

Original Research Article

<https://doi.org/10.20546/ijcmas.2021.1001.038>

## Assessing Diversity of Sesame Genotypes Using Cluster Analysis and Principal Component Analysis

Deepak Gupta<sup>1\*</sup>, Suresh Muralia<sup>2</sup>, Vikas Khandelwal<sup>3</sup> and Anju Nehra<sup>1</sup>

<sup>1</sup>Department of Plant Breeding and Genetics,  
SKN College of Agriculture, Jobner, India

<sup>2</sup>Agriculture Research Station, Navgaon, India

<sup>3</sup>ICAR-AICRP on Pearlmillet, Jodhpur, India

\*Corresponding author

### ABSTRACT

#### Keywords

Sesame, PCA, Genetic diversity, Genotypes, Inter cluster distance, Intra cluster distance

#### Article Info

Accepted:  
04 December 2020  
Available Online:  
10 January 2021

The present research work comprises twenty one genotypes of sesame were evaluated in a randomized complete block design (RCBD) with three replications at Agricultural Research Station, Navgaon (Alwar) during *Kharif* 2019. The presence of genetic divergence among these lines was observed by Mahalanobis's  $D^2$  statistic. All the accesses were grouped into five distinct clusters. The highest number of genotypes appeared in cluster I (14) followed by cluster II (3) and cluster III (2) while the lowest numbers of entries were reported in cluster IV (1) & V (1). Principal component analysis (PCA) indicated that the three principal components (PC1 to PC3) showed 67.43 per cent of the total variability. The first principal component had high positive loading for seed yield per plant, capsules bearing length and seeds per capsule which contributed more to the diversity. The second principal component had high loading for days to maturity, number of primary branches per plant and 1000-seed weight. The genotypes *viz.*, RT-346, GT-10 and EC-351880 were identified as the most diverse genotypes, based on component scores and these genotypes could be utilized as parents in hybridization programme.

### Introduction

Sesame is one of the most ancient oilseed crop and called queen of oilseed crops due to its stabilized keeping quality contributed by high degree of resistance to oxidation and rancidity (Bedigian and Harlan, 1986) and nearly 73% of the oil is used for edible purposes and preferred for cooking due to zero cholesterol, 8.3% for hydrogenization and 4.2% for industrial purposes in the manufacture of paints, pharmaceuticals and

insecticides because of its stability, anti-bacterial, anti-viral, anti-fungal and anti-oxidant properties (Pornparn *et al.*, 2009).

Tanzania produces highest production of sesame (14.6%) of the world followed by Myanmar (12.78%) and India (12.4%). In India, the productivity is very low particularly in Rajasthan. The major reasons for low productivity are lack of fast-adapting cultivars, capsule shattering, uneven ripening, poor crop stand establishment, lower fertilizer

responses, profuse branching, low harvest index, indeterminate growth habit and susceptibility to diseases (Tripathy *et al.*, 2019).

Sesame has been treated as less input intensive crop, the role of breeding improved varieties has been considered as a promising approach. Variability in terms of genetic divergence for agronomic traits is the key component of breeding programme for broadening the gene pool of sesame. Genetic distance estimates can be estimated by different methods i.e. principal component analysis (PCA), cluster analysis and discriminate analysis. Principal Component Analysis (PCA) analyzes a data in which observations are described by several inter-correlated quantitative dependent variables (Abdi and Williams, 2010) while, cluster analysis is a convenient method for identifying homogenous groups of objects called clusters. It is particularly useful in characterizing individual accessions, to detect duplications of genetic materials in germplasm collections, and as a general guide in selecting parents for crossing in breeding programs.

To boost the yield potential and genetic gain under low or no input condition, it is necessary to explore the existing genetic diversity. The present study was carried out to ascertain the nature and magnitude of genetic divergence among 21 sesame genotypes and to select better accessions for sesame improvement programme.

