MULTIVARIATE ANALYSIS OF MAIZE F_2 POPULATIONS TO MEASURE RACIAL DIFFERENTIATION

Orlando Jose Martinez

Inst. of Statistics Mimeo Series # 133

TABLE OF CONTENTS

LIST OF TABLES

Page

INTRODUCTION

Characterization of the myriad forms of maize in the Western Hemisphere began in Mexico about 1945. Since then, nearly 12,000 collections of indigeneous varieties and their grouping into over 280 races of maize have been described. Despite these vast genetic resources, many of the races had disappeared because of replacement by improved hybrids and varieties of maize, modern agricultural technology, and market demands for uniformity. Thus, plant breeders and researchers involved in hybrid or varietal development have the responsibility not only to improve maize, but also to colect, classify, and preserve for future use those genetic resources which have been accumulated through thousands of years of domestication.

Classification of maize based on scientific evidence is more than an academic exercise. It is important for the study of the evolution of maize, and it provides basic knowledge with respect to localization of the best genetic sources for future breeding programs. It contributes to the interchange of the germ-plasm throughout the world. Additionally, a classification would be of still greater value if information pertinent to the genetic divergence among any two races were available. Mendel's law indicates that if two parents differ from each other by only one gene, the relative frequency of the parental types in the F_2 generation is one half. If two genes are involved, that frequency is two in sixteen, and so on for more genes. This type of inheritance suggests that the F_2 generation of a racial cross provides a basis to evaluate the degree of racial divergence among parental types.

The objective of this thesis was to investigate the usefulness of various measures of statistical distance between races of maize, relative to their F_2 generation, and thereby provide an estimate of genetic divergence between the pairs of races involved in the cross.

REVIEW OF LITERATURE

Description of the Races of Maize

One of the earlier studies to classify maize was done in 1899 by E. L. Sturtevant, who published a monograph, "Varieties of Corn". He catalogued the variability of maize into six main groups, five of which were based on the composition of the endosperm of the kernel. That classification was used almost without modification for about fifty years. An explanation for this is that the classification continued to be useful. Sturtevant's five kernel characteristics, dent, flint, flour, sweet and pop, are recognized in commerce but have little botanical significance (Mangelsdorf, 1974).

Anderson and Cutler (1942) pointed out that Sturtevant's classification is largely artificial, since it is based almost entirely on characteristics of the endosperm. They stressed that a natural classification should be based upon the entire genotype, rather than upon traits controlled by a single locus such as Sturtevant's endosperm characteristics. With that idea, Anderson (1946, 1947) did preliminary surveys of the maize races of Mexico and Guatemala.

These studies influenced the work of Wellhausen et al. (1952), who presented a systematic and comprehensive study on the maize races of Mexico. During a period of seven years, 2000 samples were collected throughout Mexico. This collection was intensively studied with respect to geographical distribution, vegetative characters of the plant; characters of the tassel; characters (external and internal) of the ear, and physiological, genetic, and cytological characters.

From these studies it was possible to group the varieties into races and describe the racial characteristics. Wellhausen and his collaborators described 25 distinct races of maize in Mexico.

The success of the Mexican studies stimulated the concept that all Western Hemisphere maize should be subjected to similar analyses. Several programs for the collection and classification of maize started in various Latin American countries. As a result, a series of publications was issued by the National Academy of Sciences - National Research Council (of the United States) as follows: Central America (Wellhausen et al., 1957), Cuba (Hatheway, 1957), Colombia (Roberts et al., 1957), Brazil and other eastern countries of South America (Brieger et al., 1958), West Indies (Brown, 1960), Bolivia (Ramírez et al., 1960), Peru (Grobman et al., 1961), Chile (Timothy et al., 1961), Ecuador (Timothy et al., 1963) and Venezuela (Grant et al., 1963)

Later, Parker and Pastorini (1965) collected 388 varieties of maize of Chile and classified them into eight groups. A general description of each group was done, its distribution by zones and varieties was given, and the respective plant and ear characteristics were presented. Hernandez and Alanis (1972) described five more races from maize samples from the Sierra Madre Occidental Mountains of Mexico, where no sampling has been done before. Rodríguez et al. (1968) defined the term "Racial Complex" as a group of races which has common discriminant characters of morphological, biological and/or geographical nature. Based upon this concept, races of maize of Bolivia were grouped into seven different racial complexes. Paterniani and Goodman (1977) restudied much of the maize from Brazil and adjacent areas and grouped those accessions into 91 populations belonging to 19 races and 15 subraces. In this study collections (not mentioned in Brieger et al., 1968) from the Huamahuaca Valley of Argentina were listed. Goodman (1978) presented a brief survey of the races of maize with respect to racial complexes of current worldwide economic importance. He also summarized the current status of racial studies with maize, and suggested the use of isozyme frequencies for grouping races of maize. He introduced the concept of morphological studies of racial F_2 populations as a complementary approach of studying interrelationships. The most comprehensive summary of the races of maize is that of Brown and Goodman (1977), whose descriptions encompass most of the maize races found in the Western Hemisphere.

In the past thirty-five years, from some 12,000 collections of indigenous varieties in the Western Hemisphere, over 280 races of maize have been described. It was expected that these racial descriptions would serve as a reasonable starting point for further studies on maize evolution; however, studies of this nature have been limited. Moreover, there have been difficulties in maintenance and preservation of these germplasms (Timothy and Goodman, 1979).

Recombination Frequencies in F_2 Generations

In studying evolution among races of corn, the question of how to measure the degree of divergence between populations arises. In addressing this problem, various statistical techniques have been applied to several different types of populations (parental populations, the F_2 population of a cross, a synthesized population). Anderson (1939) studied the total recombination in F_2 and later generations from interspecific crosses of tobacco. He stressed the importance of linkage as a mechanism to cause restriction on the total possible recombination among characters in individuals of F_2 and advanced generations. Anderson (1949), Dempster (1949) and Smith (1950) carried out both theoretical and experimental studies to evaluate recombinant individuals in natural and simulated populations. These studies led to additional investigations in three broad categories.

(1) Studies on the origin and evolution of Corn Belt maize by Anderson and Brown (1947, 1948, 1950, 1952), using archeological and historical evidence, showed that the common dent corns of the United States Corn Belt originated principally from a mix of the Northern Flints and Southern Dents. These common dent corns were developed by American farmers and plant breeders during the nineteenth century as a result of large-scale, systematic crossing and recrossing of the two races (the Northern Flints and the Southern Dents). Anderson and Brown (1952) demonstrated similar results genetically and cytologically.

(2) The use and application of multivariate statistical analysis in classification. Anderson (1949) proposed the use of metroglyphs (pictorialized scatter diagrams) as a technique for displaying associations in which several characters can be considered simultaneously to give a perception of the variation. For each individual, a graphical representation called a glyph may be obtained in which each individual is pictured as a circle in a two-way scatter diagram. The position of each ray emanating from that circle corresponds to each additional character. The length of the ray corresponded to the magnitude of that character for that individual. Anderson used these pictorialized diagrams in his preliminary studies of races of maize in Mexico and Guatemala to analyze the variation of the associated characters among the races collected.

