59.13**	43.61	36.50*	57.30	43.62
DM	115.53**	107.69	107.44	100.24*	113.30	109.78
GFP	59.40	48.59*	63.84	63.79	56.00	65.58**
PH	88.29**	84.90	67.19	49.71*	61.80	55.16
BPP	4.99**	4.54	3.96	1.83*	4.30	3.72
PPP	68.24	62.45	41.50	27.87*	70.60**	42.80
SPP	2.61	2.84**	2.70	2.67	2.60	2.50*
PL	3.59	3.54	3.73**	3.43	3.30*	3.38
GY	3228.96**	3023.18	2006.83	923.27	782.90	578.30*
HSW	14.56	12.99*	17.44	17.98**	15.71	16.62
HI	0.4167	0.4300**	0.3598	0.2570*	0.2573*	0.2667
OC	20.94	20.05	22.03	23.04**	22.60	18.18*
PC	43.04	45.59**	41.83	40.14	41.00	38.38*

* and ** = lowest and highest value of cluster mean, DF = days to 50% flowering, DM = days to maturity, GFP = Grain Filling Period, PH = Plant Heigh, BPP = Number

Principal Component Analysis

The first four principal components (PC1 to PC4) accounted for 77.98% of the total variation among 81 soybean genotypes (Table 11, Fig 4). Components PC1, PC2, PC3 and PC4 with Eigenvalues of 5.88, 1.73, 1.49, and 1.04 contributed 45.25%, 13.29%, 11.44%, and 8.00% of the total variation, respectively (Table 11). A similar observation was made by Sileshi *et al.* (2019b) and Vianna *et al.* (2013), who reported that the first four principal components accounted for 82% and 71.6% of the total variation among soybean genotypes, respectively..

The first principal component (PC1), accounting for the highest variation, was strongly associated with days to 50% flowering, plant height, grain yield, pods per plant, harvest index, and branches per plant, with correlation values of 0.372, 0.357, 0.356, 0.355, and 0.336, respectively (Table 11). This indicated that the population with greater PC1 value is considered high yielding and formed by having long- days to 50% flowering and plant height, and more grain yield, pods per plant, and harvest index. Therefore, selection for traits under the first principal component may be desirable and should be carefully considered. Supporting this, Iqbal *et al.* (2008), Vianna *et al.* (2013), and Marconato *et al.* (2016) noted that quantitative traits significantly contributed to the first three principal components, accounting for over 70% of the total variation. Furthermore, in line with the current findings, these authors highlighted that the key contributors to the first principal component (PC1) were grain yield, harvest index, plant height, and the number of pods per plant.

The second principal component (PC2), accounting for 13.29% of the variance, was primarily explained by pod length, grain filling period, hundred seed weight, days to maturity, and protein content, with correlation values of 0.521, 0.468, 0.399, 0.328, and -0.297, respectively (Table 11). Marconato *et al.* (2016) found

that PC2 explained 20.30% of the variance, mainly due to the grain-filling period. Similar results for protein content were reported by Miladinovic *et al.* (2006), while Iqbal *et al.* (2008) emphasized days to maturity and hundred seed weight. The third principal component (PC3) was primarily influenced by seeds per pod (0.575) and oil content (0.482), while the fourth principal component (PC4) was mainly associated with oil content (0.655) and seeds per pod (-0.568) (Table 11). Likewise, Sileshi *et al.* (2019b) noted significant contributions of seeds per pod to PC3.

The first two principal components, PC1 (45.25%) and PC2 (13.29%), were grouped based on the bi-plot analysis on a two-dimensional plane, with trait and genotype distributions visualized in Fig 4. Among the 81 soybean genotypes, the highest principal component scores were observed across four components (Table 11), which can serve as indicators for selection based on the variability explained by each PC. High scores for specific components reflect elevated trait values in the corresponding genotypes. The maximum score of PCs for specific components indicates high values for the traits in those specific genotypes. The results indicated that genotypes G-58, G-19, G-23, G-31, G-46, G-38, G-61, G-14, and G-9 exhibited high values for days to 50% flowering, plant height, branches per plant, pods per plant, grain yield, and harvest index, as reflected in PC1. This aligns with Sileshi *et al.* (2019b) regarding days to 50% flowering and plant height. In PC2, genotypes Pawe-03, Pawe-01, G-41, G-78, G-8, G-15, and G-34 exhibited high values for days to 50% flowering, grain filling period, pod length, and hundred seed weight. For PC3, genotypes G-74, Pawe-03, G-14, G-24, G-28, G-1, G-57, and G-34 exhibited high values for days to maturity, grain filling period, seeds per pod, pod length, and oil content. In PC4, genotypes G-59, G-50, G-49, G-30, G-80, Gizo, G-34, and G-55 displayed high values for seeds per pod and oil content.

