MEYVECİLİK ARAŞTIRMA İSTASYONU MÜDÜRLÜĞÜ
FRUIT RESEARCH STATION

ISSN: 2148-0036
Yıl /Year: 2015
Cilt(Sayı)/Vol.(Issue): 2(1)
Sayfa/Page: 35-41
Araştırma Makalesi
Research Article

Meyve Bilimi Fruit Science

Geliş Tarihi (date of arrival): 08.10.2014
Kabul Tarihi (date of acceptance): 19.12.2014

Evaluation of Some Carob (*Ceratonia siliqua* L.) Genotypes in Silifke (Mersin, Turkey) Province by Cluster Analysis

Saim Zeki BOSTAN[1], Yeliz KAŞKO ARICI[2]

[1]Horticultural Department, Agricultural Faculty, Ordu University 52200 Ordu, Turkey
[2]Biometry-Genetics Unit, Department of Animal Science, Agricultural Faculty, Ordu University 52200 Ordu, Turkey
szbostan@hotmail.com (Sorumlu Yazar)

Abstract

This study was aimed to the evaluation of some carob (*Ceratonia siliqua* L.) genotypes by cluster analysis. In this study, the fruits harvested from 17 wild carob genotypes were evaluated in 2011 and 2012 years in Silifke (Mersin, Turkey). In these genotypes, pod weight, pod length, pod width, pod thickness, fruit stalk length, fruit stalk thickness, seed number, seed weight, pulp ratio, seed width, seed length, seed thickness, leaf width, leaf length, leaf petiole length, leaf petiole thickness, total soluble solids, pH and titratable acidity were determined. Relationship among these characters of the carob genotypes was analyzed by Principal Component Analysis (PCA) and Hierarchical Cluster analysis (HCA). Single linkage (SLINK) technique and Euclidean distance were used for clustering. Clusters in the cluster dendrogram were identified based on similarities. The dendrogram obtained from HCA showed that these genotypes of carob were collected in 5 clusters. According to HCA, 17 carob genotypes were similar in rates ranging from 46% to 89%. While the highest similarity was found between 1 and 16 genotypes, the most distant genotype was 8. Cluster I consisted of 1, 3, 4, 9, 11, 13, 14, 15 and 16 genotypes. 5, 7 and 17 were placed in cluster II; 10 and 12 were placed in cluster III; 2 and 6 were placed in cluster IV. Genotype 8 was considered as singular.

Key Words: Carob, *Ceratonia siliqua*, cluster analysis, principal component analysis, genotype

Silifke'de (Mersin, Türkiye) Yetişen Bazı Keçiboynuzu (C*eratonia siliqua* L.) Genotiplerinin Kümeleme Analizi ile Değerlendirilmesi

Özet

Bu çalışma, bazı keçiboynuzu (*Ceratonia siliqua* L.) genotiplerinin kümeleme analizi ile değerlendirilmesi amacıyla yürütülmüştür. Çalışmada 2011 ve 2012 yıllarında, Silifke (Mersin, Türkiye) ilçesinde doğal olarak yetişmekte olan 17 genotipten hasat edilen meyveler değerlendirilmiştir. Genotiplerde meyve ağırlığı, meyve boyu, meyve eni, meyve kalınlığı, meyve sapı uzunluğu, meyve sapı kalınlığı, tohum sayısı, tohum ağırlığı, meyve eti oranı, tohum eni, tohum boyu, tohum kalınlığı, yaprak eni, yaprak boyu, yaprak sapı uzunluğu, yaprak sapı kalınlığı, suda çözünür kuru madde miktarı, pH ve titre edilebilir asitlik özellikleri belirlenmiştir. Genotiplerin bu özellikleri arasındaki ilişkiler, temel bileşenler analizi ve hiyerarşik kümeleme analizi (HKA) ile analiz edilmiştir. Kümelemede tek bağlantı tekniği (SLINK) ve öklit uzaklığı kullanılmıştır. Küme dendogramındaki kümeler, benzerliklerine göre tanımlanmıştır. HKA sonucunda elde edilen dendogram, keçiboynuzu genotiplerinin 5 kümede toplandığını göstermiştir. HCA'ya göre 17 genotip %46'dan %89'a değişen oranlarda benzer bulunmuştur. En yüksek benzerlik, 1 ve 16 numaralı genotipler arasında bulunurken en uzak, 8 numaralı genotip olmuştur. I. küme 1, 3, 4, 9, 11, 13, 14, 15 ve 16 numaralı genotiplerden oluşmuştur. 5, 7 ve 17 II. kümeye, 10 ve 12 III. kümeye yerleşirken 2 ve 6 IV. kümeye yerleşmiştir. 8 numaralı genotip ise tekil olarak kümelenmiştir.

