Environment Conservation Journal 24 (3): 79-86, 2023

Journal homepage: https://www.environcj.in/

Environment Conservation Journal

ISSN 0972-3099 (Print) 2278-5124 (Online)

Principal component analysis among vegetable soybean genotypes (*Glycine max* L. Merrill)

Devi Sri Dunna ✉

Department of Genetics and Plant Breeding, College of Agriculture, Central Agricultural University, Imphal, Manipur, India

Nanita Devi Heisnam

Department of Genetics and Plant Breeding, College of Agriculture, Central Agricultural University, Imphal, Manipur, India

Renuka Devi Thokchom

Department of Genetics and Plant Breeding, College of Agriculture, Central Agricultural University, Imphal, Manipur, India

Bireswar Sinha

Department of Plant Pathology, College of Agriculture, Central Agricultural University, Imphal, Manipur, India

Okendro Singh

Department of Basic Sciences, College of Agriculture, Central Agricultural University, Imphal, Manipur, India

ARTICLE INFO

Received : 18 August 2022
Revised : 29 December 2022
Accepted : 22 January 2023

Available online: 11 April 2023

Key Words:
Eigen values
Genotypes
Miracle crop
Screeplot

ABSTRACT

In the present study, 33 soybean genotypes were selected to study the association between the fourteen quantitative characters under study and to assess the magnitude of divergence between 33 genotypes for 14 traits during *Kharif,* 2021. For all measured traits, the results revealed significant variation among tested entries. The principal component analysis (PCA) was conducted for the 1quantitative characters and only 7 showed > 1 Eigen value and showed about 81.748% of total variation and remaining 7 components having a eigen value less than 1.00 and contributed only 18.252% of total Variation. Among the traits studied PC1 showed 20.7235 while, PC2 to PC 14 exhibited 15.205%, 12.69%, 9.476%, 9.142%,7.271%, 7.241%, 6.171%, 4.089%, 2.865%, 2.185%, 1.506%, 0.897% and 0.536% variability, respectively. Scree plots explained the percentage of variance. A high PC score for a specific genotype shows high value for those variables like EC 915989, EC 915900, EC 915993, EC 915975, EC 915903, EC 915898 and EC 915959 in PC1 and this indicated that these genotypes have high values for traits such as 100 Fresh pod weight and 100 Fresh seed weight.

Introduction

Protein malnutrition is common among the poor in India, and it is expected to worsen as the population grows by 2025. There is a need for a versatile crop that can alleviate protein malnutrition while also being suitable for the Indian cropping system. Glycine max (vegetable soybean) (L.) Merill is known in Japan as edamame and in China as maodou. Edamane is a large-seeded vegetable soybean with a sweet flavour, bright green colour, light hilum, and a soft texture. While the pods and plants are still green, the pods are harvested when they are completely seeded. Samuel Bowen, a sailor, introduced soybean to the United States in 1765, and it quickly became a popular crop in the south by the twentieth century (Brachfeld and Choate, 2007). Farmers in Asia, where edamame is a popular vegetable, is harvested at the full-seed stage, with a larger, sweet, nutty, and mildly flavoured seed (Zhang *et al.,* 2015). Soybean was consumed as a vegetable in the Far East as early as the second century BCE because of its nutritional and medicinal properties (Shanmugasundaram, 2001; Shurtleff and Aoyagi, 2009). The current challenge in Edamame breeding is to develop improved cultivars that outperform in terms of vegetable pod yield. At present varieties that are bred for vegetable purpose are limited in India. In any crop improvement effort, the first step is the collection of germplasm and the evaluation for magnitude variability of economic traits of interest.

Corresponding author E-mail: *dunna.devisri@gmail.com*
Doi:https://doi.org/10.36953/ECJ.14532442

To overcome the scarcity of resources I used the IISR's elite vegetable soybean varieties with the assistance of AICRP, soybean. Soybean, as a promising vegetable legume crop in India, has a lot of room for genetic improvement in terms of different quantitative and qualitative traits besides widening the genetic base of the crop through mutagenesis. Vegetable Soybean is one of the most nutritious leguminous crop. It is also known as the "miracle crop" because of its exceptional qualities. The magnitude of the association between yield and other components is very crucial.

