Cluster Analysis of Morpho-Economic Characteristics in Some Subspecies and F_1 Hybrids Belonging to the *G.hirsutum* L. Species

Qodirjon Komiljonovich Ashirov[1], Bakhtiyar Khushbakovich Amanov[2], Nilufar Muydinovna Tursunova[2,*]

[1]Andijan Institute of Agriculture and Agrotechnologies, Andijan, Uzbekistan
[2]Chirchik State Pedagogical University, Tashkent, Uzbekistan

Abstract This article demonstrated that some subspecies and varieties belonging to *G.hirsutum* L. species were divided into 2 cluster groups based on morpho-economic characteristics, and F_1 hybrids were divided into 3 cluster groups. Based on the initial sources, it was determined that the Euclidean distance ranged from 0,2-2,9 units, while in F_1 hybrids, this distance ranged from 0,3-26,1 units. During the experiment, it was found that individual selection based on each characteristics when evaluating the morpho-economic traits of initial sources and F_1 plants often does not lead to accurate conclusions, cluster analysis, on the other hand, allows for the selection of samples that are close to each other based on the relationship between these characteristics and multiple traits. Furthermore, when selecting a form for the initial source, it is possible to use hybrids within a cluster group, rather than a single variety, to transfer a specific trait.

Keywords Cotton, Subspecies, Variety, Hybrid, Generation, Genotype, Cluster, Valuable economic traits, Raw cotton in one boll, Fiber length, Fiber yield, 1000 seed weight

1. Introduction

In these days, cotton cultivation areas are increasing because its high profitability and growing demand are beneficial for agriculture. Cotton cultivation technologies and methods are constantly improving, which increases its productivity. This technological crop is widely used in various industries around the world. One of the main advantages of cotton is its fiber, which is used in the production of yarn, fabric, and other products. Cotton also has seeds with high nutritional value and oil content, making it very important for the feed industry. Therefore, cotton is a crop that plays a significant role in the global economy.

While cotton was previously grown mainly for fiber production, its other values are currently being explored. Through technologies and specific research, new prospects related to cotton are being developed, which further increases its relevance. Research on genetic modification of cotton, creating pest-resistant varieties, and adapting them to climate change is one of the most pressing issues in these days. Scientists have also demonstrated through their research that it is possible to obtain new forms and valuable sources with beneficial traits and characteristics through interspecific and intraspecific hybridization [2,3,4,6,7,10,13,15].

Furthermore, scientists have determined that the fertility efficiency of hybrid bolls and seed from hybrid bolls in intraspecific and interspecific hybridization of cotton species, the positive or negative outcome of the results, depends on the phylogenetic proximity or distance of the species involved in the hybridization and the physiological characteristics of the initial sources [12].

In research studies, a variety of *G.hirsutum* cotton was crossed with cultivated diploid species having A_1, A_2 genomes, resulting in the production of sterile triploids. These hybrids, after chromosome doubling using colchicine to overcome sterility *G.hirsutum* L. x *G.herbaceum, G.arboreum* L. x *G.hirsutum* L. amphiploids were crossed with *G.hirsutum* L., resulting in forms close to the valuable economic traits of the *G.hirsutum* L. species. In recent years, as a result of research conducted in our country, the fertility and seed results of hybrid bolls were studied based on intraspecific and interspecific hybridization of *G.herbaceum* L. and *G.arboreum* L. intraspecific diversity. Transgressive forms with valuable economic traits were identified and recommended as initial sources in practical breeding research [14].

It has been noted that the genetic patterns identified in hybrids of one group of cotton varieties based on quantitative traits may differ in hybrids of another group of varieties, and

the genetic-breeding indicators relate to the degree of dominance, epistasis, combining ability, heritability, and others [1].

On the other hand, the placement of the first fruiting branch at the lower node, a morphological trait of cotton, is considered a sign of early maturity for varieties, and a strong correlation between early maturity and heat resistance has been noted [9,11].

Besides, to assess the genetic diversity of yield traits of cotton, it was found that genetic variability among genotypes is significantly expressed using the cluster analysis method, which indicates that yield can be increased by using genotypes from different clusters in hybridization [8].

For 40 genotypes of species *G.hirsutum* L., 15 quantitative traits were observed to separate into 7 cluster groups. Clusters I and III each contained 8 genotypes, and cluster VII contained 7 genotypes. Based on hierarchical cluster analysis, all genotypes were grouped into 7 clusters, where clusters VII each contained 13 genotypes and cluster IV contained 8 genotypes. The random distribution of genotypes indicates a lack of parallelism between geographic and genetic diversity [5].