(3) The use and applicaiton of multivariate techniques in introgression. Anderson's studies (1949, p. 33) stated, "if we think of all characters of one species being represented at one of the apices of a multidimensional cube and all the characters of the other species at the opposite apex, then the recombinations that we get in the F_2 form a narrow spindle through the center of the cube." Goodman (1966) developed multivariate measures of the spindle width (recombination value of an F_2 individual) and analyzed the structure of introgressive populations of cotton. Namkoong (1966) used a discrimination function to measure introgression in two pines species. Smouse (1972) used canonical analysis for describing multiple species hybridization.

Finally, this study is based upon the length of the linkage spindle formulated by Anderson and uses several generalizations of Dempster's (1949) parental-combination variance (i.e., the variance along the major axis of Anderson's linkage spindle) to provide estimates of racial divergence of races of maize.

Chromosome Knobs and Their Use in Classification of Maize

The usefulness of knobs as a racial characteristic and as a valuable criterion for judging relationships was recognized by Longley (1938). Wellhausen et al. (1952), in the description of races of maize of Mexico, considered chromosome knob number as a racial characteristic.

Similar types of studies were done later by Roberts et al. (1957) for races in Colombia, by Ramírez et al. (1960) for races in Central America, by Brown (1960) for races in the West Indies, by

Grobman *et al*. (1961) for races in Peru and by Timothy *et al*. (1961, 1963) for races in Chile and Ecuador.

The measure of relationships between races of maize has been clarified by detailed data analysis on chromosome knob constitution, e.g., knob size, knob position, frequencies of different knob types, and geographical distribution of specific kinds of knobs (McClintock, 1959). Studies of this type were done by McClintock (1960) for races of maize of Guatemala and Mexico, and also for races of maize of Bolivia, Chile and Ecuador, for which results were given by Ramírez *et al*. (1960) and Timothy *et al*. (1961, 1963). Further studies on chromosome knobs in Latin American races of maize have been done by Kato and Longley (1965). Later, Kato and Blumenschein (1967) attempted to determine the different distinct knob complexes, their centers of origin, and the major migraiton paths of these knob complexes. They described eight complexes: Mesa Central Complex, Tuxpeno Complex, Zapalote Chico Complex, Small Knob Complex, Southern Guatemala Complex, Venezuela Complex, Andean Complex and Northwest Caribbean Secondary Complex.

Recently, McClintock (1977) presented a brief summary of extensive data accumulated over a period of years. That data provides the best information yet published with respect to the significance of chromosome constitution in tracing the origin and migration of races of maize in the Americas, and on evolution of new races resulting from introgression between maize originating in different centers.

Statistical Techniques in Numerical Taxonomy

The term numerical taxonomy is defined by Sneath and Sokal (1973, p. 167) as "the grouping by numerical methods of taxonomic units into taxonomic groups on the basis of their character states." In the present study this grouping of taxonomic units (races of maize) into categories implies a classification which demands the use of multivariate techniques. To obtain groupings by numerical techniques, Sokal and Sneath (1963) have summarized measures of association into three categories: coefficients of association, coefficients of correlation, and measures of distance. Among these categories, the measures of distance have the property of being intuitively appealing; moreover, they have been used frequently in biology with reasonable success. Some of these measures of distance are defined and described below.

Euclidean distance. This measure was called the taxonomic distance by Sokal and Sneath (1963); it measures the distance between two points in p-dimensional Euclidean space. If μ_{ik} represents the mean for the character k of the population i, the squared Euclidean distance between population i and population j is defined as

$$D^2_{ij} = \sum_{k=1}^{p} (\mu_{ik}-\mu_{jk})^2 \quad .$$

Notice that this measure does not consider the correlation among the characters studied, thus it would be most useful in those studies in which the characters involved are essentially independent. Usually a standardization is made, say, $\mu^*_{ik} = \mu_{ik}/\sigma_k$, where σ_k is the standard deviation for character k.

The Euclidean distance is usually estimated by

$$\hat{D}_{ij}^2 = \sum_{k=1}^{p} (\bar{z}_{ik}-\bar{z}_{jk})^2 ,$$

where

$\bar{z}_{ik} = \bar{x}_{ik}/s_k$,

$\bar{x}_{ik}$ is the estimated mean for character k of the population i

and s_k is the estimated standard deviation for character k among populations.

Generalized distance. This measure of distance was developed by Mahalanobis (1936). Several authors in the recent literature (Sokal and Sneath, 1963, Sneath, 1976, Blackith and Reyment, 1971) have employed this method to estimate relative distances between populations. The squared Mahalanobis' generalized distance between taxi i and j is defined by

$$D_{ij}^2 = (\underline{\mu}_i-\underline{\mu}_j)'\Sigma^{-1}(\underline{\mu}_i-\underline{\mu}_j)$$

where $\underline{\mu}_i$ is the vector of character means for taxon i and Σ is the common within population covariance matrix (Mahalanobis, 1936). The squared Mahalanobis' distance between taxa i and j can be expressed equally well (Appendix A.1) by

$$D_{ij}^2 = (\underline{\mu}_i^*-\underline{\mu}_j^*)'\mathbb{R}^{-1}(\underline{\mu}_i^*-\underline{\mu}_j^*)$$

where $\underline{\mu}_i^*$ is the standardized vector of means for taxa i and $\mathbb{R}$ is the common within population correlation matrix.

Then, the squared Mahalanobis distance is usually estimated by

$$\hat{D}_{ij}^2 = (\bar{\underline{z}}_i - \bar{\underline{z}}_j)'R^{-1}(\bar{\underline{z}}_i - \bar{\underline{z}}_j) ,$$

where $\bar{\underline{z}}_i$ is the estimated standardized vector of means for taxa i and R the sample pooled within population correlation matrix.

Principal components. The method of principal component analysis consists in transforming the set of variables $x_1, x_2, \ldots, x_p$ to a new set $y_1, y_2, \ldots y_p$, satisfying the following conditions:

1. Each y_i is a linear combination of the x's, say

$$y_i = a_1x_1 + a_2x_2 + \ldots + a_px_p$$

2. If y_j is another linear combination of the x's, say

$$y_j = b_1x_1 + b_2x_2 + \ldots + b_px_p$$

then we want $\underline{a}'\underline{a} = 1$, $\underline{b}'\underline{b} = 1$, $\underline{a}'\underline{b} = 0$. This last condition leads to the y's being uncorrelated.

3. Of all possible combinations of this type, y_1 has the greatest variance, y_2, the second highest variance and so on (Morrison, 1967).

A geometrical interpretation is that principal components result in rotation of the coordinate axes to a new coordinate system with certain statistical properties such as the new set of variables are uncorrelated, and their variances are the nonzero characteristic roots of the covariance matrix of the original set of variables.

The technique of principal components has been introduced here for several reasons. It can be shown (Appendix A.2) that the squared Mahalanobis' distance between taxi i and j is also expressed by

$$D_{ij}^2 = (\underline{\xi}_i - \underline{\xi}_j)' \Lambda^{-1} (\underline{\xi}_i - \underline{\xi}_j) \ ,$$

where $\underline{\xi}_i$ is the vector of principal component means for taxon i; i.e., $\underline{\xi}_i$ is the mean vector for taxon i based upon the new set of variables obtained by the principal components technique; and Λ is a diagonal matrix whose elements are the characteristic roots of the matrix Σ. Thus the relation between principal components analysis and the Mahalanobis' generalized distance analysis is direct.

Now let $\underline{\xi}_i = (\xi_{i1}, \xi_{i2}, \ldots, \xi_{ip})$, then the squared Mahalanobis' generalized distance can also be expressed by

$$D_{ij}^2 = \sum_{k=1}^{p} (\xi_{ik} - \xi_{jk})^2 / \lambda_k \ ,$$

where λ_k are the characteristic roots of Σ .