Broschat (1979) considered PCA as powerful technique for data reduction which removes interrelationships among components. In this PC biplot analysis, the aggregation of traits and genotypes has presented in Fig 4. Genotypes G-27, G-59, G-32, G-28, and G-30 are recommended for enhancing grain yield. For improving days to maturity, G-78, G-40, G-16, G-44, and G-60 are optimal, while G-34, G-13, Gozela, AFGAT, Hawassa-04, and G-49 are key for enhancing hundred seed weight. Other genotypes are categorized for improving various traits. However, G-74, G-69, G-37, G-39, G-58, G-9, G-23, and Andinet, located in the fourth quadrant, did not exhibit any standout traits. Vianna *et al.* (2013) and Marconato *et al.* (2016) similarly highlighted key traits that can be exploited using principal component analysis.

Table 11. Eigenvectors, variance explained and Eigenvalues of the first four PCs of soybean genotypes evaluated over location, 2019/2020

	Eigenvectors			
Traits	PC1	PC2	PC3	PC4
Days to 50% flowering	0.372	-0.061	-0.026	0.024
Days to maturity	0.283	0.328	-0.362	-0.107
Grain filling period (days)	-0.197	0.468	-0.384	-0.159
Plant height (cm)	0.357	0.123	0.071	0.021
Branch per plant	0.311	0.273	-0.073	0.082
Pod per plant	0.355	0.077	-0.092	-0.028
Seed per pod	0.043	0.098	0.575	-0.568
Pod length (cm)	0.080	0.521	0.357	-0.238
Grain yield (kg ha $^{-1}$)	0.356	0.057	0.016	0.227
Hundred seed weight(g)	-0.283	0.399	0.007	0.177
Harvest index (%)	0.336	-0.0001	0.049	0.187
Oil content (%)	-0.064	0.205	0.482	0.655
Protein content (%)	0.247	-0.297	0.098	-0.166
Eigenvalue	5.88	1.73	1.49	1.04
Explained variance (%)	45.25	13.29	11.44	8.00
Cumulative variance (%)	45.25	58.54	69.98	77.98

Table 12. Traits having values greater than 0.3 in each PCs over the combined location, 2019/2020

PC1	PC2	PC3	PC4
Days to 50% flowering	Days to maturity	Days to maturity	Seed per pod
Plant height	Grain filling period	Grain filling period	Oil content
Branch per plant	Pod length	Seed per pod	-
Pod per plant	Hundred seed weight	Pod length	-
Grain yield	-	Oil content	-
Harvest index -	-	-	

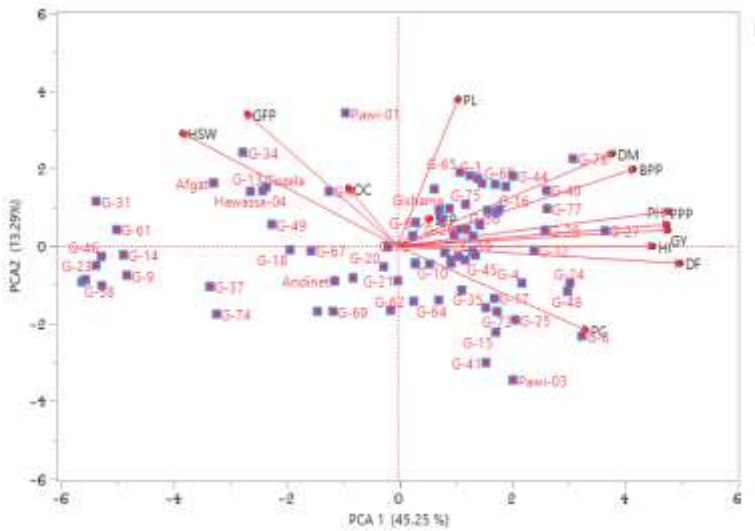

Fig 4. Biplot of PC1 and PC2 illustrates the relationships between genotypes (red circles, n=81) and traits (black).