2. The Object and the Methods of the Research

Cluster analysis, a biostatistical method, is a software program that encompasses various algorithms used to classify objects under study. By applying these operations, the initial commonality of selected objects is divided into clusters, or groups of similar objects. The complexity of the cluster method lies in the fact that the objects studied for research are multidimensional. That is, they are described by not one, but several indicators, and their grouping into clusters is carried out in multidimensional space.

In our experiment, we used the Statgraphics computer program to cluster varieties, employing Euclidean distance as a measure of genetic proximity and Ward's method as the clustering technique. The Statgraphics computer program is a convenient tool for statistical data analysis, allowing us to determine mean values, standard deviations, standard errors, confidence levels, and other statistical information.

Theoretical basis of cluster analysis: As is known from multivariate statistics, the natural measure of distance between objects is the Mahalanobis distance. Euclidean distance is a special case of the Mahalanobis distance. Geometric interpretation of Euclidean distance between two points (traits) x and y: The Euclidean distance (d_{12}) between genotypes (x_1, y_1, and x_2, y_2) obtained for the two analyzed traits is determined using the Pythagorean theorem:

$$d_{12} = \sqrt{(x^1 - x^2)^2 + (y^1 - y^2)^2}$$

If n traits are measured in the plants of two populations (where n is the number of traits), then the following formula

arises for n-dimensional space:

$$d_{12}^2 = \sum_{R=1}^{n} \sqrt{(x_1^{(R)} - x_2^{(R)})^2} \qquad (1)$$

Here, $x_1^{(R)}$ and $x_2^{(R)}$ represent the average value of the R trait in the first and second populations.

If the units if measurement for the studied traits are different, then the Euclidean distance becomes meaningless. Therefore, the trait values are divided by the deviation from the square root of the pooled variance for both populations.

$$d_{12}^2 = \sum_{R=1}^{n} \left(\frac{x_1^{(R)} - x_2^{(R)}}{\sigma^{(R)}} \right)^2 \qquad (2)$$

It is also worth noting that when Euclidean distance is used as an indicator determining the degree of genetic proximity of samples, the following limitations exist: multivariate observations are extracted from a normally distributed parent population; it is important that the $x^{(1)}$, $x^{(2)}$, $x^{(n)}$ traits are equally significant in classifying genotypes.

Due to the fact that more than two populations were taken to determine morpho-economic traits in the initial sources and F_1 hybrids, it was necessary to group the genotypes according to their proximity to each other. Therefore, the task was to determine the different levels of physiological indicators of these varieties and hybrids, and to combine the traits between them according to their proximity, and for this purpose, cluster analysis was used. In our research, to perform cluster analysis, 5 different morpho-economic traits of the hybrid obtained by crossing the subspecies *G.hirsutum* L. were analyzed. The morpho-economic traits of these varieties and F_1 hybrids, such as plant growth period, cotton weight per boll, fiber length, fiber yield, and 1000 seed weight, were determined and statistical analysis was performed.

On the other hand, to cluster the studied initial sources and F_1 hybrids, we used the Statgraphics computer program, employing Euclidean distance as a measure of genetic proximity and Ward's method as the clustering technique.

3. The Results of the Research

The minimum number of cluster groups combining the analyzed initial sources and F_1 hybrids was 3, and the maximum number was 9. The indicators of traits such as plant growth period, cotton weight per boll, fiber length, fiber yield, and 1000 seed weight were determined for these varieties and F_1 hybrids, and cluster analysis was performed (Tables 1, 2).

As a result of studying the characteristics mentioned above, two groups clusters were identified and analyzed in the initial sources that differ in the set of characteristics and retain their personalizations in the genotype and hybrid generations, and three groups clusters were identified and analyzed in the F_1 plants (Figure 1).

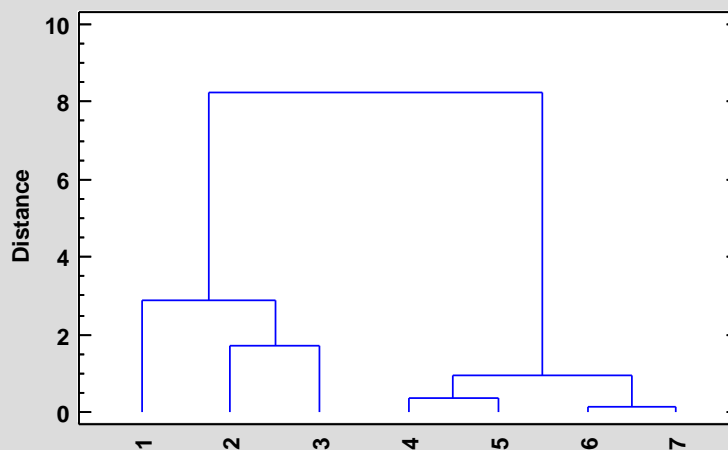**Table 1.** Clustering of initial sources based on morpho-economic traits