Then, the Mahalanobis' squared generalized distance can be estimated by

$$\hat{D}_{ij}^2 = (\bar{\underline{y}}_i - \bar{\underline{y}}_j)' D^{-1} (\bar{\underline{y}}_i - \bar{\underline{y}}_j)$$

$$= \sum_{k=1}^{p} (y_{ik} - y_{jk})^2 / d_k$$

where $\bar{\underline{y}}_i = (y_{i1}, y_{i2}, \ldots, y_{ip})$ is the estimated vector of principal component means for taxon i, and D is a diagonal matrix whose elements, d_k, are the characteristic roots of the sample covariance matrix S.

From this expression it is clear that D^2_{ij} can be inflated if any of the characteristic roots, the d_k, approximate zero. This raises the question: how many principal components should be used? Several alternatives have been proposed, some of them are discussed by Goodman (1972).

If we assume that $\underline{x}$ has a multivariate normal distribution with mean zero and covariance matrix Σ, then the principal components have a geometrical interpretation. In fact, the expression $\underline{x}'\Sigma^{-1}\underline{x}$ specifies an ellipsoid in p-dimensional Euclidean space (Morrison, 1967). Thus it seems reasonable that the geometrical property of the principal components fits Anderson's linkage spindle. We can consider the F_2 generation of a particular racial cross and use the principal components technique on this generation to obtain distances between the parents relative to the F_2 population. Plots of the first few principal components often have been successfully used to group and describe biological populations (Jolicoeur and Mosiman, 1960; Bird and Goodman, 1977; Hussaini *et al.*, 1977; Cervantes *et al.*, 1978).

Cluster Analysis. Cluster analysis is a technique used to find groupings, say, of n units (objects, races of maize, experimental units) such that the units within groups are more similar (in correlation, distance, or some other metric sense) than the units in different groups or clusters. Various techniques have been developed for enumerating the clusters (Sokal and Sneath, 1963). If the groups or clusters are obtained as a nested grouping, then a graphical representation of the grouping can be done, and this representation is called a dendogram.

Canonical Analysis. This technique is similar to principal components analysis. It consists of transforming the set of variables $x_1, x_2, \ldots, x_p$ to a new set (canonical variates) $y_1, y_2, \ldots, y_p$; in this case the linear combination is obtained by maximizing the ratio of the variance among populations to the variance within populations. The first axis corresponds to the direction of greatest variability among the means of the p groups relative to the within variation. The second is orthogonal to the first and follows the direction of the next greatest variation, and so on (Blackith and Reyment, 1971). As in principal components, plots of the first few canonical variables have been used in classification of populations (Blackith and Reyment, 1971; Smouse, 1972; Hussaini *et al.*, 1977).

Multivariate Analysis of Races of Maize

In the description and classification of the Latin American races of maize, summarized in a preceding section, many different plant characters were measured by several different groups of scientists. Common to all of these studies was the fact that the groupings were not made on the basis of all characters simultaneously. However, the usefulness and significance of multivariate analysis techniques to measure variation between biological populations had been established by several workers (Fisher, 1936; Rao, 1960; Sneath and Sokal, 1963; Blackith and Reyment, 1971; Sneath, 1976, etc.).

The first study in maize considering overall morphological similarities was given by Edwards (1966) for races of maize from Southern and Southeastern Europe. Data from 80 varieties and 34 morphological characters were analyzed by four different clustering methods. The degree of consistency between the grouping formed by subjective visual methods and by cluster analysis indicated that numerical methods are suitable for racial classification of maize.

Goodman (1967) used Mahalanobis' generalized distance as a measure of grouping 15 races from the subtropical region of Southeastern South America. The results suggested general agreement with those obtained using conventional taxonomic procedures. Mochizuki and Okuno (1967) proposed principal components analysis to classify 57 flint lines from Japan. Squared distances between lines were calculated from the first four principal components (out of ten). Based upon these distances, the lines were classified into 14 varieties and the varieties were classified into four varietal groups. In general, the classification agreed with one based on the conventional methods. In another study Goodman (1968) used multivariate analysis of variance to measure morphological similarity among 15 races. The multivariate approach used was to compare the mean vectors for each pair of races using the residual product matrix from a randomized complete block design. The effects of transformations and within plot sampling were investigated. It was shown that these had very little effect on the relative distances when the commonly employed Mahalanobis' distance technique was used.

Salhuana (1969) used Mahalanobis' distance to evaluate the morphological differences among 15 collections of maize from Peru and 101 F_1's generated from these collections. Tentatively, he concluded that Mahalanobis' distance was useful for measuring the overall divergence between populations, but that it does not provide an accurate prediction of heterosis in crosses. Bird (1970) used his own collection of 1000 ears, in addition to data presented by Grobman et al. (1961), to study Peruvian maize. He aimed to describe cluster types, define character axes which reduced the variation to more comprehensive patterns, and relate the types of variation to the environment. Using a factor analysis technique, eleven factors were described. Eighteen clusters of types with distinctive forms were described. Goodman (1972) used a modification of Mahalanobis' distance to study 25 races of maize from Mexico. A dendogram of the Mexican races was in general agreement with relationships given by Wellhausen et al. (1952).

Bird and Goodman (1977) and Goodman and Bird (1977) studied ear characters for 219 Latim American races and subraces of maize. They used principal components and cluster analyses to group the races. Fourteen groups of races were delineated, and the characteristics of the groups and their interrelationship were described. Goodman and Bird (1977) pointed out that only two of the 14 groups of races have been widely used in breeding programs and that the races which are currently considered good sources of germ plasm are concentrated in the Caribbean dents and flints. Cervantes et al. (1978) used the average parental effects (general combining abilities), the interaction effect of the parents (specific combining abilities) and the genotype by environment interactions for the classification of the races of maize in Mexico. Their classification based upon a dendogram calculated from the average parental effects was in close agreement with the conclusions reached by Wellhausen et al. (1952) and Goodman (1972), with the exception of one race, Maiz Dulce, which they grouped with the Mexican Conicos (Wellhausen et al., 1952, had suggested that Maiz Dulce was of South American origin). They also concluded that the classifications based upon the average parental effects and upon genotype by environmental interactions were better than the classification based upon specific effects.

Multivariate Classification of Other Crops

Studies to measure similarities among racial groups using multivariate methods have been done in many different crops. Morishima and Oka (1960) studied the pattern of interspecific variation of 16 species of the genus *Oryza*. Using factor analysis, the Roschevicz's Section Sativa could be divided into two groups, *Oryza sativa* with four species and *Oryza officinalis* with six other species. The inner structure of these groups was also shown in cluster diagrams. Murty and Pavate (1962) did studies of improved varieties of *Nicotiana tabacum* L.; using multivariate analysis, thirteen varieties were examined with respect to genetic diversity and prediction of genetic advance. On the basis of Mahalanobis' distances, the 13 varieties were classified into four clusters. A discriminant function was obtained to select the best variety, taking into consideration genotypic and phenotypic values of fourteen characters.

Forty self-compatible forms of Brassica campestris were studied over a period of two years by Murty and Qadry (1966). On the basis of Mahalanobis' distances, the forty populations could be grouped into nine clusters. They found that genetic diversity between populations may not be associated with geographic distribution.