### **Materials and Methods**

Twenty one sesame genotypes were grown in a randomized block design (RBD) with three replications during *Kharif* seasons 2019 at Agricultural Research Station, Navgaon (Alwar) with standard cultural practices. Each plot consists of five rows of 4 m length and

30 cm apart. The experimental area occupied was quite uniform with respect to topography and fertility. The Agriculture Research Station, Navgaon-Alwar is situated at North East Corner of Rajasthan between 76°7'-28°2' N latitude. The average rainfall of the zone is 500 mm. The crop was grown under normal crop season. Ten quantitative characters i.e. days to 50 per cent flowering, days to maturity, plant height (cm), number of primary branches per plant, number of capsules per plant, capsules bearing length (cm), number seeds per capsule, 1000-seed weight, harvest index and seed yield per plant (g) were taken. The data collected were subjected to multivariate analysis utilizing Mahalanobis  $D^2$  statistic as suggested by Mahalanobis (1936) and Rao (1952) using statistical software TNAUSTAT. Accessions were grouped into five clusters following Tocher's method as suggested by Rao (1952). The necessity of the principal component analysis (PCA) for measuring the degree of divergence has been established by SPSS 16.0 version.

### **Results and Discussion**

In present investigation, the analysis of variation found considerable level of variability among different genotypes of sesame for all the characters (Table 1). The largest coefficient of variation was observed for harvest index (11.41%), number of capsules per plant (11.24%) and seed yield per plant (10.54%). The variation was found low for days to 50% flowering and days to maturity. This variability among the sesame genotypes could attribute to the inherent genetic differences or adaptation to local environmental condition.

The success from hybridization programme is expected from the combination of genetically diverse parents which have role in the heterotic expression and segregation potential

of crosses (Tripathy *et al.*, 2016). In the present study, 21 genotypes were grouped into five clusters based on Tocher's method (Table 2) on the basis of genetic diversity as measured by Mahalanobis D<sup>2</sup>-values.

Out of these, cluster I was the largest cluster comprising of 14 genotypes followed by cluster II consisting of 3 genotypes and cluster III including two genotypes. Rest two clusters were monogenotypic suggesting the existence of high degree of heterogeneity among the genotypes. This type of grouping

of different genotypes is imperative important for breeding program since, it clearly indicates the significance of a given character in any group.

The dendrogram also observed the relative magnitude of resemblance among the different clusters (Fig. 1). This supports the findings of Parameswarappa *et al.*, (2010) and Ahadu (2012) though the number of clusters differed depending on the number of genotypes and characters studied.

**Table.1** Variations in quantitative traits of 21 genotypes of sesame

Characters	Mean ± S.Em.	Coefficient of variance (%)	Range	
			Minimum	Maximum
Days to 50% flowering	52.68±0.99	3.25	49.00	56.33
Days to maturity	88.33±0.87	1.70	84.00	92.33
Plant height (cm)	96.37±5.15	9.25	82.67	114.33
Number of primary branches per plant	3.97±0.22	9.80	2.27	4.86
Number of capsules per plant	38.22±2.48	11.24	24.00	50.58
Capsules bearing length (cm)	38.03±2.16	9.84	23.87	46.58
Seeds per capsule	46.32±2.42	9.04	33.33	62.50
1000-seed weight (g)	2.81±0.13	8.07	2.11	3.33
Harvest index	21.96±1.45	11.41	16.34	27.05
Seed yield per plant (g)	3.03±0.18	10.54	2.32	4.59

**Table.2** Clustering of 21 genotypes for various characters in sesame

Clusters	Number of genotypes	Composition of cluster
<b>I</b>	14	<b>ES-22, ES-322, ES-237, ES-384-1-84, EC-362405, EC-351880, ES-26, ES-45. EC-370959, EC-359002, GT-3, ES-307, GT-10 and Pragati</b>
<b>II</b>	3	<b>ES-231-3-84, RT-346 and ES-46</b>
<b>III</b>	2	<b>ES-264 and TKG-22</b>
<b>IV</b>	1	<b>RT-351</b>
<b>V</b>	1	<b>MT-175</b>

**Table.3** Intra and inter cluster average distances among clusters of sesame genotypes following Tochers method

Clusters	I	II	III	IV	V
<b>I</b>	<b>27.77</b>	42.77	51.27	87.27	<b>80.81</b>
<b>II</b>		<b>34.27</b>	58.31	74.58	<b>129.32</b>
<b>III</b>			<b>32.88</b>	61.45	<b>97.17</b>
<b>IV</b>				<b>0.00</b>	<b>155.73</b>
<b>V</b>					<b>0.00</b>