Cluster numbering	Source numbering	Clustered groups
		Subspecies and varieties
I	1; 2; 3	*G.hirsutum* L. subs. *mexicanum* var.*microcarpum* f.*palmeri*, G.*hirsutum* L. subs. *punctatum*, G.*hirsutum* L. subs. *punctatum* var. *gambia*
II	4; 5; 6; 7	*G.hirsutum* L. subs. *euhirsutum* "Sulton" variety, "Omad" "Genofond-2" "Kelajak" varietes

Table 2. Morpho-economic traits of initial sources

Cluster numbering	Growing season	Cotton weight per boll	Fiber length	Fiber yield	1000 seed weight
I	191,77	2,28	23,63	27,8	92,67
II	113,53	6,18	34,45	35,73	124,88

Table 3. Morpho-economic indicators of F_1 hybrids

Cluster number	Growing season	Cotton weight per boll	Fiber length	Fiber yield	1000 seed weight
I	157,63	4,3	30,61	33,79	106,82
II	147,25	5,1	34,83	32,97	116,73
III	148,39	4,73	33,48	35,12	123,79

Figure 1. Dendrogram of cluster separation based on morpho-economic traits of initial sources

The first cluster group, in which the morpho-economic characteristics of the initial sources were studied, including the subspecies *G.hirsutum* L. subs. *mexicanum* var.*microcarpum* f. *palmeri*, *G.hirsutum* L. subs. *punctatum*, *G.hirsutum* L. subs. *punctatum* var. *gambia*, demonstrated lower indicators for all characteristics of the sample compared to the varieties and hybrids in the other cluster.

The second cluster group, consisting of 4 varieties, primarily exhibited higher indicators compared to the first cluster group (with subspecies) in terms of plant growing season, cotton weight per boll, fiber length, fiber yield, and 1000 seed weight. These varieties include the cultivated varieties: *G.hirsutum* L. subs. *euhirsutum* "Sulton", *G.hirsutum* L. subs. *euhirsutum* "Omad", *G.hirsutum* L. subs. *euhirsutum* "Genofond-2", *G.hirsutum* L. subs. *euhirsutum* "Kelajak".

Study of morpho-economic characteristics in F_1 hybrid plants shown below: the first cluster group, consisting of the hybrids G.hirsutum L. subs.*mexicanum* var.*microcarpum* f.*palmeri* x "Kelajak", *G.hirsutum* L. subs.*microcarpum* f.*palmeri* x "Genofond-2", *G.hirsutum* L. subs.*mexicanum* var.*microcarpum* f.p*almeri* x "Omad", *G.hirsutum* L. subs.*mexicanum* var.*microcarpum* f.*palmeri* x "Sulton", *G.hirsutum* L. subs.*punctatum* x "Genofond-2", *G.hirsutum* L. subs.*punctatum* var.*gambia* x "Kelajak", *G.hirsutum* L. subs.*punctatum* var.*gambia* x "Genofond-2" , G.hirsutum L. subs.*punctatum* var.*gambia* x "Omad", *G.hirsutum* L.

subs.*punctatum* var.*gambia* x "Sulton", exhibited lower morpho-economic indicators compared to the second and third cluster groups in terms of growing season, cotton weight boll, fiber length, and 1000 seed weight. However, in terms of fiber yield, it demonstrated higher morpho-economic indicators than the second cluster group and lower indicators than the third cluster group (Tables 3, 4; Figure 2).

The second cluster group, consisting of the hybrids "Kelajak" x *G.hirsutum* L. subs.*mexicanum* var.*microcarpum* f.*palmeri*, "Sulton" x *G.hirsutum* L. subs.m*exicanum* var.

microcarpum f.*palmeri*, *G.hirsutum* L. subs.*punctatum* x "Kelajak", "Kelajak" x *G.hirsutum* L. subs.*punctatum*, "Genofond-2" x *G.hirsutum* L. subs.*punctatum*, *G.hirsutum* L. subs.*punctatum* x "Omad", showed higher morpho-economic indicators compared to the first and third cluster groups in terms of plant growing season (147.25 days), cotton weight per boll (5.1 g), and giber length (34.83 mm). However, in terms of fiber yield (32.97%), it exhibited lower morpho-economic indicators compared to the other F_1 hybrid plants in the first and second cluster groups (Tables 3, 4; Figure 2).