Somayajulu et al. (1970) employed Mahalanobis' distances to measure the genetic diversity among 67 wheat strains. The estimates were obtained for three different environments representing different levels of fertility. When the 67 populations were grouped into clusters for each environment, the number of clusters varied from environment to environment, as did the populations within clusters. However, while instability of grouping was evident when the strains involved exhibit a low magnitude of genetic divergence, it did not occur for clusters representing the highest order of divergence.

Jeswany et al. (1970) used Mahalanobis' distance to measure the nature of genetic divergence in 100 types of flax (Linum usitatissimun) from seven geographic regions in the world, while Ram and Panwar (1970) employed Mahalanobis' distance and canonical analysis to measure genetic divergence among races of rice from China, Japan, Taiwan and India. Ram and Panwar (1970) found that genetic diversity seemed to be associated with geographic diversity; however, when Menndirata et al. (1971) employed Mahalanobis' distance among 30 indigeneous and exotic varieties of sorghum, they found that, in general, geographic diversity was not related to genetic diversity. Govil and Murty (1973) used 24 varieties from a world collection of sorghum representing different geographical areas and taxonomic groups; they also found that the clustering pattern of varieties did not follow their geographical distribution.

Chaundhary and Singh (1975) used Mahalanobis' distance to measure genetic divergence among four Indian and four exotic barley varieties and their 56 F_1 generations (including reciprocals). The 64 populations were grouped into ten clusters. They found that the clustering pattern of the hybrids was somewhat influenced by their parentage, while an exotic cultivar was uniquely divergent from the remaining varieties. Hussaini et al. (1977) used principal components analysis and canonical variates analysis to study the nature of variation and to classify a world collection of finger millet. Twelve groups were identified by plotting the first two principal components; similar results were obtained from canonical variates analysis. Baum (1977) presented a systematic and comprehensive study on wild and cultivated oats, representing 27 well-recognized species from 26 countries of the world. A set of 5300 varieties was studied, and 28 characters were considered, including some cytological information such as number of chromosomes and type of polyploidy. Possible classifications using a number of clustering procedures based on subsets of characters were provided.

Standardized Euclidean distance, Mahalanobis' distance, principal components, and canonical variates are techniques which employ standardization of the characters used. To estimate these distances generally the within group variances or within group covariance matrix has been used. However, in the case of self pollinated species, the within covariance matrix will largely represent environmental variance. Hence, distances based upon such matrices (Murty and Pavate, 1962; Somayajulu et al., 1970; Ram and Panwar, 1970; Chaundhary and Singh, 1975) may not provide the most realistic picture with regard to genetic divergence (Goodman, 1969).

MATERIALS AND METHODS

Experimental Material

The data utilized in this study were obtained from 47 samples of races of corn from Latin America and from 30 F_2 populations obtained from crosses among certain pairs of these races. Parents and F_2's were grown during a period of six years, from 1971 to 1976, at two locations: the Genetics Nursery of North Carolina State University at Raleigh, North Carolina, and at the winter nursery facilities of the Agricultural Alumni Seed Improvement Association (Purdue University) near Homestead, Florida.

Table 1 lists the 47 races, collection numbers, and numbers of plants used. Table 2 lists the pedigree for each F_2 population and the number of plants used. The grouping of the populations is explained in the discussion.

The material grown at Homestead was planted in a completely random design with plot sizes ranging from 5 to 30 plants, with separate designs for separate families. The distance between rows was 36 inches and the plants within rows were planted 12 inches apart. The number of plots, plot size and planting dates are given in Appendix B, Table 1.

The F_2 material grown at Raleigh was planted in a completely randomized design with single plant plots, the distance between rows was 48 inches and within rows 12 inches. The parental materials were grown with the same spacing, but with three replicate plots of thirty plants each. Planting dates and numbers of plots are provided in Appendix B, Table 1.

Table 1. Races of corn, collection numbers, and numbers of plants utilized in the study.

Number	Race name	Collection	Number of plants
1	Amagaceño	Composite	25
2	Andaquí	Composite	47
3	Aragüito	Ven 628	47
4	Cacahuacintle	Mex 7	39
5	Canilla	Ven 874	43
6	Camba	Composite	34
7	Capio	Ant 318	18
8	Capio	Composite	16
9	Cariaco	Cor 338	24
10	Cariaco Col.	Composite	46
11	Cateto Assis Brasil	RGS XIV	51
12	Cateto Grande	MT I	60
13	Chapalote	Sin 2	52
14	Chaparreño	Composite	39
15	Chococeño	Cho 314	31
16	Chulpi	Ecu 434	17
17	Chuncho	Hco 63	11
18	Clavo	Cho 311	17
19	Confite Morocho	Composite	31
20	Coroico Amarillo	Composite	24
21	Dente Branco Rio Grandense	RGS X	60
22	Enano	Bov 1143	28
23	Guaribero	Composite	39
24	Güirua	Composite	41
25	Harinoso de Ocho Occidentales	Nay 29	32
26	Kcello	Bov 948	29
27	Kcello	Ecu 704	50
28	Lenha	RGS XX	32
29	Mochero	Lbq. 5	22
30	Morado	Bov 567	35
31	Moroti Precoce	Bol I	51
32	Nal-Tel	Yuc 7	15
33	Negrito Col.	Composite	22
34	Olotón	Gua 639	17
35	Olotón	Gua 686	29
36	Pira Col.	Composite	42
37	Piricinco	SM9	57
38	Pollo	Cun 401	38
39	Pororo	Composite	31
40	Reventador	Nay 15	34

Table 1. Continued.

Number	Race name	Collection	Number of plants
41	Reventador	Nay 39	25
42	Salpor	Gua 476	46
43	Tabloncillo Perla	Nay 16	27
44	Tepecintle	Chs 76	43
45	Vandeño	Chs 31	38
46	Zapalote Chico	Oax 48	34
47	Zapalote Chico	Oax 50	45
Total			1622

Table 2. Racial crosses (F_2 generation) of corn utilzed in this study and numbers of plants per cross.

Number	Racial Cross	Number of plants
	Group 1	
1	Harinoso de Ocho O. x Tabloncillo Perla	41
2	Negrito x Moroti Precoce	36
3	Chococeño x Aragüito	131
4	Confite Morocho x Chapalote	72
5	Cateto Assis Brasil x Cateto Grande	57
6	Cariaco x Chulpi	79
7	Capio x Salpor	115
8	Salpor x Cacahuacintle	67
9	Zapalote Chico Oax 48 x Zapalote Chico Oax 50	53
	Group 2	
10	Mochero x Chuncho	69
11	Chapalote x Reventador (Nay 15)	50
12	Piricinco x Morado	120
13	Amagaceño x Olotón Gua 639 (Cross 1)	44
14	Amagaceño x Olotón Gua 639 (Cross 2)	39
15	Chaparreño x Camba	73
16	Cariaco x Güirua	76
17	Tepecintle x Vandeño	53
	Group 3	
18	Nal-Tel x Reventador (Nay 15)	50
19	Pira x Pororo	46
20	Andaqui x Guaribero	67
21	Amagaceño x Olotón (Gua 686)	67
22	Chapalote x Piricinco	75
23	Cacahuacintle x Capio	62
24	Pollo x Kcello (Ecu 704)	71
25	Nal-Tel x Reventador (Nay 39)	33
	Group 4	
26	Kcello Ecu 704 x Kcello Bov 948	55
27	Clavo x Canilla	99
28	Confite Morocho x Enano	65
29	Tabloncillo Perla x Lenha	53
30	Dente Branco x Chapalote	48
Total		1946

The F_2 generation was obtained by sibling the F_1; crosses were made without regard to direction of crossing.

