# Multivariate Analysis of Seed Yield-Related Traits in Safflower (*Carthamus tinctorius*) Genotypes

*Pushpa H.D.[1]\*, Helan Baby Thomas[2], Anil Kumar A.[1], Mangesh Y. Dudhe[1],*
*Praduman Yadav[3] and Anjani Kamali[1]*
*[1]Division of Genetics and Plant Breeding,*
*ICAR-Indian Institute of Oilseeds Research, Rajendranagar, Hyderabad (Telangana), India.*
*[2]Division of Plant Biotechnology,*
*ICAR-Indian Institute of Oilseeds Research, Rajendranagar, Hyderabad (Telangana), India.*
*[3]Division of Crop Production,*
*ICAR-Indian Institute of Oilseeds Research, Rajendranagar, Hyderabad (Telangana), India.*

*(Corresponding author: Pushpa H.D.\*)*

**ABSTRACT:** Safflower is highly regarded for its nutritional and pharmaceutical properties. Despite this, the area under safflower crop has been declining over the years, due to its low productivity. Genetic diversity plays an important role in the improvement of crop. Exploiting the existing genetic diversity within the safflower germplasm is crucial in identifying the genotypes with seed yield and wider adaptability. Several statistical methods have been employed for the estimation of diversity among different genotypes. Multivariate analysis is one of the most often used techniques for determining genetic diversity between genotypes. In the present study, a total of 16 safflower advanced breeding lines were evaluated under rainfed conditions during *rabi* 2021 at ICRISAT-Hyderabad. Data were recorded on nine quantitative traits.  The total variation among genotypes was divided into nine Principal Components (PCs). The first three PCs with Eigen values>1 contributed to 73.34% of the total variability among genotypes. Traits with maximum values in PC1 were number of effective capitula and seed yield. PC2 with a 22.82% variation contribution was highly associated with seed yield, number of seeds per capitula, and days to maturity.  In PC3 major contributors were plant height and number of effective capitula. Traits *viz.*, plant height, number of effective capitula, number of seeds per capitula, days to maturity and seed yield contributed significant positive component loading to these principal components (PCs). multivariate analysis divided the 16 genotypes into three clusters. The findings from present study revealed that selection for seed yield in safflower may be based on the traits *viz.*, number of effective capitula, number of branches and 100 seed test weight.

**Keywords:** Correlation, Cluster, Oilseed, Principal Component Analysis, Seed yield.

## INTRODUCTION

Safflower (*Carthamus tinctorius* L.) is an important oilseed crop cultivated since time immemorial. It is a herbaceous annual plant30–150 cm in height, heavily branched, mostly spiny, and cultivated during the winter/spring seasons (Bérvillé *et al.,* 2005). Different parts of the plant have a wide range of uses, including food coloring, flavoring, biofuel, edible oil, pharmaceutical applications, and bird feed. The crop is highly valued for its nutritional and pharmaceutical properties and is grown in over 60 nations around the globe (Singh and Nimbkar 2006). In India, during 2021-22, safflower is cultivated in an area of 75,010 ha with production 57,260 tonnes and productivity 763 kg/ha (FAOSTAT 2022).

The main objective in any safflower breeding program is to develop high yielding cultivars coupled with high oil content. Identifying and characterizing the traits that affect safflower seed and oil yield is crucial in the genetic improvement of these attributes (Golparvar, 2011). Yield is a complex quantitative character that is directly or indirectly dependent on the interactions of various other traits (Tonk *et al.,* 2011). Therefore, it is crucial to comprehend the extent to which these traits can influence yield, particularly in a breeding program that seeks to develop a high-yielding genotype (Bidgoli *et al.,* 2006).

Several statistical methods are used for understanding the effect of different traits on crop yield. Determining the correlation coefficients helps in analyzing the relationships between yield and its associated traits, and this has a very important role in selecting the breeding materials (Zafarnaderi *et al.,* 2013). Principal component analysis (PCA) is a multivariate statistical technique that focuses on simplifying complex datasets (Slavkovic *et al.,* 2004). Cluster analysis is an agglomerative hierarchical method that classifies variables into groups and cluster variables (Katar, 2013). Considering all these factors, a multivariate

analysis study was conducted on 16 safflower advanced breeding lines to understand the genetic variability among them and to determine the traits that affect seed yield and its related traits.