Material and Methods

R studio with R software version 4.2.1 was used to perform the analysis in which different packages like princomp, promp, FactoMineR were used in the process. First the data structure was prepared in excel and later the data was imported , scaled and adjusted, The scaled data then using princomp and prcomp packages PCA structures was concluded and 3D plots were prepared, later on using FactoMineR rotated component were prepared and the conclusions were drawn. In the present study, 33 soybean genotypes were selected to study the association between the fourteen quantitative characters under study and to assess the magnitude of divergence between 33 genotypes for 14 traits during *Kharif,* 2021. Catell suggests using a stree plot (PCA) to calculate the eigenvalue. Jeffers proposed the selection of PCs with eigen values >1. R utilised to perform PCA by equation given here under. PCA is a well-known dimension reduction method that seeks linear combinations of X columns with the lowest variance, or equivalently, the highest information (Massy, 1965; Jolliffe, 1986). It is frequently used in chemo metrics to provide the most concise data representation. The singular value decomposition of X is one of several standard approaches for determining the principal components. In chemometrics, the nonlinear iterative partial least-squares (NIPALS) algorithm is commonly used to estimate the principal components (Wold, 1966). This experiment was carried out at Andro research farm, CAU, Imphal East during kharif 2021. Geographically the CAU Research Farm, Andro is situated in the Imphal East district of Manipur at the 24° 76.42' North latitude and 94° 05.19' East longitudes at an altitude of 808m above Mean Sea Level which comes under the Eastern Himalayan Region (II) and the agro-climatic zone Sub-Tropical Zone (NEH-4) of Manipur (Agro meteorological observatory, ICAR-RC NEH Region, Manipur Centre, Lamphelphet, Imphal). The soil type of the experimental field was clay in texture with acidic soil reaction having pH value ranging from 5.5 to 5.6. It has high available potassium but low available nitrogen and phosphorous.

Results and Discussion

Principle Component Analysis (PCA) is a well-known data analysis technique that is the most commonly used multivariate technique for reducing the dimensionality of a data set. According to Rotaru *et al*. (2012), using the PCA method in agricultural studies is beneficial. The present investigation was performed for 14 traits of vegetable soybean. Out of 14 principal components in which 7 components showed > 1.00 Eigen value and showed about 81.748% of total variation. However, remaining 7 components having a eigen value less than 1.00 and contributed only 18.252% of total Variation. Among the traits studied PC1 showed 20.7235 while, PC2 to PC 14 exhibited 15.205%, 12.69%, 9.476%, 9.142%,7.271%, 7.241%, 6.171%, 4.089%, 2.865%, 2.185%, 1.506%, 0.897% and 0.536% variability, respectively. Among Seven principal components first four components showed major contribution (58.093%) towards total variation. Eigen value and variance associated with each principal component decreased gradually and cumulative variability increase gradually which is shown as screen plot. The result or principal components are depicted as a factor loading score. A more or less same trend was observed by Srikanth Thippani *et al*. (2017), Jakhar and Arjun (2018). These results are presented in the Table 1 and depicted for Variables in Fig. 1, Individuals in Fig.2 and Biplot in Fig.3.Scree plots are created by graphing the principal component numbers (X-axis) and the percentage of variation explained (Y-axis).The Principal Component 1 showed 20.7 percent variability with eigen value 2.901 which then declined gradually. From the graph, it is clear that the maximum variation was observed in Principal Component 1(Fig. 4). In PC1 the traits that had high positive component were 100 fresh

Table 1: Principal components (PCs) analysis for metric traits in soybean genotypes