The characters considered in this study are:

1. Ear length, in centimeters;
2. Ear diameter, measured as the maximum diameter at the midpoint of ear length.
3. Number of rows of kernels, near the midpoint of the ear;
4. Kernel length, the average length, in millimeters, of 10 kernels near the middle of the ear.
5. Kernel thickness, the average thickness, in millimeters, of 10 consecutive kernels in a row, near the midpoint of an ear.

The choice of these characters was based on the criteria given by Goodman and Paterniani (1969). With the exception of ear length, all characters have values of the ratio.

$$r = [\hat{\sigma}_r^2/(\hat{\sigma}_y^2 + \hat{\sigma}_e^2)] \geq 3.0$$

where $\hat{\sigma}_r^2$ and $\hat{\sigma}_y^2$ are estimated components of variance due to differences among races and differences among environments, respectively, and $\hat{\sigma}_e^2$ is the mean square error due to races by environments.

Statistical Methods

Since the purpose of this study was to investigate the utility of various methods to estimate divergence in maize, six procedures were utilized to measure that divergence. For the construction of some of those measures of divergence, standardizations and standard errors with respect to the F_2 populations were used. The more customary procedure is to use parental standardizations and standard errors. Some justifications for the methods used here are given next, but more details are in the discussion chapter.

Euclidean Distance.

A squared Euclidean distance between race i and race j was calculated from

$$D1 = \sum_{k=1}^{5} (\bar{x}_{ik} - \bar{x}_{jk})^2 / s_k^2 \ ,$$

where

$\bar{x}_{ik}$ = mean of the character k of race i ,

s_k = standard deviation for character k of the F_2 .

Generalized Mahalanobis' distance.

The squared Mahalanobis' distance between race i and race j was estimated from

$$D2 = \sum_{k=1}^{5} (y^*_{ik} - y^*_{jk})^2 / d^*_k \ ,$$

where

$y^*_{ik} = \underline{b}'_k \underline{z}^*_i$,

$\underline{b}'_k$ = the k^{th} characteristic vector of R^*, the pooled correlation matrix of race i and j,

$\underline{z}^*_i$ = the standardized (with respect to the pooled parental variances) vector of means for race i,

d^*_k = the k^{th} characteristic root of R^* .

Generalized distance.

A squared generalized distance between race i and race j was estimated from

$$D3 = \sum_{k=1}^{5} (y_{ik} - y_{jk})^2/d_k ,$$

where

$y_{ik} = \underline{a}'_k \underline{z}_i$

a'_k = the k^{th} characteristic vector of R, the correlation matrix of the F_2 ,

$\underline{z}_i$ = the standardized (with respect to the F_2) vector of means for race i,

d_k = the k^{th} characteristic root of R.

Modified generalized distance.

Considering only those characteristic roots with the property $d_k \geq 1$, a modified squared generalized distance was calculated from

$$D4 = \sum_{k=1}^{q} (y_{ik} - y_{jk})^2/d_k \; ; \; d_k \geq 1 \; , \; 1 \leq q \leq 5 \; .$$

Approximate Dempster's Distance.

The geometrical property of the multivariate normal distribution, as mentioned before, fits Anderson's (1939) linkage spindle. Consequently, one generalization of Dempster's parental-combination variance (measure of tendency of F_2 individuals to be similar to one or the other parent with respect to p characters simultaneously) would be the maximum variance when p-orthogonal characters are considered simultaneously in an F_2 population. An estimate of that variance is the first characteristic root of the correlation matrix of the F_2. Then a squared distance measure between the parents relative to that estimate is

$$D5 = (y_{i1} - y_{j1})^2/d_1$$

where

$y_{i1} = \underline{a}_1' \underline{z}_i$,

$\underline{a}_1'$ = the 1st characteristic vector of R, the correlation matrix of the F_2 ,

$\underline{z}_i$ = the standardized (with respect to the F_2) vector of means for race i ,

d_1 = the 1st characteristic root of R.

Dempster's Distance.

The first characteristic root obtained from the correlation matrix of an F_2 population will be an estimate of Dempster's parental-combination variance whenever the direction of the maximum variance coincides with the direction of the straight line which can be drawn between the centroids of the parents. For those populations which do not have that desirable property (and most will not, although the differences may be minor), another estimator will be needed. Such an estimator can be calculated as follows: consider a straight line drawn between the parental centroids in the p-dimensional space and consider the F_2 population as a set of n points in that space, which may or may not fall close to this line. Then we want to orthogonally project each of those points onto the line. Once the points are projected onto the line, their variation would correspond to Dempster's parental-combination variance. The formal procedure is given in Appendix A.3. In that case Dempster's parental-combination variance becomes

$$s^2 = 1 + 2 \sum_{k<k'}^{p} m_k^* m_{k'}^* \, r_{kk'} \ ,$$

where the

$$m_k^* = \frac{T_{ik} - T_{jk}}{\left[\sum_{k=1}^{p} (T_{ik} - T_{jk})^2 \right]^{\frac{1}{2}}} \ , \quad k = 1,\ldots,p$$

are the direction cosines of the line between the parents i and j and

$r_{kk'}$ is the correlation coefficient between characters k and k' of the F_2 population

$$T_{ik} = \frac{\bar{x}_{ik}}{s_k} , \; k = 1,\ldots,p ,$$

where

$\bar{x}_{ik}$ = the mean of parent i for the k^{th} character, and

s_k = the standard deviation of the F_2 for the k^{th} character.

Finally, the squared distance between the race i and race j relative to Dempster's parental-combination variance is

$$D6 = \sum_{k=1}^{5} (\bar{z}_{ik} - \bar{z}_{jk})^2/s^2 ,$$

where

$\bar{z}_{ik}$ = standardized (with respect to the F_2 generation standard deviation) mean for character k of race i.

RESULTS

The estimated means and variances for the five characters of the 47 Latin American races of corn are presented in Appendix B, Table 2. Similarly, Appendix B, Tables 3 through 17 contain the means, variances and correlation coefficients for the 30 F_2 populations in the study.

Table 3 lists standardized values of the six squared distances considered, i.e., Euclidean distance (D1), Mahalanobis' generalized distance (D2), Generalized distance (D3), Modified generalized distance (D4), Approximate Dempster's distance (D5), and Dempster's distance (D6). For comparative purposes, the original values of the six measures were standardized such that the maximum value of each was 50 and the minimum value was 1; the original values of the six squared distances are provided in Appendix B, Table 18. Standard errors were obtained for the squared Euclidean distance (D1), squared Mahalanobis' distance (D2) and squared generalized distance (D3); upper bounds for the standard errors of the modified squared generalized distance (D4) and the squared approximate Dempster's distance (D5) were calculated. An approximate standard error of the squared Dempster's distance (D6) was also obtained. All these values are provided in Appendix B, Table 19. The derivations are given in Appendix A.4. Spearman correlation coefficients among the six distances were calculated (Table 4). All the correlation coefficients were positive, ranging from $r = 0.95$ between the Euclidean distances, D1, and Dempster's distances, D6, to $r = 0.15$ between generalized distances, D3, and the approximate Dempster's distances, D5.