## MATERIALS AND METHODS

**Experimental material.** In the present study, sixteen high-yielding advanced breeding lines along with two national checks (A-1 and PBNS-12) (Table 1) were evaluated during the *rabi*-2021. The breeding lines were developed through crossing using elite parental lines *viz.*, DSF-317, DSF-22-15, DSF-134-15 and DSF-2305-15 which were high seed yielding and high oil content inbreds and advanced till the $F_7$ generation through pedigree breeding method.

**Table 1: List of genotypes used in the study.**

| Sr. No. | Genotypes |
|---------|-----------|
| 1. | ISF-106-15 |
| 2. | ISF111-15 |
| 3. | ISF-112-15 |
| 4. | ISF-113-15 |
| 5. | ISF-115-15 |
| 6. | ISF-120-15 |
| 7. | ISF-122-15 |
| 8. | ISF-128-15 |
| 9. | ISF-129-16 |
| 10. | ISF-132-16 |
| 11. | ISF-135-16 |
| 12. | ISF-136-16 |
| 13. | ISF-137-16 |
| 14. | ISF-138-16 |
| 15. | ISF-142-16 |
| 16. | ISF-143-16 |
| | A-1 (National check) |
| | PBNS-12 (Varietal check) |

**Experimental design.** The trial was conducted in the experimental fields of ICRISAT, Hyderabad (17.5111° N, 78.2752° E). Sowing was done in black soil. The experiment was conducted in three replications using the random block design method. Each genotype was sown in five rows with each row having a 5 m row length with a spacing of 45 × 10 cm. Recommended agronomic and plant protection practices were followed to ensure a good crop stand. Five plants of each genotype were randomly tagged and data on nine quantitative traits (Table 2) were recorded.

**Table 2: Quantitative traits recorded in the study.**

| Sr. No. | Traits | Abbreviation | Unit |
|---------|--------|--------------|------|
| 1. | Plant height | PHT | cm |
| 2. | Days to 50% flowering | DFF | number |
| 3. | Number of branches | NB | number |
| 4. | Number of effective capitula | NEC | number |
| 5. | Number of seeds per capitula | NSC | number |
| 6. | Days to maturity | DM | number |
| 7. | 100 seed weight | TSWT | g |
| 8. | Oil content | OC | % |
| 9. | Seed yield per plant | SYLD | g |

The oil analysis from seed was done using the Soxhlet Extraction method. The niger oil was obtained by chemical extraction, in which 10 g of decorticated seeds were crushed to extract the oil with hexane solvent for 4 h by Soxhlet type extractor. The oil extract obtained was dried under low pressure at 70 °C in a rotary evaporator until the solvent is eliminated (Hosni *et al.*, 2022). The weight of the oil was recorded, and the oil content was obtained using the formula:

$$\text{Oil content (\%)} = \frac{(\text{weight of flask with oil - weight of empty flask})}{\text{weigth of ground seeds}} \times 100$$

**Statistical analysis.** A statistical analysis was performed on the aggregate mean values. Genetic diversity assessment among the genotypes were carried out using principal component analysis (PCA), cluster and correlation using R-3.5.2 win software. The contribution of different traits to total variation for all the genotypes were determined by PCA and the first two principal components that had the highest contribution to the variability were used to make biplot. PCA and Cluster analyses were performed by ggdendro package in R v3.5.2 (Andrie de Vries and Brian D. R., 2022).

