# Research Article

# Principal Component and Cluster Analyses as Tools in the Assessment of Genetic Diversity for Late Season Cauliflower Genotypes

**Saba Aleem[1]\*, Mehvish Tahir[2], Iram Sharif[3], Muqadas Aleem[4], Muhammad Najeebullah[2], Ali Nawaz[1], Amina Batool[1], Muhammad Imran Khan[1] and Waheed Arshad[1]**

[1]*Barani Agricultural Research Station, Fatehjang, Pakistan;* [2]*Vegetable Research Institute, Ayub Agricultural Research Institute, Faisalabad, Pakistan;* [3]*Cotton Research Station, Ayub Agricultural Research Institute, Faisalabad, Pakistan;* [4]*National Center for Soybean Improvement, Key Laboratory of Biology and Genetics and Breeding for Soybean, Nanjing Agricultural University, Nanjing 210095, China.*

**Abstract** | Genetic diversity is the baseline of any breeding program as determining the variability of germplasm is an indispensable assist for crop improvement strategies. Multivariate analyses such as cluster and principal component analysis measures the amount of genetic diversity in respect of several charaters and asseses the relative contribution of different traits to the total variation. In the present study, nineteen genotypes were evaluated for presence of genetic diversity through principal component and cluster analyses. The experiment was designed in randomized complete block design with two replications at the experimental area of Vegetable Research Institute, Faisalabad, Pakistan in 2019. In the principal component analysis, first three principal components had eigenvalue >1 and contributed 80.86% of the variation among the genotypes. Traits i.e. plant weight, curd weight, curd yield, leaf length, plant height to extreme, and the number of leaves contributed significant positive component loading to these principal components (PCs). Biplot analysis among the first two PCs found the Casper RZ, AA Cauli-08, and Snow queen as diverse genotypes. Cluster analysis based on Euclidian distance grouped the genotypes into 3 distinct clusters. Some of the genotypes that have narrow genetic base were grouped into a similar cluster. Based on these results, it may be concluded that some of the genotypes are higly diverse while most of the genotypes are similar in nature. Genotypes from the distinct cluster should be used for obtaining diverse recombinants in segregating generations, exploiting heterosis, and broaden the genetic base of the cauliflower germplasm.

## Introduction

Cauliflower (*Brassica oleracea* L. Var. *botrytis*) is a cool-season vegetable and is mostly cultivated for its white curd. It evolved from wild cabbage (*Brassica oleracea* L. 2n=18, CC), known as cole worts, through mutation, selection, and adaptation (Purugganan *et al.*, 2000). Cauliflower was firstly originated in Cyprus and then was moved to other regions such as Syria, Turkey, Egypt, Italy, Spain, North-Western Europe, America, and the USA. Through domestication in these areas, different forms of cauliflower such as italian, cornish, notherns, roscoff, angers, indian, erfurt and snow ball were evolved in the 18th and 19th century (Swarup and Chatterjee, 1972). These days, cauliflower is mostly cultivated in China, India,

USA, Spain and Italy. In Pakistan, the mean yield of cauliflower is 17.0 t/ha which is 2.7 t/ha less compared to the largest producing country China (FAO, 2017).

To boost the yield of cauliflower, exploitation of genetically diverse genotypes with desirable traits is required for the development of high yielding hybrids and varieties. For which estimation of genetic diversity of the germplasm is required as it is the baseline of any breeding program and provides a great array of diverse genotypes that could be used for the development and evaluation of breeding material and new varieties with desirable characteristics (Vanlalneihi *et al.*, 2019). Knowledge about genetic diversity helps the breeders to identify diverse parents for creating a segregating population with maximum variability and for introgressing desirable genes such as biotic and abiotic stress resistance and wider adaptability from diverse material to available genetic material (Dangi *et al.*, 2018; Evgenidis *et al.*, 2011; Javed *et al.*, 2017; Tuhina-Khatun *et al.*, 2015). Comprehensive information regarding genetic diversity also helps in developing superior genotypes to be used in hybrid breeding as parents (Singh *et al.*, 2014).