**Table.4** Mean values for yield and its components in various clusters

Character	Clusters					Contribution (%)
	I	II	III	IV	V	
Days to 50% flowering	53.00	51.77	51.83	49.00	56.33	3.80
Days to maturity	88.76	88.67	86.67	84.00	89.00	6.19
Plant height (cm)	94.00	104.22	96.50	94.00	108.00	2.86
Number of primary branches per plant	4.12	4.40	3.63	2.67	3.03	12.38
Number of capsules per plant	40.89	35.89	31.73	35.07	24.00	10.95
Capsules bearing length (cm)	38.11	36.80	43.40	44.16	23.87	4.28
Seeds per capsule	44.50	48.49	57.95	55.00	33.33	8.57
1000-seed weight (g)	2.81	3.21	2.41	3.17	2.11	20.00
Harvest index	21.72	24.15	21.02	26.17	16.34	16.67
Seed yield per plant (g)	2.82	3.51	3.25	4.59	2.41	14.29

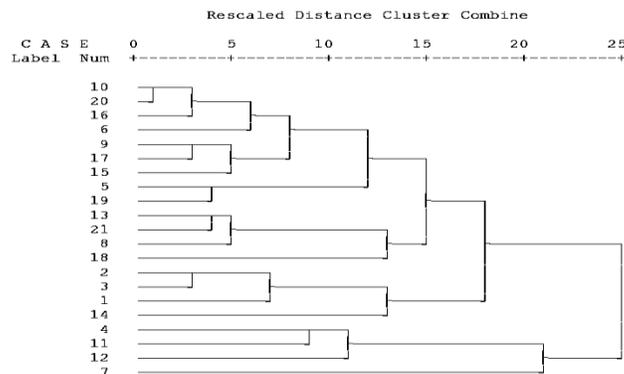
**Table.5** Eigen value and percent of total variation for various principal components

Component	Eigen values	Variance explained (%)	Cumulative (%)
<b>PC1</b>	3.529	35.29	<b>35.29</b>
<b>PC2</b>	1.896	18.96	<b>54.25</b>
<b>PC3</b>	1.323	13.23	<b>67.48</b>
<b>PC4</b>	0.958	9.58	<b>77.06</b>
<b>PC5</b>	0.723	7.23	<b>84.29</b>
<b>PC6</b>	0.587	5.87	<b>90.16</b>
<b>PC7</b>	0.407	4.07	<b>94.23</b>
<b>PC8</b>	0.322	3.22	<b>97.45</b>
<b>PC9</b>	0.144	1.44	<b>98.89</b>
<b>PC10</b>	0.111	1.11	<b>100</b>
<b>PC11</b>	<b>3.529</b>	<b>35.29</b>	<b>35.29</b>

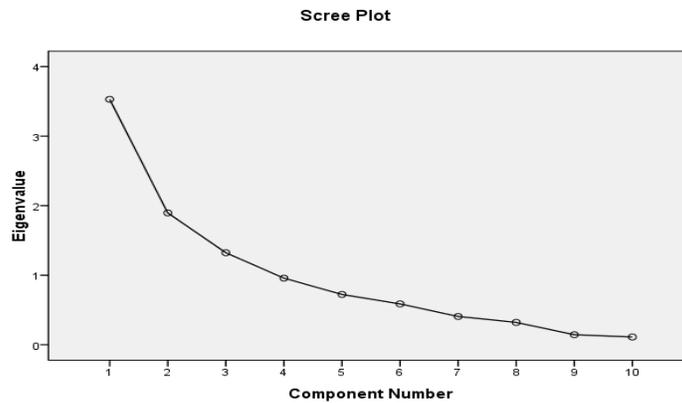
**Table.6** Factor loadings of nine characters with respect to different PC's (Principal components)

Characters	PC1	PC2	PC3
Days to 50% flowering	-0.804	-0.188	0.050
Days to maturity	-0.713	0.433	0.198
Plant height (cm)	-0.180	-0.291	0.668
Number of primary branches per plant	-0.377	0.657	0.267
Number of capsules per plant	0.186	0.756	-0.157
Capsules bearing length (cm)	0.786	0.255	-0.196
Seeds per capsule	0.760	-0.125	-0.154
1000-seed weight (g)	0.320	0.655	0.112
Harvest index	0.499	0.153	0.712
Seed yield per plant (g)	0.787	-0.227	0.396

