Agriculture, Forestry and Fisheries
2022; 11(2): 67-75
http://www.sciencepublishinggroup.com/j/aff
doi: 10.11648/j.aff.20221102.11
ISSN: 2328-563X (Print); ISSN: 2328-5648 (Online)

Principal Component and Cluster Analyses for Quantitative Traits in Black Cumin (Nigella Sativa)

Yewubdinber Lemma[1], Wassu Mohammed[2], Sintayehu Adimas[1]

[1]Crop and Horticulture Directorate, Ethiopian Biodiversity Institute, Addis Ababa, Ethiopia

[2]School of Plant Sciences, Haramaya University, Dire Dawa, Ethiopia

Email address:

yewu.lemma@gmail.com (Y. Lemma), wasmoha@hotmail.com (W. Mohammed), sintayahu.2010@gmail.com (S. Adimas)

To cite this article:

Yewubdinber Lemma, Wassu Mohammed, Sintayehu Adimas. Principal Component and Cluster Analyses for Quantitative Traits in Black Cumin (Nigella Sativa). *Agriculture, Forestry and Fisheries.* Vol. 11, No. 2, 2022, pp. 67-75. doi: 10.11648/j.aff.20221102.11

Received: January 28, 2022; **Accepted**: March 8, 2022; **Published**: March 18, 2022

Abstract: Breeding for high yield crop needs information on the nature and magnitude of variation in the available materials, relationship of yield with other agronomic characters and the degree of environmental influence on the expression of these components characters. Multivariate analyses such as cluster and principal component analysis measures the amount of genetic variability in respect of several characters and assesses the relative contribution of different traits to the total variation. This study was conducted with the aim of identifying better performing black cumin genotypes and related traits with the help of principal component analysis and cluster analysis of major quantitative traits of the crop *In principal component analysis, The first four principal component axes (PCAs) accounted 70.62% of the total variability in which PCA1 contributed 28.43%, and PCA2, PCA3, and PCA4 exhibited 18.91%, 13.30%, and 9.98% contribution to the total variability, respectively.* Cluster analysis based on Euclidian distance grouped the genotypes into 10 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 highly 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 variety, and broaden the genetic base of the black cumin germplasm.

Keywords: Principal Component, Cluster Analyses, Nigella Sativa, Genetic Divergence

1. Introduction

Nigella sativa L. is commonly known as black seed or black cumin. It is an annual plant which belongs to a family of Rananculacea. The species is originated in Egypt and East Mediterranean. The crop is cultivated in many countries in the world like Middle Eastern, South Europe, Turkey and Syria. Several scholars have expressed conflicting views on the origins of black cumin [1, 2].

Black cumin seed contains mucilage, alkaloids, essential amino acids, tannins, resins, crude fiber, saponins, minerals, vitamins, fatty oils and essential oils [3]. Black cumin oil is excellent oil being used in nutritional and pharmaceutical applications due to some active components present such as tocopherol, phenolic compounds, and thymoquinone believed to be responsible for beneficial effects to human health including antioxidant [4]. The pharmacological properties of black cumin and its ingredients had been investigated by in vitro and in vivo studies conducted on human and laboratory animals. The investigations of the seed extracts reveal abroad spectrum of activities including galactagogic and insect repellant effects [5, 6] and is essentially used for wound-healing cases and improvement of respiratory and nervous systems [7].

The major resource of plant breeders is the genetic variability in gene pool accessible to the crop of interest [8]. The successes of crop improvement programs are highly reliant on the efficient manipulation of that genetic variability. Grouping the existed genotypes based on morphological characters enables breeder to exploit existed genetic resources for further breeding programs. It is real that knowing relationship among the existed landraces enables breeders to develop extreme phenotypes for cultivar development. It is also important to knowing the major traits contributing for the total observed variations among the landraces. This entails

breeders to focus on specific characters to develop varieties for specific environmental conditions allowing answering priority community problems. Therefore this study was organized to estimate the magnitude of genetic distance and to identify the major traits contributing for the observed variations among the studied genotypes.

This study was conducted with the aim to identify better performing black cumin genotypes with the help of principal component analysis and cluster analysis of major quantitative traits of the crop. The identification of such genotypes with superior traits could help in making the selection of good performer convenient and the planning of further breeding strategies effective.

2. Materials and Methods

2.1. Description of the Study Site

The study was conducted at Debre-Zeit Agriculture Research Center (DZARC) in 2018/19 main cropping season (rainy season). Debre-Zeit is located at 08° 44' N latitude and 38° 58'E longitude at a distance of 47 km from Addis Ababa on the Southeast direction at an altitude of 1860 m.a.s.l. The site received an annual rainfall of 1151.6 mm and area has minimum and maximum temperature of 19.03 and 26.91°C, respectively.

2.2. Experimental Materials and Design

A total of 25 genotypes were evaluated. The genotypes were collected from Oromiya Regional State by Ethiopian Biodiversity Institute. Genotypes were evaluated in 5 x 5 simple lattice designs. The seeds were drilled in rows spaced 30 cm each other and a distance of 15 cm between plants in each row. Each genotype was assigned in a plot of 3 m long and 1.80m wide (3 x 1.8 m = 5.4 m^2). The spacing between two plots within the same block was 50cm and the spacing between blocks and replications was 1 and 2 m, respectively. The layout and randomization were done as per the standard procedure set by [9].