Characters	Principal component	Eigen values	percentage of variance	cumulative % of variance
Days to 50% flowering	PC1	**2.901**	20.723	20.723
Plant height(cm)	PC2	**2.129**	15.205	35.928
Days to Harvest	PC3	**1.777**	12.69	48.617
Pod length(cm)	PC4	**1.327**	9.476	58.093
No. of seeds/pod	PC5	**1.28**	9.142	67.236
No. of pods/plant	PC6	**1.018**	7.271	74.506
Fresh pod yield/plant(g)	PC7	**1.014**	7.242	81.748
100 Fresh pod weight(g)	PC8	0.864	6.172	87.92
100 Fresh seed weight(g)	PC9	0.572	4.089	92.009
100 Dry seed weight(g)	PC10	0.401	2.866	94.875
Crude sugar content	PC11	0.306	2.185	97.06
Crude protein content	PC12	0.211	1.506	98.566
Hydration rate	PC13	0.126	0.897	99.463
Cooking time(min)	PC14	0.075	0.537	100

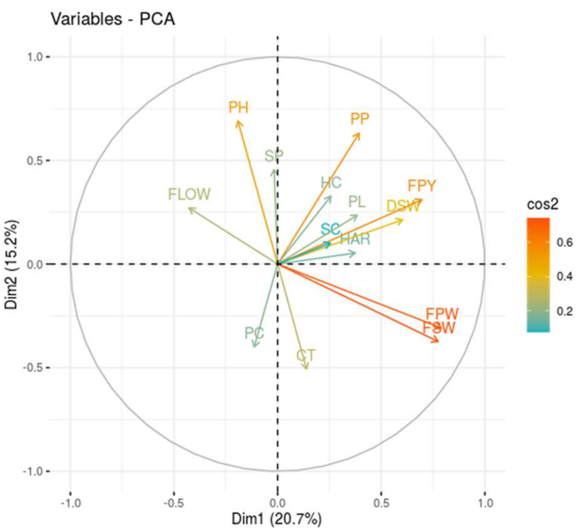

Figure 1: PCA graph of different traits for first two principal components

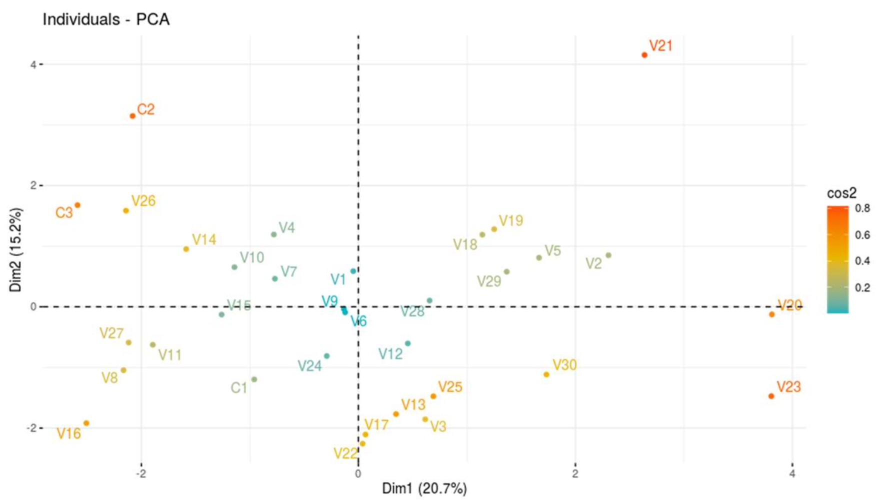

Figure 2: PCA graph for germplasm lines across first two principal components

seed weight (0.952) followed by 100 fresh pod weight(0.920), that accounted that these traits were mostly related to PC1. In PC2 the traits that had high positive component were No. of pods/plant(0.933) and Fresh pod yield/plant (0.865). In PC3 that had high positive component were Days to 50% flowering (0.730) and plant height (0.548). In PC4 that had high positive component were 100 Dry seed weight(0.614) and Hydration rate(0.883). In PC5 that had high positive component were No. of seeds/plant (0.890) and days to Harvest(0.590) . In PC6 that had high positive component was Crude sugar content(0.904), In PC7 that had high positive component was pod length (0.849). The rotated component matrix components were presented in Table 2 and the traits having eigen values for each PC were represented in Table 3 and Figure 5.

