Rev. Brasil. Genet. 18, 1, 99-103 (1995)
(Brazil. J. Genetics)

Use of numerical taxonomy to compare "Mocó" Cotton with other Cotton species and races

José de Alencar Nunes Moreira, Elêusio Curvelo Freire, José Wellington dos Santos and Robson Macêdo Vieira

ABSTRACT

'Mocó" Cotton (*Gossypium hirsutum* r. *marie galante* Hutch), is grown in Northeastern Brazil, but its origin is unknown. The objective of our study was to compare mocó with other cotton species and races to clarify its origin and botanical classification. The study was conducted on *G. hirsutum* r. *marie galante* Hutch (exotic), *G. barbadense* var. *brasiliense*, the "verdão" (green) types of *G. hirsutum* L., and the cotton plants generally known under the mocó name and represented by cream fiber, Francisco Raimundo; "mocó das Serras" (highland mocó) from the Seridó Valley, "mocozinho", and, finally, improved mocó. The variables chosen for study were leaf, bract, and flower traits measured in five plants selected at random before flowering. The methodology consisted of the use of cluster analysis, principal coordinates, canonical discriminant analysis and linear discriminant functions. The results favor the hypothesis that the mocó cotton lineage consists of several trunks rather than a single one, and that the *"marie galante"* race is one of them.

INTRODUCTION

Mocó cotton (*Gossypium hirsutum* r. *marie galante* Hutch), grown in Northeastern Brazil, is of unknown origin. Current thinking about the origin of mocó follows two major lines. The autochthonous theory proposes that mocó originated from the Seridó region in the state of Rio Grande do Norte and that it's recently derived from an ancestral type. The second theory suggests that mocó was derived from cotton plants introduced at different times in the Seridó region. Stephens (1967) agrees with the second theory and proposes that mocó belongs to the *marie galante* race which originated in the Antilles. According to Stephens, the mocó race may have been introduced into Northeastern Brazil by the Dutch from the Antilles, via west Africa.

Freire and Moreira (1991) disagree with the latter hypothesis and propose that representatives of this cotton plant may form an individualized group separate from the *marie galante* race and from *G. barbadense* L.

The present investigation is a continuation of the study by Freire and Moreira (1991), to clarify the relation of mocó to other cotton species and races and contribute to to mocó's botanical classification.

MATERIAL AND METHODS

The materials were obtained from the "Banco Ativo de Germoplasma" (Active Germplasm Bank, BAG) maintained by the "Centro Nacional de Pesquisa

Centro Nacional de Pesquisa do Algodão (CNPA), EMBRAPA, Caixa Postal 174, 58107-720 Campina Grande, PB, Brasil. Send correspondence to J.A.N.M.

de Algodão (National Center for Cotton Research) in the Experimental Field of Patos (State of Paraiba) of this Institution. The field is located in the physiographic region of Seridó, which is the natural habitat of mocó cotton in Northeastern Brazil. The region is characterized by a predominance of dark, non-calcic soils, arid climate, and annual rainfall of 660 mm.

The materials studied were previously described by Freire and Moreira (1991). The accessions were classified into the following four groups: (1) *G. hirsutum* r. *marie galante* Hutch (exotic), (2) *G. barbadense* var *brasiliense*, (3) "verdões" (green) types belonging to *G. hirsutum* L., and those generally known as mocó. The mocó group was divided into five subgroups, cream fiber mocó (4), Francisco Raimundo (5), highland mocó (6), from the Seridó Valley, "mocozinho" (7), and improved mocó (8).

Cotton plants of each type and subtype were planted in 10 m progeny rows with no replications and distributed at random in the trial area. Five plants were selected randomly from each group and subgroup before flowering and morphological measurements of leaf, bract, flower traits and indices were taken from Freire and Moreira (1991).

The methodology used to determine taxonomic classification included cluster analysis, principal coordinates, canonical discriminant analysis and linear discriminant functions. For cluster analysis, we started from the matrix and calculated the Mahalanobis D^2 distances by the method of Morrison (1976). For clustering, we used the algorithm of the single linkage method with the D^2 distances according to the sequence described by Dunn and Everitt (1982). The data thus generated were used to construct the dendrogram, which illustrates the relationships among the cotton plants under study (Figure 1).

