Journal of Oilseed Brassica, 9 (1) : 45-48, January 2018

Evaluation of advanced breeding lines of Indian mustard using principal component analysis

Manoj Kumar Duppala, Beena Nair*, KS Gowtham and Arti Rathod

AICRP on Mustard, College of Agriculture, Nagpur, Maharashtra, India
**Corresponding author: beenanair2007@rediffmail.com*
(Received: 16 October 2017; Revised: 20 November 2017; Accepted: 13 December 2017)

Abstract

The efficiency of a breeding program depends mainly on the direction of the correlation between yield and its component traits, and the relative importance of each trait contributing to seed yield. The interrelationship between eighteen quantitative and seven qualitative characters in 70 genotypes of Indian mustard were computed. Principal component analysis was utilized to examine the variation and estimated the relative contribution of various traits towards the total variability. In this study, component 1 had contribution from the traits such as siliqua beak length, siliqua texture, leaf hairiness, leaf colour, dentation of leaf margin, flower petal colour and seed colour which accounted 27.6% of the total variability. Leaf length, leaf width, plant height, main shoot length, primary branches plant^{-1}, number of siliqua plant^{-1}, seed yield plot^{-1}, seed yield plant^{-1} contributed 14.8% to the total variability in component 2. The remaining variability of 9.4%, 8.2%, 5.5%, 4.6%, and 4.4% was consolidated in components 3, 4, 5, 6 and 7 respectively, by various traits *viz.*, spikelet fertility, single plant yield, grain length and number of productive tillers. The cumulative variance of 74.5% of total variation among 24 characters was explained by the first seven axes. Thus, the results of principal component analysis revealed the high level of genetic variation and the traits contributing for the variation were identified. Factor analysis was used for understanding the data structure and traits relationship. Hence, this population can be utilized for trait improvement in breeding programs for the traits contributing to major variation.

Key words: *Divergence, factor analysis, genetic mustard, principal component analysis*

Introduction

In India, agriculture is one of the main sources of national income and occupation. Besides providing food to nation, contributes to market of industrial goods and earns foreign exchange. Agricultural development is an integral part of overall economic development. Food availability is a necessary condition for food security. India is more or less self sufficient in cereals but deficit in pulses and oilseeds. Due to increasing changes in consumption patterns, there is a need to increase crop diversification and improved allied activities. The growth of edible oil consumption and increasing population coupled with limited availability of oilseeds and shifting of acreage to other crops have resulted in continuous demand supply gap for edible oil which is being met by imports. In India, the oil obtained from rapeseed- mustard accounts for 2/ 3rd edible oil consumption in the country. The projected demand for oilseeds in India by 2020 is around 34 million tons which is to be met by rapeseed-mustard. The genetic variability has been the central dogma of plant breeding. An understanding of the genetic behaviour of various agro-morphological and yield contributing traits is required for efficient selection of genotypes and breeding for trait specific genotypes (Meena *et al*., 2017). Multivariate analytical techniques, which simultaneously analyze multiple measurements on each individual under investigation, are widely used in analysis of genetic diversity irrespective of the dataset. PCA is defined as "a method of data reduction to clarify the relationships between two or more characters and to divide the total variance of the original characters into a limited number of uncorrelated new variables" (Wiley, 1981). This will allow visualization of the differences among the individuals and identify possible groups. The reduction is achieved by linear transformation of the original variables into a new set of uncorrelated variables known as principal components (PCs). The first step in PCA is to calculate eigen values, which define the amount of total variation that is displayed on the PC axes. The first PC summarizes most of the variability not summarized by the first PC and uncorrelated with the first, and so on (Jolliffe, 1986). Because PCs are orthogonal and independent of each other, each PC reveals different properties of the original data and may be interpreted independently. In this way, the total variation in the original data set may be broken down into components that are cumulative. The proportion of variation accounted for by each PC is

expressed as the eigen value divided by the sum of the eigen values. The eigenvector defines the relation ofthe PC axes to the original data axes. Factor analysis was used to determine structural factors related to plant growth traits and yield components in Indian mustard (*B. juncea* L.).