Conclusion

The combined analysis of variance showed highly significant differences among the tested genotypes. In addition, the genotype × location interaction effects were also highly significant for most of the traits studied, suggesting that the genotypes responded differently across locations, likely due to environmental variations. High genotypic coefficients of variation (GCV), heritability (H^2), and genetic advance as a percent of mean (GAM) were observed for branches per plant (24.91, 80.89, and 46.15%) and pods per plant (26.46, 71.90, and 46.22%), respectively, while grain yield showed high GCV and GAM (23.10 and 34.87%). This finding confirms the potential for improvement through phenotypic selection.

The 81 soybean genotypes were grouped into six clusters based on their similarities, with the maximum inter-cluster distance observed between clusters II and VI. This suggests that genotypes from these clusters could be ideal candidates for a crossing program, by taking into account other qualitative traits. In the principal component analysis, the first four principal components accounted for 77.98% of the total variation among the tested genotypes, indicating that the evaluated traits capture a substantial portion of the overall diversity, thereby confirming the potential for future improvement through directional selection and hybridization. It is recommended that intercrossing genotypes from genetically diverse clusters, specifically clusters II and VI, along with those showing superior mean performance could be effective in a soybean improvement program while considering other qualitative traits. The current study has identified high-performing genotypes that should be prioritized for improving soybean productivity. However, it is advisable to incorporate molecular techniques to validate and strengthen these recommendations.

Acknowledgments

The assistance obtained from the Amhara Agricultural Research Institute (ARARI), Gondar Agricultural Research Center (GARC), crop research programs of GARC and Metema and West Armachiho research stations are highly appreciated.

References

Adetiloye, I.S., Ariyo, O.J., Alamu, O. and Osewa, S.O., 2020. Agronomic potential and genetic diversity of 43 accession of tropical soybean (*Glycine max* (L) Merrill). International Journal of Plant Research, **10 (2): 33-39.**

Adie, M.M. and Krisnawati, A., 2017. Characterization and clustering of agronomic characters of several soybean genotypes. Nusantara Bioscience, **9(3): 237-242.**

Aditya, J.P., Bhartiya, P. and Bhartiya, A., 2011. Genetic variability, heritability and character association for yield and component characters in soybean (*Glycine max* (L.) Merrill). Journal of Central European Agriculture, **12(1):27-34.**

Allard, R.W. 1960. Principles of plant breeding. *John Willey and Sons. Inc. New York, 166-169.Inc. New York*, **pp.166–169**.

Arshad, M., Aslam, M. and Irshad, M. 2009. Genetic variability and character association among morphological traits of mungbean, vigna radiata l. Wilczek genotypes. Journal of Agricultural Research, **47(2): 368-1157.**

Ayalew, D., Tesfaye, K., Mamo, G., Yitaferu, B. and Bayu, W., 2012. Variability of rainfall and its current trend in Amhara region, Ethiopia. African Journal of Agricultural Research, **7(10):1475-1486.**

Baraskar, V.V., Kachhadia, V.H., VachhanI, J.H., Barad, H.R., Patel, M.B. and Darwankar, M.S., 2014. Genetic variability, heritability and genetic advance in soybean (*Glycine max* (L.) Merrill). Electronic Journal of Plant Breeding, **5(4):802-806.**

Bhatia, R., Dey, S.S. and Kumar, R., 2017. Genetic divergence studies in tulip (Tulipa gesneriana L.). Indian Journal of Horticulture, **74(4):562-567.**

Bisht, M., Singh, K., Bhatt, P., Kunduru, B. and Chourasia, K.N., 2018. Studies on variability parameters in early generation of soybean (*Glycine max* (L.) Merrill). International Journal of Chemical Studies, **6(1): 208-211.**

Broschat, T.K., 1979. Principal component analysis in horticultural research. Hort. Sci., **14: 114-117.**

Burton, G.W. and De Vane, D.E., 1953. Estimating heritability in tall fescue (Festuca arundinacea) from replicated clonal material. Agronomy Journal, **45(1): 478-8.**

Central Statstics Agency (CSA). 2022. *Report on area and production of major crops*. Addis Ababa, Ethiopia. Volume I, statistical bulletin 59.

Chandrawat, K.S., Baig, K.S., Sarang, D.H., KiihneDumai, P.H., Dhone, P.U. and Kumar, A., 2015. Association analysis for yield contributing and quality parameters in soybean. Int. J. Envion. Sci, **7(2):113-118.**

Ciampitti, I.A., de Borja Reis, A.F., Córdova, S.C., Castellano, M.J., Archontoulis, S.V., Correndo, A.A., Antunes De Almeida, L.F. and Moro Rosso, L.H., 2021. Revisiting biological nitrogen fixation dynamics in soybeans. Frontiers in Plant Science, **12(3):727021**.