The extent of genetic diversity could either be estimated through univariate and multivariate analyses. In recent times, multivariate analyses such as cluster, principal component analysis (PCA), and principal coordinate analysis (PCoA) have become popular to check similarities and differences among genotypes regarding multiple traits. Recently, in many crops such as cotton, rice, tomato, maize, wheat, sugar cane, ginger, bitter gourd, garlic, and onion PCA and cluster analyses have been widely used for estimation of genetic diversity (Baranwal *et al.*, 2013; Chakma *et al.*, 2012; Dangi *et al.*, 2018; Evgenidis *et al.*, 2011; Jarwar *et al.*, 2019; Pahadi *et al.*, 2017; Ravishanker *et al.*, 2013; Sharma *et al.*, 2018; Singh *et al.*, 2014; Tahir *et al.*, 2013). Studies are available in cauliflower regarding genetic diversity estimates of genotypes for mid-season (Chatterjee *et al.*, 2018; Yousef *et al.*, 2018; Zhu *et al.*, 2018b). However, the genetic diversity of genotypes for the late season cultivaion is not available.

Therefore, the present study was conducted to estimate genetic variation among late-season cauliflower genotypes using principal component and cluster analyses. This study provides useful information regarding genetic diversity and characterization of genotypes for desirable traits which is helpful for future cauliflower breeding.

## Materials and Methods

### *Experimental materials*

Nineteen cauliflower genotypes were used to assess the extent of phenotypic variation and genetic diversity for late season. The genotypes were obtained from different private seed companies. The detailed information of all genotypes is presented in Table 2.

### *Experimental design and measurement of traits*

This experiment was conducted at the experimental area of Vegetable Research Institute, Faisalabad, Pakistan (73–74° E and 30–31.5° N). Sowing of nursery was done on October 15, 2019. The trial was laid out according to RCB design with two replications keeping a plot size of 7 × 1.5 m. Seedlings of all the genotypes were transplanted in the field on November 16, 2019 keeping planting geometry of P×P=45 cm and R×R=75 cm. Standard cultural practices and plant protection measures were carried out regularly. Data regarding thirteen traits, presented in Table 1 were recorded and used for further analyses. Five plants of each replication were tagged and used for data collection of all morphological parameters.

**Table 1:** *Studied morphological parameters along with their SI units.*

| Sr. No. | Morphological characters | Abbreviation | Unit |
|---|---|---|---|
| 1 | Plant weight | PW | kg |
| 2 | Stem length | SL | cm |
| 3 | Number of leaves | NL | Number |
| 4 | Leaf length | LL | cm |
| 5 | Leaf width | LWI | cm |
| 6 | Plant height to extreme | PHE | cm |
| 7 | Plant height to apex | PHA | cm |
| 8 | Plant width | PWI | cm |
| 9 | Curd height | CH | cm |
| 10 | Curd diameter | CD | cm |
| 11 | curd weight | CW | kg |
| 12 | Curd Yield | CY | Tonnes/hacter (ton/ha) |
| 13 | Maturity Days | MD | Days |

Leaf length, stem length, leaf width, plant height to extreme, plant height to apex, plant width, curd height, and curd diameter were measured by a ruler.

The curd weight was measured by electronic balance. Curds of all the plants of the experimental unit area were harvested at maturity and curd yield was calculated. Maturity days were calculated from the date of transplanting to curd harvest.

**Table 2:** *Studied genotypes along with their supplier.*

| Sr. No. | Genotypes name | Seed supplier |
|---|---|---|
| 1 | HCF-971 A | Kanzo Pvt. Ltd |
| 2 | HCF-961 A | Suncrop Pvt. Ltd |
| 3 | Contessa (Dp-101002) | Roghay Crop Science |
| 4 | Elgon | Roghay Crop Science |
| 5 | Snow Queen | Tri Star Seed Company |
| 6 | TCF-609 | Tara Crop Science Pvt. Ltd |
| 7 | Benazir | Green Gold Agri Seeds Pvt Ltd |
| 8 | Gul Bahar | CGN Seed Co. Ltd |
| 9 | Easy Top | Evereast Pakistan Pvt. Ltd |
| 10 | Bishop RZ | More Green Zarai Market |
| 11 | Gulfam | Tri Star Seed Company |
| 12 | AA Cauli H-82 | Ali Akbar Seeds Pakistan |
| 13 | AA Cauli-08 | Ali Akbar Seeds Pakistan |
| 14 | Winner-SS#96 | Seed and Seed Processing |
| 15 | Casper RZ | More Green Zarai Market |
| 16 | Winner-CF#105 | Seed and Seed Processing |
| 17 | Winner-CF#94 | Seed and Seed Processing |
| 18 | Diamond CA-03 | Suncrop Pesticides (Pvt) Ltd |
| 19 | Diamond CA-07 | Suncrop Pesticides (Pvt) Ltd |

