

# ASSESSMENT OF GENETIC DIVERSITY AMONG SOYBEAN (*Glycine max* (L.) Merr.) GENOTYPES MAKING USE OF AGRO-MORPHOLOGICAL BASED ON NUTRITIONAL QUALITY TRAITS

KUJANE, K.<sup>1\*</sup> – SEDIBE, M. M.<sup>1</sup> – MOFOKENG, M. A.<sup>2</sup>

<sup>1</sup>Central University of Technology Free State, Department of Agriculture, Private Bag X 29539, Bloemfontein 9301, South Africa

<sup>2</sup>Agricultural Research Council-Grain Crops, Private Bag X 1251, Potchefstroom, South Africa

\*Corresponding author  
e-mail: Kujane2429@gmail.com

(Received 29<sup>th</sup> Mar 2021; accepted 19<sup>th</sup> Jul 2021)

**Abstract.** Present study aims to assess the genetic diversity of 30 soybean genotypes based on agronomic and quality parameters using near infrared spectroscopy. Oribi, Crawford, BD 1601, Egret, B 66 S 41, B 66 S 387, Dundee, B 66 S 37 and Santa Rosa flowered earlier than B 66 S 256, Solar 12 and B 66 S 8. Yeluanda, 15/06/2012, B 66 S 31 had the highest fibre contents and height. DB 1601 had the highest number of seeds per pod. Oribi had the highest number of branches with PR 165-52 and B 66 S 24 having the highest yields. Dundee and N69-2774 were associated with high oil content while Ex RHOD, R 5-4-2 M, N69-2774 and DB 1601 were associated with both ash and stearic acid. Crawford, Egret and B 66-387 accessions were associated with palmitic acid. B 66 S 8 had the highest oleic acid content. PR 165-52 and B 66 S 8 contained the highest amounts of oil. Multivariate analysis helped us understand important agronomic and genetic traits when selecting soybean genotypes.

**Keywords:** agronomic, multivariate analysis quality, parameters, principal component analysis, variation, traits

## Introduction

Genetic diversity based on agronomic traits is one of the oldest and most commonly used methods. It holds the advantage of providing a direct, simple, rapid and inexpensive way of characterising varieties. Phenotypic characterisation is generally viewed as the best determinant of taxonomic classification and agronomic value of crop plants and this makes it the most classical approach when coming to characterisation (Cholastova and Knotova, 2012). Phenotypic characterisation is also important for processes such as development, production and marketing of varieties. Govindarao (2010) stated that it is important to register a new cultivar that is distinct from other existing cultivars in markets where plant breeders' rights exist due to the fact that different cultivars could be identified based on phenotypic descriptors.

The phenotypic characteristics that are commonly used to assess genetic variability in soybean include plant morphology, seedling, seed quality and seed morphological characteristics. Numerous studies have explored the significance of phenotypic characterisation in estimating genetic diversity in soybean (Chen and Nelson, 2004; Ibidunni et al., 2020). A study was conducted by Hamzekhanlu et al. (2011) have studied 34 mutant lines including one control cultivar and detected the variability for number of grains per plant, number of pods per plant, number of leaves per plant, number of nodules per plant, nodule dry weight, 100 seed weight, plant dry weight (shoot dry weight), root

dry weight, harvest index and seed yield per plant. The genotypes were then clustered into four groups.

Furthermore, Manjaya and Bapat (2008) carried out a study observing the genetic variation during the characterisation of 55 soybean varieties using phenotypic traits that included: days to 50% flowering, days to maturity, plant height and number of branches per plant, number of pods per plant, number of seeds per plant, 100 seed weight and yield per plant. Antalikova et al. (2008) used 52 morphological and agronomic characters to find variability for the traits measured on the 52 studied genotypes. In another study involving phenotypic characterisation of 139 soybean genotypes Iqbal et al. (2008), revealed quite a number of significant differences between all the traits that were assessed.

The Asian Vegetable Research and Development Centre (AVRDC), United States and Pakistan studied the genetic diversity of 92 soybean genotypes and they found high coefficient of variations (CVs) coupled with wide ranges on leaf area (44.8%), number of branches per plant (31.7%), pods per plant (29.5%), 100 seed weight (39.0%) and grain yield per plant (46.6%). The findings showed a high level of diversity among the studied genotypes. Interestingly, the genotypes were classified into three distinct groups with the Pakistan germplasm forming its own cluster. Ojo et al. (2012) performed a similar study on 42 genotypes and found seven clusters. The study revealed that the 100 seed weight, number of pods per plant, pod yield per plant and seed yield per plot accounted for the greatest phenotypic variation, implying that there was broad diversity.

Phenotypic traits play a very important role in crop improvement and genetic diversity studies although they may be altered or influenced by the environment. Selection based on these traits is still widely practiced and will continue to play a significant role in estimating diversity among various genotypes using ANOVA in crop research. Results exhibiting high CVs and significant differences present high scope for selection. Furthermore, the clustering patterns obtained from phenotypic data, in respect of the number of clusters generated and genotypes contained in a cluster help to show diversity and the relatedness. The objectives of this study were to determine the presence of genetic diversity among the soybean genotypes using agro-morphological and nutritional quality traits.