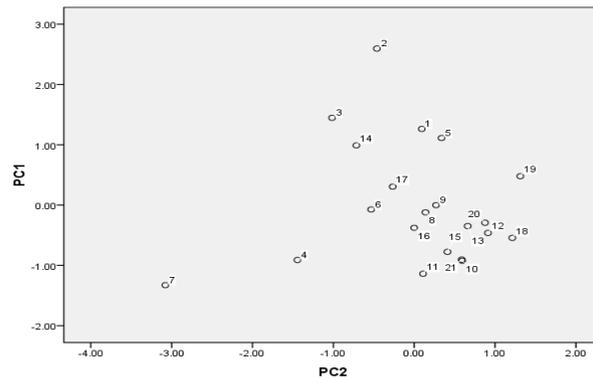
**Fig.1** Dendrogram based on 10 characters observed in sesame genotypes



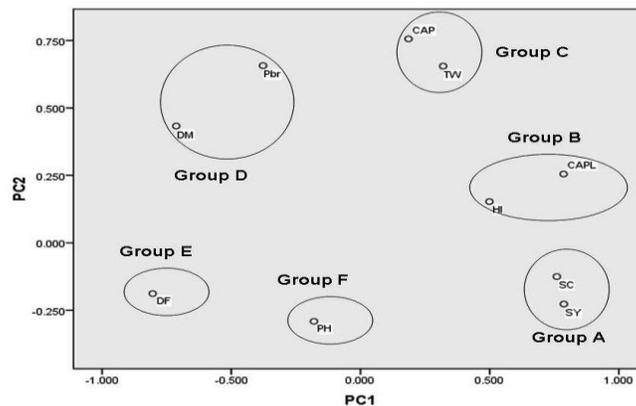
**Fig.2** Scree plot showing the Eigen value variation for ten quantitative traits of sesame



**Fig.3** Genotypes plotted for PC 1 and PC 2



**Fig.4** Plot of the first two PCAs indicating relation among various sesame traits



DF-Days to 50% flowering, DM-Days to maturity, PH-Plant height (cm), Pbr-Number of primary branches/plant, CAP- Number of capsules/plant, CAPL-capsules bearing length, SC-Number of seeds per capsule, TW-1000-seed weight (g), HI- Harvest index and SY-Seed yield per plant (g).

Inter and intra cluster  $D^2$  values were worked out from divergence analysis (Table 3) and revealed that the inter cluster distances were greater than the intra cluster distances indicating wider genetic diversity among the genotypes of different groups. An assessment of cluster showed that clusters were heterogeneous within themselves and between each other based on major character. The intra cluster distance ranged from 0.00 to 34.27 and the inter cluster distance ranged from 42.77 to 155.73, indicating that the selected genotypes were highly divergent. Maximum inter cluster divergence was observed between clusters IV and V (155.73) followed by clusters II and V

(129.32) indicating greater the distance between two clusters, wider the genetic diversity among the genotypes and good recombinants can be realized by mating between the genotypes, RT-351, ES-231-3-84, RT-346, ES-46 and MT-175 in a definite fashion. The lowest inter cluster divergence (42.77) was noticed between cluster I and II indicating that the genotypes included in them were closely related and hence, may not be emphasized upon to be used in hybridization programme. Accessions falling under distant clusters could be hybridized to get the higher heterotic responses. Crosses involving parents belonging to the most divergent clusters

would be expected to manifest maximum heterosis and wide variability of genetic architecture. Similar results were obtained by Hika *et al.*, (2015) and Dash *et al.*, (2018). Such grouping of genotypes from different genetic origin may be attributed to the unidirectional selection practiced by the breeder in a particular location (Singh and Bains, 1968). Murty and Arunachalam (1966) have reported that genetic drift and selection, both natural and artificial in different environments could cause greater diversity among genotypes than geographic distances. Probably due to above two factors, the genotypes originating from the different crosses have been clustered together.

The cluster mean values were calculated over genotypes for ten agro-morphological characters including seed yield per plant in sesame, which revealed that a wide range of variation (Table 4) existed among clusters. The genotype viz., RT-351 in cluster IV had maximum cluster mean values for the characters seed yield per plant (4.59 g), harvest index (26.17), capsule bearing length (44.16 cm), short plant height (94.0 cm), early flowering (49 days) and early maturity (84.0 days) and genotypes viz., ES-264 and TKG-22 in cluster III had maximum mean for seeds per capsule (57.95). These genotypes may be important for further breeding program to increase seed yield meanwhile cluster I could serve as a parent material in the improvement of capsule number as it possess highest capsule number than any other cluster. The variation observed in cluster mean also points to the degree of variability by Veni *et al.*, (2008) and Peluzio *et al.*, (2012).