*Statistical analysis*
Genetic diversity assessment among the genotypes was carried out by PCA and cluster analysis. Contribution of the traits to the total variations for all the genotypes was checked by PCA and the first two principal components that had highest contribution to variability were used for making biplot; a graphical display helpful for checking diversity among genotypes. While, cluste analysis was used to determine the similarity and differences among genotypes using Euclidean distance for distance measuring. Euclidean distance is commonly used as this parameter best reflect the difference that exist among the genotypes (Carter *et al.*, 1989; Kendall, 1975). Unweighted pair group method with arithmetic mean (UPGMA) method was used in cluster analysis as UPGMA provides more accurate grouping information and commolnly used in sorghum for the estimation of genetic variations (Agrama and Tuinstra, 2003; Uptmoor *et al.*, 2003). PCA and Cluster analyses were performed by XLSTAT using its analyzing data tab and further principal component analysis for PCA and agglomerative hierarchical clustering for cluster analysis.

## Results and Discussion

*Principal component analysis*
To check the redundancy among the studied traits, correlation coefficient analysis was performed. Bold values of correlation coefficient indicates that correlation/redundancy exists between the variables so; it's possible to reduce the observed variables into a smaller number of principal components (Table 3).

In the present study, selected 19 genotypes were characterized according to 13 phenotypic traits by PCA analysis. Total variation was divided in to 13 principal components. The first three PCs, had eigen values more than 1 and explained the 80.87% variation cumulatively were selected as these attributes contribute more to variation than remaining ones (Figure 1).

**Table 3:** *Correlation matrix showing correlation among studied phenotypic traits.*

| Variables | MD | PW | CW | CY | SL | NL | LL | LWD | PHE | PHA | PWD | CH |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| PW | 0.84 | | | | | | | | | | | |
| CW | 0.90 | 0.96 | | | | | | | | | | |
| CY | 0.89 | 0.96 | 0.99 | | | | | | | | | |
| SL | 0.42 | 0.41 | 0.39 | 0.41 | | | | | | | | |
| NL | 0.41 | 0.47 | 0.50 | 0.53 | 0.31 | | | | | | | |
| LL | -0.10 | 0.11 | 0.04 | 0.04 | 0.55 | 0.24 | | | | | | |
| LWD | 0.32 | 0.30 | 0.35 | 0.36 | 0.51 | 0.15 | 0.47 | | | | | |
| PHE | 0.18 | 0.31 | 0.26 | 0.26 | 0.78 | 0.30 | 0.84 | 0.45 | | | | |
| PHA | 0.29 | 0.38 | 0.37 | 0.37 | 0.65 | 0.34 | 0.50 | 0.53 | 0.73 | | | |
| PWD | 0.26 | 0.43 | 0.38 | 0.39 | 0.64 | 0.38 | 0.75 | 0.28 | 0.89 | 0.59 | | |
| CH | 0.40 | 0.46 | 0.42 | 0.46 | 0.48 | 0.04 | 0.33 | 0.56 | 0.51 | 0.68 | 0.59 | |
| CL | 0.68 | 0.68 | 0.69 | 0.70 | 0.14 | 0.52 | 0.02 | 0.20 | 0.10 | 0.24 | 0.30 | 0.37 |
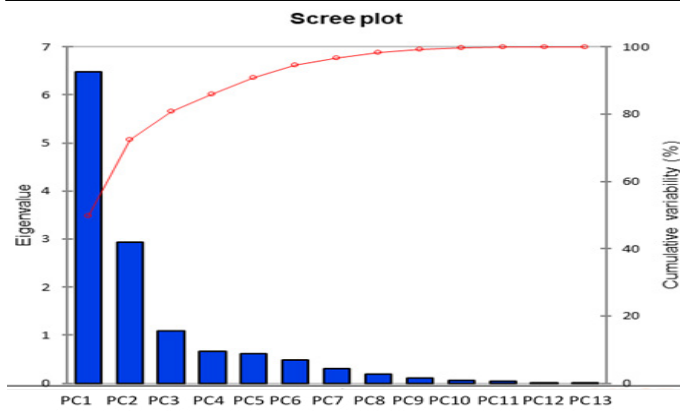
**Figure 1:** *Scree plot presenting PCs with their cumulative variability.*