## Materials and methods

### *Study site*

The experiment was conducted at the Agricultural Research Council – Grain Crops, South Africa located at 26°44'43.16"S–27°04'47.71"E with an altitude of 1340 metres above sea level. Thirty soybean genotypes that are sourced and maintained by the Agricultural Research Council were grown in under controlled conditions in a growth chamber until the 4<sup>th</sup> leaf stage. Day and night temperatures of the growth chamber were kept constant at 29 °C, and plants were irrigated to field capacity every fourth day depending on soil moisture depletion. Two seeds of each accession were planted in a 5 litre pot containing locally obtained loamy soil. This procedure was replicated three times for each genotype. *Table 1* shows the list of accessions obtained from Agricultural Research Council-Grain Crops gene bank used in the study. Soybean seed was analysed for nutritional quality using the DA 7250 (Perten Instruments) which checked for the following nutritional quality traits: moisture, ash, fibre, linoleic acid, linolenic acid, palmitic acid, stearic acid, and oleic acid.

**Table 1.** A list of accessions obtained from Agricultural Research Council-Grain Crops gene bank used in the study

No.	Origin/place of collection	Accession name	Growth habit	No.	Origin/place of collection	Accession name	Growth habit
1	Unknown (ARC-GCI)	69 S 7	Indeterminate	16	Zimbabwe	Oribi	Determinate
2	Unknown (ARC-GCI)	B 66 S 31	Indeterminate	17	Unknown (ARC-GCI)	ND 85	Determinate
3	Unknown (ARC-GCI)	Lee Ex RHOD	Indeterminate	18	USA	AGS 239	Determinate
4	China	Columbia M 8 A	Semi-determinate	19	Unknown (ARC-GCI)	61 S 156	Determinate
5	Unknown (ARC-GCI)	IBIS	Indeterminate	20	Brazil	Santa Rosa	Indeterminate
6	USA	R-5-4-2 M	Indeterminate	21	Unknown (ARC-GCI)	B 66 S 41	Determinate
7	Unknown (ARC-GCI)	Egret	Indeterminate	22	Unknown (ARC-GCI)	Egret	Determinate
8	Unknown (ARC-GCI)	15/06/2012	Indeterminate	23	Unknown (ARC-GCI)	Crawford	Determinate
9	Unknown (ARC-GCI)	Dundee	Indeterminate	24	Unknown (ARC-GCI)	B 66 S 37	Indeterminate
10	Unknown (ARC-GCI)	Solar 12	Determinate	25	Unknown (ARC-GCI)	B 66 S 387	Indeterminate
11	USA	Hawkeye	Indeterminate	26	Unknown (ARC-GCI)	B 66 S 24	Indeterminate
12	USA	N69-2774	Indeterminate	27	Unknown (ARC-GCI)	B 66 S 256	Indeterminate
13	Asia	Maksura	Determinate	28	USA	Kahala	Semi-determinate
14	USA	DB 1601	Determinate	29	Unknown (ARC-GCI)	B 66 S 8	Indeterminate
15	USA	Yeluanda	Indeterminate	30	USA	PR 165-52	Indeterminate

### **Experimental layout and management**

This study was conducted during the 2017/2018 growing season. The field has a well-drained sandy loam soil that had a pH ranging between 5.3 and 5.5. Selective pre-emergence herbicide was applied immediately after planting, subsequently; post-emergence herbicide was applied 31 days after sowing. Thirty soybean genotypes were planted in single rows of 3 m, with intra-row and inter-row spacing of 75 cm and 10 cm, respectively, using randomized complete block design, replicated three times. Plants were irrigated to field capacity once or twice a week using sprinklers depending on the soil moisture.

### **Data collection**

Data were collected on the 4<sup>th</sup> of January 2018 according to the Standard Key Descriptor Lists for Characterizations for soybean (IBPGR, 1984).

### **Agronomic attributes**

Agronomic parameters were measured according to soybean descriptor list (IBPGR, 1984). The number of days to 50 percent flowering that was calculated as days from

planting to when 50 percent of the plants in each plot have flowered, days to maturity which are days calculated as days from planting to the day when 90 percent of the pods within the plot have dried, plant height was measured from the ground surface to the tip of the growing point using meter ruler over five randomly selected plants at maturity and recorded in centimetres, number of pod per plant which are all the pods per plant harvested, counted and averaged over three plants, pod length of three randomly selected pods per line were measured using a ruler and average length per pod expressed in centimetres, number of seeds per pod are the total number of seed in each pod was counted and averaged over three pods, hundred seed weight are hundred randomly selected good seeds counted and weighed in grams using a digital scale, seed weight is the after threshing of dried pods from each net plot and the seeds were weighed using digital weighing scale and expressed in grams, nodes at flowering are number of nodes at 50 percent flowering stage and lastly, the branch number per plant which are the number of branches counted from the main stem per plant.

### *Qualitative attributes*

The following qualitative traits were measured according to soybean descriptor list (IBPGR, 1984):

<b>Descriptors</b>	<b>Scale</b>
Flower colour was recorded using the Munsell Colour Chart using the following	1 = White 2 = Yellow 3 = Red 4 = Purple
Stem types were recorded as follows;	3 = Determinate 5 = Semi-determinate 7 = Indeterminate
Leaflet shape was recorded as follows;	3 = Narrow 5 = Intermediate 7 = Broad
Pubescence was recorded	As either "present" or "absent" on the pods
Pubescence density was recorded as follows;	3 = Sparse 5 = Semi-sparse 7 = Normal 9 = Dense (stem/leaves)
Pubescence colour were recorded as follows;	1 = Grey 2 = Light brown 3 = Brown (tawny)
Pubescence type was recorded as either	1 = Erect 2 = Semi-appressed 3 = Appressed 4 = Curly 5 = Retrorse tip
Corolla colour was recorded as follows; Data was collected from 3 to 5 randomly selected plants within each row	1 = White; 2 = Purple throat; 3 = Purple

### *Data analysis*

Analysis of variance was conducted using PROC GLIMMIX, SAS version 9.4 (PROC GLIMMIX SAS Institute 2013). The Tukey's Student Range Test was used to separate means that were significantly different at P=0.05 as described by Steel and Torrie (1980).