Table 1. *List and description of black cumin genotypes.*

No.	Accession code	Collection area		Coordinate		Altitude
		Zone	District	Latitude	Longitude	(m.a.s.l)
1	20785	Arssi	Sherka	07-35-49-N	39-34-20-E	2313
2	30749	Bale	Ghiner	07-03-50-N	40-40-07-E	1816
3	30752	Bale	Ghiner	07-05-13-N	04-39-22-E	1902
4	30753	Bale	Ghiner	07-06-29-N	40-38-26-E	1957
5	30754	Bale	Ghiner	07-06-29-N	40-38-26-E	1957
6	30755	Bale	Ghiner	07-10-50-N	40-43-02-E	Notknown
7	30759	Bale	Ghiner	07-09-21-N	40-41-07-E	1951
8	30760	Bale	Ghiner	07-09-21-N	40-41-07-E	1951
9	30762	Bale	Ghiner	07-09-21-N	40-41-07-E	1951
10	30763	Bale	Ghiner	07-09-21-N	40-41-07-E	1951
11	30764	Bale	Ghiner	07-09-21-N	40-41-07-E	1951
12	30765	Bale	Ghiner	07-09-49-N	40-38-52-E	2008
13	30766	Bale	Ghiner	07-09-49-N	40-38-52-E	2008
14	30767	Bale	Ghiner	07-04-32-N	40-34-25-E	1653
15	30768	Bale	Ghiner	07-04-32-N	40-34-25-E	1653
16	30769	Bale	Ghiner	07-04-32-N	40-34-21-E	1653
17	30770	Bale	Goro	07-04-32-N	40-34-25-E	1653
18	30771	Bsale	Goro	07-04-32-N	40-34-25-E	1653
19	30772	Bale	Goro	07-04-32-N	40-34-25-E	1653
20	30773	Bale	Goro	07-04-32-N	40-34-25-E	1653
21	30781	Bale	Goro	07-04-32-N	40-34-25-E	1653
22	30782	Bale	Goro	07-04-32-N	40-34-25-E	1653
23	30783	Bale	Goro	07-04-32-N	40-34-25-E	1653
24	30784	Bale	Goro	06-54-24-N	40-39-27-E	1748
25	30786	Bale	Goro	06-54-24-N	40-39-27-E	1648

2.3. Data Collection

The data were collected both from whole plot and ten randomly selected samples of plants and the data were recorded on the following parameters: Days to 50% emergence, days to flowering, days to 50% flowering, flowering duration, days to 90% maturity, plant height, thousand grain weight, grain yield number of capsule, number of seed per capsule, number of primary and secondary branch, oleoresin and essential oil content. Oleoresins and essential oils extraction and determination was made following the standard methods of [10]. The percentage oil yield was calculated as suggested by [11] as follow.

$$\text{Percentage yield of oil (\%)} = \frac{\text{mass of oil}}{\text{mass of the seed sample}} * 100\%$$

2.4. Data Analysis

2.4.1. Principal Component Analysis

Principal component analysis (PCA) was computed to find out the characters, which accounted more to the total variation. The data was standardized to mean zero and variance of one before computing principal component analysis. The principal component based on the correlation matrix was calculated using SAS software version 9.0 [12].

2.4.2. Genetic Distance and Clustering

Euclidean distance (ED) was computed from all data collected for black cumin accessions after standardization (subtracting the mean value and dividing it by the standard deviation) as:

Pythagorean Theorem formula (Source: [13])

$$EDjk = \sqrt{\sum_{i=1}^{n}(Xij - Xik)^2}$$

Where $EDjk$ = distance between accessions j and k; Xij and Xik = phenotype traits values of the ith character for genotypes j and k, respectively; and n = number of phenotype traits used to calculate the distance. The distance matrix from phenotype traits was used to construct dendrogram based on the Unweighted Pair-Group Method with Arithmetic Means (UPGMA). The results of cluster analysis were presented in the form of dendrogram. In addition, mean ED was calculated for each genotype by averaging of a particular genotype to the other 49 genotypes. The calculated average distance (ED) was used to estimate which genotype(s) are closest or distant to others.

3. Results and Discussion

3.1. Principal Component Analysis

Principal component analysis results of 12 quantitative traits are presented in Table 2. The first four Principal component axes (PCA1 to PCA4) accounted varied percentage of total variance that ranged from 9.98% to 28.43% and accounted for 70.62% of the total variation with eigenvalues ranged from 1.20 to 3.41. The PCA1 contributed 28.43% to the total variability among genotypes, while PCA2, PCA3, and PCA4 exhibited 18.91%, 13.30%, and 9.98% contribution to the total variability, respectively.. Principal component analysis (PCA) is one of the multivariate statistical techniques which are a powerful tool for investigating and summarizing underlying trends in complex data structures [14]. This analysis reflects the importance of the largest contributor to the total variation at each axis for differentiation [15].

The first principal component axis (PCA1) contributed relatively higher proportion to total variability. Days to first flowering (24.44%), days to 50% flowering (23.50%), days to 90% maturity (21.64%) and plant height (9.02%) had higher contribution to PCA1. Number of secondary branches (29.72%) followed by number of primary branches, number of capsule and number of seed per capsule with had 23.54%, 21.52% and 10.99%, respectively, had larger contribution to PCA2. Oleoresin content of seeds (23.10%) had relatively more contribution to the total variance of PCA3 followed by plant height (18.47%), seed yield (16.71%), and number of seed per capsule (13.34%) and also thousand seed weights (39.33%), had relatively more contribution to the total variance of PCA4. The traits that had more contribution to the total variance of each PCA suggested these traits they were responsible for the differentiation of the black cumin genotypes into different groups.