## RESULTS AND DISCUSSION

*Correlation analysis.* A correlation coefficient analysis was done to see if the traits under study were redundant. Correlation coefficient values marked as bold show that there is redundancy or correlation between the variables, which makes it possible to condense the observed variables into a smaller set of principal components (Table 3). The presence of highly significant and positive correlation was observed between days to 50% flowering and plant height; days to 50% flowering and days to maturity; number of seeds per capitula and days to maturity, and seed yield per plant with 100 seed test weight. A negative but significant correlation was observed with plant height and 100 seed test weight, number of effective capitula and number of seeds per capitula; days to maturity and number of seeds per capitula; 100 seed test weight and days to 50% flowering. Results indicated that there is a relationship between oil content and seed yield was negative but non-significant, similarly between 100 seed test weight and oil content. Similar correlation results were found by Mali *et al.* (2022) in which plant height, number of branches/plant, number of effective capitulum/plant, number of seed/capitulum and test weight showed strong positive significant correlation with seed yield. A correlation between oil yield and seed yield, seed weight, number of branches per plant, number of heads per plant, and oil content has been established through a study conducted by Kose *et al.* (2018).

**Table 3: Correlation analysis of the quantitative traits studied.**

|  | PHT | DFF | NB | NEC | NSC | DM | TSWT | OC | SYLD |
|---|---|---|---|---|---|---|---|---|---|
| PHT | 1.000 | | | | | | | | |
| DFF | **0.596**\*\* | 1.000 | | | | | | | |
| NB | 0.254 | -0.225 | 1.000 | | | | | | |
| NEC | -0.279 | -0.403 | 0.436 | 1.000 | | | | | |
| NSC | 0.206 | 0.265 | -0.467 | **-0.513**\* | 1.000 | | | | |
| DM | 0.255 | **0.470**\* | -0.253 | **-0.591**\*\* | **0.521**\* | 1.000 | | | |
| TSWT | **-0.548**\* | **-0.499**\* | -0.045 | 0.009 | -0.102 | 0.142 | 1.000 | | |
| OC | 0.325 | 0.285 | 0.178 | -0.196 | 0.315 | 0.142 | -0.381 | 1.000 | |
| SYLD | -0.190 | -0.398 | 0.182 | 0.425 | 0.068 | 0.182 | **0.598**\*\* | -0.173 | 1.000 |
| (PHT: Plant Height; DFF: Days to fifty percent; NB: Number of branches; NEC: Number of effective capitula; NSC: Number of seeds per capitula; DM: Days to maturity; TSWT: 100 Seed weight; OC: Oil content; SYLD: Seed yield per plant) | | | | | | | | | |

***Principal Component Analysis.*** The principal component analysis for selected 16 genotypes with nine phenotypic traits divided the total variation into nine principal components. The first three PCs had eigenvalues of more than one and explained the 73.34% variation cumulatively were selected as these attributes contribute more to variation than the remaining one (Fig. 1). Eigenvalues, percentage of variability, and cumulative contribution of each component to variation, as well as each character's contribution to the corresponding PCs are shown in Table4. PC1 explained 35.74% of the total variability. Traits with maximum values in PC1 were number of effective capitula and seed yield. PC2 with 22.82% variation contribution was highly associated with seed yield, number of seeds per capitula and days to maturity. In PC3 major contributors were plant height and number of effective capitula. Traits *viz.*, plant height, number of effective capitula, number of seeds per capitula, days to maturity and seed yield contributed significant positive component loading to these principal components (Fig. 2).