Cornelious, B.K. and Sneller, C.H., 2002. Yield and molecular diversity of soybean lines derived from crosses of northern and southern elite parents. Crop science, **42(2): 642-647.**

Deshmukh, S.N., Basu, M.S. and Reddy, P.S., 1986. Genetic variability, character association and path coefficients of quantitative traits in virginia bunch varieties of groundnut. Indian Journal of Agricultural Sciences, **56(12): 816-821**.

Dong, Y.S., Zhao, L.M., Liu, B., Wang, Z.W., Jin, Z.Q. and Sun, H., 2004. The genetic diversity of cultivated soybean grown in China. Theoretical and Applied Genetics, **108(2): 931-936.**

Dong, Y.S., Zhuang, B.C., Zhao, L.M., Sun, H. and He, M.Y., 2001. The genetic diversity of annual wild soybeans grown in China. Theoretical and Applied Genetics, **103(1): 98-103.**

Ethiopian Agriculture Authority (EAA). 2023. Plant variety release, protection and seed quality control directorate. Crop variety register issue no. 24. Addis Ababa, Ethiopia

FAOSTAT. 2022. *FAO statistics database on the World Wide Web*. Available at: http://www.fao.org/faostat/en/#data/QC/visualize (Accessed 4 May 2023)

Getnet, B.E., 2018. Genetic variability, heritability and expected genetic advance as indices for selection in soybean [*Glycine max* (L.) Merrill] varieties. American Journal of Life Sciences, **6(4): 52-56.**

Gomez, K.A. and Gomez, C.M., 1984. Statistical procedures for agricultural research. 2nd Edition, John Wiley and Sons, New York, 680 p.

Govindaraj, M., Vetriventhan, M. and Srinivasan, M., 2015. Importance of genetic diversity assessment in crop plants and its recent advances: an overview of its analytical perspectives. Genetics research international, **1(1): 431487.**

Guleria, H., Kumar, P., Jyoti, B., Kumar, A., Paliwal, A. and Paliwal, A., 2019. Genetic variability and correlation analysis in soybean (*Glycine max* (L.) Merrill) genotypes. International Journal of Chemical Studies Abbreviation, **7(1): 1928-1932.**

Gupta, A., Mahajan, V., Khati, P. and Srivastva, A.K., 2010. Distinctness in Indian soybean (*Glycine max*) varieties using DUS characters. Indian Journal of Agricultural Sciences, **80(12), 1081.**

Hakim, L. and Suyamto, S., 2017. Gene action and heritability estimates of quantitative characters among lines derived from varietal crosses of soybean. Indonesian Journal of Agricultural Science, **18(1): 25-32.**

Hammer, Ø., Harper, D.A. and Ryan, P.D. 2020. Paleontological statistics (PAST) software package for education and data analysis, version 4.03. *Palaeontologia electronica*, **4(1): 9.**

Hanson, C.H., Robinson, H.F. and Comstock, R.E., 1956. Biometrical studies of yield in segregating populations of Korean lespedeza 1. Agronomy journal, **48(6): 268-272.**

Hartley, H.O., 1950. The maximum F-ratio as a short-cut test for heterogeneity of variance. Biometrika, **37(4):308-312.**

Hartman, G.L., West, E.D. and Herman, T.K., 2011. Crops that feed the World 2. Soybean—worldwide production, use, and constraints caused by pathogens and pests. Food Security, **3: 5-17.**

IAR. 1982. Progress report of field crops research for the 1981/82 cropping season. Ethiopian Institute of Agricultural Research.

IBPGR (International Board for Plant genetic Resources). 1984. Descriptor for Soybean. **84(183): 50**.

Iqbal, Z., Arshad, M., Ashraf, M., Mahmood, T. and Waheed, A., 2008. Evaluation of soybean [*Glycine max* (L.) Merrill] germplasm for some important morphological traits using multivariate analysis. Pakistan Journal of Botany, **40(6): 2323-2328.**

Jain, R.K., Joshi, A., Chaudhary, H.R., Dashora, A. and Khatik, C.L., 2018. Study on genetic variability, heritability and genetic advance in soybean (*Glycine max* (L.) Merrill). Legume Research-An International Journal, **41(4): 532-536.**

Jeffers, J.N., 1967. Two case studies in the application of principal component analysis. Journal of the Royal Statistical Society: Series C (Applied Statistics), **16(3): 225-236.**

Jha, A., Shrivastava, A.N. and Mishra, S., 2016. Principal component analysis in advanced genotypes of soybean (*Glycine max* (L.) Merrill)" during Kharif-2014. Advances in Life Sciences, **5(9): 3508-3513.**

Jing R, Vershinin A, Grzebyta J, Shaw P, Smýkal P, Marshall D, Ambrose MJ, Ellis TN, Flavell AJ. 2010. The genetic diversity and evolution of field pea (Pisum) studied by high throughput retrotransposon based insertion polymorphism (RBIP) marker analysis. BMC Evolutionary Biology. **10:1-20.**

JMP. 2018. JMP version 14 pro statistical software. SAS Institute Inc., Cary, NC, 1989–2021.