Table 2: Principal Components for yield and its contributing characters of Vegetable soybean

	RC1	RC2	RC3	RC4	RC5	RC6	RC7
Days to 50% flowering	-0.071	0.016	**0.730**	-0.373	0.144	-0.179	-0.154
Plant height(cm)	-0.433	0.312	**0.548**	0.183	-0.219	0.136	0.256
Days to Harvest	0.303	-0.085	-0.048	0.250	**0.590**	-0.142	0.241
Pod length(cm)	0.066	0.103	-0.014	0.074	0.168	0.125	**0.849**
No. of seeds/pod	-0.192	0.108	0.080	-0.038	**0.890**	0.050	0.040
No. of pods/plant	-0.034	**0.933**	0.080	0.035	0.127	-0.038	0.069
Fresh pod yield/plant(g)	0.413	**0.865**	-0.006	0.041	-0.069	0.057	0.032
100 Fresh pod weight(g)	**0.920**	0.124	-0.044	0.070	0.009	0.039	0.131
100 Fresh seed weight(g)	**0.952**	0.132	-0.060	0.047	-0.085	0.061	0.010
100 Dry seed weight(g)	0.107	0.229	-0.338	**0.614**	0.098	0.166	0.333
Crude sugar content	0.122	-0.049	0.024	-0.021	-0.084	**0.904**	0.179
Crude protein content	0.215	-0.245	0.168	-0.279	-0.381	-0.558	0.443
Hydration rate	0.037	-0.020	0.160	**0.883**	0.060	-0.038	-0.053
Cooking time(min)	-0.009	-0.013	-0.868	-0.186	-0.002	-0.064	-0.042

Rotated Component Matrix; Extraction by Principal Component Analysis; Rotation by Varimax with Kaiser Normalization

Table 3: Rotated component matrix interpretation for the traits having values >0.5 in each PCs

	PC1	**PC2**	**PC3**	**PC4**	**PC5**	**PC6**	**PC7**
TRAITS	100 Fresh pod weight (g)	No. of Pods /plant	Days to 50% Flowering	100 Dry seed weight(g)	Days to Harvest	Crude Sugar Content	Pod length(cm)
	100 Fresh seed weight(g)	Fresh Pod Yield /plant(g)	Plant Height(cm)	Hydration Rate	No. of seeds/pod		

The table shows the top PC scores for all traits and all 33 soybean genotypes in these seven components. These scores, the intensity of which can be determined by the variability explained by each of the principal components, can be used to suggest precise selection indices. A high PC score in a specific component for a specific genotype indicates that the variables in that genotype have high values. EC 915989, EC 915900, EC 915993, EC 915975, EC 915903, EC 915898 and EC 915959 in PC1 and this indicated that these genotypes have high values for traits such as 100 Fresh pod weight and 100 Fresh seed weight. The highest PC score of EC 915898 followed by EC 915978, EC 915975 and EC 915974 in PC2 was mainly related with No. of pods/plant and Fresh Pod yield/plant which are mainly yield attributing traits. A more or less same trend was observed by Bello *et al.* (2012) where he obtained comparable results by using principal component analysis to