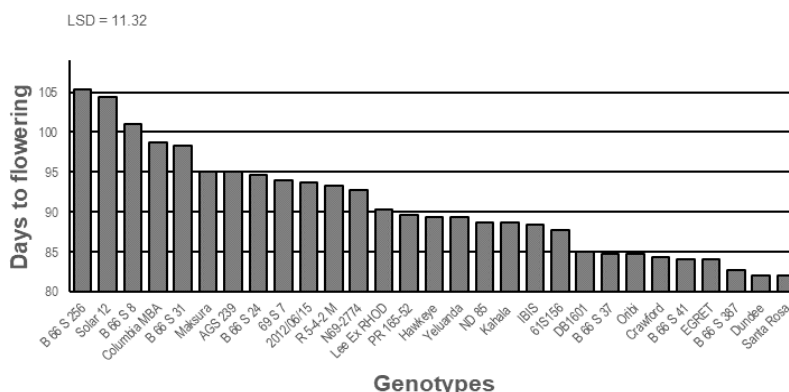
The principal component analysis (PCA) based on the linear correlation between variables and loading factors was used in multivariate analysis. Agromorphological data were subjected to multivariate data analysis using principal component analysis (PCA – XLSTAT, 2015) to identify and evaluate the groupings between the variables following the description.

## Results

### Agro-morphological diversity

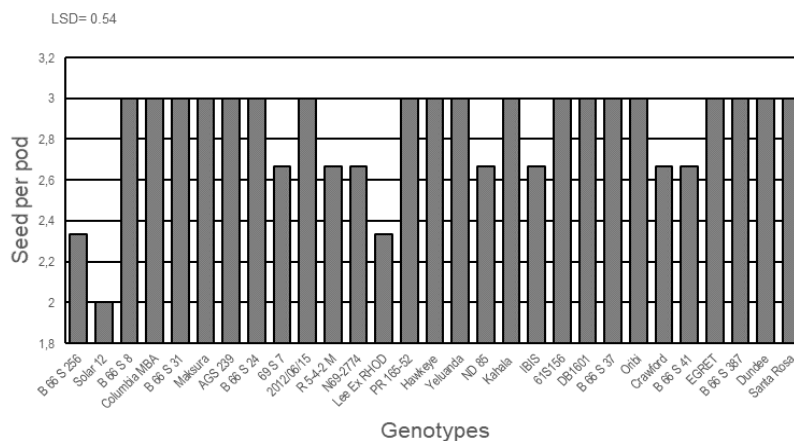
#### Quantitative data

The early flowering genotypes, BD 1601, B 66 S 37, Oribi, Crawford, B 66 S 41, Egret, B 66 S 387, Dundee and Santa Rosa, took less than 85 days to flower. In contrast, B 66 S 256, Solar 12 and B 66 S 8 took more than 100 days to flower. Lee Ex RHOD, PR 165-52, Hawkeye, Yeluanda, ND 85, Kahala, IBIS and 61 S 156 genotypes were intermediate and took 86 to 90 days to flower (*Figure 1*).



**Figure 1.** Number of days to 50% flowering of 30 genotypes of soybean

About 63% of the evaluated genotypes had three seeds per pod, only Solar 12 had only two seeds per pod (*Figure 2*) and the remaining genotypes had less than two seeds per pod.



**Figure 2.** Seed number per pod of 30 soybean genotypes evaluated

### Qualitative data

The qualitative data of the thirty soybean genotypes grown in 2017/18 season is presented in *Table 2*. All genotypes studied matured with erect pubescence type. The studied soybean genotypes were dominated by 86.7% of dense and 13.3% of normal pubescence density. Most of the genotypes had 60% of the grey and 20% of brown and light brown pubescent colours. The corolla colours varied between purple and white with 73% of the genotypes being purple and the remainder were white. The genotypes showed different leaf shapes including narrow, intermediate and broad leaves with 46.7% of the genotypes having intermediate leaves, 43.3% had narrow leaves whereas 10% of the genotypes had broad leaves.