Jo, H., Lee, J.Y., Cho, H., Choi, H.J., Son, C.K., Bae, J.S., Bilyeu, K., Song, J.T. and Lee, J.D., 2021. Genetic diversity of soybeans (*Glycine max* (L.) merr.) with black seed coats and green cotyledons in Korean germplasm. Agronomy, **11(3): 581**.

Johnson, H.W., Robinson, H.F. and Comstock, R.E., 1955. Estimates of genetic and environmental variability in soybeans. Agronomy Journal, **47(7): 314–318.**

Khan, M.S.A., Karim, M.A., Haque, M.M., Karim, A.J.M.S. and Mian, M.A.K., 2014. Variations in agronomic traits of soybean genotypes. SAARC J. Agri., **12(2): 90-100.**

Khojely, D.M., Ibrahim, S.E., Sapey, E. and Han, T., 2018. History, current status, and prospects of soybean production and research in sub-Saharan Africa. The Crop Journal, **6(3): 226-235.**

Kumar, S., Kumari, V. and Kumar, V., 2020. Genetic variability and character association studies for seed yield and component characters in soybean [*Glycine max* (L.) Merrill] under North-western Himalayas. Legume Research-An International Journal, **43(4): 507-511**.

Liu, X., He, J., Wang, Y., Xing, G., Li, Y., Yang, S., Zhao, T. and Gai, J., 2020. Geographic differentiation and phylogeographic relationships among world soybean populations. The Crop Journal, **8(2): 260-272.**

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

Malek, M.A., Rafii, M.Y., Shahida Sharmin Afroz, M., Nath, U.K. and Mondal, M.M.A., 2014. Morphological characterization and assessment of genetic variability, character association, and divergence in soybean mutants. The Scientific World Journal, **1(1): 968796.**

Mapope, N. and Dakora, F.D., 2016. N2 fixation, carbon accumulation, and plant water relations in soybean (*Glycine max* (L.) Merrill) varieties sampled from farmers' fields in South Africa, measured using 15N and 13C natural abundance. Agriculture, Ecosystems & Environment, **221: 174-186.**

Marconato, M.B., Pereira, E.M., Silva, F.M., Bizari, E.H., Pinheiro, J.B., Mauro, A.O. and Unêda-Trevisoli, S.H., 2016. Genetic divergence in a soybean (*Glycine max*) diversity panel based on agro-morphological traits. Genetics and molecular research, **15: 1-10.**

Martens, H. and Naes, T. 1989. Multivariate Calibration; John Willey & Sons. *Inc.: New York*, **4(6): 441–504.**

Mesfin, H.H., 2018. Path analysis, genetic variability and correlation studies for soybean (*Glycine max* (L.) Merill) for grain yield and secondary traits at Asosa, Western Ethiopia. Greener Journal of Plant Breeding and Crop Science, **6(3): 35-46.**

Minitab. 2019. Minitab statistical software version 19.0. Minitab Inc, State College,PA. URL http://www.minitab.com/.

Naik, S.M., Madhusudan, K., Motagi, B.N. and Nadaf, H.L., 2016. Diversity in soybean (*Glycine max*) accessions based on morphological characterization and seed longevity characteristics. Progressive Research–An International Journal, **11(3): 377-381.**

Narayanan, S.S., 1993. Biometrical techniques in plant breeding. Kalyani Publishers. Kalyani publishers, Ludhiana, India: **pp-195.**

National Metrology Institute of Ethiopia (NMIE). 2020. The national metrology institute of Ethiopia's metrology data.