The characters contributing maximum to divergence needs greater emphasis for deciding on the clusters for the purpose of selection of parents in the respective cluster for hybridization. Out of ten characters studied, five characters namely 1000-seed

weight (20%) followed by harvest index (16.67%), seed yield per plant (14.29%), number of primary branches per plant (12.38%) and number of capsules per plant (10.95%) contributed 74.29% towards total divergence (Table 4). Therefore, these characters should be given importance during selection of parents for hybridization and also in the follow-up selection of segregating populations. Similar observation was reported by Abate and Mekbib (2015) and Kiranmayi *et al.*, (2016). Thus, contribution of characters to divergence depends on the number of characters studied and the influence of the environment on the expression of characters.

PCA is a well-known method of dimension reduction that can be used to reduce a large set of variables to a small set that still contains most of the information in the large set (Jolliffe, 1986). Only the first 3 PCs showed Eigen values more than one, and they cumulatively explained 67.48% variability (Table 5). The first PC explained 35.29 per cent of the total variation, and the remaining 2 PCs explained 18.96 and 13.23 per cent variation, respectively. The first one absorbed and accounted for maximum proportion of total variability in the set of all PCs and the remaining ones accounted for progressively lesser and lesser amount of variation. Similar result has been reported earlier by Singh *et al.*, (2018) in sesame. Eigen values of 10 principal components are shown within the scree plot (Fig. 2).

The analysis without rotation of axes failed to load all the variables signifying that it could not offer much information regarding the idea of correlation between the variables and the principal components. Varimax Rotation was applied and this resulted in loading of all the variables on different principal components. Factors' loadings of different variables thus obtained are presented in Table 6. It revealed that the first principal component (PC1)

which accounted for the highest variation (35.29%) was mostly related traits are seed yield per plant, capsules bearing length and seeds per capsule. The second principal component (PC2) was dominated by traits *viz.*, days to maturity, number of primary branches per plant and 1000-seed weight while, PC3 consisted mainly of plant height and harvest index. Similar results were reported by Shim *et al.*, (2016) and Singh *et al.*, (2018).

In present investigation the results of hierarchical cluster analysis and PCA confirmed the findings of each other. Principal factor scores (PF scores) for all the 21 genotypes were estimated for all the three PCs. These scores may be utilized to construct precise selection indices based on variability exhibited by each of the principal factor. The genotypes were plotted for PC1 and PC2 which together explained 67.48 per cent variability and included the major seed yield characters (Fig. 3 & 4). Positive side of PC1 indicated the genotypes giving high seed yield per plant and number of seeds per capsule which were indicative as 2 (RT-351), 3 (TKG-22), 14 (ES-264) and 17 (ES-384-1-84). Similarly, positive values of PC 2 were 8 (ES-22), 10 (ES-45), 11 (ES-46), 12 (ES-231-3-84), 13 (ES-237), 15 (ES-307), 18 (ES-359002), 20 (EC-370959) and 21 (EC-362405) and these genotypes were combining for early maturing and number of primary branches per plant. The genotypes 1 (RT-346), 5 (GT-10) and 19 (EC-351880) scored high and positively both for PC1 and PC2 i.e. the genotypes giving number of capsules per plant, 1000-seed weight, harvest index and capsules bearing length contributing characters. These studies also suggest that it was possible to reduce large number of variables into only three principal factors and identify different lines better for different combinations of characters. Hence, indirect selection for seed yield based on component

traits may lead to create better genetic recombinants for improving yield and yield attributing characters in sesame.