Materials and Methods

The total 64 recombinant lines and six checks of Indian mustard were undertaken to find out the extent of variation among recombinant inbred lines and their deviation from check genotypes. The genotypes were evaluated during the *Rabi* season of 2016-17 using alpha lattice design with three replications at Agricultural Botany Section, College of Agriculture, Nagpur, India. This consisted of 7 blocks with 10 plots per block. Each plots consisted of three rows 3m long and 45 cm apart. The distance between plants on each row was 10cm, which were sufficient for statistical analysis. Crop management factors like land preparation, crop rotation, fertilizer, and weed control were followed as recommended for local area. All the plant protection measures were adopted to make the crop free from insects. Observations were recorded on five randomly selected plants in each replication for twenty five variables including eighteen quantitative and seven qualitative characters i.e. number of lobes leaf^{-1}, leaf length (cm), leaf width (cm), petal length (mm), petal width (mm), days to 50 % flowering (on plot basis), days to maturity (on plot basis), plant height (cm), main shoot length (cm), primary branches plant^{-1}, siliqua length (cm), siliqua beak length (mm), number of siliqua plant^{-1}, no of seeds siliqua^{-1}, 1000 seed weight (g), seed yield plot^{-1} (kg ha^{-1}) and seed yield plant^{-1} (g) and oil content (%), leaf hairiness, leaf colour, dentation of leaf margin, flower petal colour, siliqua texture, siliqua angle with main shoot and seed colour. PCA and PFA statistics was used in this study for computing genetic divergence for eighteen quantitative characters and seven qualitative characters. Principal factor analysis was carried out using principal component method. The principal component analysis and principal factor analysis was calculated by using online data analysis, NARS statistical portal provided by IASRI, New Delhi, India.

Results and Discussion

Principal component analysis (PCA) for the data revealed that, the first seven principal components had eigen values greater than one and altogether explained 74.45 % of the total accumulated variability (table 1). The first principal component had maximum eigen value i.e. 6.61. The second, third, fourth, fifth, sixth and seventh principal components had the eigen values of 3.6, 2.2, 2.0, 1.3, 1.1 and 1.05 respectively. The first principal component accounts for maximum proportion of total variability among all variables and remaining components account for progressively lesser amount of variation. The first principal component accounted for maximum variability i.e. 27.6 %. The second, third, fourth, fifth and sixth principal components explained 14.8 %, 9.4 %, 8.2 %, 5.6 %, 4.6 % and 4.4 % of the total variation, respectively. In the present study, principal factor analysis was carried out using principal component method. Initially the data were analyzed without any rotation but it failed to load all the variables. The failure of principal factor analysis without rotation to draw sensible conclusions prompted to go for analysis with rotation. Varimax method of orthogonal rotation (Kaiser, 1958) was utilized to rotate the factor axes. To select the relevant traits in various principal factors, the correlation values (>0.12) were considered as relevant for that principal factor.

Table 1: Total variance of different principal components

Principal Component	Eigen Value	Per cent Variability	Cumulative % Variability
1	6.6	27.6	27.6
2	3.6	14.8	42.4
3	2.2	9.4	51.8
4	2.0	8.2	59.9
5	1.3	5.6	65.5
6	1.1	4.6	70.0
7	1.1	4.4	74.5

Data revealed that all the 24 variables had high loadings on different principal factors and none of them was left after rotation of the principal factor axes. Moreover, it clearly grouped the similar type of variables by loading them together on a common principal factor. The data presented in table 2 revealed clearly that the first principal factor (PF) showed high loadings for siliqua beak length (0.449), siliqua texture (0.559), leaf hairiness (0.697), leaf colour (0.569), dentation of leaf margin (0.905), flower petal colour (0.907) and seed colour (0.434). The second principal factor enabled high loadings for leaf length (0.676), leaf width (0.550), plant height (0.691), main shoot length (0.463), primary branches plant^{-1} (0.202), number of siliqua plant^{-1} (0.763), seed yield plot^{-1} (0.770), seed yield plant^{-1} (0.834) could be designated as yield factor . These variables *viz.*, number of lobes leaf^{-1} (0.520), plant height (0.104), seed yield plot^{-1} (0.409), seed yield plant^{-1} (0.247), were found to have highly positive loadings but not the highest on PF3. The fourth principal factor had variables like number of lobes leaf^{-1} (0.549), days to maturity (0.109) and 1000 seed weight (0.541), whereas, PF5 showed high loading for siliqua length (0.255), no of

seeds siliqua^{-1} (0.313) and siliqua angle with main shoot (0.798). The sixth principal factor exhibited high loading of petal length (0.633). The traits like petal width (0.521) and days to 50 % flowering (0.136) showed factor loading on PF7.