Anahtar Kelimeler: Keçiboynuzu, *Ceratonia siliqua*, kümeleme analizi, temel bileşenler analizi, genotip

1. Introduction

The carob tree has been grown since antiquity in most countries of the Mediterranean basin, usually in mild and dry places with poor soils. The carob tree is an important component of the Mediterranean vegetation and its cultivation in marginal and prevailing calcareous soils of the Mediterranean region is important environmentally and economically (Batlle and Tous, 1997).

Turkey is one of the homeland regions of carob (Pekmezci et al., 2008). Various wild carob genotypes have been grown in the Mediterranean and Aegean regions of Turkey (Gübbük et al., 2010).

In the carob selection breeding studies, many pomological characters have been considered such as pod, seed and leaf traits.

In a research that was conducted in Mediterranean and Aegean regions of Turkey, 4 superior

domesticated and 10 wild types among a total of 54 types were selected in terms of fruit and seed features (Pekmezci et al., 2005). In the other research that was conducted in East Mediterranean region of Turkey, 5 superior domesticated and 8 wild types among a total of 34 types were selected in terms of fruit and seed features (Pekmezci et al., 2008).

In selection breeding studies, even method seems easy, application is very inconvenient and requires lots of attention. In that studies, a vast number of characters according to breeding objective can be worked on; in this situation, much time and workforce can be needed. Therefore, to know some relationships between characters will decrease workload by providing working on less character. For this purpose, some studies have been conducted on carob genotypes (Barracosa et al., 2007; Konate et al., 2007; Barracosa et al., 2008; Naghmouchi et al., 2009; Sidina et al., 2009; Tetik et al., 2011).

Multivariate statistical analysis refers to multiple advanced techniques for examining relationships among multiple variables at the same time. Researchers use multivariate procedures in studies that involve more than one dependent variable, more than one independent variable or both. This type of analysis is desirable because researchers often hypothesize that a given outcome of interest is affected or influenced by more than one thing (Johnson and Wichern, 2007).

Multivariate statistical techniques which simultaneously analyze multiple measurements on each individual under investigation are widely used in plant breeding programme. Among the different multivariate techniques, cluster analysis (CA) and principal component analysis (PCA) are great potential for classification of problems (Mekonnen et al., 2014).

CA is a multivariate method which divides data into groups (clusters) such that similar data objects belong to the same cluster and dissimilar data objects to different clusters. The resulting data partition improves data understanding and reveals its internal structure. Partition clustering algorithms divide up a data set into clusters or classes, where similar data objects are assigned to the same cluster whereas dissimilar data objects should belong to different clusters. In other words, cluster analysis is an exploratory data analysis tool for organizing observed data such as people, brands, events, companies, countries, etc. into meaningful taxonomies, groups or clusters, which maximizes the similarity of cases within each cluster and maximizes the dissimilarity between clusters or groups that are initially unknown (Rencher, 2002; Johnson and Wichern, 2007; Alpar, 2011).

Two types of clustering algorithms are nonhierarchical and hierarchical. In nonhierarchical clustering, such as the k-means algorithm, the relationship between clusters is undetermined. Hierarchical clustering (HCA) repeatedly links pairs of clusters until every data object is included in the hierarchy. With both of these approaches, an important issue is how to determine the similarity between two objects, so that clusters can be formed from objects with a high similarity to each other. Commonly, distance functions, such as the Manhattan and Euclidian distance functions, are used to determine similarity (Anderson, 2003; Johnson and Wichern, 2007; Özdamar, 2011).

PCA is a multivariate technique that analyzes a data table in which observations are described by several inter-correlated quantitative dependent variables. Its goal is to extract the important information from the table, to represent it as a set of new orthogonal variables called principal components, and to display the pattern of similarity of the observations and of the variables as points in maps (Anderson, 2003; Özdamar, 2011)

Objective of this research was to evaluate the some carob (*Ceratonia siliqua* L.) genotypes by PCA and HCA for important pomological characters. According to these relationships, cluster analysis is made and groups have been created in terms of similar characters. Thus, it was aimed to describe the characters as in brief information and interpret them.