Eigenvalues, variability %, and cumulative contribution of each component to variation along with the contribution of each character to respective PCs are presented in Table 4. PC1 explained 49.81% of the total variability and majorly influenced by plant weight, curd weight, and curd yield. PC2 with 22.62% contribution was highly associated with two traits: leaf length and plant height to extreme. In PC3, the number of leaves had a major impact. Traits such as plant weight, curd weight, curd yield, leaf length, plant height to extreme, and number of leaves contributed positively to these PCs cumulatively.
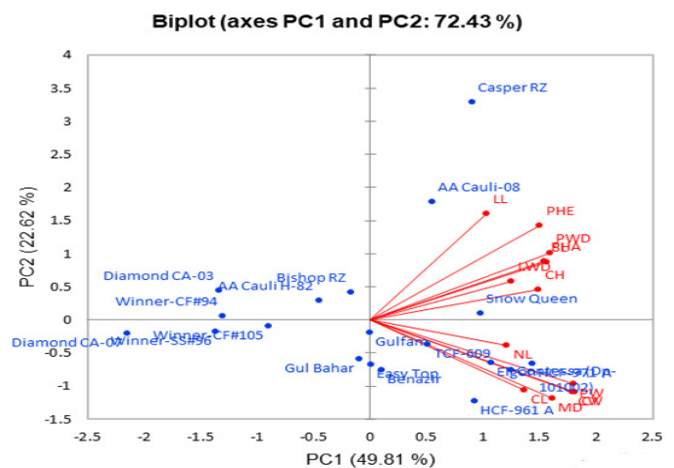
**Table 4:** *Principal component analysis for 13 morphological traits.*

|  | PC1 | PC2 | PC3 |
|---|---|---|---|
| Eigenvalue | 6.476 | 2.941 | 1.096 |
| Variability (%) | 49.814 | 22.620 | 8.432 |
| Cumulative % | 49.814 | 72.434 | 80.866 |
| **Traits** | **Eigen vectors** | | |
| Maturity days (days) | 0.294 | -0.320 | -0.112 |
| Plant weight (kg) | 0.328 | -0.261 | 0.009 |
| Curd weight (kg) | 0.326 | -0.295 | -0.017 |
| Curd yield (t/ha) | 0.331 | -0.293 | -0.023 |
| Stem length (cm) | 0.284 | 0.234 | -0.015 |
| Number of leaves (pieces) | 0.221 | -0.106 | 0.610 |
| Leaf length (cm) | 0.187 | 0.433 | 0.222 |
| Leaf width (cm) | 0.227 | 0.157 | -0.455 |
| Plant height to extreme (cm) | 0.273 | 0.384 | 0.152 |
| Plant height to apex (cm) | 0.281 | 0.240 | -0.146 |
| Plant width (cm) | 0.290 | 0.272 | 0.258 |
| Curd height (cm) | 0.272 | 0.125 | -0.482 |
| Curd length (cm) | 0.250 | -0.289 | 0.123 |

*Characters with high coefficients in PC axes considered more important, thus eigen values above 0.30 are shown.*

### Biplot analysis

To check the diversity among selected genotypes they were plotted on biplot regarding the first two PCs that had Eigenvalue greater than two and contributing 72.43% variability. Genotypes that are closely located on biplot, perceived as alike when rated on given attributes. More the distance between the point of origin and genotype, more diverse will be the genotypes from others. Regarding the first two PCs, genotypes were differentiated into five diverse groups (Figure 2). On biplot, two genotypes i.e. Casper RZ and AA-Cauli-08 clogged far away from the origin and were considered as diverse from the others (Figure 2). Remaining genotypes were clustered into three groups. Genotypes i.e. Gulfam, Gulbahar, Easy Top, TCF-609, and Benazir were clustered together and closer to origin as well hence, these genotypes are less diverse and have less breeding value. Genotypes i.e. Elgon, Contessa DP (101002), HCF-961A and HCF-971A formed the third group while genotypes Diamond CA-03, Winner CF#105, Winner SS #96, Diamond CA-07, Winner CF#94, AA- Cauli H-182 and Bishop Rz were differentiated from rest of the genotypes and clustered in fourth group. The fifth group had only one genotype i.e. snow queen (Figure 2).