**Table 2.** Qualitative traits of 30 soybean genotypes recorded in the study

Genotype	Maturity period	Flower colour	Stem determination	Pubescence	Pubescence type	Pubescence density	Pubescence colour	Corolla colour	Leaflet shape
<b>69 S 7</b>	Early	Purple	Indeterminate	Absent	Erect	Dense	Grey	Purple	Narrow
<b>B 66 S 31</b>	Early	White	Indeterminate	Absent	Erect	Dense	Grey	Purple	Intermediate
<b>Lee Ex RHOD</b>	Early	White	Indeterminate	Absent	Erect	Dense	Brown	White	Narrow
<b>Columbia M8A</b>	Early	Purple	Semi-determinate	Absent	Erect	Dense	Brown	Purple	Broad
<b>IBIS</b>	Early	Purple	Indeterminate	Absent	Erect	Normal	Grey	Purple	Narrow
<b>R 5-4-2 M</b>	Early	White	Indeterminate	Absent	Erect	Dense	Grey	White	Intermediate
<b>Egret</b>	Early	Purple	Indeterminate	Absent	Erect	Dense	Grey	Purple	Narrow
<b>15/06/2012</b>	Early	Purple	Indeterminate	Absent	Erect	Dense	Grey	Purple	Intermediate
<b>DUNDEE</b>	Early	Purple	Indeterminate	Absent	Erect	Dense	Brown	Purple	Narrow
<b>Solar 12</b>	Early	Purple	Determinate	Absent	Erect	Dense	Grey	Purple	Intermediate
<b>Hawkeye</b>	Early	White	Indeterminate	Absent	Erect	Normal	Light brown	Purple	Narrow
<b>N69-2774</b>	Early	White	Indeterminate	Absent	Erect	Normal	Grey	White	Narrow
<b>Maksura</b>	Early	White	Determinate	Absent	Erect	Dense	Grey	Purple	Narrow
<b>DB 1601</b>	Early	Purple	Determinate	Absent	Erect	Dense	Grey	Purple	Broad
<b>Yeluanda</b>	Early	Purple	Indeterminate	Absent	Erect	Dense	Grey	Purple	Intermediate
<b>Oribi</b>	Early	White	Determinate	Absent	Erect	Dense	Grey	White	Intermediate
<b>ND 85</b>	Early	White	Determinate	Absent	Erect	Normal	Light brown	White	Narrow
<b>AGS 239</b>	Early	Purple	Determinate	Absent	Erect	Dense	Brown	Purple	Intermediate
<b>61 S 156</b>	Early	White	Determinate	Absent	Erect	Dense	Grey	Purple	Intermediate
<b>Santa Rosa</b>	Early	Purple	Indeterminate	Absent	Erect	Dense	Light brown	Purple	Narrow
<b>B 66 S 41</b>	Early	Purple	Determinate	Absent	Erect	Dense	Brown	Purple	Narrow
<b>Egret</b>	Early	Purple	Determinate	Absent	Erect	Dense	Grey	Purple	Intermediate
<b>Crawford</b>	Early	Purple	Determinate	Absent	Erect	Dense	Light brown	Purple	Intermediate
<b>B 66 S 37</b>	Early	Purple	Indeterminate	Absent	Erect	Dense	Grey	White	Intermediate
<b>B 66 S 387</b>	Early	White, Purple	Indeterminate	Absent	Erect	Dense	Brown	White	Narrow
<b>B 66 S 24</b>	Early	White	Indeterminate	Absent	Erect	Dense	Grey	White	Intermediate
<b>B 66 S 256</b>	Early	Purple	Indeterminate	Absent	Erect	Dense	Grey	Purple	Intermediate
<b>Kahala</b>	Early	White	Semi-determinate	Absent	Erect	Dense	Light brown	Purple	Intermediate
<b>B 66 S 8</b>	Early	Purple	Indeterminate	Absent	Erect	Dense	Grey	White	Narrow
<b>PR 165-52</b>	Early	White	Indeterminate	Absent	Erect	Dense	Light brown	Purple	Narrow

### Diversity of nutritional quality traits

The nutritional quality traits of soybean genotypes were analysed using analysis of variance. Significant differences ( $P \leq 0.05$ ) were observed among the soybean genotypes based on ash content and oleic acid. Ash content varied between 4.8 and 6.6% whereas oleic acid ranged between 15.4 and 27.8% (Figure 3). The genotypes AGS 239 and Yeluanda had relatively higher ash content. Higher oleic acid contents were recorded for genotypes B 66 S 387, followed by B 66 S 8 and Hawkeye and the lowest was genotype DB 1606 (Figure 4).