The principal coordinate analysis was based on the same matrix with the Mahalanobis D^2 distances used in cluster analysis. The matrix was transformed to another F by the method of Sneath and Sokal (1973). After transformation, the eigenvalues and eigenvectors of the matrix were obtained and the principal coordinates were calculated (Figure 2).

Discriminant analysis and discriminant linear functions were applied to the data for fiber length, petal length, bract size index, and staminal column length according to the method of Freire and Moreira (1991). These variables were those of highest discriminant power identified by using the Stepdisc procedure of the Statistical Analysis System (SAS) Library (1985). The Discrim procedure of this program was also used for discriminant analysis and to calculate the functions used to separate mocó from the other cottons studied.

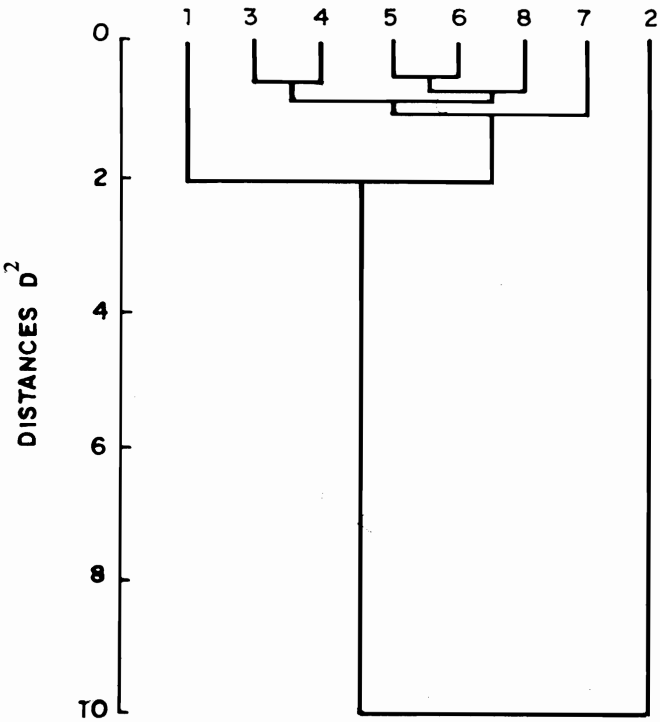

Figure 1 - Dendogram showing the relationships among cotton types. 1. *marie galante*; 2. barbadense; 3. green types; 4. Cream fibers; 5. Francisco Raimundo; 6. highland mocó; 7. mocozinho; 8. improved mocó.

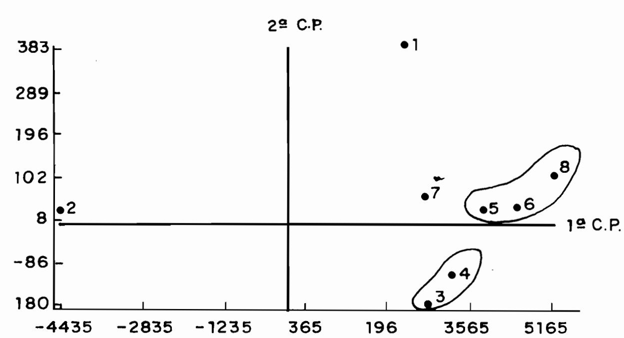

Figure 2 - Distribution of cotton accessions over the two principal coordinates based on Mahalanobis D^2 distances. 1. *marie galante*; 2. barbadense; 3. green types; 4. Cream fibers; 5. Francisco Raimundo; 6. highland mocó; 7. mocozinho; 8. improved mocó.

RESULTS AND DISCUSSION

The traits studied varied considerably among the various groups of cotton plants.