2. Materials and Methods

This research was carried out during 2011 and 2012 years in Silifke county (Mersin province, Turkey) province (Figure 1.). Geographic location of Silifke county located is latitude: 36° 22' 31" N; longitude: 33° 55' 59" E.

The natural populations of *Ceratonia siliqua* distributed in different places in the province

Figure 1. Map of Silifke county (Mersin province, Turkey)

were collected for this study. Each of 17 genotypes in total 350.000 m^2 area were sampled randomly collected from tree. In the samples pomological characters were determined for statistical analyses: pod weight, pod length, pod width, pod thickness, fruit stalk length, fruit stalk thickness, seed number, seed weight, pulp ratio, seed width, seed length, seed thickness, leaf width, leaf length, leaf petiole length, leaf petiole thickness, total soluble solids, pH and titratable acidity.

2.1. Statistical Analysis

In order to find the main variation trends among important pomological characters in the carob genotypes and to evaluate their correlation, the data were analyzed by multivariate statistical analyses viz principal component analysis, cluster analysis and Pearson's Correlation.

First, Pearson's correlation coefficient was computed and then PCA was applied to the 19 pomological characters. Since the aim of PCA is to reveal common principles in the data, we pooled all sampling two years and used genotypes means in the multivariate analysis procedures. PCA, based on a covariance matrix was performed and Varimax orthogonal rotation of the extracted component axes was done to facilitate the assembling of tested variables into specific principal components (PCs). The adequacy of the PCA was determined with the Kaiser-Meyer-Olkin (KMO). Finally, CA subjected to a hierarchical clustering algorithm (HCA) based on Euclidean distances and using Single linkage (SLINK) techniques.

All data analyses were performed by using IBM SPSS version 22 and Minitab version17 statistical software.

3. Results and Discussion

The overall mean values for all pod, fruit, seed and leaf characters (variables) measured and their standard deviations were presented in Table 1.

PCA requires that there be some correlations greater than 0.30 between the variables included in the analysis. For this set of variables, there are several correlations in the matrix greater than 0.30, satisfying this requirement (Tabachnick and Fidell, 2007).This was shown in the correlation matrix in Table 2. Furthermore, the KMO measure was

Table 1. Means, standard deviations and ranges of pomological characters determined for pods, fruits, seeds and leafs of *C. siliqua*

Characters	n	Mean±S.D.	Range
Pod Weight (g). V1	442	17.35±5.56	28.66
Pod Length (cm). V2	442	15.20±2.70	17.50
Pod Width (mm). V3	442	21.70±2.74	14.30
Pod Thickness (mm). V4	442	8.46±3.51	70.94
Fruit Stalk Length (mm). V5	314	10.84±2.46	20.43
Fruit Stalk Thickness (mm). V6	314	3.19±0.66	6.66
Seed Number. V7	440	10.56±2.76	16.00
Seed Weight (g). V8	440	1.97±0.59	3.18
Pulp Ratio (%). V9	425	97.52±1.06	6.88
Seed Width (mm). V10	441	7.15±0.55	6.35
Seed Length (mm). V11	441	9.78±0.73	4.43
Seed Thickness (mm). V12	441	4.03±0.51	7.18
Leaf Width (mm). V13	102	32.72±5.35	26.23
Leaf Length (mm).V14	102	51.76±8.71	47.91
Leaf Petiole Length (mm). V15	102	3.60±1.06	5.51
Leaf Petiole Thickness (mm). V16	102	1.28±0.24	1.56
Total Soluble Solids (%). V17	91	8.71±0.56	2.50
Ph. V18	91	5.61±0.61	4.78
Titratable Acidity (g l^{-1}). V19	44	1.91±0.26	28.66

Table 2. Pearson coefficient correlation between pomological characters determined for pods. fruits. seeds and leafs of *C. siliqua*