Table 2. *Contribution of 12 quantitative traits in first four principal component axes for 25 black cumin genotypes evaluated at Debre-Zeit in 2018/19.*

Trait	PCA1	PCA2	PCA3	PCA4
Day to first flowering	0.913	-0.095	-0.303	0.044
Day to 50% flowering	0.895	-0.277	-0.274	0.076
Day to 90% maturity	0.859	-0.316	-0.298	-0.083
Plant height (cm)	0.555	0.07	0.543	-0.35
Number of primary branches	0.452	0.731	-0.109	0.132
Number of secondary branches	0.073	0.821	-0.129	-0.344
Number of capsules/plant	0.394	0.699	-0.037	0.188
Number of seeds per capsule	0.132	0.499	0.461	0.384
Thousand seed yield (g)	0.294	-0.183	0.409	0.686
Seed yield (kg ha^{-1})	0.322	0.015	0.516	-0.50
Oleoresin (%)	0.328	-0.071	0.607	-0.033
Essential oil (%)	0.225	-0.061	0.018	-0.141
Eigen value	3.41	2.27	1.6	1.2
Contribution to variability (%)	28.43	18.91	13.3	9.98
Cumulative %	28.43	47.34	60.64	70.62

PCA1, first principle component axis; PCA2, second principle component axis; PCA3, third principle component axis; and PCA4, fourth principle component axis, and number in parenthesis indicated the contribution of trait to the PCAs.

The biplot to genotype-trait that showed the contributions of genotypes and traits to PCA1 and PCA2 are presented in Figure 1. The genotypes and the quantitative traits were revealed on a biplot to clearly visualize their associations and differences. PCA1 and PCA2 biplot explained the 47.34% of total variability among the genotypes, showing that day to first flowering, days to 50% flowering, day to 90% maturity, plant height, number secondary branch, number of primary branch, and number of capsule per plant were considered as the most discriminating traits. The genotypes that were found on the right top quadrant were closely associated and characterized by high number of primary branch, number of secondary branch and number of capsule. The genotypes separate on the bottom right quadrant were associated with late maturing genotypes. The genotypes concentrated around the origin had similar genetic characteristics, while the genotypes that were found far from the origin are considered as unrelated genotypes [16, 17].

The biplot to genotype-trait that indicated the contributions of genotypes and traits to PCA3 and PCA4 are presented in Figure 2. These two PCA account 23.28% to the total variability among genotypes, showing thousand seed weight, number of seed per capsule, oleoresin, plant height and seed yield were the most contributing traits. These PC3 and PC4 biplot provided important information regarding the similarities as well as the pattern of differences among the black cumin genotypes and of the interrelationships between traits.

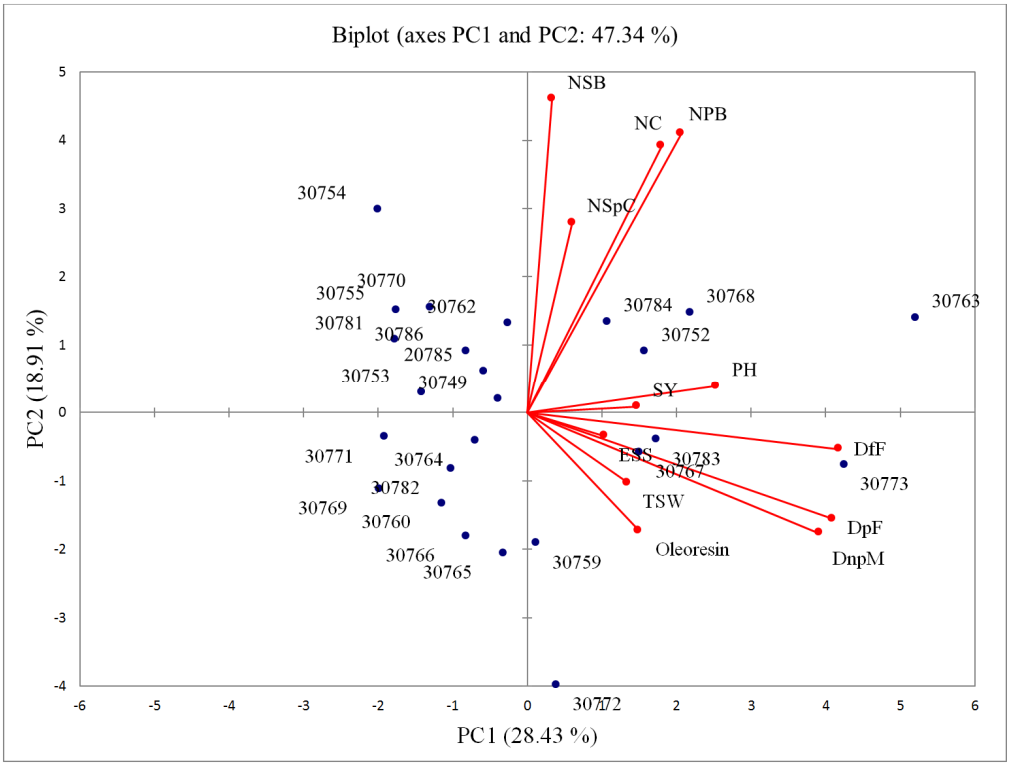

Figure 1. *Biplot of genotype –trait indicated contribution of genotypes and traits to PCA1 and PCA2 for 12 traits of 25 black cumin genotypes evaluated at Debre-Zeit 2018/19.*

DfF = Days to first Flowering, DpF = Days to 50% Flowering, DnpM=Days to 90% Maturity, TSW= Thousand Seed Weight, SY=Seed Yield, PH=Plant Height, NPB=Number of Primary Branch, NSB= Number of Secondary Branch, NC=Number of Capsule, NSpC=Number of Seed per Capsule.