Table 3. Standardized values of Euclidean distance (D1), Mahalanobis' distance (D2), generalized distance (D3), modified generalized distance (D4), approximate Dempster's distance (D5), and Dempster's distance (D6) for the racial crosses in this study.

No.	Racial Cross	Squared Distances*					
		D1	D2	D3	D4	D5	D6
	Group 1						
1	Harinoso de O. x Tabloncillo P.	1.0	1.0	1.0	1.8	1.9	1.0
2	Negrito x Moroti Precoce	1.0	3.5	2.6	1.7	1.7	1.6
3	Chococeño x Aragüito	2.9	12.0	8.4	1.2	1.1	2.9
4	Confite Morocho x Chapalote	4.0	19.9	11.4	1.0	1.0	4.6
5	Cateto Assis Br. x Cateto Grande	2.4	2.4	2.5	3.1	3.1	2.3
6	Cariaco x Chulpi	2.1	11.7	3.9	3.7	2.0	3.0
7	Capio x Salpor	2.5	4.3	4.4	3.5	1.1	4.4
8	Salpor x Cacahuacintle	3.7	11.5	9.0	3.7	1.7	3.6
9	Zapalote Chico Oax 48 x Zapalote Chico Oax 50	2.6	3.4	4.3	3.2	2.0	3.7
	Group 2						
10	Mochero x Chuncho	2.2	4.9	3.2	4.9	2.1	2.9
11	Chapalote x Reventador (Nay 15)	2.1	7.0	3.1	4.5	2.1	3.6
12	Piricinco x Morado	2.0	4.9	3.1	5.6	1.2	3.7
13	Amagaceño x Olotón Gua 639 (2)	2.9	5.6	2.3	4.2	3.9	3.1
14	Amagaceño x Olotón Gua 639 (1)	3.2	5.6	4.8	3.0	4.9	3.9
15	Chaparreno x Camba	3.1	4.4	4.4	5.0	2.9	4.4
16	Cariaco x Güirua	4.8	14.9	9.8	4.9	3.2	5.3
17	Tepecintle x Vandeño	3.9	18.9	8.9	7.2	2.7	4.7
	Group 3						
18	(Nal-Tel) x Reventador Nay 15	7.9	17.6	16.0	3.6	2.3	10.7
19	Pira x Pororo	7.4	28.5	15.6	3.2	1.8	11.2
20	Andaqui x Guaribero	5.8	13.7	16.6	7.6	1.2	9.9
21	Amagaceño x Olotón (Gua 686)	4.9	7.1	5.3	9.2	5.4	6.1
22	Chapalote x Piricinco	9.5	19.4	11.3	5.7	3.8	12.2
23	Cacahuacintle x Capio	7.2	35.8	14.4	10.2	1.1	12.8
24	Pollo x Kcello (Ecu 704)	7.0	11.5	11.2	13.8	4.2	9.0
25	Nal-Tel x Reventador (Nay 39)	8.1	16.5	9.3	10.0	6.1	10.8

Table 3 (continued).

No.	Racial Cross	Squared Distances* D1	D2	D3	D4	D5	D6
	Group 4						
26	Kcello Ecu 704 x Kcello Bov 948	7.5	12.9	8.0	12.3	7.3	10.2
27	Clavo x Canilla	7.4	12.5	9.7	15.5	6.2	16.7
28	Confite Morocho x Enano	15.0	50.0	22.4	27.0	7.0	16.5
29	Tabloncillo Perla x Lenha	29.9	40.7	43.4	24.9	21.3	38.2
30	Dente Branco x Chapalote	50.0	45.0	50.0	50.0	50.0	50.0

* Standardized to range from 1 to 50.

Table 4. Spearman correlation coefficients among Euclidean distance (D1); Mahalanobis' distance, (D2); generalized distance, (D3); modified generalized distance, (D4); approximate Dempster's distance, (D5); and Dempster's distance, (D6).

	Spearman correlation coefficients					
	D1	D2	D3	D4	D5	D6
D1	1.0	0.85**	0.85**	0.66**	0.53**	0.95**
D2		1.0	0.89**	0.47**	0.23	0.81**
D3			1.0	0.44**	0.15	0.84**
D4				1.0	0.68**	0.73**
D5					1.0	0.45**
D6						1.0

** Significant at the 1% level.

DISCUSSION

Statistical Viewpoint

This thesis is mainly concerned with the study of racial divergence among races of maize from Latin America. Six procedures or measures have been considered here for measuring such divergence. Among these Euclidean distance (D1) is a measure that does not consider the correlation among the taxonomic characters and, hence, is more adequate in studies where the characters are essentially independent. However, as is observed in Appendix B, Tables 3 to 17, there is at least one significant correlation among the characters for each of the F_2 populations. For this reason, the D1 distance was not used here with the purpose of making inference as to whether or not two populations are closely related; its use is as a comparison measure. A standardization was used for this measure in an effort to give equal weighting to each character.

The squared Mahalanobis' generalized distance, D2, was the second measure used. This measure as well as the third measure, D3, generalized distance as mentioned before, has the property of increasing as the number of characters increases or when one or more characteristic roots approach zero. Hence, this measure is also used here as a comparison measure. Mahalanobis' generalized distance was originally defined as the generalized distance between the means of two groups, based on the common covariance-matrix (see Appendix A.1). The problem of inequality of covariance matrices on classification and estimation of genetic divergence among some races of corn from Latin America has been considered by Goodman (1967) and Salhuana (1969).

Despite the fact that multivariate tests indicated that the matrices used in those studies were not homogeneous, essentially the same relative distances for the races were obtained when different covariance matrices were used.

In this study, rather than using the within-race covariance matrix (to construct D1. D3. D4. D5. and D6), the F_2 covariance was used. Some of the reasons are: (1) Mendelian inheritance suggests that in the F_2 generation of a racial cross, the probability of recovering the parental types is inversely related to their degree of genetic differentiation; (2) the variability of the F_2 generation of a racial cross provides a test of the degree of divergence of the two races involved: the more diversity among the parents, the more variability of the F_2 generation; (3) a set of points of an F_2 generation can be described by the linkage spindle proposed by Anderson (1944). If a multivariate distribution is assumed, this distribution fits the linkage spindle, and the length of the spindle is proportional to the Dempster's parental-combination variance. Hence, measures of distance between the parents relative to the F_2 population can be constructed and are appropriate for measuring racial differentiation.

For the purpose of making inferences as to whether or not two populations show close relationship, only three of the six procedures will be used: modified generalized distance, D4; approximate Dempster's distance, D5; and Dempster's distance, D6. To obtain the standard errors, it was assumed that the characters in this study have multivariate normal distributions. However, multivariate normality is not required for the computations of the distances themselves. It is observed in Appendix A.3 that normality is not required to obtain Dempster's parental combination-variance and hence D6. Furthermore, the measure has the advantage that it can be calculated from the means, variances, and correlations of the populations studied without matrix inversions or extraction of characteristic roots and vectors. However, in the case of Dempster's distance D5, the multinormality condition is needed for the maximum characteristic root (d_1) of the covariance-matrix of the F_2 generation to be an exact measure of the average length of Anderson's (1949) "linkage spindle". D5 also requires calculation of eigenvalues and eigenvectors, which is a more complex technique than the means and correlation coefficients needed for the estimation of D6. The two measures D5 and D6 are different, but closely related: if the direction of the maximum variance coincides with the direction of the straight line which is drawn between the centroids of the parents, then the D5 and D6 values are equal. D5 can result in very low values (regardless of the geometric distance between the parents) if, by chance, the F_2 linkage spindle (which in extreme cases may resemble a spheroid) is oriented in a direction essentially perpendicular to the line between the parental centroids.