	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12	V13	V14	V15	V16	V17	V18
V2	.69**																	
V3	.86**	.46																
V4	.51*	.23	.41															
V5	.60*	.35	.71**	.31														
V6	.45	.42	.44	-.03	.59*													
V7	-.01	.18	-.37	-.29	-.08	.22												
V8	.14	.28	-.07	-.36	.27	.33	.73**											
V9	.81**	.42	.89**	.56*	.55*	.24	-.56*	-.27										
V10	-.06	.23	.18	-.06	.36	.07	-.32	.14	.06									
V11	.59*	.21	.80**	.16	.58*	.30	-.39	-.01	.66**	.16								
V12	.19	.40	-.08	.26	.17	.03	.50*	.49*	-.17	.24	-.28							
V13	-.09	.17	-.24	.10	-.01	.10	.317	.05	-.22	-.25	-.37	.18						
V14	.50*	.36	.40	.35	.08	.30	-.04	-.18	.44	-.45	.14	-.28	.43					
V15	-.31	-.09	-.24	-.27	-.05	.30	.06	-.16	-.27	-.08	-.41	-.27	.45	.23				
V16	-.04	.01	.01	-.22	.19	.49*	.05	.19	-.03	.09	.07	.13	.10	-.30	.26			
V17	.45	.17	.30	.42	.06	-.13	-.26	-.18	.54*	-.21	.14	-.18	-.24	.25	-.22	-.16		
V18	.09	-.21	.16	.00	.36	.14	-.05	-.19	.12	-.23	.12	-.16	.02	-.12	.39	.16	.13	
V19	-.33	-.43	-.32	.09	-.03	-.14	-.05	-.04	-.24	-.03	-.09	-.32	-.11	-.12	-.08	-.19	.19	-.12

* Correlation is significant at the 0.05 level.
** Correlation is significant at the 0.01 level.

0.675. It thus required the result of PCA in order to reduce information bias in the final clusters caused by these information redundancies. As the scree plot (Figure 2) and Component plot in rotated space (Figure 3) indicate there are three factors in the model of the PCA.

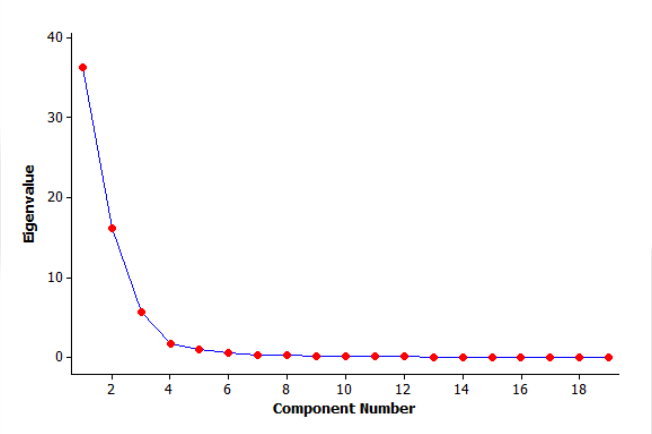

Figure 2. Scree plot of the PCA

The scree option in the PCA statement produces a scree plot that illustrates the rate of change in the magnitude of the eigenvalues for the factors. The rate of decline tends to be fast for the first few factors but then levels off. The "elbow", or the point at which the curve bends, is considered to indicate the maximum number of factors to extract. The Figure 2 illustrated an example of a rather idealistic scree plot, where a clear elbow occurred at the fourth factor, which had an eigenvalue right around 1. The scree plot suggested a maximum of four factors in the PCA.

From Figure 2, component number greater than or equal to three has very slight change in eigenvalues. Only the first four components have eigenvalues greater than 1. Therefore, from the scree plot and latent root criterion, the first four components (PCs) were selected for analysis.

Results from the PCA presented in Table 3, revealed that only four of the nineteen principal components had eigenvalues greater than 1.while, the first three axes with eigenvalues of 36.27, 16.14 and 5.70 respectively, jointly accounted for 93.7% of the total variation among the genotypes.

Table 3. Eigenvalues. percent and cumulative variances of the four most important characters from PCA

	PC1	PC2	PC3	PC4
Eigenvalues	36.265	16.143	5.703	1.709
Percent variance	0.585	0.260	0.092	0.028
Cumulative variance	0.585	0.845	0.937	0.965

The first four principal axes together explained above 96% of the total variation among the 19 characters that described the 17 genotypes. The major characters described by the first three principal axes are presented in Table 4. PC1 principal component axis was mainly loaded by pomological characters. Clearly the first factor of the initial solution was much more important than the other. PC4 - PC19 was discarded because of very low variance contributions.