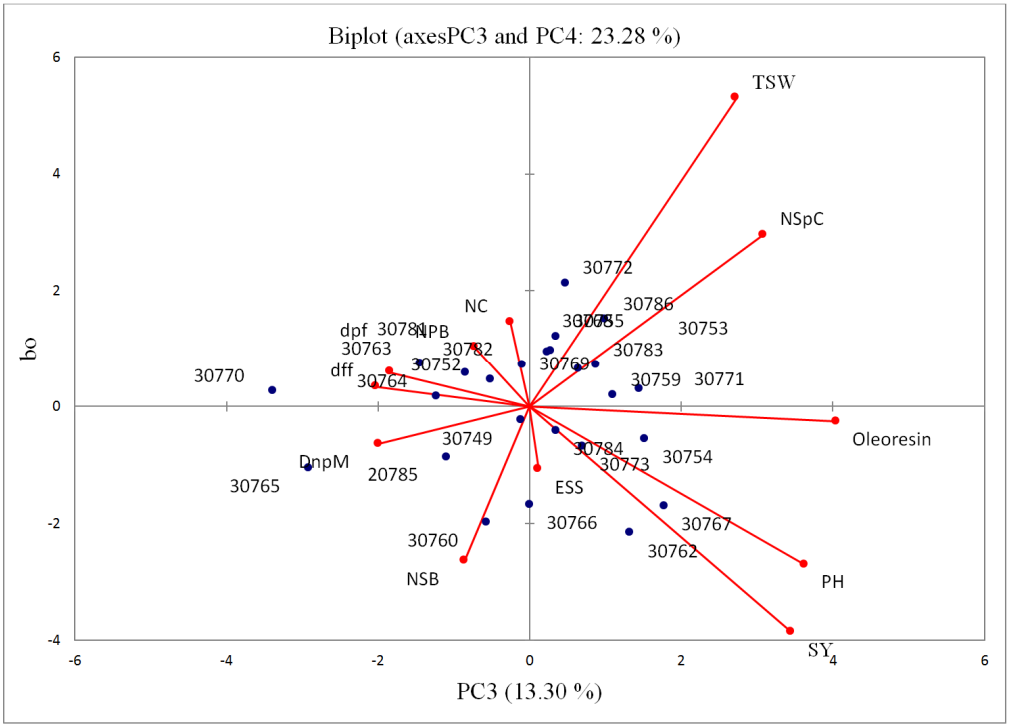

Figure 2. *Biplot of genotype –trait indicated contribution of genotypes and traits to PCA3 and PCA4 for 12 traits of 25 black cumin genotypes evaluated at Debre-Zeit 2018/19.*

DfF = Days to first Flowering, DpF = Days to 50% Flowering, DnpM=Days to 90% Maturity, TSW= Thousand Seed Weight, SY=Seed Yield, PH=Plant Height, NPB=Number of Primary Branch, NSB= Number of Secondary Branch, NC=Number of Capsule, NSpC=Number of Seed per Capsule.

The genotypes scattered in all four quadrants on the axes, indicating that there were a wide genetic variability for the traits studied. The genotypes that overlapped and closer to each other in the principal component axes had similar genetic makeup. However, genotypes which are far from each other could be considered as genetically distinct [16, 17]. The black cumin genotypes in the top right quadrant were closely associated with thousand seed yield and number of seed per capsule. The bottom right quadrant consists of the genotypes that are closely related with high oleoresin content, plant height and grain yield.

3.2. Genetic Divergence Analysis

3.2.1. Genetic Distances Among Black Cumin Genotypes

The genetic distances for all possible pairs of 25 black cumin genotypes are showed in Table 3. The genetic distances of genotypes ranged from 1.94 to 8.27 with the mean, standard deviation, and coefficient of variation of 4.69, 1.32 and 28.15%, respectively (Table 4 and Figure 3). The highest genetic distances (Euclidean distance) was computed between 30770 and 30773 (8.27) followed by between 30754 and 30763 (8.02) and 30769 (8.01). The lowest genetic distances (Euclidean distance) was estimated between 30769 and 30782 (1.94) followed by between 30752 and 30768 (1.99) (Table 3). A total of 121 (40.33%) and 87 (29%) pair of genotypes had Euclidean distances in the range between 3.37 and 4.69, and 4.69 and 6.02, respectively. On the other hand, 49 (16.33%) and 43 (14.33%) pair of genotypes had Euclidean distances >6.017 which was higher (>overall mean ED + SD) than the overall mean ED of genotypes and <3.375 which was lower (< overall mean ED - SD), respectively (Figure 3).

The mean genetic distance of each black cumin genotype was calculated to generate information about the most distant and closest genotypes to the 25 black cumin genotypes (Table 4). Depending on the mean Euclidean distance, 30763 (6.63) followed by 30773 (6.33) had higher mean distances and 30770 (5.69) and 30755 (4.73) had mean Euclidean distances above the overall genotypes mean distance. The remaining genotypes had mean genetic distances lower than the overall mean genetic distance of genotypes indicating the genotypes were not distant to other genotypes. The result suggested that crossing among the genotypes that showed more distant to others to combine the desirable traits in progenies.

The extent of diversity present between genotypes determines the extent of improvement gained through selection and hybridization. The more divergent the two genotypes are the more will be the probability of improving through selection and hybridization. The genotypes with high genetic distance can be used in hybridization programs for obtaining the progenies for important economic traits [18]. The selection of parents for hybridization-based on genetic diversity has been emphasized in previous studies [19].