PCA is used to reduce redundancy by breaking down the phenotypic traits into multiple principal components and giving you the chance to choose PCs that have characteristics that are more significantly contributing to variation. Principal component analysis helps identifying three main components that account for 56.5% of the total variations. In a study with 122 safflower genotypes, Jabbari *et al.* (2021) identified that the first and second principal components account for 29.5% and 15.9% of the total variation, respectively. The genotypes were divided into four groups by principal components analysis (PCA) and the genotypes in the first groups have higher grain yield than others. Ali *et al.* (2020) used principal component analysis to identify that the days to flower initiation, days to 50% flowering, days to flower completion, seed yield per plant, capitula per plant, branches per plant, seeds per capitulum, and capitulum diameter were the major contributors to the observed genetic variability in the evaluated 96 safflower panel. Seed yield per plant reflected a significant and positive correlation with capitula per plant, branches per plant, and capitulum diameter. In a study on five spring safflower genotypes, by using PCA, Katar (2013) found that key characteristics for obtaining high oil yielding safflower genotypes in cluster analysis were seed weight, oil content, and seed yield per plant. A study conducted by Mozaffari and Asadi (2006) reported a positive association between plant height, head diameter, and oil content for the first principal component. Ahmadzadeh (2010) discovered that three main components with eigenvalues greater than one contributed 72.92 percent of the total variability in a study using 30 spring safflower genotypes.
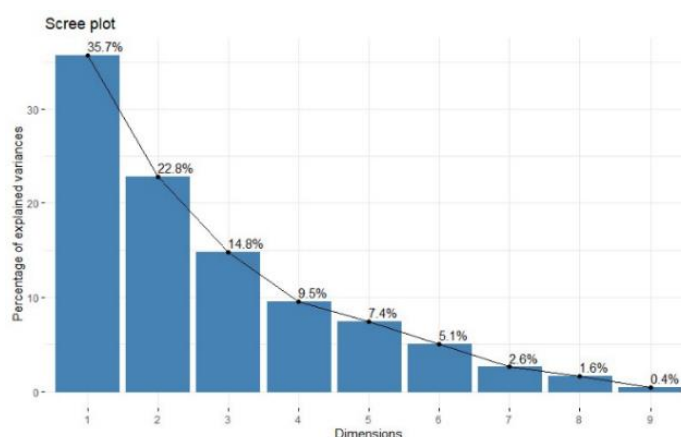


**Fig. 1.** Scree plot.

**Table 4: Principal component analysis for seed yield related traits in safflower advanced breeding lines.**

|  | PC 1 | PC 2 | PC 3 |
|---|---|---|---|
| **Eigenvalue** | 3.22 | 2.05 | 1.33 |
| **Variance %** | 35.74 | 22.83 | 14.77 |
| **Cumulative variance %** | 35.74 | 58.57 | 73.34 |
| **Traits** | **Eigen vectors** | | |
| **PHT** | -0.268 | 0.085 | **0.918** |
| **DFF** | -0.081 | -0.001 | 0.1 |
| **NB** | 0.077 | -0.043 | 0.189 |
| **NEC** | **0.776** | -0.288 | **0.29** |
| **NSC** | -0.205 | **0.365** | -0.037 |
| **DM** | -0.172 | **0.365** | -0.008 |
| **TSWT** | 0.026 | 0.057 | -0.077 |
| **OC** | -0.08 | 0.017 | 0.139 |
| **SYLD** | **0.484** | **0.799** | 0.039 |
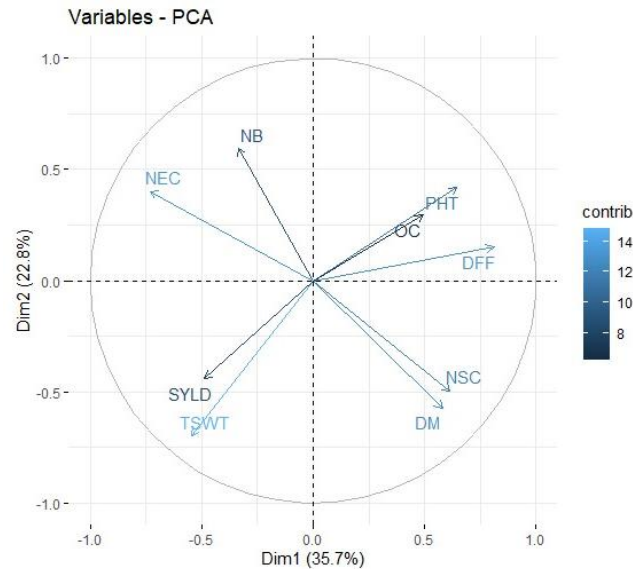


**Fig. 2.** PCA variables.

***Biplot analysis.*** In order to assess the diversity within the chosen genotypes, a biplot was created using the first two PCs with Eigenvalues greater than one, which accounted for 58.57% of the variability. Genotypes that are close to each other on a biplot are seen as similar based on the characteristics assigned. The more genetic diversity there is among other genotypes, the greater the distance between the genotype and the site of origin. Distinct groupings of genotypes were identified for the first two PCs (Fig. 3). On biplot, two genotypes i.e. ISF111-15 and ISF-137-16 (2 & 13) clogged far away from the origin and were considered as diverse from the others (Fig. 3).