Marie galante and *G. barbadense* var. *brasiliense* seem to be isolated from each other and from the other cotton plants, based on cluster analysis (Figure 1). Thus, *marie galante* forms an entity separate from all others, especially when compared with the mocó types. Two other distinct groups were found; one consisting of verdão and cream fiber mocó and the second of Francisco Raimundo mocó and Seridó Valley mocó. Improved mocó was linked to the latter group, whereas mocozinho remained isolated.

The classification by cluster analysis agrees with that obtained based on group means, starting from Euclidean distances (Freire and Moreira, 1991). The difference resides in the position of verdão and mocozinho. The remaining results were in agreement in both studies, indicating the similarity of classification.

Principal coordinate analysis showed that the first two eigenvalues accounted for almost 90% of the variability. For this reason, the two principal coordinates were chosen to represent the new variable, starting from the first two eigenvectors.

The cotton plants were plotted along the two axes representing the first and second principal coordinates (Figure 2). Three distinct clusters are evident by this technique: 1) Francisco Raimundo mocó, Seridó mocó and improved mocó; 2) mocozinho, and 3) verdão and cream fiber mocó. However, *marie galante* and *G. barbadense* appeared to be isolated and distinct from each other and in relation to the other clusters. A repetition of the clustering pattern was expected because of the magnitude of variability in the first two eigenvalues of the distance matrix.

The first cluster includes those accessions that are related to one another. Francisco Raimundo mocó was selected by Captain Francisco Raimundo, farmer from Rio Grande do Norte, from introductions from the Seridó Valley in the same state. Francisco Raimundo must be related to the mocó types grown in this region, as well as the mocó types from the Seridó Valley located in the same cluster as Francisco Raimundo. The similarity of the two accessions and the improved mocó can be explained in the same manner, since improved mocó is the product of selection from the Seridó region. Therefore, clustering in both cases reflects the phylogenetic relationships involving the three cotton types because of the common trunk from which they originated.

There was a separation of the cluster (5, 6 and 8) formed by the mocó types from that formed by the *marie galante* race, as also found by Freire and Moreira (1991).

Analysis of Figure 2 also reveals that the position of *G. barbadense* is distinct in relation to the mocó types in the first cluster.

The results disagree with the assumption of Boulanger and Pinheiro (1972) that mocó originated from crosses occurring continuously and at random between *G. barbadense* and *G. hirsutum*.

Figure 3 illustrates the arrangement of the points corresponding to the different cotton accessions on the plane formed by the first and second canonical axes. The positions repeated the relationships among cotton accessions obtained by cluster and principal coordinate analysis. There was a definite separate position of *G. barbadense* in relation to the remaining accessions and the concentration of points corresponding to *Marie galante* and to the mocó types designated Francisco Raimundo, Seridó Valaley mocó and improved mocó. There were however, two overlapping zones: the first refers to mocozinho in the *Marie galante* cluster of points, and the second refers to mocozinho in the cluster of points corresponding to cream fiber mocó.

The percentages of the classification within and outside their respective groups obtained by canonical discriminante analysis are included in Table I. The percentages of classification of *Marie galante*, *G. barbadense*, verdão and Francisco Raimundo mocó in their own group was 100%. Thus, on the basis of the variables studied, these appear to be distinct, since they were not confused with the remaining ones. The Seridó valley mocó, however, overlapped with Francisco Raimundo