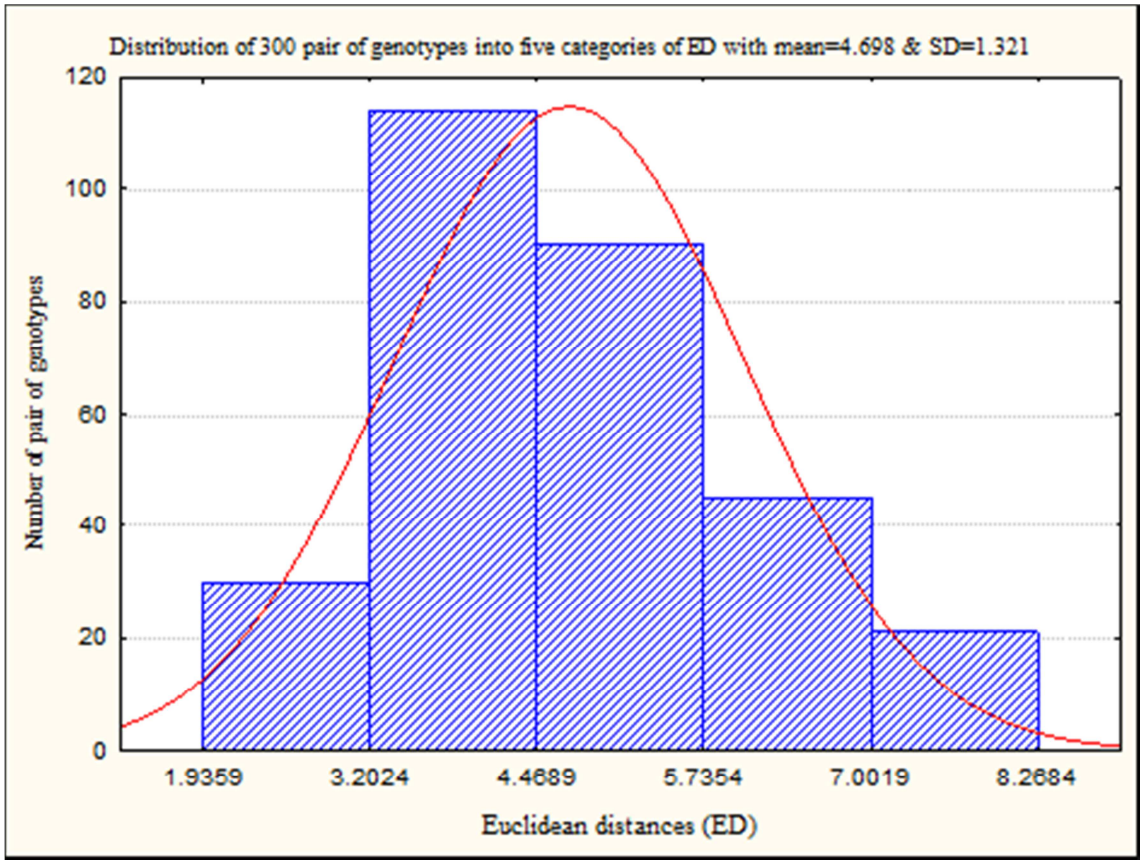

Figure 3. *Distribution pair of black cumin genotypes into five categories of Euclidean distances.*

Table 3. *Euclidean distances of 25 black cumin genotypes based on 12 quantitative traits evaluated at Debre-Zeitin 2018/2019.*

	30749	**30752**	**30753**	**30754**	**30755**	**30759**	**30760**	**30762**	**30763**	**30764**	**30765**	**30766**
20785	3.95	3.80	5.22	5.93	5.13	4.12	3.50	4.74	7.21	4.78	6.33	4.98
30749		2.61	2.67	4.59	3.88	3.60	3.38	3.38	6.25	3.21	5.41	3.67
30752			3.90	5.23	4.56	4.04	4.51	3.99	4.09	3.62	5.64	4.60
30753				3.44	2.77	4.12	4.47	3.78	7.00	2.83	5.10	3.97
30754					3.58	6.25	5.73	4.02	8.02	5.04	6.99	5.76
30755						5.22	4.96	5.03	7.56	4.25	6.06	5.51
30759							3.64	4.48	7.10	4.41	5.92	4.00
30760								4.56	7.86	4.25	4.76	2.97
30762									6.75	4.04	6.34	4.05
30763										6.18	7.10	7.50
30764											3.35	3.22
30765												4.14

Table 3. *Continued.*

	30767	**30768**	**30769**	**30770**	**30771**	**30772**	**30773**	**30781**	**30782**	**30783**	**30784**	**30786**
20785	5.62	4.78	4.56	4.97	5.08	6.41	6.37	3.13	4.41	5.31	4.33	4.16
30749	3.85	3.83	3.34	4.58	3.47	4.83	5.92	2.77	2.44	3.70	2.93	3.77
30752	4.03	1.99	4.54	4.73	4.53	5.29	4.83	3.52	3.62	2.80	2.02	3.61
30753	4.34	4.30	3.03	5.19	2.32	5.09	6.95	3.15	2.17	3.68	3.54	3.42
30754	5.35	5.30	5.08	5.90	3.57	7.82	7.84	4.37	4.60	5.52	4.10	3.74
30755	5.58	4.52	3.23	5.26	3.75	6.51	7.06	3.56	3.02	4.83	4.20	3.39
30759	3.67	4.62	3.76	6.20	4.00	3.56	5.85	4.50	3.52	3.48	4.16	4.37
30760	4.26	5.61	3.33	5.22	4.00	5.12	6.00	4.21	3.29	4.96	4.31	4.55
30762	3.59	4.18	4.96	6.23	3.86	6.80	6.81	4.20	4.54	4.10	3.05	4.80
30763	6.09	3.71	8.01	7.34	7.85	7.64	5.14	7.17	6.85	4.78	4.86	6.97
30764	4.90	4.09	3.35	4.84	3.29	4.96	6.90	3.20	2.68	3.64	3.71	4.28
30765	5.94	6.22	5.07	5.84	5.36	5.74	7.10	5.73	4.30	5.40	5.64	6.29
30766	3.92	5.41	3.52	5.99	3.36	4.67	6.73	4.66	3.53	3.98	4.16	4.89
30767		4.23	5.09	6.55	4.39	5.30	4.82	5.80	4.34	2.95	2.96	4.96
30768			5.05	5.72	4.97	6.04	4.97	4.46	4.30	2.50	2.26	4.03
30769				5.40	2.55	4.27	6.95	3.45	1.94	4.34	4.40	3.34
30770					5.85	7.03	8.27	4.44	4.98	5.92	4.72	5.29
30771						4.82	7.08	3.57	2.58	4.08	3.89	2.78
30772							6.54	5.94	3.98	4.47	5.79	5.38
30773								7.25	5.89	5.24	5.27	6.19
30781									3.03	4.94	4.07	3.24
30782										3.76	3.76	3.11
30783											2.39	4.12
30784												3.44