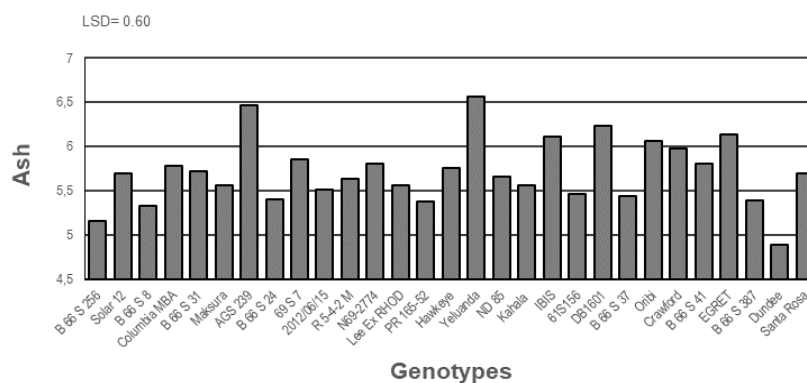


Figure 3. Ash content percentage of 30 genotypes of soybean evaluated

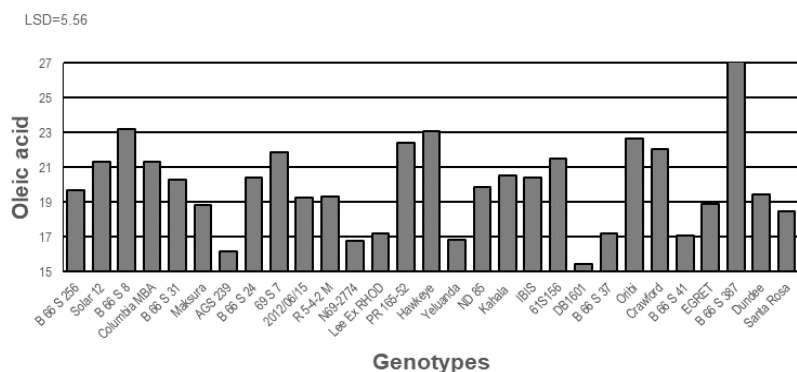


Figure 4. Oleic acid content of 30 genotypes of soybean evaluated

### Principal component analysis

Principal component analysis (PCA) of agro-morphological and quality traits of soybean. Table 3 shows the agronomic data that were subjected to a principal component analysis, which revealed that three most important PCs contributed 21.3%, 14.9% and 9.1% to the total variation. The traits that contributed the most variation in the first PC were number of branches per plant, pod number per plant, pod weight before threshing and yield per plant for agro-morphological traits. In the second PC, the traits that were responsible for contributing the most variation were oil, oleic acid, protein, ash, linoleic acid and stearic acid. The largest contributors to variation in the third PC were moisture content, protein content, days to 50% flowering and plant height.

**Table 3.** Factor loadings of the three PCs based on agronomic and quality traits

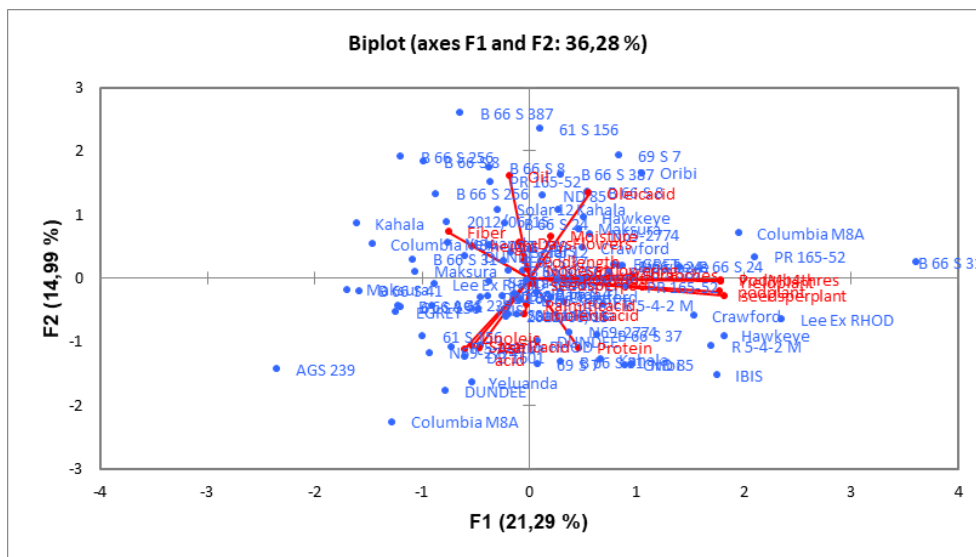
Traits	Factor loadings		
	F1	F2	F3
DFW	-0.048	0.288	-0.499
PHT	-0.278	0.264	-0.450
BNP	0.505	0.023	-0.011
PNP	0.932	-0.114	0.055
PBT	0.941	-0.016	0.045
SNP	0.017	-0.049	0.119
SDP	0.953	-0.147	0.019
YDP	0.937	-0.029	0.010
PDL	-0.022	0.140	-0.066
PDW	0.175	-0.006	-0.390
Protein	0.243	-0.578	-0.537
Oil	-0.096	0.841	0.271
Moisture	0.108	0.346	0.711
Ash	-0.245	-0.580	0.202
Fiber	-0.393	0.379	-0.276
Linoleic acid	-0.318	-0.588	0.234
Linolenic acid	-0.023	-0.300	-0.348
Oleic acid	0.293	0.701	-0.224
Palmitic acid	-0.014	-0.219	0.056
Stearic acid	-0.281	-0.562	0.159
NFW	-0.003	0.066	0.228
<b>Eigenvalue</b>	4.470	3.148	1.920
<b>Variability (%)</b>	<b>21.287</b>	<b>14.992</b>	<b>9.142</b>
<b>Cumulative (%)</b>	21.287	36.279	45.422

Days to 50% flowering=DFW; Plant height=PHT; Number of branches/plant=BNP; Pod number/plant=PNP; Pod weight before threshing=PBT; Seed number per/pod=SNP; Seed number/plant=SDP; Yield/plant (g) =YDP; Pod length (mm) =PDL; Pod width (mm) =PDW; Nodes at flowering=NFW

### Principal component biplots

Figure 5 shows that number of branches per plant and pod weight before threshing were highly significant and highly correlated with plant height, days to 50% flowering, nodes at flowering, pod length as well as oil, fiber, oleic acid and moisture; whereas they were negatively correlated with pod number per plant, seed per plant, ash, linoleic acid, linolenic acid, palmitic acid, stearic acid and ash. However, was also positively and significantly associated with pod width. Protein was highly significant and positively correlated with palmitic acid, linolenic acid, linoleic acid as well as stearic acid but negatively correlated fiber. It was also significant and positively correlated with seed number per pod. Seed number per pod was highly significant and negatively associated with pod weight per plant, days to 50% flowering, but positively correlated with pod number per plant. Seed number per pod was highly and positively associated with seed number per plant and protein. Pod number per plant was significant and positively correlated with protein. Pod number per plant before threshing was highly and significantly associated with pod width.