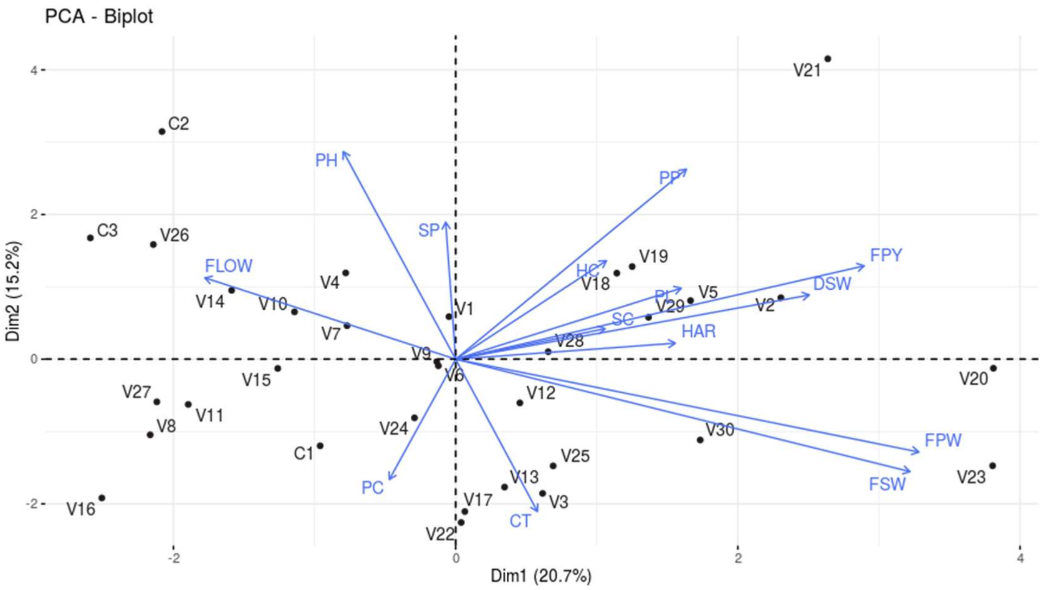

Figure 3: Distribution and grouping of germplasm lines across first two principal components.

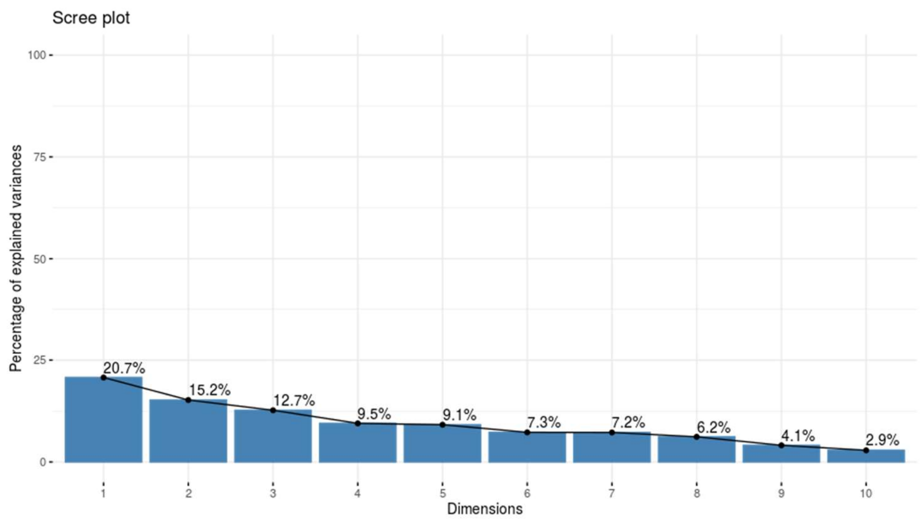

Figure 4: Scree Plot

evaluate 56 soybean genotypes to determine the percent contribution of variation by each trait to seed yield. The most important factors in genotype variation were discovered to be days to Harvest, plant height, and No. of pods/plant. Days to Harvest, plant height, and No. of pods/plant are all variables with a high weighting in the first PC, which accounted for 39.83% of total variation. PC2 explained an additional 20.52% of the total variation. These two factors accounted for 60.36% of all variable variation. As a result, genetic improvement of these traits will eventually result in higher seed yield and more or less same trend was observed by Ojo *et al*. (2012) where he investigated and reported three major components: PC1 had 52% variability, PC2 had 26% variability, and PC3 had 10% variability among genotypes for the traits studied. The No. of pods/plant, pod length, pod yield/plant, 100-seed weight, and seed yield/plot were all highly correlated with the first principal component. Days to 50% flowering and days to Harvest were the two components that were most