Table 4. *Range and mean Euclidean distance of 25 black cumin genotypes estimated from 12 quantitative traits as evaluated at Debre-Zeit in 2018/19.*

Genotype	**Minimum**	**Maximum**	**Mean**	**SD**	**CV (%)**
20785	3.13	7.21	4.95	1.00	20.11
30749	2.44	6.25	3.83	1.00	25.98
30752	1.99	5.64	4.00	0.95	23.69
30753	2.17	7.00	4.02	1.27	31.55
30754	3.44	8.02	5.32	1.35	25.40
30755	2.77	7.56	4.73	1.26	26.58
30759	3.48	7.10	4.52	1.02	22.51
30760	2.97	7.86	4.56	1.07	23.44
30762	3.05	6.81	4.68	1.11	23.72
30763	3.71	8.02	6.63	1.25	18.82
30764	2.68	6.90	4.13	1.02	24.72
30765	3.35	7.10	5.66	0.91	16.09
30766	2.97	7.50	4.55	1.13	24.88
30767	2.95	6.55	4.69	0.97	20.72
30768	1.99	6.22	4.46	1.09	24.44
30769	1.94	8.01	4.27	1.35	31.51
30770	4.44	8.27	5.69	0.93	16.41
30771	2.32	7.85	4.21	1.35	32.07
30772	3.56	7.82	5.58	1.11	19.86
30773	4.82	8.27	6.33	0.95	15.06
30781	2.77	7.25	4.35	1.25	28.75
30782	1.94	6.85	3.78	1.15	30.37

Genotype	Minimum	Maximum	Mean	SD	CV (%)
30783	2.39	5.92	4.20	0.97	23.15
30784	2.02	5.79	3.92	0.99	25.20
30786	2.78	6.97	4.34	1.09	25.04
Overall	1.94	8.27	4.70	1.32	28.15

3.2.2. Clustering of Genotypes

The Euclidean distance matrix of pair of genotypes estimated from quantitative traits was used to construct dendrograms based on the Unweighted Pair-group methods with Arithmetic Means (UPGMA). The 25 black cumin genotypes were grouped into ten distinct clusters at cut point 3.38 Euclidean distance (mean Euclidean distance of genotypes- standard deviation) with 43.32% and 56.68% variation within and between clusters, respectively, as variance decomposition for the optimal classification (Figure 4).

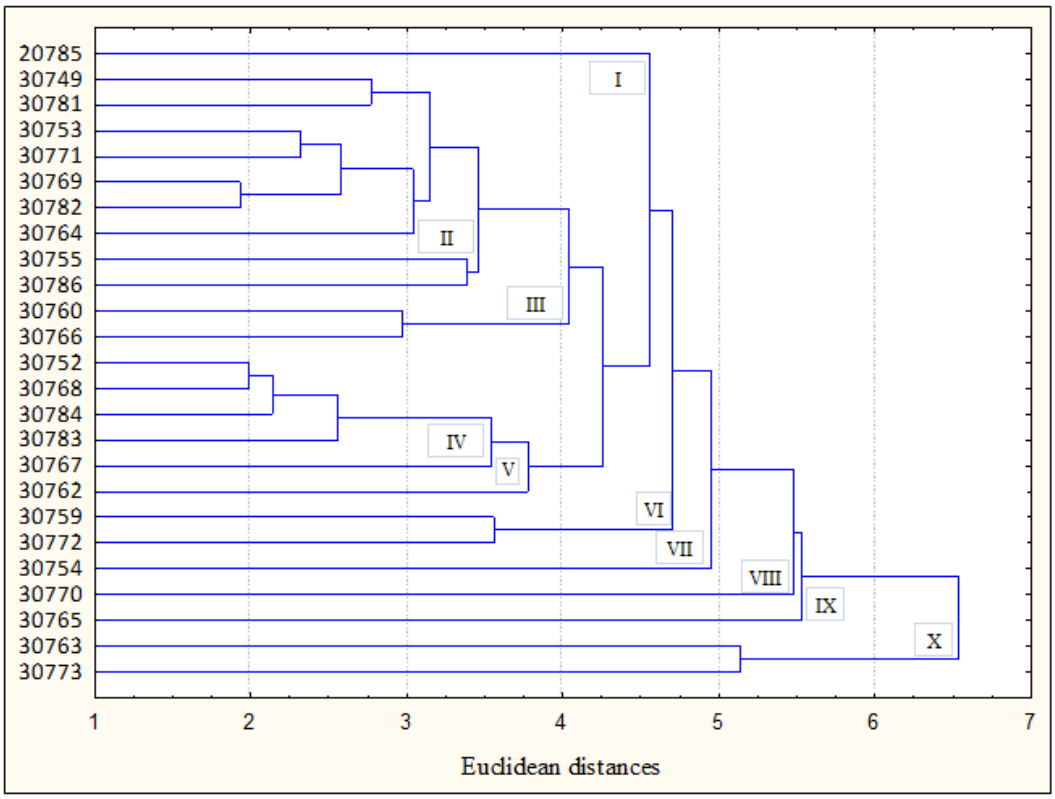

Figure 4. *Dendrogram depicting dissimilarity of 25 black cumin genotypes by Unweighted Pair Group Method with Arithmetic Means (UPGMA) clustering method from Euclidean distances matrix estimated from 12 quantitative traits.*

Cluster II was the largest and consisted of nine genotypes (36%) of the total genotypes. Cluster IV consisted of five (20%) and other three Cluster (III, VI, and X) each consisted of two (8%) genotypes, while the five Cluster (I, V, VII, VIII, and IX) constructed only by one (4%) genotype. The result suggested that the genotypes grouped under same cluster had similarity for many traits but dissimilarity to other genotypes in other clusters with one or more traits. Cluster analysis sequestrates genotypes into clusters which exhibit high homogeneity within a cluster and high heterogeneity between clusters [20].