The modified generalized distance, D4, was constructed using those principal components whose characteristic roots were greater than or equal to 1.0. While there is not an analytical procedure to select the components to be used in taxonomical analyses, usually those with characteristic roots above 1.0 seem most likely to be of biological significance, while those with low values often describe secondary or casual variation. Similar procedures (using those $d_k \geq 1$) have been used reasonably successfully in psychological applications (Rummel, 1970), and in races of maize by Goodman (1972) and Goodman and Bird (1977).

The reason for the use of the correlation matrix to estimate the modified generalized distance, D4, is that principal components calculated from means and covariance matrices are not independent of the choice of scale (if the variables are measured in different units, as in this case, then the result of a principal components analysis depends on the units of measurement (Morrison, 1967)). Thus, a standardization of all variables to variance 1 is preferable for most principal components analyses.

The use of the correlation matrix to estimate the modified generalized distance D4 was justified in the above paragraph; however, to estimate the generalized distance, D3, either the correlation matrix or the covariance matrix may be used (D3 is also invariant under linear transformations, the procedure is analogous to that given in Appendix 1, in that the F_2 covariance matrix is used instead of the common parental covariance matrix). For most purposes, Dempster's distance, D6, seems to be the most suitable measure of distance among two races relative to their F_2 generation, as it requires the fewest assumptions.

Some other properties of the D4, D5, and D6 measures can be illustrated by considering the following three racial crosses in Table 3: Dente Branco x Chapalote, and Tabloncillo Perla x Lenha in Group 4 and Zapalote Chico Oax 48 x Zapalote Chico Oax 50 in Group 1. There is not a close geographical or ancestral relationship between Dente Branco and Chapalote. Chapalote is an ancient, indigeneous race of Mexico with small, cigar-shaped ears and small flinty seeds. It is typical of Northwestern Mexico (Wellhausen, *et al*., 1952). Dente Branco has many rows, rather long ears, and white, dented kernels. It is typical of southern Brazil and was possibly introduced there from the United States about 100 years ago (Paterniani and Goodman, 1977). Similarly, there is not a close relationship between Tabloncillo Perla and Lenha. The first is a prehistoric race from Northwestern Mexico, with medium length, narrow ears that are slightly tapering at both ends (Wellhausen, *et al*., 1952). Lenha is a race of unknown origin from southernmost Brazil and has thick, cylindrical, short and many rowed ears (Brieger, *et al*., 1958). These racial crosses serve as checks in the following sense: any reasonable measure of racial divergence would have very large values for these two crosses. There is little or no relationship among the parents, and the measure must have the property of detecting racial divergence when such divergence actually exists. In Table 3 the measures satisfied this property for the two crosses described above. Indeed, Dente Branco x Chapalote had the largest value of 50 squared units (s.u.) for D1, D3, D4, D5, and D6, while the values for Tabloncillo Perla x Lenha range from 43.4 to 21.3 s.u. for D3 and D5, respectively.

The racial cross Zapalote Chico Oax 48 x Zapalote Chico Oax 50, was deliberately chosen as an example of two parents which are closely related. These are two different collections of the same race and were collected from the same geographical area, the state of Oaxaca in Southern Mexico. Their D4, D5, and D6 values were 3.19, 2.14, and 3.74, respectively, with an average of 2.99 s.u. These values suggest that the three measures D4, D5, and D6 also have the property of expressing little divergence when such divergence is known to be small.

The field designs used in this experiment (a mixture of a completely randomized design with single plant plots and various randomized block designs) probably were not the most adequate since the total number of entries was 77(30 F_2 populations and 47 collections of maize) and these were grown in two separate locations. Hence, the choice of more appropriate experimental designs (complete randomized block or incomplete block design) at a single location probably would somewhat improve the precision of the estimates of differentiation among maize races.

Biological Viewpoint

The cross Zapalote Chico Oax 48 x Zapalote Chico Oax 50, as indicated, represents two closely related collections; the average value of D4, D5, and D6 was 2.99 s.u. with a standard deviation of 0.87. From this, 3.0 s.u. can be considered as an arbitrary upper limit of close relationship among racial crosses. For comparative purposes the crosses have been arbitrarily grouped as follows: Group 1, the set of crosses for which the average value of D4, D5, and D6 is less than or equal to 3.0 s.u.; Group 2, the set of crosses for which the average of D4, D5, and D6 is larger than 3.0 s.u. but less than or equal to 5.0 s.u.; Group 3, the set of crosses for which the average of D4, D5, and D6 is larger than 5.0 s.u. but less than or equal to 10.00 s.u.; Group 4, the set of crosses for which the average of D4, D5, and D6 is larger than 10.00 s.u. This divides the 30 racial crosses into four groups of about the same size each.

Group 1, Closely Related Races. The racial crosses for this group are: Harinoso de Ocho x Tabloncillo Perla; Negrito x Moroti Precoce; Chococeño x Aragüito; Confite Morocho x Chapalote; Cateto Assis Brasil x Cateto Grande; Cariaco x Chulpi; Capio x Salpor; Capio x Cacahuacintle and Zapalote Chico Oax 48 x Zapalote Chico Oax 50. All these crosses have the average value of D4, D5, and D6 smaller than 3.0 s.u., suggesting close relationships among the races invovled in each cross.

Harinoso de Ocho x Tabloncillo Perla had D4, D5, and D6 values of 1.83, 1.89, and 1.0 s.u., respectively; on this basis it would appear reasonable to conclude that the two races are closely related. A similar conclusion has been reached by Wellhausen, et al. (1952), who have suggested Harinoso de Ocho to be an ancestor of Tabloncillo Perla, and by Goodman and Bird (1977), who classified the two races into a subgroup of the Caribbean Dents. Goodman (1972), Brown and Goodman (1977) and Cervantes, et al. (1978), using different techniques, have all indicated close relationships between these two races.

The second member of Group 1 was Moroti Precoce x Negrito with D4, D5, and D6 values of 1.65, 1.74, and 1.62 s.u. which suggest a close relationship among the two races. A possible relationship among them was reported by Roberts et al. (1957), who mentioned a personal communication from F. G. Brieger suggesting that a race from Brazil might be a common ancestor of Negrito and Cariaco from Colombia. However, in the booklet on the races of maize from Brazil by Brieger et al. (1958), no relationship among Negrito and Moroti Precoce is mentioned. Paterniani and Goodman (1977) indicated that Moroti Precoce is an ancient race and probably is the result of selection by

Guarani Indians hundreds of years ago. Goodman and Bird (1977) did not find any relationship among the two races and classified them in two different groups.

The racial cross Chococeño x Aragüito had D4, D5, and D6 values of 1.21, 1.12, and 2.90, indicative of close relationship. Chococeño from Colombia has short, thick, conical ears, and has been hypothesized to have originated by hybridization of maize and tripsacum (Roberts et al., 1957). Aragüito from Venezuela also has short, stubby, strongly conical ears, and was possibly introduced from Central America (Grant et al., 1963). The low values of the three distances are in agreement with the findings of Brown and Goodman (1977), where it was indicated that the Chococeños and Aragüitos in general have the same ear shape. However, with respect to many other characteristics, the two races are very different: Aragüito's plants are short, slender, and very early (55 days to flowering; Grant, et al., 1963). Chococeño's plants are very tall with many tillers, relatively late (100 days to flowering), and highly tripsacoid (Roberts et al., 1957). The contrasting values in days to flowering would make hybridization among the two races improbable under normal conditions. However, the relationships among the races in this study are based solely on ear morphology. A more critical analysis of this cross would require the study of additional characteristics. When the collections of maize were assembled, emphasis was placed upon both internal and external ear characteristics, which were thought to be polygenic rather than simply inherited characters; moreover, ears are relatively easy to collect, measure and preserve.