Table 4: Principle component scores of soybean genotypes

Genotypes	RC1	RC2	RC3	RC4	RC5	RC6	RC7
AGS-ACC	-0.034936266	-0.379835128	-0.467201909	-0.007735112	-1.117690966	**1.479109168**	-2.429748321
Karune	-1.769132537	0.960255876	**1.297266461**	-0.221807336	0.903606341	-0.328352164	0.164778903
NRC 105	-0.966812829	-0.264202699	**1.902818963**	-0.307204623	-0.149691149	0.798211927	-0.73811453
EC 915895	-0.446972606	-0.068169658	0.806440466	**1.128795093**	-2.248440251	-0.325948877	**1.673123674**
EC 915898	**1.07578808**	**2.997804337**	0.192475565	0.350291669	-1.38676415	-0.844012357	-1.919461732
EC 915900	**1.363486211**	-1.344361777	-0.026359629	-0.216408623	0.795493399	-0.218135888	-0.037426951
EC 915902	-1.59657889	-0.601153726	-0.428118991	**1.915537538**	0.15313189	-0.761906867	**1.164923006**
EC 915903	**1.106227369**	-0.898520134	**1.050558202**	0.063047482	**1.47926242**	**2.0392975**	**1.174943156**
EC 915908	-1.000668843	-0.285914648	-1.086688415	0.755091374	-0.937076293	0.757942136	0.653304897
EC 915909	-0.258439847	0.004712798	0.598476403	0.0449178	0.238422026	-1.08850357	0.219813443
EC 915910	-1.473336901	-0.282943278	-1.741754841	-1.294384486	0.612558905	-0.316078698	-0.44992662
EC 915913	0.298515533	-0.196840527	0.60209444	-1.410144618	0.54903056	-0.008837286	**1.437404564**
EC 915919	-1.149498287	-0.669979485	0.163803645	0.645493199	-1.247133923	**1.883033343**	0.240839458
EC 915923	0.098047376	-0.161963542	**1.318050678**	-1.534888122	-1.21039105	-0.465513686	-0.046830658
EC 915924	-0.320028755	-0.21723012	-1.135074507	**1.357971156**	-1.007563275	0.5386164	-0.67301945
EC 915926	0.566775447	-0.353364205	-0.828100863	-0.977254738	-0.385581909	0.344363432	0.107241867
EC 915933	-0.37442973	0.083554796	**1.479172542**	-0.048700622	-0.270350176	-0.763271126	-0.380608327
EC 915937	-0.113553486	0.711816547	0.799200896	-1.780233709	-0.865852719	-0.916687314	0.432203382
EC 915945	-0.881236355	-0.731194447	-1.044196048	-2.172353589	-0.774492903	**1.29370759**	-0.793771441
EC 915949	0.279425072	0.047741601	-1.22123404	-1.254246128	-0.620977732	-1.280718723	0.750685155
EC 915959	**1.000638448**	-0.081931258	**1.673470675**	0.047368994	0.468984202	**1.14378003**	0.638458025
EC 915974	0.45145444	**1.078074466**	0.15210382	0.180848295	**1.688104553**	-0.480732331	-0.631596913
EC 915975	**1.229244739**	**1.652171045**	-0.944252016	-0.530489897	-0.385272809	-0.217104803	**2.287670891**
EC 915978	-0.644085431	**2.615079176**	-0.093589366	**1.145877072**	**1.344311578**	**1.648575087**	0.033829023
EC 915983	**1.004980277**	-1.293798241	-0.275675882	**1.051511108**	-0.109960249	-2.095959501	-0.666331928
EC 915989	**2.168505837**	-0.006629361	-0.93535979	**1.010066419**	-0.026584073	0.907789773	-0.391560195
EC 915993	**1.310893064**	-0.985048535	**1.253142594**	0.26805993	0.54224033	-0.237953307	-1.648451464
EC 916000	0.639809924	-0.313563948	-0.849909787	0.234459825	-0.002465246	-0.655825632	-0.432816595
EC 916009	-1.45619745	-0.186345578	0.707036369	**1.247736982**	0.457340155	-1.333178508	-0.566585375
EC 916022	-0.846913635	-1.30444743	-0.562801952	-0.366015819	**1.88872081**	-0.683925384	-0.563568824
EC 916025	0.042030113	0.543444518	-0.600400953	-0.654649725	0.520485728	0.597787118	0.041572639
EC 916032	-0.284459508	0.82785335	-1.484181466	0.125032737	**1.669256453**	-0.210284224	0.219169705
EC 916039	0.981459423	-0.895070786	-0.271211261	**1.204410473**	-0.564660475	-0.199283257	**1.129857535**