Table 4. Eigen analysis of the correlation matrix of the first three principal components

Characters	PC1	PC2	PC3
Pod Weight (g)	0.282	**-0.517**	**-0.548**
Pod Length (cm)	0.055	-0.051	-0.144
Pod Width (mm)	0.128	**-0.339**	-0.235
Pod Thickness (mm)	0.062	-0.063	-0.095
Fruit Stalk Length (mm)	0.038	-0.126	**-0.333**
Fruit Stalk Thickness (mm)	0.013	-0.012	-0.028
Seed Number	0.005	0.094	-0.130
Seed Weight (g)	-0.006	0.002	-0.049
Pulp Ratio (%)	0.047	-0.111	-0.051
Seed Width (mm)	-0.015	-0.010	-0.025
Seed Length (mm)	0.011	-0.077	-0.038
Seed Thickness (mm)	-0.005	0.005	-0.050
Leaf Width (mm)	**0.355**	**0.747**	**-0.517**
Leaf Length (mm)	**0.875**	-0.069	**0.453**
Leaf Petiole Length (mm)	0.021	0.066	0.037
Leaf Petiole Thickness (mm)	-0.004	0.003	-0.017
Total Soluble Solids (%)	0.013	-0.039	0.010
pH	-0.003	-0.004	-0.035
Titratable Acidity (g l^{-1})	-0.007	0.006	0.026

According to Table 4 and Figure 3, PC1 included the variables "leaf width" and "leaf length"; PC2 includes the variables "pod weight", "pod width" and "Leaf Width"; PC3 included the variables "pod weight", "fruit stalk length", "leaf width" and "leaf length" (n>300; >0.3). The components explained 93.7% of the total variance in the variables which was included on the components.

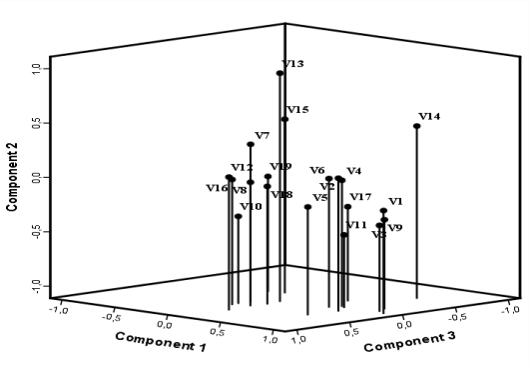

Figure 3. Component plot of PCA in rotated space

HCA was utilized to investigate the similarities and dissimilarities among the genotypes with respect to fruit, seed and leaf characters. The HCA was made the variables obtained from PCA.

Memberships of the clusters and their average distance levels obtained from HCA were determined and presented in Table 5.

Table 5. Memberships of the clusters and their average distance levels obtained from HCA

Clusters	Memberships	Average distance from centroid
I	1. 3. 4. 9. 11. 13. 14. 15. 16	3.09
II	5. 7. 17	3.71
III	10. 12	3.02
IV	2. 6	1.92
V	8	0

The analysis of "genetic similarity" through the HCA was shown in Figure 4. Cluster diagram based on Euclidean dissimilarity using SLINK. Clusters in the cluster dendrogram were identi-

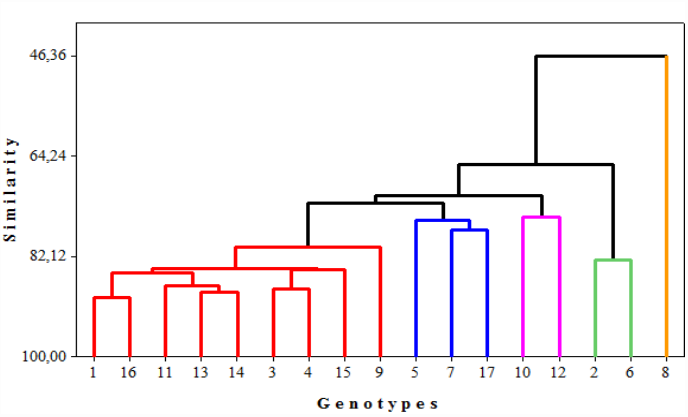

Figure 4. Between the cluster similarity dendrogram of carob genotypes

fied based on similarities. The dendrogram obtained from HCA showed that these genotypes of carob were collected in 5 clusters. According to HCA, 17 carob genotypes were similar in rates ranging from 46 to 89%. While the highest similarity was found between 1 and 16 genotypes, the most distant genotype was 8. Cluster I consisted of 1, 3, 4, 9, 11, 13, 14, 15 and 16 genotypes. 5, 7 and 17 were placed in cluster II; 10 and 12 were placed in cluster III; 2 and 6 were placed in cluster IV. Genotype 8 was considered as singular. HCA leads to identify five clusters confirming the PCA results.