**Figure 2:** *Biplot among first two PCs representing genetic diversity.*

### Cluster analysis

The principal component analysis clearly explained that some of the studied genotypes were diverse from others while many are similar and positioned near to each other on biplot. But clear-cut grouping of these genotypes not occurred by PCA. So to differentiate the genotypes on the basis of similarity and differences, they were subjected to cluster analysis based on Euclidean distance among the phenotypic characters. Cluster analysis grouped the 19 genotypes into 3

distinct clusters as presented in dendogram (Figure 3). Cluster 1 was the largest cluster and contained ten genotypes HCF-971 A, HCF-961 A, Contessa (Dp-101002), Elgon, Snow Queen, TCF-609, Benazir, Gul Bahar, Easy Top, and Winner-CF#94. Cluster 2 had seven genotypes, Bishop RZ, Gulfam, AA Cauli H-82, Winner-SS#96, Winner-CF#105, Diamond CA-03, Diamond CA-07. Cluster 3 contained only two genotypes AA Cauli-08 and Casper RZ (Table 5).
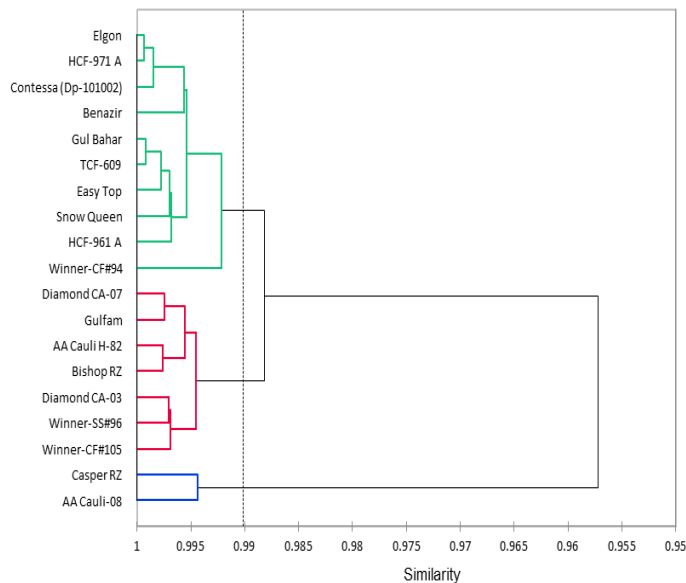


**Figure 3:** *UPGMA dendogram based on Euclidean distance matrix constructed from 19 phenotypic traits data of 19 cauliflower genotypes.*

Cluster means also showed significant differences for all the studied traits as shown in Table 5. Highest mean value for maturity days (176.50), plant weight (1.41 kg), curd weight (0.78 kg), curd yield (23.49 t/ha), and curd length (22.17 cm) was seen in cluster one. In cluster 2, minimum mean values were recorded for all the studied traits. Cluster 3 contains the highest value for stem length (21.83 cm), leaf length (72.25 cm), leaf width (29.33 cm), plant height to extreme (93.16 cm), plant height to apex (397.83 cm) and plant width (122.50 cm) (Table 6).

Genetic distance between the parents is hypothesized as the origin of heterosis and correlated with diverse segregants in segregating generations (Dangi *et al.*, 2018; Kumar *et al.*, 2017; Rathinavel, 2019). Comprehensive knowledge regarding genetic diversity is required to decide breeding strategies. Here, the genetic diversity among the cauliflower genotypes was studied by using morphological traits. Cauliflower morphological traits are complex and show redundancy as most of these traits were intercorrelated (Zhu *et al.*, 2018b).

**Table 5:** *Clustering of genotypes according to given morphological traits.*

| Cluster 1 | Cluster 2 | Cluster 3 |
|---|---|---|
| HCF-971 A | Bishop RZ | AA Cauli-08 |
| HCF-961 A | Gulfam | Casper RZ |
| Contessa (Dp-101002) | AA Cauli H-82 | |
| Elgon | Winner-SS#96 | |
| Snow Queen | Winner-CF#105 | |
| TCF-609 | Diamond CA-03 | |
| Benazir | Diamond CA-07 | |
| Gul Bahar | | |
| Easy Top | | |
| Winner-CF#94 | | |