Table 4. Cluster of F_1 hybrids based on morpho-economic traits

| Cluster numbering	F_1 numbering	Cluster-grouped varieties, F1
I	1; 7; 3; 17; 19; 5; 21; 11	*G.hirsutum* L. subs.*mexicanum* var.*microcarpum* f.*palmeri* x "Kelajak", *G.hirsutum* L. subs.*mexicanum* var.*microcarpum* f.*palmeri* x "Genofond-2", *G.hirsutum* L. subs.*mexicanum* var.*microcarpum* f.*palmeri* x "Omad" variety, *G.hirsutum* L. subs.*mexicanum* var.*microcarpum* f.*palmeri* x "Sulton" variety, *G.hirsutum* L. subs. *punctatum* x "Genofond-2" variety, *G.hirsutum* L. subs. *punctatum* var.*gambia* x "Kelajak" variety, *G.hirsutum* L. subs. *punctatum* var.*gambia* x "Genofond-2" variety, *G.hirsutum* L. subs.*punctatum* var.*gambia* x "Omad" variety, *G.hirsutum* L. subs.*punctatum* var.*gambia* x "Sulton" variety
II	2; 10; 13; 8; 12; 9	"Kelajak" x *G.hirsutum* L. subs.*mexicanum* var.*microcarpum* f.*palmeri*, "Sulton" variety x *G.hirsutum* L subs.*mexicanum*var.*microcarpum* f.*palmeri*, *G.hirsutum* L. subs. *punctatum* x "Kelajak" variety, "Kelajak" variety x *G.hirsutum* L. subs. *punctatum,* "Genofond-2" variety x *G.hirsutum* L. subs. *Punctatum, G.hirsutum* L. subs. *punctatum* x "Omad" variety
III	4; 20; 14; 16; 18; 6; 15; 22; 24	"Genofond-2" x G.hirsutum L. subs.*mexicanum* var.*microcarpum* f.*palmeri*, "Omad" x *G.hirsutum* L. subs.*mexicanum* var.*microcarpum* f.*palmeri*, "Omad" variety x *G.hirsutum* L. subs. *punctatum, G.hirsutum* L. subs. punctatum x "Sulton" variety, "Sulton" variety x *G.hirsutum* L. subs.*punctatum, G.hirsutum* L. subs.*punctatum* var.*gambia* x "Kelajak" variety, "Kelajak" variety x *G.hirsutum* L. subs.*punctatum* var.*gambia,* "Omad" variety x *G.hirsutum* L. subs.*punctatum* var.*gambia,* "Sulton" variety x G.hirsutum L. subs.*punctatum* var.*gambia*
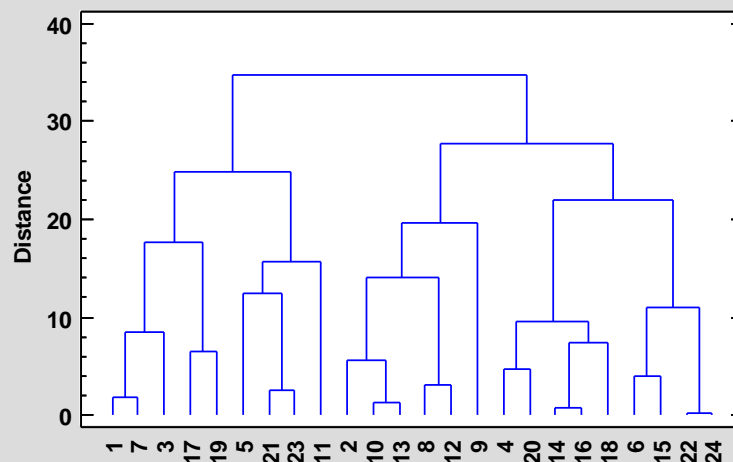
Figure 2. Dendrogram of F_1 hybrids clustered by morpho-economic traits

The third cluster group, consisting of hybrids such as "Genofond-2" x *G.hirsutum* L. subs.*mexicanum* var.*microcarpum* f.*palmeri*, "Omad" x *G.hirsutum* L. subs.*mexicanum* var.*microcarpum* f.*palmeri*, "Omad" x *G.hirsutum* L. subs.*punctatum*, *G.hirsutum* L. subs.*punctatum* x "Sulton", "Sulton" x *G.hirsutum* L. subs.*punctatum*, *G.hirsutum* L. subs.*punctatum* var.*gambia* x "Kelajak", "Kelajak" x *G.hirsutum* L. subs.*punctatum* var.*gambia*, "Omad" x *G.hirsutum* L. subs.*punctatum* var.*gambia*, and "Sulton" x *G.hirsutum* L. subs.*punctatum* var.*gambia*, demonstrated high morpho-economic indicators in terms of fiber yield (35.12%) and 1000 seed weight (123.79 g), while showing lower results compared to the second cluster group in terms of growth period (148.39 days), cotton weight per boll (4.73 g), and fiber length (33.48 mm) (Figure 2).