**Figure 5.** A Principal Component Biplot depicting agro-morphological and nutritional quality traits of 30 soybean accessions

#### Correlation analysis among phenotypic and quality traits

The phenotypic and quality traits were analysed using correlation coefficients. The association of the traits were reported based on the 5% significance level. Plant height was positively correlated with days to flowering. Pod number per plant was positively correlated with plant height and branch number per plant. Pod weight before threshing recorded a positive correlation with plant height, branch number per plant and pod number per plant. Seed number per pod was positively associated with plant height, branch number per plant and pod number per plant. Seed yield per plant was positively associated with plant height, branch number per plant, pod number per plant and seed number per pod. Pod width recorded a negative correlation with seed number per pod. Number of nodes at flowering was negatively correlated with palmitic acid. Oil content recorded a positive correlation with days to flowering and a negative one with protein. Moisture content showed a negative association with protein content and a positive association with oil content. Ash was positively correlated with plant height, seed yield per plant and oil content, and negatively associated with pod weight before threshing. Fiber recorded a positive correlation with plant height, number of branches per plant, seed number per plant, seed yield per plant and oil content and also showed a negative correlation with pod number per plant and pod weight before threshing. Linoleic acid was negatively correlated with days to flowering and positively correlated with oil and ash. Linoleic showed a negative association with moisture content and positive association with linoleic acid. Oleic acid correlated positively with seed yield per plant, oil content and ash and negatively with linoleic acid. Palmitic acid correlated positively with ash and fibre and negative with linoleic and oleic acids. Stearic acid showed a positive correlation with oil content and linoleic acid and a negative correlation with oleic acid. On leaf shape 43.33% were narrow, 46.67 intermediates and 10% were indeterminates. Pubescence density varied at 86.67% for the erect types and 13.33% for the appressed types. Sixty percent of the stems were determinate, 6.67% semi- determinate and 33.33% indeterminate.

### *Relationship between the traits*

The majority of soybean cultivars that are grown and consumed throughout the world today exhibit yellow or white seed coats, whereas the majority of known accessions of the wild progenitor, *G. soja*, have black or, rarely, brown seed coats. Soybean cultivars should also have chemical composition and qualitative traits that are in need and particularly easy to grow and maintain especially for beneficial aspects either for personal consumption or financial benefits.

The principal component biplot analysis was not only able to identify specific traits contributing to yield but also to identify the out-performing varieties, as this helps to select new lines with these traits. Soybean varieties with higher plant height, more oil content, more branches and pods, grains, fibre, and early flowering days obtained higher yield. Dundee and N69-2774 were associated with high oil content. According to the PCA biplots Lee Ex RHOD, R 5-4-2 M, N69-2774 and DB 1601 were associated with both ash and stearic acid as illustrated on *Figure 5*. Crawford, Egret and B 66 S 387 accessions were associated with palmitic acid. Yeluanda, 15/06/2012, B 66 S 31 had the highest fibre and height. Maksura, N69-2774 and Crawford were associated with both moisture and days to flowering. B66 S8 had the highest oleic acid content. DB 1601 had the highest number of seeds per pod. PR 165-52 and B 66 S 8 contained the largest oil content. High node set at flowering was observed in 15/06/2012 while high pods per plant were observed in Crawford and Lee Ex RHOD. B 66 S 387 and Oribi had the highest number of branches while PR 165-52 and B 66 S 24 had the highest yields per plant.

### **Discussion**

Genetic diversity of germplasm collections can be determined using a traditional cluster analysis; which is easy and effective. This type of method studies the genetic diversity by forming core subsets while grouping accessions according to similar characteristics into one homogenous category. Relationships of accessions are grouped according to similar units and thus make it easy to understand them better and easily translatable. The evaluation of genetic diversity among germplasm is highly required in any hybridization program as it aids in improving the genes of the most diverse lines as well as promoting the use of genetic variations. It is imperative for a breeder to obtain information on genetic diversity and relationships among breeding materials in order to improve the efficiency of a crop, this helps with efficiently managing and conserving germplasm resources. De Chavez et al. (2017) recently studied the diversity among soybean accessions in Philippines reporting a vast diversity. Another study conducted by Khatab et al. (2012) reported the presence of genetic diversity among soybean genotypes using ago-morphological descriptors. Hamzekhanlu et al. (2011) studied 34 mutant lines including one control cultivar and found variability for ten quantitative traits in soybean. A study finding significant differences among all the assessed phenotypic traits was reported by Iqbal et al. (2008).

The assessment of morphological traits with the aim of studying the genetic diversity and classification of existing germplasm material is a traditional method that is low level but can also be a powerful taxonomic tool that can be used to group germplasm prior to their characterisation particularly using more precise marker technologies. Li et al. (2020) observed soybean varieties with higher plant height, more nodes of main stem, branches, pods, grains, and 100-grain weight, or longer growth periods may have higher yield. This concurs with Achina et al. (2019) who observed that the significant attributes include the

number of pods per plant, seeds per pod, and seed weight, which determine the seed yield. From the research, DB 1601 had the highest number of seeds per pod while PR 165-52 and B 66 S 24 had the highest yields per plant.