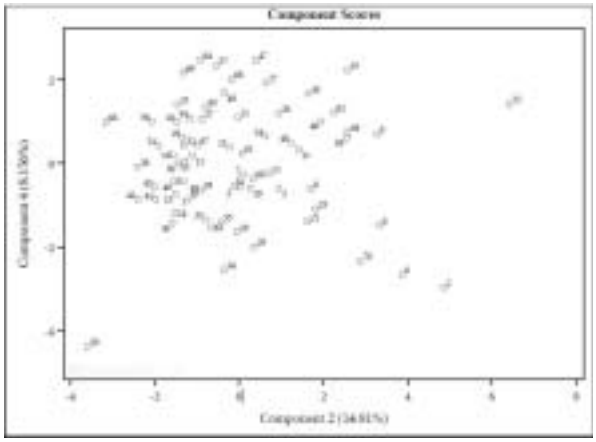

Fig 1: Distribution of mustard genotypes based on principal factor 2

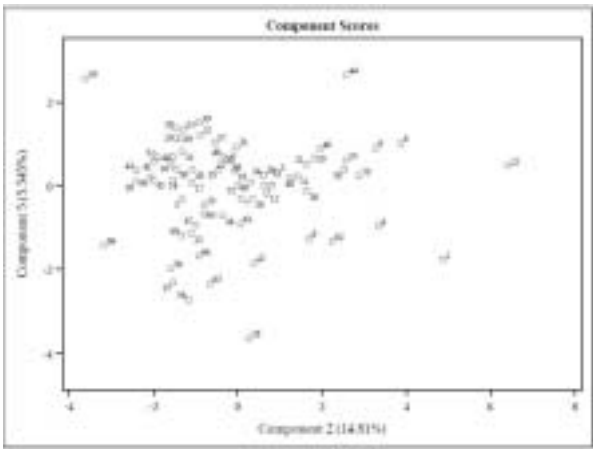

Fig 2: Distribution of mustard genotypes based on principal factor 2 and 5

Note: component = Principal factor and component score = Factor loading

In Fig. 1, all the genotypes were plotted for PF2 (seed yield and its important components) and PF4 (days to maturity and 1000 seed weight). The sequence order of corresponding genotypes from extreme right are ACN 162, ACN 141, ACN 146, ACN 144, ACN 149, Kranti, ACN 184, ACN 199, ACN 202, ACN 206, ACN 186, ACN 159, ACN 148, ACN 161, ACN 170 and ACN 151 which were found high yielding, stood out towards the positive portion of PF2 axis in the plot, whereas The sequence order of corresponding genotypes from extreme top are ACN 208, ACN 187, ACN 201, BIO 902, ACN 167, ACN 170 and ACN 151 which had days to maturity and 1000 seed weight clustered towards the positive side of PF4 axis. The genotypes which found place towards the positive end of both the factors are supposed to be superior collectively both for seed yield, days to maturity and 1000 seed weight, the traits of interest in oilseed crops. On the basis of present investigation, genotypes ACN 199, ACN 206, ACN 184, ACN 149 and ACN 203 have been identified superior for both the characters collectively.

In Fig. 2, all the genotypes were plotted for PF2 (seed yield and its important components) and PF5 (siliqua length and no of seeds siliqua^{-1}). The sequence order of corresponding genotypes from extreme right are ACN 162, ACN 141, ACN 146, ACN 144, ACN 149, Kranti, ACN 184, ACN 199, ACN 202, ACN 206, ACN 186, ACN 159, ACN 148, ACN 161, ACN 170 and ACN 151 which were found high yielding, stood out towards the positive portion of PF2 axis in the plot, whereas the sequence order of corresponding genotypes from extreme top are ACN 184, ACN 203, ACN 179, ACN 1967, ACN 150, ACN 175, ACN 165, ACN 169, ACN 146, ACN 201, ACN 171, ACN 149, ACN 186 and ACN 162 which had days to maturity and1000 seed weight clustered towards the positive side of PF5 axis. The genotypes which found place towards the positive end of both the factors are supposed to be superior collectively both for seed yield, siliqua length and no of seeds siliqua^{-1}, the traits of interest in oilseed crops. On the basis of present investigation, genotypes ACN 184, ACN 146, ACN 149, ACN 186, and ACN 152 have been identified superior for both the characters collectively.

The data thus revealed, the following genotypes are ACN 141, ACN 146, ACN 149, ACN 161, ACN 162 and ACN 206 were found high yielding, stood out towards the positive portion of PF axis. The clear cut grouping of similar type of variables by getting loaded on common principal factor elaborates the successful transformation of 24 inter related variables into 6 to 7 independent principal factors explaining about 70 % of the variability of the original set. The genotypes selected are the superior genotypes and these may be identified for collective improvement in mustard improvement programme and thus may be further evaluated in initial varietal trial to identify the best genotypes. Similar results of yield and yield contributing characters for the second principal factor have also been reported by Neeru *et al*. (2015), Rameeh (2014), Belete (2011) and Alemayehu *et al*. (2002).