4. Conclusions

Our research revealed that the studied varieties were divided into 2 cluster groups based on morpho-economic traits, and the F_1 hybrids were divided into 3 cluster groups. It was found that the Euclidean distance for the initial sources ranged from 0.2 to 2.9 units, while for the F_1 hybrids, this distance ranged from 0.3 to 26.1 units. During the experiment, it was observed that evaluating the morpho-economic traits of the initial sources and F_1 plants by selecting each trait individually often does not lead to accurate conclusions. Cluster analysis, on the other hand, allows for the selection of samples that are close to each other in terms of several traits and their interrelationships. This makes it possible to use not just one variety, but a group of hybrids within a cluster to transfer a specific trait when selecting a form for the initial source.

REFERENCES

[1] Abzalov M.F. Interaction of genes in cotton *Gossypium hirsutum* L. // Izd. "FAN" AN RUz. Tashkent, 2008. - P. 106-112.

[2] Amanov, B., Muminov, K., Samanov, S., Abdiyev, F., Arslanov, D., & Tursunova, N. (2022). Cotton introgressive lines assessment through seed cotton yield and fiber quality characteristics.

[3] Amanov B.Kh., Nabiyeva N.N., Tutushkina N.V. Phylogenetic relationships of some polymorphous cotton species of the section Magnibracteolata. Modern biology and genetics. 2023; №1 (3), 40-53.

[4] Amanov, B., Muminov, K., Samanov, S., Abdiev, F., Arslanov, D., & Tursunova, N. (2022). Cotton introgressive lines assessment through seed cotton yield and fiber quality characteristics. -P. 321-330.

[5] Asha R., Lal Ahamed M, D. Ratna Babu and P. Anil Kumar "Multivariate Analysis in Upland Cotton (*Gossypium hirsutum* L.)" Madras Agric. J., 100 (4-6): 333-335, June 2013.

[6] Deyi Shao, Tao Wang, Haiping Zhang, Jiam EI Zhu, Feiyu Tang Genetic variability and Heritability Study in F_2 Segregants of Desi Cotton for Yield and its Component Traits. // Journal Botany. - Pakistan. 2016. - № 48. - P. 1945-1949.

[7] El-Zanaty A.M., K.F.M. Salem, R.M. Esmail. Detection of Genetic Diversity in Egyptian cotton (*G.barbadense* L.) varieties using RAPD markers and morphological traits. // Nature and Science. 2012. - №10. - P.123.

[8] Heilegiorgis D., Mesfin M., Genet T. Genetic divergence analysis on some bread wheat genotypes grown in Ethiopia Journal of Central European Agriculture, 2011. №12 (2), -P. 344-352.

[9] Jatoi W.A., M.J.Baloch, A.Q.Panhwar, N.F.Veesar, Panhwar S.A. Charaterization and identification of early maturing upland cotton varieties. // Sarhad Journal Agric. 2012. - №26. -P. 993-996.

[10] Jawahar T.G., Patil B.R. Genetic variability and Heritability Study in F_2 Segregants of Desi Cotton for Yield and its Component Traits. // International Journal of Current Microbiology and Applied Sciences. 2017. - №6. - P. 2679-2684.

[11] Lakho A.R., Baloch M.J., Bhutto H.U., Chang M.S., Tunio G.H., Solangi M.Y. Impact of early maturity on seed cotton yield and some economic characters in upland cotton. // Sindh Balochist Journal Plant Scienc. 2007. -№3. - P.48-52.

[12] Linskens G.F. Inhibition reaction during incompatible pollination and its overcoming // Plant Physiology. – 1973. – V. 20. – Issue 1. – Pp. 192-202.

[13] Murtagh F., Legendre P. Ward's Hierarchical Agglomerative Clustering Method: Which Algorithms Implement Ward's Criterion? Journal of Classification 2014. 31: 274-295.

[14] Muminov Kh.A. Use of A_1, A_2 genomic types in the enrichment of genotypes of cultivars of *G.hirsutum* L. // Diss.doc.bio. sci. (DSc). Tashkent. 2017. - Pp. 14-16.

[15] Nabieva N.N. Phylogenetic relationships of some polymorphic cotton species belonging to the section *Magnibracteolata.* // Thesis of diss.cand. bio.sci. Tashkent. 2010. - Pp. 10-12.