In this study, the nutritional quality traits showed diversity among the studied accessions where the oil and protein contents were within the range recorded in other studies (Shi et al., 2010). The oil content showed a range from 13.16% to 15.41%, while protein ranged between 33.29% and 35.71% with Dundee and N69-2774 having the highest high oil content. In order to select accessions with good qualities for hybridization and conservation, breeders will find the diversity among the accessions very beneficial. In a study conducted by Mazid et al. (2013), it was highlighted that the analysis of genetic diversity is very crucial for identifying parents that helps to achieve long-term selection gain.

The study revealed that traits that contributed to the most variation were number of branches per plant, pod number per plant, pod weight before threshing, seed number per plant, and yield per plant, protein and oleic acid. High number of pods per plant was observed in Crawford and Lee Ex RHOD. High oleic acid content was recorded for genotype B 66 S 387, followed by B 66 S 8 and Hawkeye. All genotypes studied matured early with erect pubescence type. The studied soybean genotypes were dominated by 86.7% of dense and 13.3% of normal pubescence density. Most of the genotypes had 60% of the grey and 20% of brown and light brown pubescent colours. The corolla colours varied between purple and white with 73% of the genotypes being purple and the remainder were white. Breeders in the direct and indirect programmes can be provided with information of the traits that were tested for correlation and traits that are desirable in order to select and use these genotypes with significance concurrently. Yield improvement programs thrive better when a breeder understands the relationship between the component of traits and yield as this will aid in making the best selection of desirable genotypes for these programs.

Plant height, days to flowering, plant branches, pods weight before threshing, and pod length concur with the results reported by a study carried out by Malek et al. (2014) where there were positive significant correlations. Machikowa et al. (2007) argue that early maturing varieties tend to have lower yield while late flowering is associated with higher soybean yield. Therefore, the extension of days to flowering of current early varieties, as well as the development of new varieties with a longer vegetative period, may result in higher yielding varieties. These results mean that when selecting genotypes that are high yielding for a breeding programme, these characters should be given more preference and emphasis as the best selection criteria. Kato et al. (2019) observed that the seed yields of semi-determinate and indeterminate lines were higher than that of the determinate ones; and that of the semi-determinate lines was marginally lower than that of the indeterminate lines. The research also highlighted that lodging score of semi-determinate varieties was smaller than that of indeterminate varieties because the main stem length of the semi-determinate varieties was shorter than that of the indeterminate varieties. Anand and Torrie (1963) stated that seed yield always showed a positive correlation with other desirable yield traits which indicates that the increase in one trait would result in the increase of the other; that is, simultaneous increase or decrease of both traits would be easy. Soybean genotypes showing high yields indicate a strong positive correlation between seed yield and other traits and would be fairly easy to identify together with high number of pods per plant.

The statistical method that is commonly used in populating genetics in order to identify the genetic distribution and structure across a geographical and ethnic is also known as the principal component analysis (PCA) (McVean, 2009). It is further explained that the reason for this type of analysis is to evaluate each and every variable that forms part of the variation that is available among the studied genotypes (Mofokeng and Mashingaidze, 2018). This type of method aids in differentiating the important and less important traits within a studied group by grouping them accordingly.

## Conclusion

Using agro-morphological and quality traits, some of the most important soybean cultivars were observed. These included Dundee and N69-2774 having the highest high oleic acid and oil contents. DB 1601 had the highest number of seeds per pod while PR 165-52 and B 66 S 24 had the highest yields. There were positive correlations showed between most of the assessed traits in relation to one another and this will be very helpful in assisting in the combined improvement of these traits by selecting ones that were found to have a positive and high correlation, as well as easily measurable phenotypic traits, although most were found to have highly significant and positive correlation with seed yield per plant. The nutritional quality traits also varied significantly among the accessions. In order to achieve a successful selection of parents for breeding and transgressive segregation, studying and knowing the presence of genetic diversity can be very useful. The aim of the present investigation was to characterise the genetic diversity of the available germplasm, determine gene action controlling grain yield and estimate the breeding gains that have been realized since the inception of the breeding programmes. The specific objectives of the study were successfully accomplished as shown above in the study.

Soybean is one of the most important leguminous crops grown globally. Understanding the genetic diversity and its interaction with the environment is of paramount importance in developing cultivars considering farmer's preferred traits. The multivariate analysis was able to reveal the relationship between the genetic diversity, agronomic and nutritional composition of selected soybean genotypes. For a successful breeding programme to function a complete understanding of the genetic diversity of the crop is required. Better knowledge of the genetic similarities and dissimilarities of breeding material could aid breeders and curators in maintaining genetic diversity, sustain long-term selection gain and conserve the germplasm. Monitoring the genetic diversity among genetic resources of elite breeding material could make crop improvement more efficient by the directed accumulation of favoured alleles thus reducing the amount of material to be screened.

**Acknowledgements.** This project was supported by the Central University of Technology, Free State Research Grant Scheme, National Research Foundation fund, and Agricultural Research Council- Grain Crops for the use of their facilities.

## REFERENCES

- [1] Achina, T., Twumasi, P., Quain, M. D. (2019): Use of morphological and molecular profiling for the establishment of relatedness between commercially viable and elite soybean. – European Journal of Biotechnology and Genetic Engineering 6: 1-17.