3.2.3. Cluster Mean Analysis

The mean values of the 12 quantitative traits in each cluster are presented in Table 5. Cluster I consisted of one genotype (20785) which was characterized by the highest essential oil and had number secondary branches, number of capsules and oleoresin greater than the average of the genotypes. Cluster II was the largest consisted of 36% of genotypes distinguished by having mean values greater than overall mean values of genotypes for thousand grain weight, number of primary branches, number secondary branches, number of capsules, number of seeds per capsule and oleoresin but lower mean values for all other traits than overall mean values of genotypes. The two clusters also consisted of early flowering and maturity genotypes than the average flowering and maturity period of the genotypes, however, both clusters have not consisted genotypes with high mean seed yield. Therefore, genotype in cluster I could be used as gene(s) source for high essential oil and genotypes in both clusters could be used as the source of genes for early maturity, growth traits, and oleoresin.

Cluster III consisted of two genotypes (30760 and 30766), cluster IV, V and VI contained 5, 1 and 2, genotypes, respectively. The four clusters consisted 10 (40%) of

genotypes having seed yield and seed oleoresin content greater than overall mean values of genotypes and particularly, the genotype (30762) which constructed solitary cluster V had the highest seed yield. However, all clusters had lower mean values than genotypes overall mean value for essential oil. Each cluster also consisted of genotypes having higher mean values than mean values of genotypes for other agronomic traits. For instances, cluster IV had mean values higher than mean of genotypes for all traits except days to 90% maturity, and cluster IV had better plant height, number of primary and secondary branches, and number of capsule than average of genotypes performances. Therefore, selection could be possible in these clusters for high seed yield, seed oleoresin content, early maturity and for other growth traits, and if further improvement is required, it is possible to make crosses with genotype in Cluster I to combine essential oil content of seeds.

Cluster VII, VIII and IX consisted of each one genotype and two genotypes were included in Cluster X. Cluster VII consisted of genotype with tallest plants, highest number secondary branches, and number of seeds per capsule, and higher mean values for number primary branches and oleoresin, early flowering and maturity than average genotypes performances. Cluster X had the highest seed oleoresin content and higher mean for all traits except seed yield and essential oil than mean values of genotypes. Cluster VIII and IX consisted genotypes with higher mean values for number primary and secondary branches than average genotypes performances, but genotype in cluster VIII was early flowering and maturity while genotype in cluster IX was late flowering and maturity as compared to the average performances of genotypes.

From cluster mean analysis, the presence of significant genetic diversity among the evaluated in black cumin genotypes suggests an opportunity for improvement of grain yield as well as other yield components through hybridization of genotypes from different clusters and subsequent selection from the segregating generations.

Table 5. *Cluster mean values for 12 quantitative traits of 25 black cumin genotypes evaluated at Debre-Zeit in 2018/19.*

Trait	Cluster										Overall		
	I	II	III	IV	V	VI	VII	VIII	IX	X	Min	Max	Mean
Dff	62.0	59.4	61.8	65.3	58.5	63.0	55.5	63.5	63.5	71.3	55.5	71.3	63.4
Dpf	78.5	77.1	78.5	82.6	76.0	81.5	71.5	80.0	83.5	90.0	71.5	90.0	80.8
DnpM	132.5	133.0	137.0	139.3	134.0	139.0	127.0	135.5	145.5	152.0	127.0	152.0	139.5
PH (cm)	45.1	47.3	49.5	52.3	53.0	48.9	55.3	44.4	48.7	55.2	44.4	55.3	49.8
NPB	4.6	4.5	3.9	4.8	4.8	4.1	4.9	4.6	4.6	5.1	3.9	5.1	4.5
NSB	4.7	4.2	4.4	4.6	4.8	2.8	5.1	4.8	4.1	4.6	2.8	5.1	4.0
NC	9.4	8.5	7.0	8.9	9.3	7.5	8.9	8.8	5.8	10.0	5.8	10.0	7.9
NSpC	42.3	46.9	41.5	47.5	43.6	44.7	51.1	46.8	38.4	48.5	38.4	51.1	44.7
TSW (g)	2.4	2.6	2.4	2.7	2.5	2.9	2.4	2.4	2.3	2.6	2.3	2.9	2.6
SY (kg ha^{-1})	2243.0	2150.0	2403.6	2510.1	2873.3	2383.4	2210.8	2004.2	1863.3	2258.3	1863.3	2873.3	2368.3
OLE (%)	31.6	32.1	32.3	31.8	32.4	32.7	30.7	20.8	30.6	35.2	20.8	35.2	28.0
ESS (%)	2.7	1.0	1.4	1.3	1.1	1.7	0.7	1.4	0.5	1.6	0.5	2.7	1.6

Dff = Days to first Flowering, Dpf= Days to 50% flowering, DnpM=Days to 90% maturity, TGW= Thousand seed weight, SY=seed Yield, PH=Plant Height, NPB=Number of Primary Branch, NSB= Number Secondary Branch, NC=Number of Capsule, NSpC=N1umber of Seed per Capsule, oleoresin and essential oil, Min=minimum, Max=maximum.