**Table 6:** *Phenotypic characterization of three clusters.*

| Characters | Cluster 1 | Cluster 2 | Cluster 3 |
|---|---|---|---|
| Maturity days (days) | 176.50 | 137.86 | 140.00 |
| Plant weight (kg) | 1.41 | 0.81 | 0.85 |
| Curd weight (kg) | 0.78 | 0.42 | 0.44 |
| Curd yield (t/ha) | 23.49 | 12.83 | 13.05 |
| Stem length (cm) | 17.93 | 13.38 | 21.83 |
| Number of leaves (pieces) | 20.57 | 19.05 | 20.33 |
| Leaf length (cm) | 53.33 | 52.51 | 72.25 |
| Leaf width (cm) | 27.55 | 25.29 | 29.33 |
| Plant height to extreme (cm) | 69.70 | 62.52 | 93.17 |
| Plant height to apex (cm) | 30.27 | 26.28 | 37.83 |
| Plant width (cm) | 100.18 | 92.86 | 122.50 |
| Curd height (cm) | 16.77 | 15.33 | 19.17 |
| Curd length (cm) | 22.17 | 19.95 | 19.83 |

Principal component analysis can reduce this redundancy by simplifying the phenotypic traits into several principal components and providing an opportunity to select PCs with traits that are contributing greater to variation (Smith, 2002). From a total of 13 PCs, only the first 3 PCs were contributing more to the total variation. (Kumar *et al.*, 2017) found the first 5 PCs as most informative when they evaluate 57 cauliflower genotypes through PCA. In the present study, PCA revealed that plant weight, curd weight, curd yield, leaf length, plant height to extreme, and number of leaves were more prevalent traits than others in all 3 selected PCs and contributed 80.86% to variation. As these traits contributed more to the selected PC's variation so, are the main traits for evaluating gentic variation of cauliflower genotypes. The results of PCA are congruent with the (Kumar *et al.*, 2017), as plant weight, curd weight, and curd yield

traits have the highest variations and more prevalent in the first five PCs variation. A similar observation was also found by (Zhu *et al.*, 2018b) while evaluating 165 cauliflower inbred lines.

PCA is also used to identify genetically diverse genotypes. The first two PCs were used for plotting biplot. (Zhu *et al.*, 2018b) also used this criterion for selecting PCs that are contributing more to variation. Biplot between PC1 and PC2 differentiated the genotypes based on genetic variation exist between them as all studied nineteen genotypes were differentiated into five groups on PC1 and PC2. Few genotypes were positioned far away from each other while some positioned near to each other showing that these genotypes had narrow genetic base.

Cluster analysis is another important multivariate analysis that is useful for finding the phylogenetic relationship and choosing genetically diverse genotypes for getting desirable recombinants. In this study, 19 cauliflower genotypes were grouped into 3 clusters (Figure 2, Table 3). Cluster 3 was far away from cluster 1 so the genotypes of this cluster are the highly diverse. Similarly, genotype Elgon was far away from AA-cauli-08 in the dendrogram, so both these genotypes were considered highly diverse. It was observed that the genotypes i.e. Elgon, HCF-971A, CONTESSA-DP101002, TCF-609, Benazir, Gulbahar, Snow queen, and Winner-CF#94 imported by different seed companies were similar to each other and genetically less divers. Similarly, the genotypes of the same seed company were also much similar and grouped into one cluster such as Seed and seed processing genotypes Winner-CF#105 and Winner-SS#96 and Suncorp (Pvt) Ltd genotypes named Diamond CA-03 and Diamond CA-07. These results showed the narrow genetic base of these exotic cauliflower genotypes. The narrow genetic base of cauliflower genotypes that were collected from similar regions was also observed by (Yousef *et al.*, 2018; Zhao *et al.*, 2014; Zhu *et al.*, 2018a, b)

Characterization of clusters regarding all the morphological traits showed that cluster 1 had late maturity, highest curd weight, yield, and vegetative growth. Cluster 2 genotypes were the type of early maturing, having lowest curd, plant weight, stem length leaf length, leaf width, plant height to extreme, plant height to apex, plant width, and curd height. Cluster 3 showed medium maturing habit, curd

weight, and medium yield than other clusters. This suggested that improvement in cauliflower yield and curd quality characters could be succeeded by selecting parents from the diverse cluster. As an example, based on cluster mean data, it is predicted that a cross between genotypes of cluster 2 and cluster 1 might results in transgressive segregants for curd weight, and short maturity traits. Choosing genotypes from diverse clusters was expected to exploit maximum level heterosis in hybrid breeding and these genotypes might be used for creating variation in segregating conditions. But, the narrow genetic base of the studied genotypes might be creating a serious concern regarding their utilization in breed program.