## References

- Abate, M. and Mekbib, F. (2015). Study on genetic divergence in low-altitude sesame (*Sesamum indicum* L.) germplasm of Ethiopia based on agro morphological traits. *Journal of Advanced Studies in Agricultural, Biological and Environmental Sciences*. 2(3):78-90.
- Abdi, H. and Williams, L.J. (2010) Principal Component Analysis. Wiley Interdisciplinary Reviews: Computational Statistics, 2, 433-459.
- Ahadu, M.A. 2012. Phenotypic variability, divergence analysis and heritability of characters in sesame (*Sesamum indicum* L.) genotypes. *Nature Science*. 1: 117-26.
- Bedigian, D. and Harlan, J.R. 1986. Evidence for cultivation of sesame in the ancient world. *Economic Botany*. 40: 137-154.
- Dash, M., Haibru, G., Kabi, M., Pradhan, B., Lenka, D. and Tripathy, S.K. (2018). Morphology-based genetic diversity analysis of sesame germplasm for yield and capsule characters. *International Journal of Current Microbiology and Applied Sciences*. 7(9):1817-1826.
- Hika, G., Geleta, N. and Jalet, Z. (2015). Genetic variability, heritability and genetic advance for the phenotypic traits in sesame (*Sesamum indicum*. L) populations from Ethiopia. *Sci. Tech. & Arts Res. J.*, 4(1): 20-26.
- Jolliffe, I.T. (1986). Principal Component Analysis. Springer, New York.
- Kiranmayi, S.L., Roja, V., Padmalatha, K., Sivaraj, N. and Sivaramakrishnan, S. (2016). Genetic diversity analysis in sesame (*Sesamum indicum* L.) using morphological, biochemical and

- molecular techniques, *Int. J. Appl. Biol. & Pharmaceutical. Tech.*, 7(1):95-110.
- Mahalanobis, P.C. (1936). On the generalised distance in statistics. *Proceed. National Inst Sci, India.* 2: 49-55.
- Murty, B.R. and Arunachalam, V. (1966). The nature of genetic divergence in relation to breeding system in some crop plants. *Ind. J. Genet.*, 26: 188-198.
- Paremeshwarappa, S.G., Palakshappa, M.G., Salimath, P.M. and Paremeshwarappa, K.G. 2010. Studies on genetic variability and character association in germplasm collection of sesame (*Sesamum indicum* L.). *Karnataka J. Agric. Sci.*, 22(2): 252-254.
- Peluzio, J.M., Pires, L.P.M., Cancellier, L.L., Afferi, F.S., Colombo, G.A., Teixeira, Junior, T. (2012). Genetic divergence among soybean cultivars in irrigated lowland in the state of Tocantins. *Ciencia Rural*, 42(3): 395-400.
- Pornparn, S., Siree, S., Wimolrat, D. and Nongnuch, D. (2009). Fertilizers for organic sesame. *J. Food Agril-India*, special issue, Pp. 197-204.
- Rao, C.R. (1952). *Advanced Statistical Methods in Biometrical Research*, John Wiley and Sons Inc., New York.
- Shim, K.B., Shin, S.H., Shon, J.Y., Kang, S.G., Yang, W.H. and Hew, S.G. (2016). Classification of a collection of sesame germplasm using multivariate analysis. *J. Crop Sci. Biotech.*, 19: 151-155.
- Singh, A., Bisen, R. and Tiwari, A. (2018). Assessing inter: Relationship of sesame genotypes and their traits using cluster analysis and principal component analysis. *Int. J. Chem. Studies*, 6(1): 2151-2153.
- Singh, R.V. and Bains, S.S. (1968). Genetic divergence for ginning outturn and its components in upland cotton. *Ind. J. Genet.*, 28: 262-269.
- Tripathy, S.K., Kar, J. and Sahu, D. (2019). Advances in Sesame (*Sesamum indicum* L.) Breeding. In *Advances in Plant Breeding Strategies: Industrial and Food Crops*; Al-Khayri, J.M., Jain, S.M., Johnson, D.V., Eds.; Springer: Cham, Switzerland, pp. 577–635.
- Tripathy, S.K., Mishra, D.R., Senapati, N., Nayak, P.K., Dash, G.B. and Mohanty, S.K. (2016). Assessment of genetic divergence in sesamum based on morpho-economic characters. *Int. J. Curr. Res.*, 8(3): 27283-27287.
- Veni, B.K., Rani, B.P. and Ramana, M.V. (2008). Genetic divergence studies in soybean [*Glycine max* (L.) Merrill]. *Soybean Res.*, 6: 77-80.

#### **How to cite this article:**

Deepak Gupta, Suresh Muralia, Vikas Khandelwal and Anju Nehra. 2021. Assessing Diversity of Sesame Genotypes Using Cluster Analysis and Principal Component Analysis. *Int.J.Curr.Microbiol.App.Sci.* 10(01): 304-312. doi: <https://doi.org/10.20546/ijcmas.2021.1001.038>