4. Conclusion

The first four Principal component axes (PCA1 to PCA4) accounted varied percentage of total variance that ranged from 9.98% to 28.43% and accounted for 70.62% of the total variation with eigenvalues ranged from 1.20 to 3.41.. Days to first flowering, days to 50% flowering, days to 90% maturity and plant height to PCA1 and number of secondary branches, number of primary branches, number of capsule and number of seed per capsule to PCA2 contributed relatively higher proportion to total variability. Oleoresin content of seeds and thousand seed weight had relatively more contribution to the total variance of PCA3 and PCA4, respectively. This suggested that these traits were responsible for the differentiation of the black cumin genotypes into different groups.

The 25 black cumin genotypes were grouped into 10 distinct clusters from Euclidean distance matrix and dendrograms constructed based on the Unweighted Pair-group methods with Arithmetic Means. The clusters consisted of high performance genotypes for varied traits suggested selection of genotypes is possible for high performances of different traits or crossing of genotypes to combine the desirable traits of genotypes to other genotypes lacking better performances.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that there are no conflicts of interest.

Acknowledgements

The authors wish to thank the Africa Center of excellence for climate smart agriculture and biodiversity conservation.

References

[1] A. K. Datta, A. Saha, A. Bhattacharya et al., "Black cumin (Nigella sativa L.)", A review. Journal of Plant Development Sciences, vol. 4, no. 1, pp. 1-43, 2012.

[2] A. Ahmad, A. Husain, M. Mujeeb et al., "A review on therapeutic potential of *Nigella sativa*," *A miracle herb*, vol. 3, no. 5, pp. 337-352, 2013.

[3] S. V. Tembhurne, S. Feroz, B. H. More, D. M. Sakarkar, "A review of the therapeutic potential of Nigella sativa (kalonji) seeds", Journal of Medicinal Plants R*esearch,* vol. 8, no. 3, pp. 167-177, 2014.

[4] E. Rezaei-Chiyaneh, S. M. Seyyedi, E. Ebrahimian, SS. Moghaddam, CA. Damalas, "Exogenous application of gamma-aminobutyric acid (GABA) alleviates the effect of water deficit stress in black cumin (Nigella sativa L.)", Ind Crops Prod, no. 112, pp. 741–748, 2018.

[5] M. F. Ramadan, "Nutritional value and applications of Nigella sativa essential oil: a mini review", Journal of Essential Oil Research, vol. 27, no. 4, pp. 271-275, 2018.

[6] F. A. Saleh, N. El-Darra, K. Raafat, I. El-Ghazzawi, "Phytochemical analysis of Nigella sativa L. utilizing GC-MS exploring its antimicrobial effects against multidrug-resistant bacteria", Pharmacogn Journal, vol. 10, no. 1, pp. 99-105, 2018.

[7] S. N. Saxena, S. S. Rathore, Y. Diwakar, et al., "Genetic diversity in fatty acid composition and antioxidant capacity of Nigella sativa L. genotypes," *LWT- Food Science and Technology*, no. *78*, pp. 198-207, 2017.

[8] Thottappily G, Crouch JH, Quin FM. "Overview of DNA marker research at IITA": Proceedings of the workshop on DNA markers at IITA. Ibadan Ed by J. H. Crouch and Tenkuoano A pp. 3-8, 1996.

[9] Cochran, W. G. and Cox, G. M., "Completely randomized, randomized block and Latin square designs", *Experimental designs*, 117-142, 1957.

[10] AOAC, "Official Methods of Analysis. Association of Official Analytical Chemists", 18th Edition, Washington, DC. 2005.

[11] Rao, B. R., Kaul, P. N., Syamasundar, K. V. and Ramesh, S. "Chemical profiles of primary and secondary essential oils of palmarosa (Cymbopogonmartinii (Roxb.) Wats var. motia Burk.)", *Industrial Crops and Products,* 21 (1): 121-127, 2005.

[12] SAS software. "SAS Institute Inc", Cary, NC, USA, 2004.

[13] Sneath, P. H. A., and Sokal, R. R. "Numerical Taxonomy: the principles and practice of numerical classification," *Fransisco USA W. F Freeman*, Pp. 573, 1973.

[14] Legendre, P., and Legendre, L. "Numerical Ecology," 2nd edition. Amsterdam. Elsevier, 853p, 1998.

[15] Sharma JR. "Statistical and Biometrical Techniques in Plant Breeding," New Age International (P) Limited Publishers, New Delhi. pp. 432, 1998.

[16] Gerrano AS, Jansen van Rensburg WS, Adebola PO. "Nutritional composition of immature pods in selected cowpea [Vigna unguiculata (L.) genotypes in South Africa" *Australian Journal of Crop Science*, 11: 134-141, 2017.

[17] Yan and Frégeau-Reid, 2018). Yan, W. and Frégeau-Reid, J. "Genotype by Yield Trait (GYT) Biplot," a novel approach for genotype selection based on multiple traits. *Scientific reports*, vol. *8,* no. 1, p. 8242, 2018.

[18] Singh, R. P., Gangadharappa, H. V. and Mruthunjaya, K.. "Cuminum cyminum–A popular spice," An updated review. *Pharmacognosy Journal*, vol. *9,* no. 3, 2017.

[19] Dubey, P. N., Singh, B., Mishra, B. K., Kant, K. and Solanki, R. K. "Nigella (Nigella sativa L.): a high value seed spice with immense medicinal potential," *Indian Journal of Agricultural Science*, no, 86, p., 967-979, 2009.

[20] Jaynes, D. B., Kaspar, C., Colvin T. S., and James D. E. "Cluster analysis of spatiotemporal corn yield patterns in an Iowa field," *Agronomy Journal.* doi: 10.2134/agronj2003.0574, 2003.