Neelima, G., Mehtre, S.P. and Narkhede, G.W., 2018. Genetic variability, heritability and genetic advance in soybean. Int. J. Pure App. Biosci, **6(2): 1011-1017.**

Oliveira, M.M., Sousa, L.B., Reis, M.C., Junior, E.S., Cardoso, D.B.O., Hamawaki, O.T. and Nogueira, A.P.O., 2017. Evaluation of genetic diversity among soybean (*Glycine max*) genotypes using univariate and multivariate analysis. Genet Mol Res, **16(2): 1-10.**

Rahman, M.M., Rasul, M.G., Bashar, M.K., Syed, M.A. and Islam, M.R., 2011. Parent selection for transplanted aman rice breeding by morphological, physiological and molecular diversity analysis. Libyan Agriculture Research Center Journal International, **2(1): 29-35.**

Ramteke, R., Kumar, V., Murlidharan, P. and Agarwal, D.K., 2010. Study on genetic variability and traits interrelationship among released soybean varieties of India (*Glycine max* (L.) Merrill). Electronic Journal of Plant Breeding, **1(6): 1483-1487.**

Rasyad, A., Adiwirman, D.I.R. and Rahmad, D. 2017. Variability of yield components and grain quality in several populations of soybean (*Glycine max* (L.) Merrill). p. 191-198. *In: Proceedings of PERIPI*-International Seminar, 2nd October 2017. Bogor, Indonesia

Reni, Y.P. and Rao, Y.K., 2013. Genetic variability in soybean (*Glycine max* (L) Merrill). International Journal of Plant, Animal and Environmental Sciences, **3(4): 35-38.**

Robinson, H.F., Comstock, R.E. and Harvey, P.H., 1949. Estimates of heritability and the degree of dominance in corn. Agronomy Journal, **41: 353-59.**

Sharma, B., Singh, B.V., Kamendra Singh, K.S., Pusphendra, P., Gupta, A.K. and Gupta, M.K., 2005. Genetic divergence in Indian varieties of soybean (*Glycine max* (L.) Merrill). Soybean Research, **3: 9-16**.

Shilpashree, N., Devi, S.N., Manjunathagowda, D.C., Muddappa, A., Abdelmohsen, S.A., Tamam, N., Elansary, H.O., El-Abedin, T.K.Z., Abdelbacki, A.M. and Janhavi, V., 2021. Morphological characterization, variability and diversity among vegetable soybean (*Glycine max* L.) genotypes. Plants, **10(4): 671.**

Sileshi, Y. 2019a. Estimation of variability, correlation and path analysis in soybean (*Glycine max* (L.) Merr.) Genotypes at Jimma, South Western Ethiopia. Journal of Natural Sciences Research, **9(7): 22-29**.

Sileshi, Y., Gedebo, A., Tesfaye, A. and Hailemariam, M., 2019b. Contribution of morphological traits to the total variability in soybean (*Glycine max* (L.)) Genotypes in Western Parts of Ethiopia. Acad. Res. J. Agri. Sci. Res. **7(7): 408-413.**

Singh, P.K. and Shrestha, J., 2019. Evaluation of soybean (*Glycine max* (L.) Merrill) genotypes for agro-morphological traits using multivariate analysis. Nepalese Journal of Agricultural Sciences, **18: 291- 428**

Singh, R.K. and Choudhary, B.D. 1985. Biometrical methods in quantitative genetic analysis. Kalyani Pub. Ludhiana, New Delhi, Revised Ed., **p.318.**

Tesfaye, A., Arega, A., Atero, B., Degu, T. and Hailemariam, M., 2018. Progress of soybean (*Glycine max* (L.) Merrill) breeding and genetics research in Ethiopia: a review. Ethiop. J. Crop Sci, **6(3): 129-152.**

Vianna, V.F., Unêda-Trevisoli, S.H., Desidério, J.A., Santiago, S.D., Charnai, K., Ferreira Júnior, J.A., Ferraudo, A.S. and Mauro, A.D., 2013. The multivariate approach and influence of characters in selecting superior soybean genotypes. African Journal of Agricultural Research, **8(30): 4162-4169.**

Viotto Del Conte, M., Carneiro, P.C.S., Vilela de Resende, M.D., Lopes da Silva, F. and Peternelli, L.A., 2020. Overcoming collinearity in path analysis of soybean (*Glycine max* (L.) Merr.) grain oil content. Plos one, **15(5): e0233290**.

Winsor, S., 2021. Record-Setting Soybeans: What CCAs Should Know? Crops & Soils, **54(4): 11-17.**

Yirga, M., Sileshi, Y., Tesfaye, A. and Hailemariam, M., 2022. Genetic Variability and Association of Traits in Soybean (*Glycine max* (L.) Genotypes in Ethiopia. Ethiopian Journal of Crop Science, **9(2): 49-74.**