In the study by Ebrahimi *et al.* (2023), the biplot was divided into eleven sectors, and the sector of genotype G80 was the winner for most traits. Genotype G58 followed by genotypes G30, G33 and G72, were the most favorable genotypes in regard to seed yield, while regarding this trait as a reference, seeds per main capitulum was identified as the most related trait which is followed by number of seeds per lateral capitulum and stem diameter.
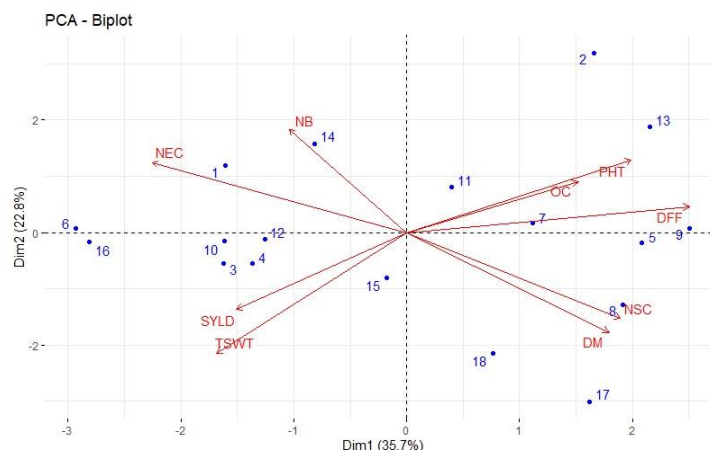


**Fig. 3.** Biplot explaining the PCA analysis (Arrows indicate the strength of the trait influence on the first two PCs.).

*Cluster analysis.* Principal component analysis provided a clear explanation for why some genotypes under study were different from others, while many others were similar and positioned close to one another on the biplot. However, PCA was unable to classify these genotypes into any distinct groups. Therefore, cluster analysis based on Euclidean distance across the phenotypic features was used to separate the genotypes based on similarities and differences. Cluster analysis grouped the 16 genotypes into 3 distinct clusters as presented in dendrogram (Fig. 4). Cluster I was the smallest cluster and contained only two genotypes (ISF-111-15, ISF-137-16). Cluster II was the largest cluster with nine genotypes *viz.,* ISF113-15, ISF106-15, ISF120-15, ISF143-16, ISF138-16, ISF142-16, ISF112-15, ISF132-16, ISF136-16. Cluster III had 7 genotypes *viz.,* A1, PBNZ-12, ISF115-15, ISF129-16, ISF128-15, ISF122-15, ISF135-16.
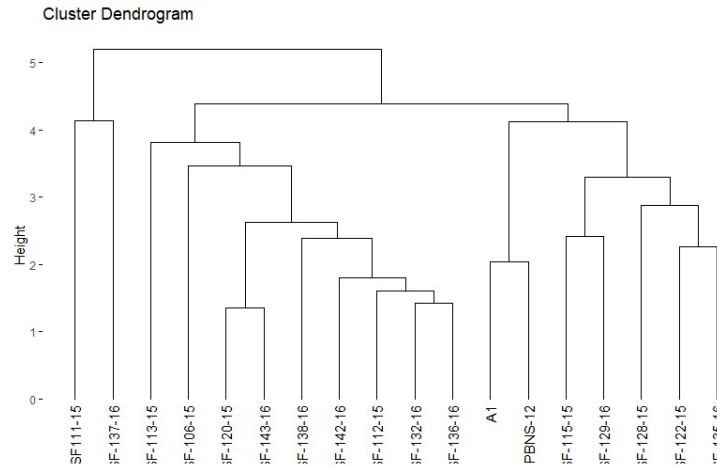