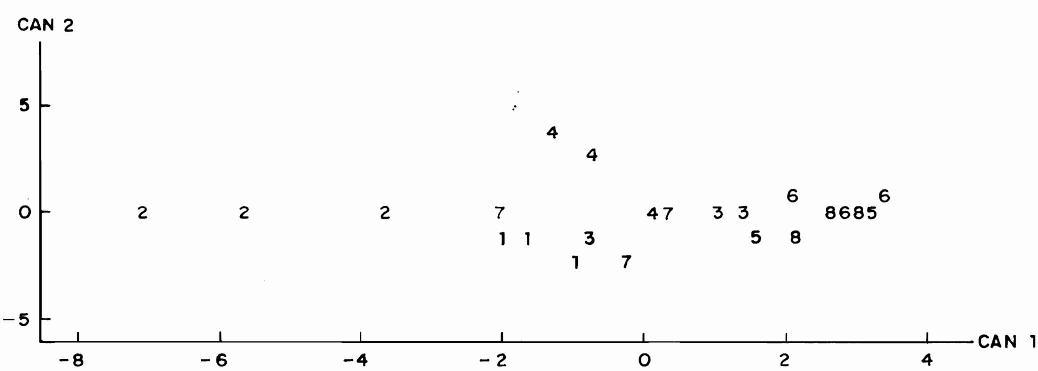

Figure 3 - Position of the cotton races and species on the plane obtained by the first and second canonical variables. 1. *marie galante*; 2. barbadense; 3. green types; 4. Cream fibers; 5. Francisco Raimundo; 6. highland mocó; 7. mocozinho; 8. improved mocó.

in one third of cases, and improved mocó overlapped with the Seridó Valley, also in one third of cases.

The latter result was expected when one considers the similarity among the mocó types studied. In this respect, canonical discriminant analysis reinforced what had been demonstrated with the use of the previous tech- niques, when the above mocó types formed an individual group separate from *marie galante* and *G. barbadense*.

The discrepancy resided in the positions of cream fiber mocó and of mocozinho. These accessions, did not show a well-defined position on the plane formed by the two canonical axes in discriminant analysis (Figure 3). This is because mocozinho overlaps with *Marie galante* in one third of cases and cream fiber mocó overlaps with mocozinho also in one third of cases (Table I).

Because these accessions were never confused with Francisco Raimundo mocó, Seridó Valaley mocó, improved mocó, and verdão, they can probably be considered to be differentiated from the remaining accessions (Table I).

Mocozinho was the only cotton type showing a certain similarity to *marie galante* (Table II). In contrast, the Francisco Raimundo, Seridó and improved mocó, consistently revealed a difference from *marie galante* by the three numerical taxonomy techniques employed.

The result obtained appears to support the hypothesis that mocó did not originate from a single

Table I - Observation numbers and classification percentages within eight cotton accessions based on discriminant analysis.

Accessions	*Marie galante*	*G. barbadense*	Green types	Cream fiber	Mocó Francisco Raimundo	Highland mocó	Mocozinho	Improved mocó
	3[1]	0	0	0	0	0	0	0
Marie galante	100.00[2]	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0	3	0	0	0	0	0	0
G. barbadense	0.00	100.00	0.00	0.00	0.00	0.00	0.00	0.00
	0	0	3	0	0	0	0	0
Green fiber	0.00	0.00	100.00	0.00	0.00	0.00	0.00	0.00
	0	0	0	2	0	0	1	0
Cream fiber	0.00	0.00	0.00	66.67	0.00	0.00	33.33	0.00
	0	0	0	0	3	0	0	0
Mocó Fco. Raimundo	0.00	0.00	0.00	0.00	100.00	0.00	0.00	0.00
	0	0	0	0	1	2	0	0
Highland mocó	0.00	0.00	0.00	0.00	33.33	66.67	0.00	0.00
	1	0	0	0	0	0	2	0
Mocozinho	33.33	0.00	0.00	0.00	0.00	0.00	66.67	0.00
	0	0	0	0	0	1	0	2
Improved mocó	0.00	0.00	0.00	0.00	0.00	33.33	0.00	66.67

1 - Observation no.
2 - Classification percent.

Table II - Mahalanobis distances between accessions of cotton studied.

Accessions	*Marie galante*	*G. barbadense*	Green types	Cream fiber	Mocó Francisco Raimundo	Highland mocó	Mocozinho	Improved mocó
Marie galante (1)	0.00							
G. barbadense (2)	18.87	0.00						
Green fiber (3)	10.07	37.16	0.00					
Cream fiber (4)	11.63	29.73	12.44	0.00				
Mocó Fco. Raimundo (5)	13.85	57.20	4.76	14.03	0.00			
Highland mocó	23.64	70.68	9.17	17.68	2.31	0.00		
Mocozinho	1.66	27.78	8.97	10.64	8.90	15.35	0.00	
Improved mocó	19.75	67.16	7.60	19.08	2.40	6.76	16.45	0.00

Table III - Linear discriminant functions for the cotton groups studied.