- [2] Anand, S. C., Torrie, J. H. (1963): Heritability of yield and other traits and interrelationship among traits in the F3 and F4 generations of three soybean crosses. – *Crop Science* 3: 508-511.
- [3] Antalíkova, G., Zakova, M., Benedíkova, D. (2008): Characterisation of soybean traits variability by cluster analysis. – *Agriculture* 54: 45-53.
- [4] Chen, Y., Nelson, R. L. (2004): Genetic variation and relationships among cultivated, wild and semiwild soybean. – *Crop Science* 44: 316-325.
- [5] Cholastova, T., Knotova, D. (2012): Using morphological and microsatellite (SSR) markers to assess the genetic diversity in Alfalfa (*Medicago sativa* L.). – *World Academy of Science, Engineering and Technology* 69: 1-7.
- [6] De Chavez, H., Borromeo, T., Oswald, B. N., Gentallan, R. (2017): Phenotypic diversity of soybean (*Glycine max* (L.) Merr.) accessions in the Philippines for utilization. – *Legume Research-An International Journal* 40: 9-15.
- [7] Govindarao, C. N. (2010): Characterisation of [*Glycine max* (L.) Merrill.] varieties through morphological, chemical, molecular markers and image analyser. – Dharwad University of Agricultural Sciences, 115p.
- [8] Hamzekhanlu, M. Y., Darbandi, A. I., Beiranvand, P., Hallajian, M. T., Majdabadi, A. (2011): Phenotypic and molecular analysis of M7 generation of soybean mutant lines through random amplified polymorphic DNA (RAPD) marker and some morphological traits. – *African Journal of Agricultural Research* 6: 1779-1785.
- [9] Ibidunni, S. A., Omolayo, J. A., Olabisi, A., Sunday, O. 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. doi: 10.5923/j.plant.20201002.02.
- [10] International Board for Plant Genetic Resources. (1984): Descriptors for soybean. – IBPGR Secretariat, Rome. AGPG: IBPGR/84/183.
- [11] Iqbal, Z., Arshad, M., Ashraf, M., Mahmood, T., Waheed, A. (2008): Evaluation of soybean [*Glycine max* (L.) Merrill] germplasm for some important morphological traits using multivariate analysis. – *Pakistan Journal of Botany* 40: 2323-2328.
- [12] Kato, S., Sayama, T., Taguchi-Shiobara, F., Kikuchi, A., Ishimoto, M., Cober, M. (2019): Effect of change from a determinate to a semi-determinate growth habit on the yield and lodging resistance of soybeans in the northeast region of Japan. – *Breeding Science* 69: 151-159. doi:10.1270/jsbbs.18112.
- [13] Khatab, I. A., Morsy, A. R., Fares, W. M., Kumamaru, T. (2016): Genetic diversity of soybean genotypes revealed by agro-morphological and SSR markers. – *Journal of the Faculty of Agriculture, Kyushu University* 61: 79-87.
- [14] Li, M., Liu, Y., Jiang, W., Wang, C., Yang, X., Li, D., Zhang, X., Xu, C., Zhang, Y., Li, W., Zhao, L. (2020): Identification of traits contributing to high and stable yields in different soybean varieties across three different Chinese latitudes. – *Frontiers in Plant Science* 10: 1642. <https://doi.org/10.3389/fpls.2019.01642>.
- [15] Machikowa, T., Waranyuwat, A., Burton, J. W., Laosuwan, P. (2007): Yield improvement of early maturing soybeans by selection for late flowering and early maturity. – *Science Asia* 33: 229-234.
- [16] Malek, M. A., Raffi, M. Y., Afroz, M. S. S., Nath, U. K., Mondol, M. M. A. (2014): Morphological characterization and assessment of genetic variability, character association, and divergence in soybean mutants. – *Scientific World Journal* 2014: 968796.
- [17] Manjaya, J. G., Bapat, V. A. (2008): Studies on genetic divergence in soybean, [*Glycine max* (L.) Merrill]. – *Journal of Oilseeds Research* 25: 178-180.
- [18] Mazid, M. S., Rafii, M. Y., Hanafi, M. M., Rahim, H. A., Shabanimofrad, M., Latif, M. A. (2013): Agro-morphological characterization and assessment of variability, heritability, genetic advance and divergence in bacterial blight resistant rice genotypes. – *South African Journal of Botany* 86: 15-22.
- [19] McVean, G. (2009): A genealogical interpretation of principal components analysis. – *PLoS Genetics* 5(10): e1000686. <https://doi.org/10.1371/journal.pgen.1000686>.

- [20] Mofokeng, M. A., Mashingaidze, K. (2018): Genetic Differentiation of ARC Soybean [*Glycine Max* (L.) Merrill] accessions based on agronomic and nutritional quality traits. – *Agriculture and Food Sciences Research* 5(1): 6-22.
- [21] Ojo, D. K., Ajayi, A. O., Oduwaye, O. A. (2012): Genetic Relationships among soybean accessions based on morphological and RAPDs techniques. – *Pertanika Journal of Tropical Agricultural Science* 35: 237-248.
- [22] Shi, A., Chen, P., Zhang, B., Hou, A. (2010): Genetic diversity and association analysis of protein and oil content in food grade soybeans from Asia and the United States. – *Plant Breeding* 129: 250-256.
- [23] Steel, R. G. D., Torrie, J. H. (1980): *Principles and Procedures of Statistics*. – 2<sup>nd</sup> ed. New York: McGraw-Hill.