## Conclusions and Recommendations

The genotypes explored in this study exhibited a narrow to wide range of genetic diversity for most of the phenotypic traits. The principal component analysis exhibited that some of the genotypes were highly diverse regarding the traits that have a high proportion in variation while some genotypes are similar in nature. This similar findings were also confirmed by cluster analysis where the genotypes were grouped into 3 clusters based on phenotypic data. As diverse genotypes were far away from each other in dendogram and fall in separate cluster while genotypes with narrow genetic base were grouped in similar cluster. Further, different clusters displayed variations in traits such as maturity days, curd yield, curd weight, plant weight, stem length, and other morphological traits. It is concluded that both genetic variation and characteristics of the parents must be considered while choosing them for the breeding program.

## Novelty Statement

Information regarding the genetic diversity of late season cauliflower genotypes has been presented in the paper which were not previously studied.

## Author's Contribution

**Saba Aleem:** Conceived the idea, wrote abstract, methodology, did statistics analysis, wrote conclusion.
**Mehvish Tahir:** Collected data helped in writing introduction, references.
**Iram Sharif:** Collected data, helped in writing introduction, result, discussion, and references.

**Muqadas Aleem:** Helped in writing introduction, result, discussion, and references.

**Muhammad Najeebullah and Ali Nawaz:** Technical Input at every step.

**Amina Batool:** Helped in writing result and discussion, introduction, references.

**Muhammad Imran Khan and Waheed Arshad:** Overall management of the article.

*Conflict of interest*

The authors have declared no conflict of interest.

## References

Agrama, H. and M. Tuinstra. 2003. Phylogenetic diversity and relationships among sorghum accessions using SSRs and RAPDs. Afr. J. Biotechnol., 2: 334-340. https://doi.org/10.5897/AJB2003.000-1069

Baranwal, D., V. Mishra and T. Singh. 2013. Genetic diversity based on cluster and principal component analyses for yield and its contributing characters in wheat (*Triticum aestivum* L.). Madras Agric. J., 100: 320-323.

Carter, R.L., R. Morris and R.K. Blashfield. 1989. On the partitioning of squared Euclidean distance and its applications in cluster analysis. Psychometrika, 54: 9-23. https://doi.org/10.1007/BF02294446

Chakma, S.P., H. Huq, F. Mahmud and A. Husna. 2012. Genetic diversity analysis in rice (*Oryza sativa* L.). J. Pl. Breed. Genet., 25: 31-39. https://doi.org/10.3329/bjpbg.v25i1.17010

Chatterjee, S., S. Sharma, D. Mukherjee and S. Patil. 2018. Estimation of genetic divergence in some mid late and late cauliflower (*Brassica oleracea* Var. *botrytis* L.) germplasm. Biosccan, 13: 459-462.

Dangi, R., A. Kumar and A. Khar. 2018. Genetic variability, heritability, and diversity analysis in short day tropical onion (Allium cepa). Indian J. Agric. Sci., 88: 948-957.

Evgenidis, G., E. Traka-Mavrona and M. Koutsika-Sotiriou. 2011. Principal component and cluster analysis as a tool in the assessment of tomato hybrids and cultivars. Int. J. Agron., 2011: 7 pages. https://doi.org/10.1155/2011/697879

FAO, 2017. Food and agriculture organization corporate statistical database; Food and agriculture data; Production crop.

Jarwar, A.H., X. Wang, M.S. Iqbal, Z. Sarfraz, L. Wang, Q. Ma and F. Shuli. 2019. Genetic divergence on the basis of principal component, correlation and cluster analysis of yield and quality traits in cotton cultivars. Pak. J. Bot., 51: 1143-1148. https://doi.org/10.30848/PJB2019-3(38)

Javed, M., S.B. Hussain and M. Baber. 2017. Assessment of genetic diversity of cotton genotypes for various economic traits against cotton leaf curl disease (CLCuD). Genet. Mol. Res., 16: 1-12. https://doi.org/10.4238/gmr16019446

Kendall, M., 1975. Multivariate analysis, Charles Griffin.

Kumar, V., D.K. Singh, A. Panchbhaiya and N. Singh. 2017. Genetic divergence studies in mid-season cauliflower (*Brassica oleracea* var. *botrytis* L.) through Principal Component Analysis (PCA) and D2 analysis. Int. J. Chem. Stud., 5: 989-993.

Pahadi, P., M. Sapkota, D.B. Thapa and S. Pradhan. 2017. Cluster and principal component analysis for the selection of maize (*Zea mays* L.) genotypes. Int. J. Exp. Res. Rev., 9: 5-10.