Variables	Accessions							
	1	2	3	4	5	6	7	8
Fiber length	16.47	14.05	17.80	17.06	19.60	18.80	19.11	17.17
Petal length	3.93	3.64	4.48	4.17	4.81	4.52	4.40	4.14
Bract size	-2.47	-1.73	-2.78	-2.01	-3.06	-2.95	-2.89	-2.66
Staminal column length	0.67	0.32	0.96	0.25	0.61	0.86	1.23	0.50
Constants	-318.90	-253.01	-389.19	-374.40	-451.05	-418.81	-432.97	-341.94

trunk but from different ones, among them the *marie galante* race itself. Table III showns the eight linear functions used to separate each cotton accession studied by discriminant analysis. The functions operate in such a manner that a given set of observations of the variables considered is allocated to the group that presents the highest value for that function.

RESUMO

O algodoeiro mocó (*Gossypium hirsutum* r. *marie galante* Hutch) cultivado no Nordeste brasileiro, até a presente data, é uma planta de origem desconhecida. O que se sabe acerca do seu aparecimento nesta região encontra-se no terreno de meras conjecturas. O presente trabalho tem por objetivo esclarecer a posição do mocó frente a outras espécies e raças de algodoeiro com vistas a contribuir para o esclarecimento de sua origem e classificação botânica. Foram estudados o *G. hirsutum* r. *marie galante* Hutch (exótico), o *G. barbadense* L. e os algodoeiros conhecidos com a designação geral de mocó representados pelo fibra creme, Francisco Raimundo, mocó das serras, mocozinho e, por fim, o mocó melhorado. As variáveis escolhidas basearam-se nas características das folhas, brácteas e flores e foram medidas em cinco plantas escolhidas ao acaso, antes da floração. A metodologia adotada constou do emprego da análise de conglomerados, coordenadas principais, análise discriminante canônica e funções discriminante lineares. Os resultados encontrados favorecem a tese de que o algodoeiro mocó pode não ter tido um tronco único, porém diversos, entre os quais o da própria raça *marie galante*.

REFERENCES

Boulanger, G. and **Pinheiro, D.** (1972). Evolution de La Production Cotonniere au Nord-Est Du Bresil. IV. Polymorphisme Des types de cotonniers Cultives; Origine Des Types "moco" et "verdão". *Cot. Fib. Trop. 26*: 335-353.

Dunn, G. and **Everitt, B.S.** (1982). *An Introduction to Mathematical Taxonomy*. Cambridge University Press, Cambridge, pp. 153.

Freire, E.C. and **Moreira, J.A.N.** (1991). Relações genéticas entre o algodoeiro mocó e diferentes espécies e raças de algodoeiro. *Rev. Brasil. Genet. 14*: 393-411.

Freire, E.C., Santos, M.S.S., Medeiros, L.C., Andrade, F.P. and **Santos, E.O.** (1990). *Avaliação preliminar da coleção de algodoeiro arbóreo no Nordeste do Brasil*. EMBRAPA-CNPA, Campina Grande, pp. 58. (EMBRAPA-CNPA, Documentos, 38).

Mahalanobis, P.C. (1936). On the generalized distance in statistics. *Proc. Mat. Inst. Sci. 12*: 49-55.

Morrison, D.F. (1976). *Multivariate Statistics Methods*. 2nd edn., McGrow-Hill-Kogakusha, Tokyo, pp. 415.

SAS Institute (1982). *SAS User's Guide Statistics*. North Carolina, Cary, pp. 584.

Sneath, P.H.A. and **Sokal, L.R.** (1973). *Principles of Numerical Taxonomy*. W.H. Freeman, San Francisco, pp. 573.

Stephens, S.C. (1967). Evolution under domestication of New World Cottons (*Gossypium* ssp). *Ciência e Cultura 19*: 118-134.

(Received June 6, 1990)