**Fig. 4.** Cluster dendrogram classifying the genotypes in to different clusters.

Highest mean value for plant height (90.8 cm), days of 50% flowering (81.6 days), oil content (36.9%) and number of branches (16.7 cm) was seen in cluster I. In cluster II, maximum mean values were recorded for number of effective capitula (24.5), 100 seed weight (5.81 g) and straw yield (41.3 kg). Cluster III contains the minimum value for plant height (73.7 cm), number of branches (8.8), number of effective capitula (29.9) and oil content (27.4%) (Table 5). Aleem *et al.* (2021) in his study categorized nineteen safflower genotypes in to three distinct clusters based on thirteen morphological characteristics. Similar results were also reported by Houmanat *et al.* (2021), 45 safflower accessions were delineated into four main clusters based on different physiological traits. In a study by Ojaq *et al.* (2020), cluster analysis of 127 safflower genotypes using the Ward method and based on Euclidean distance resulted in four cluster groups.

There were a large number of genotypes (45 genotypes) in the first cluster with the dominant characteristics such as yellow-coloured flower, highest plant height, mostly thorny but with a wide range of genotypes without thistle, late-flowering with the number of heads per plant, number of grains per head, 1000-grain weight and relatively low grain yield. Golami *et al.* (2018) performed cluster analysis on 64 safflower genotypes and results based on Euclidean distance and Ward method showed that there were five different clusters. The genotypes of the third cluster had high values in some important traits such as thousand seed weight, seed yield and oil content while the genotypes of the other clusters were high in number of capitulum and harvest index traits. Rameshknia *et al.* (2013) performed cluster analysis for different traits on 26 spring safflower varieties and they were classified into seven clusters.

**Table 5: Phenotypic characterization of three clusters.**

| Characters | Cluster I | Cluster II | Cluster III |
|---|---|---|---|
| Trait PH | 90.87 | 76.67 | 73.77 |
| Trait DFF | 81.67 | 80 | 80.98 |
| Trait NB | 16.67 | 12.93 | 8.87 |
| Trait NEC | 37.33 | 54.6 | 29.95 |
| Trait NSC | 21.33 | 24.5 | 24.96 |
| Trait DM | 118.67 | 119.67 | 119.93 |
| Trait 100 SW | 3.71 | 5.81 | 4.32 |
| Trait OC | 36.88 | 28.65 | 27.45 |
| Trait SY | 15.07 | 41.31 | 18.39 |

**CONCLUSIONS**

The various statistical techniques employed in this study indicated that the most beneficial characteristics for improvement of seed yield in safflower would likely be the selection based on the seed yield traits *viz.*, number of effective capitula, number of branches, and 100 seed test weight. The genotypes having high genetic diversity and more cluster distance can be used as parents in the crossing program, to recover

transgressive segregants for seed yield-related traits. The results of this study could potentially guide future safflower breeding program design and implementation.

# REFERENCES

Ahmadzade, A. R., Majidie, E., Alizade, B. and Omidi, A. H. (2010). Study the yield, yield components and morphological traits in the spring safflower (*Carthamus tinctorius* L.) using multivariate statistical approaches. *Journal of Agricultural Science, 6*, 1-8.

Aleem, S., Tahir, M., Sharif, I., Aleem, M., Najeebullah, M., Nawaz, A. and Arshad, W. (2021). Principal component and cluster analyses as tools in the assessment of genetic diversity for late season cauliflower genotypes. *Pakistan Journal of Agricultural Research*, *34*(1).

Ali, F., Yilmaz, A., Chaudhary, H. J., Nadeem, M. A., Rabbani, M. A., Arslan, Y., and Baloch, F. S. (2020). Investigation of morpho agronomic performance and selection indices in the international safflower panel for breeding perspectives. *Turkish Journal of Agriculture and Forestry*, *44*(2), 103-120.

Berville, A., Breton, C., Cunliffe, K., Darmency, H., Good, A. G. and Gressel, J. (2005) Issues of ferality or potential ferality in oats, olives, the pigeon-pea group, ryegrass species, safflower, and sugarcane. In: Gressel J (ed) Crop ferality and volunteerism: a threat to food security in the transgenic era? Taylor and Francis Books, Boca Raton

Bidgoli, A. M., Akbari, G. A., Mirhadi, M. J, Zand, E., and Soufizadeh, S. (2006). Path analysis of the relationships between seed yield and some morphological and phonological traits in safflower (*Carthamus tinctorius* L.). *Euphytica*, *148*, 261-268.