Purugganan, M.D., A.L. Boyles and J.I. Suddith. 2000. Variation and selection at the CAULIFLOWER floral homeotic gene accompanying the evolution of domesticated Brassica oleracea. Gene, 155: 855-862.

Rathinavel, K., 2019. Agro-morphological characterization and genetic diversity analysis of cotton germplasm (*Gossypium hirsutum* L.). Int. J. Curr. Microbiol. App. Sci., 8: 2039-2057. https://doi.org/10.20546/ijcmas.2019.802.237

Ravishanker, S.K., D. Baranwal, A. Chatterjee and S. Solankey. 2013. Genetic diversity based on cluster and principal component analyses for yield and quality attributes in ginger (Zingiber officinale Roscoe). Int. J. Plant Breed. Gene., 7: 159-168.

Sharma, V.R., S. Malik, M. Kumar and A. Sirohi. 2018. Morphological classification of genetic diversity of garlic (*Allium sativum* L.) germplasm for bulb and yield-related traits using principal component analysis. Int. J. Curr. Microbiol. App. Sci., 7: 2016-2022. https://doi.org/10.20546/ijcmas.2018.706.238

Singh, H., V. Singh, R. Kumar, D. Baranwal and P. Ray. 2014. Assessment of genetic diversity based on cluster and principal component analyses for yield and its contributing characters in bitter

gourd. Indian J. Hortic. 71: 55-60.

Smith, L., 2002. Tutorial on principal component analysis chapter 3.

Swarup, V. and S. Chatterjee. 1972. Origin and genetic improvement of Indian cauliflower. Econ. Bot., 26: 381-393. https://doi.org/10.1007/BF02860710

Tahir, M., H. Rahman, R. Gul, A. Ali and M. Khalid. 2013. Genetic divergence in sugarcane genotypes. Am. J. Exp. Agric., 3: 102. https://doi.org/10.9734/AJEA/2013/2283

Tuhina-Khatun, M., M.M. Hanafi, M. Rafii Yusop, M. Wong, F.M. Salleh and J. Ferdous. 2015. Genetic variation, heritability, and diversity analysis of upland rice (*Oryza sativa* L.) genotypes based on quantitative traits. BioMed Res. Int., 2015: 7 Pages. https://doi.org/10.1155/2015/290861

Uptmoor, R., W. Wenzel, W. Friedt, G. Donaldson, K. Ayisi and F. Ordon. 2003. Comparative analysis on the genetic relatedness of Sorghum bicolor accessions from Southern Africa by RAPDs, AFLPs and SSRs. Theor. Appl. Genet., 106: 1316-1325. https://doi.org/10.1007/s00122-003-1202-7

Vanlalneihi, B., P. Saha, P. Kalia, S. Singh, N. Saha, A. Kundu, N. Singh and A. Bhowmik. 2019. Genetic and principal component analysis for agro-morphological traits, bioactive compounds, antioxidant activity variation in breeding lines of early Indian cauliflower and their suitability for breeding. J. Hortic. Sci. Biotechnol., pp. 1-13. https://doi.org/10.1080/14620316.2019.1627912

Yousef, E.A., T. Mueller, A. Boerner and K.J. Schmid. 2018. Comparative analysis of genetic diversity and differentiation of cauliflower (*Brassica oleracea* var. *botrytis*) accessions from two ex situ genebanks. PLoS One, 13: e0192062. https://doi.org/10.1371/journal.pone.0192062

Zhao, Z., H. Gub, X. Sheng, H. Yub, J. Wang, J. Zhao and J. Cao. 2014. Genetic diversity and relationships among loose-curd cauliflower and related varieties as revealed by microsatellite markers. Sci. Hortic., 166: 105-110. https://doi.org/10.1016/j.scienta.2013.12.024

Zhu, S., X. Zhang, Q. Liu, T. Luo, Z. Tang and Y. Zhou. 2018a. The genetic diversity and relationships of cauliflower (*Brassica oleracea* var. *botrytis*) inbred lines assessed by using SSR markers. *PLoS One*, 13: e0208551. https://doi.org/10.1371/journal.pone.0208551

Zhu, S., X. Zhang, Q. Liu, T. Luo, Z. Tang and Y. Zhou. 2018b. Phenotypic variation and diversity of cauliflower (*Brassica oleracea* Var. *Botrytis*) inbred lines. Int. J. Agric. Biol., 20: 1041-1048.