4. Conclusions

In this study, some carob (*Ceratonia siliqua* L.) genotypes which selected from Silifke were investigated in terms of "genetic similarity". Based on HCA results, one genotype (genotype 8) was completely separated from the others. On the basis of connecting distances between parameters, five clusters were distinguished. HCA confirmed the results of PCA. While the highest similarity was found between 1 and 16 genotypes, the most distant genotype was 8. Genotype 8 was considered as singular. Thus, genotype 8 together the others can be high the possibility of creating variations in probable hybridization studies.

As a result of, usage of similar genotypes will be prevented in the breeding programs in this subject and by this way the success of breeding programs can be advanced. This study can assist geneticists and breeders to identify populations with desirable characteristics for inclusion in carob breeding programs.

References

Alpar R, 2011. Çok Değişkenli İstatistiksel Yöntemler. Detay Yayıncılık, Ankara, Turkey.

Anderson TW, 2003. An Introduction to Multivariate Statistical Analysis, 3.rd Edt. New York:John Wiley.

Barracosa P, Osorio J, Cravador A, 2007. Evaluation of Fruit and Seed Diversity and Characterization of Carob (*Ceratonia siliqua* L.) Cultivars in Algarve Region. Scientia Horticulturae 114: 250 -257.

Barracosa P, Lima MB, Cravador A, 2008. Analysis of Genetic Diversity in Portuguese *Ceratonia siliqua* L. Cultivars Using RAPD and AFLP Markers. Scientia Horticulturae, 118: 189-199.

Batlle I, Tous J, 1997. Carob tree. *Ceratonia siliqua* L. Promoting the Conservation and Use of Underutilized and Neglected Crops. 17. Institute of Plant Genetics and Crop Plant Research, Gatersleben/International Plant Genetic Resources Institute, Rome, Italy.

Gübbük H, Kafkas E, Güven D, Güneş E, 2010. Physical and Phytochemical Profile of Wild and Domesticated Carob (*Ceratonia siliqua* L.) Genotypes. Spanish Journal of Agricultural Research 8(4): 1129-1136.

Johnson RA, Wichern DW, 2007. Applied Multivariate Statistical Analysis. 6th ed. Prentice Hall, New York.

Konate I, FilaliI-Maltouf A, Berraho EB, 2007. Diversity Analysis of Moroccan Carob (*Ceratonia siliqua* L.) Accessions Using Phenotypic Traits and Rapid Markers. Acta Botanica Malacitana 32:79-90.

Mekonnen F, Mekbib F, Kumar S, Sharma TT, 2014. Phenotypic Variability and Characteristics of Lentil (Lens culinaris Medik.) Germplasm of Ethiopia by Multivariate Analysis. Journal of Agricultural and Crop Research Vol. 2(6), pp, 104-116, June.

Naghmouchi S, Khouja ML, Romero A, Tous J, Boussaid M, 2009. Tunisian carob (*Ceratonia siliqua* L.) populations: Morphological variability of pods and kernel. Scientia Horticulturae 121: 125-130.

Özdamar K, 2011. Paket Programlar ile İstatistiksel Veri Analizi-II, 8.Baskı. Eskişehir Kaan Kitabevi.

Pekmezci M, Gübbük H, Eti S, Erkan M, Onus N, Biner B, Adak N, Karaşahin I, 2005. Doğu Akdeniz Bölgesinde Yabani ve Kültür Formunda Yetişen Keçiboynuzu Tiplerinin Seleksiyonu. BAHÇE 34 (2): 73 – 82.

Pekmezci M, Gübbük H, Eti S, Erkan M, Onus N, Karaşahin I, Biner B, Adak N, 2008. Batı Akdeniz ve Ege Bölgesi'nde Yabani ve Kültür Formunda Yetişen Keçiboynuzu Tiplerinin Seleksiyonu. Akdeniz Üniversitesi Ziraat Fakültesi Dergisi 21(2): 145–153.

Rencher AC, 2002. Methods of Multivariate Analysis, 2nd. Edt. New York:John Wiley.

Sidina MM, Hansali EM, Wahid N, Ouatmane A, Boullı A, Haddiou ı A, 2009. Fruit and Seed Diversity of Domesticated Carob (*Ceratonia siliqua* L.) in Morocco. Scientia Horticulturae 123: 110-116.

Tabachnick BG, Fidelli LS, 2007.Using multivariate statistics. Boston: Pearson/Allyn & Bacon.

Tetik N, Turhan İ, Oziyci HR, Gübbük H, KarhanM, Ercişli S, 2011. Physical and Chemical Characterization of *Ceratonia siliqua* L. Germplasm in Turkey. Scientia Horticulturae 129: 583-589.