Table 2: Factor loadings of characters with respect to different principal factors (varimax rotation)

Trait / Principal factor	PF1	PF2	PF3	PF4	PF5	PF6	PF7
Number of lobes leaf^{-1}	0.074	0.147	0.520	0.549*	0.121	0.345	0.111
Leaf length (cm)	0.097	0.676*	-0.496	0.062	-0.104	-0.013	0.039
Leaf width (cm)	0.239	0.550*	-0.499	-0.004	-0.194	-0.026	0.190
Petal length (mm)	-0.545	0.014	-0.139	-0.081	-0.261	0.633*	-0.177
Petal width (mm)	-0.572	0.083	0.049	0.354	0.148	0.070	0.521*
Days to 50 % flowering	-0.765	0.058	-0.035	0.091	-0.077	-0.161	0.136*
Days to maturity	-0.881	0.041	-0.009	0.109*	-0.016	0.088	-0.051
Plant height (cm)	0.110	0.691*	0.104	0.304	0.285	0.222	0.224
Main shoot length (cm)	0.252	0.463*	-0.499	-0.156	-0.303	-0.021	0.198
Primary branches plant^{-1}	-0.684	0.202*	0.122	0.202	-0.059	-0.211	-0.048
Siliqua length (cm)	-0.590	0.111	-0.491	0.214	0.255*	0.176	0.132
Siliqua beak length (mm)	0.449*	0.284	-0.414	0.373	0.087	-0.054	-0.354
Number of siliqua plant^{-1}	-0.250	0.763*	0.397	-0.249	0.025	-0.140	-0.084
No of seeds siliqua^{-1}	-0.364	-0.033	-0.410	-0.528	0.313*	0.131	-0.052
1000 seed weight (g)	0.412	0.079	-0.116	0.541*	0.059	0.229	-0.464
Seed yield plot^{-1} (Kg ha^{-1})	-0.178	0.770*	0.409	-0.191	0.114	0.034	-0.180
Seed yield plant^{-1} (g)	-0.162	0.834*	0.247	-0.233	0.001	-0.111	-0.192
Leaf hairiness	0.697*	0.120	-0.084	0.180	-0.163	0.004	0.194
Leaf colour	0.569*	0.127	-0.105	0.267	-0.006	-0.243	0.006
Dentation of leaf margin	0.905*	0.055	0.194	0.053	0.009	-0.012	0.013
Flower petal colour	0.907*	-0.063	0.032	-0.169	0.135	0.150	0.027
Siliqua texture	0.559*	0.070	0.367	-0.151	-0.317	0.118	0.316
Siliqua angle with main shoot	0.303	-0.015	-0.118	-0.107	0.798*	-0.169	0.108
Seed colour	0.434*	0.111	-0.011	-0.513	0.161	0.432	0.081

References

Alemayehu N and H Becker. 2002. Genotypic diversity and patterns of variation in germplasm material of Ethiopian mustard (*B. carinata* A. Braun). *Genet Resour Crop Evol* **49**: 573-582.

Belete YS. 2011. Genetic variability, correlation and path analysis studies in Ethiopian mustard (*B. carinata* A. Braun) genotypes. *Int J Plant Breed Genet* **5**: 328-338.

Jolliffe IT. 1986. Principal component analysis. Springer-Verlag, Berlin.

Kaiser HF. 1958. The varimax criterion for analytic rotation in factor analysis. *Psychometrika* **23**: 187.

Meena HS, Kumar A, Singh VV, Meena PD, Ram B and Kulshrestha S. 2017. Genetic variability and inter-relation of seed yield with contributing traits in Indian mustard (*B. juncea*). *J Oilseed Brassica* **8**: 131-137.

Neeru, Thakral NK, Avtar R and Singh A. 2014 Evaluation and classification of Indian mustard (*B. juncea* L.) genotypes using principal component analysis. *J Oilseed Brassica* **6**: 167-174.

Rameeh V. 2014. Multivariate regression analyses of yield associated traits in rapeseed (*B. napus* L.) genotypes. *J Adv Agric* **1**: 1-5.

Ray K, Dutta J, Banerjee H, Biswas R, Phonglosa A and Pari A. 2014 Identification of principal yield attributing traits of indian mustard [*B. juncea* (L.) Czern & Coss.] using multivariate analysis. *The Bioscan* **9**: 803-809.

Wiley EO, Warburton ML, Xianchun X, Crossa J, Franco J, Melchinger AE. 1981. Phylogenetics: The theory and practice of phylogenetics and systematics. John Wiley, New York.