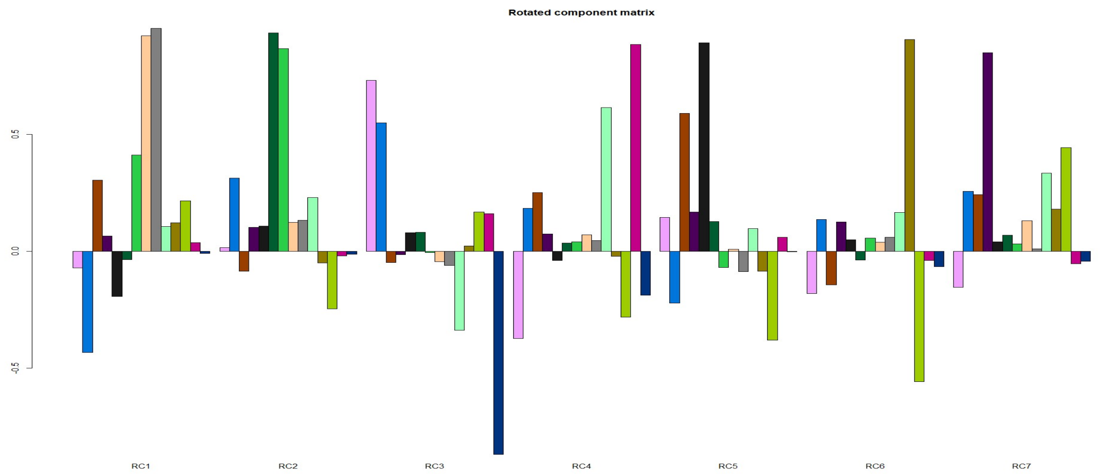

Figure 5: Rotated Component Matrix for 15 yield contributing traits of soybean.

closely related to the second main component. The highest PC score of NRC 105 followed by EC 915959, EC 915933, EC 915923, Karune, EC 915993 and EC 915903 in PC3 was mainly related with days to 50% flowering and plant height. More or less same trend was observed by Iqbal et al. (2008) where he investigated three major components, the first three of which accounted for 67.77% of the total variance. PC1 included seed yield and biological yield/plant, while PC2 included 100-seed weight and harvest index and PC3 included days to Harvest and No. of branches/plant. The highest PC score of EC 915902 followed by EC 915924, EC 916009, EC 916039, EC 915978, EC 915895, EC 915983 and EC 915989 in PC4 was

mainly related with 100 dry seed weight and Hydration Rate. The highest PC score of EC 916022 followed by EC 915974, EC 916032, EC 915903 and EC 915978 in PC6 was mainly related with Days to Harvest and No. of seeds/pod. The highest PC score of EC 915903, EC 915919, EC 915978, AGS-ACC, EC 915945 and EC 915959 in PC5 was mainly related with Crude Sugar content. The highest PC score of EC 915975, EC 915895, EC 915913, EC 915903, EC 915902 and EC 916039 in PC7 was mainly related with Pod length. More or less same trend was observed by Dubey *et al*. (2018) where he investigated fifty soybean accession for PCA in order to assess economic traits. Only five of the eighteen principal components (PCs) had more than one eigenvalue and revealed 73.44% variability among the observed traits. The variation in the first principal component was 37.13%, followed by 13.02% in the second principal component, 10.17% in the third principal component, 6.88% in the fourth principal component, and 6.24% in the fourth principal component (PC). These results are represented in the table 4.