Carter, R. L., Morris, R. and Blashfield, R. K. (1989). On the partitioning of squared Euclidean distance and its applications in cluster analysis. *Psychometrika*, *54*, 9-23.

Ebrahimi, H., Sabaghnia, N., Javanmard, A., and Abbasi, A. (2023). Genotype by trait biplot analysis of trait relations in safflower. *Agrotechniques in Industrial Crops*, *3*(2), 67-73.

Food And Agriculture Organization of The United Nations (2022). FAOSTAT. [Software, E-Resource] Retrieved from the Library of Congress,

Gholami, M., Sabaghnia, N., Nouraein, M., Shekari, F., Janmohammadi, M. (2018). Cluster Analysis of Some Safflower Genotypes using a Number of Agronomic Characteristics. *JCB, 10*(25), 159-166.

Golparvar, R. (2011). Assessment of relationship between seed and oil yield with agronomic traits in spring safflower cultivars under drought stress condition.

*Journal of Research in Agricultural Science*, *7*(2), 109-113.

Houmanat, K., Douaik, A., Charafi, J., Hssaini, L., El Fechtali, M.and Nabloussi, A. (2021). Appropriate statistical methods for analysis of safflower genetic diversity using agglomerative hierarchical cluster analysis through combination of phenotypic traits and molecular markers. *Crop Science, 61,* 4164–4180.

Jabbari, H., Fanaei, H. R., Shariati, F., Sadeghi Garmarodi, H., Abasali, M., & Omidi, A. H. (2022). Principal components analysis of some Iranian and foreign safflower genotypes using morphological and agronomic traits. *Journal of Crops Improvement, 24*(1), 125-143.

Katar, D. (2013). Determination of efficiency of yield components on oil yield per plant in safflower breeding by different statistical methods. *Global Journal of Science Frontier Research: Agriculture & Veterinary*, *13*, 11-20.

Kose, A., Hatipoglu, H. and Arslan, H. (2018). Genotype reaction and effect of the sowing time under arid ecological conditions in safflower (*Carthamus tinctorius* L.). *Fresenius Environmental Bulletin*, *27*(5A), 3577-3586.

Mali, B. B., Wadikar, P. B., Gorte, A. S. and Yadav, U. V. (2022). Relationship among traits using correlation and path coefficient analysis in safflower (*Carthamus tinctorius* L.). *Journal of Agriculture Research and Technology*, *47*(1), 39.

Mozaffari, K. and Asadi, A. A. (2006). Relationships among traits using correlation, principal components and path analysis in safflower mutants sown in irrigated and drought stress condition. *Asian Journal of Plant Sciences, 5*(6), 977-983.

Ojaq, S. M. M., Mozafari, H., Jabbari, H., and Sani, B. (2020). Evaluation of yield of safflower (*Carthamus tinctorius* L.) genotypes under semi-arid conditions. *Plant Genetic Resources*, *18*(4), 270-277.

Rameshknia, Y., Tahmasebpoor, B. and Sabbagh Tazeh, E. (2013). Investigation of the important traits of spring safflower varieties through multivariate statistical methods. *Bulletin of Environment, Pharmacology and Life Sciences*, *2*(8), 29-34.

Singh, V. and Nimbkar, N. (2006). Safflower (*Carthamus tinctorius* L.). In: Genetic Resources, Chromosome Engineering, and Crop Improvement (ed: Ram J. Singh), *6,* 167-194.

Slavkovic, L., Skrbic, B., Miljevic, N. and Onjia, A. (2004). Principal component analysis of trace elements in industrial soils. *Environmental Chemistry Letters*, *2*, 105-108.

Tonk, F.A., İlker, E.and Tosun, M. (2011). Quantitative inheritance of some wheat agronomic traits. *Bulgarian Journal of Agricultural Science, 17,* 783-788.

Zafarnaderi, N., Aharizad, S. and Mohammadi, S.A. (2013). Relationship between grain yield and related agronomic traits in bread wheat recombinant inbred lines underwater deficit condition. *Annals of Biological Research*, *4,* 7-11.