Conclusion

The principal component analysis sorted out of total fourteen traits into seven main principal components. They had more than one Eigen value and a cumulative variance of about 81.748% was identified through all seven components, with PC1 explaining the most variance. AGS-ACC, Karune, NRC 105, EC 915895, EC 915898, EC 915900, EC 915902, EC 915903, EC 915923, EC 915924, EC 915933, EC 915945, EC 915959, EC 915974, EC 915975, EC 915978, EC 915983, EC 915989, EC 915993, EC 916009, EC 916022, EC 916032, EC 916039 were identified as putative genotypes based on principle component analysis and per se performance. However, EC 915903 was found to be the best because it had a higher value for a greater number of important traits.

Conflict of interest

The authors declare that they have no conflict of interest.

References

Bello, L. L., Aondover, S., & Terkimbi, V. (2012). Studies on relationship between seed yield and yield components in soybean [*Glycine max* (L.) Merrill]. *Electronic Journal of Plant Breeding*, *3*(4), 1012-1017.

Brachfeld, A., & Choate, M. (2007). *Eat your food! Gastronomical glory from garden to gut: a Coastalfields cookbook, nutrition textbook, farming manual and sports manual.* Coastal fields. Arvada, CO, USA, pp. 275.

Dubey, N., Avinashe, H. A., & Shrivasthava, A. N. (2018). Principal component analysis in advanced genotypes of soybean [*Glycine max* (L.) Merrill]. *Plant Archives, 18*(1), 501-506.

Fehr, W. R., Caviness, C. E., Burmood, D. T., & Pennington, J. S. (1971). Stage of development descriptions for soybeans, *Glycine max* (L.) Merrill. *Crops Science*, *11*, 929-931.

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

Jakhar, N. K., & Arjun, K. (2018). Principal component analysis and character association for yield components in greengram [*Vigna radiata* (L.) Wilczek] genotypes. *Journal Pharmacogn Phytochem*istry, *7* (2): 3665-3669.

Jolliffie, I.T. (1986). *Principal Component Analysis*. Springer.

Massay, W.F. (1965). Principal components regression in exploratory statistical research. *Journal of the American Statistical Association*, *60*, 234-246.

Ojo, D. K., Ajayi, A. O., & Oduwaye, A. O. (2012). Genetic relationships among soybean accessions based on morphological and RAPDs techniques. *Pertanika Journal of Tropical Agriculture Science, 35*(2), 237 – 248.

Shanmugasundaram, S. (2001). Global extension and diversification of fresh and frozen vegetable soybean. In: Lumpkin TA and Shanmugasundaram S (Compilers), 2nd International Vegetable Soybean Conference, Washington State University, Pullman, p. 161-165.

Shurtleff, W., & Aoyagi, A. (2009). History of edamame, green vegetable soybeans, and vegetable-type soybeans (1275-2009): Extensively annotated bibliography and sourcebook. Soybean information Center, Lafayette, CA.

Srikanth, T., Eshwari, K.B., & Bhave, M. H. V. (2017). Principal component analysis for yield components in Greengram Accessions (*Vigna radiate* L.). *International Journal on Pure Applied Biosciences,* ***5*** (4), 246-253.

Wold, H. (1966). Estimation of Principal Components and Related Models by Iterative Least Squares. In: Multivariate Analysis, Krisnaiaah P.R. (Eds). Academic Press, New York, pp. 391-420.

Zhang, Q.Y., Li, Y.S., & Liu, X. B. (2015). Edible quality and its regulation in vegetable soybean (Glycine max (L.) Merril).2015 ASA, CSSA, SSSA International Annual Meetings November 15-18, 2015, Minneapolis, MN.

Publisher's Note: ASEA remains neutral with regard to jurisdictional claims in published maps and figures.