The values D4, D5, and D6 for the racial cross Confite Morocho x Chapalote were 1.0, 1.0, and 4.56 s.u., which suggest these two races are closely related. Chapalote is an ancient, indigeneous race from Mexico, and Grobman et al. (1961) hypothesized that Confite Morocho was the ancestral form of a lineage leading to Tabloncillo and Harinoso de Ocho of Mexico, Cuzco Gigante of Peru and Cabuya of Colombia. However, no relationship has been reported between Confite Morocho and Chapalote.

The D4, D5, and D6 values for the cross Cateto Assis Brasil x Cateto Grande were 3.12, 3.07, and 2.25 s.u. respectively, suggesting close relationship. This degree of association would be expected since both are subraces (from separate regions of Brazil) of the race Cateto.

The racial cross Cariaco x Chulpi had D4, D5, and D6 values of 3.68, 2.02, and 3.04 s.u., hence the two races are probably closely related. Cariaco is a floury corn from Colombia and Venezuela (this cross involved Colombian Cariaco) with short and very thick ears. Chulpi is an indigeneous sweet corn from Ecuador (and other Andean areas of South America), with short thick ears, some of which are conical and some are cylindrical. The results agree with thos postulated by Goodman and Bird (1977), who suggested that the two races belong to a broad-eared Cariaco group.

The values D4, D5, and D6 for the racial cross Capio x Salpor were 3.46, 1.09, and 4.41 s.u., respectively, implying a close relationship. Capio from Colombia is described by Roberts et al. (1957) as having very long, thick ears, with a strong taper from the base to the tip. Salpor is described by Wellhausen et al. (1957) as a race introduced into Guatemala. It has large, thick ears strongly resembling those of the Colombian Capio, and it may be that Salpor is an early form of Capio or a form of Capio introgressed by the Guatemalan race Serrano. The results found here seem to support the hypthesis that Capio is the Colombian counterpart of Salpor.

The racial cross Salpor x Cacahuacintle had D4, D5, and D6 values of 3.66, 1.71, and 3.58 s.u., which suggest a close relationship between the races. Wellhausen et al. (1952, 1957) have stated that these races are very similar in both their external and internal characteristics. This resemblance was first recognized by Anderson (1946) who stated that Cacahuacintle was probably introduced into Mexico from Guatemala. Although Goodman and Bird (1977) did not include the Guatemalan races in their study, they stated that Cacahuacintle was a very distinctive race of the Conico group and that Salpor shared the plant type of this group. They hypothesized a type of continuous variation from Cacahuacintle through Salpor and Capio, since Salpor and Cacahuacintle share the same type of plant while Capio does not. In this study, close relationship was found between Cacahuacintle and Salpor and between Salpor and Capio, but not between Capio and Cacahuacintle (the D4, D5, and D6 values for the later racial cross were 10.15, 1.07, and 12.82 s.u., respectively).

The last member of Group 1 was the cross Zapalote Chico Oax 48 x Zapalote Chico Oax 50; the use of this cross has already been discussed.

Group 2. Somewhat Related Races. The racial crosses within this group are: Mochero x Chuncho; Chapalote x Reventador; Piricinco x Morado; Amagaceño x Olotón (Gua 639, crosses 1 and 2); Chaparreño x Camba; Cariaco x Güirua, and Tepecintle x Vandeño. All of these have an average value of D4, D5, and D6 larger than 3.0 s.u. and less than or equal to 5.0 s.u.

The first member of this group was Mochero x Chuncho, whose D4, D5, and D6 values were 4.86, 2.05, and 2.93. Both races are from Peru; Mochero's plants are small with short, stubby, cylindrical ears having kernels irregularly arranged in a type of spiral. Chuncho is very tall and late in maturity with medium to long, conical to elongated ears (Grobman et al., 1961). Goodman and Bird (1977) indicated a certain degree of association among them, as they classified both races into a very broad Andean Floury Group. However, they felt that the assignment of Chuncho to the Andean Floury Group should be viewed with caution.

The cross Chapalote x Reventador had D4, D5, and D6 values of 4.54, 2.05, and 3.57 s.u., respectively, suggesting some relationship between the two races. Chapalote, as indicated, is a flint with a cigar shaped ear typical of Northwestern Mexico. Reventador is a popcorn with long and narrow ears, is also from Northwestern Mexico (Wellhausen et al., 1952). The distance values found in this study do not suggest very clear similarities for the two races; however, a very close relationship has been hypothesized by Wellhausen et al. (1952), Goodman (1972), Brown and Goodman (1977) and Cervantes et al. (1978).

The D4, D5, and D6 values for the cross Piricinco x Morado were 5.6, 1.15, and 3.7 s.u., respectively. Piricinco is an ancient, stable race from Peru, with very long ears and irregular row numbers at the base. It has a wide distribution from the eastern Andean slopes of Peru to the low areas of Bolivia and into a large area of Brazil (Grobman et al., 1961). Grobman and collaborators indicated that Cutler (1946) first described this race in Bolivia as Coroico, the name of the locality where it was first collected. Morado is from Bolivia and has long, slightly tapered ears with irregular rowing, floury endosperm and cherry or red pericarp (Ramírez et al., 1960). The distance values here seem to support the findings of Ramírez et al. (1960), who suggested that the Coroicos and Morados are probably related, and with the groupings given by Goodman and Bird (1977) and Brown and Goodman (1977), who have classified the two races into a floury Amazonian-interlocked group.

Two samples of the cross Amagaceño x Olotón Gua 639 were used to observe the agreement of the measures studied when different, independently constructed F_2 populations were evaluated. The results were in reasonable agreement in the sense that the average of D4, D5, and D6 were close: 3.97 and 3.77 s.u. for samples 1 and 2, respectively. These values support the findings of Wellhausen _et al_. (1952, 1957), who considered the Guatemalan Olotón to be related to the Colombian Montaña complex in which Amagaceño is included. Goodman and Bird (1977) and Goodman (1978) have reported close relationship between the two races, despite their geographical isolation. However, when a different collection, Olotón Gua 686, was crossed to Amagaceño, the D4, D5, and D6 values were much higher (9.19, 5.37, and 6.10 s.u., respectively); hence the similarities among the races are less apparent.

The cross Chaparreño x Camba had D4, D5, and D6 values of 5.0, 2.89, and 4.41 s.u., respectively. Chaparreño is an ancient, floury race from Peru, with short, cylindrical to globose ears, often having kernels distributed irregularly. It probably originated in the valley of Chaparra on the southwestern coast of Peru, where the race is intensively grown (Grobman _et al_., 1961). Camba is a floury race from Bolivia, described by Ramírez _et al_.(1960) as having long ears with a strong, regular taper from the base to the tip, and white, dented kernels. No relationship between the two races has been suggested. The distances found suggest neither a very close nor a very distant relationship among the races.

The cross Cariaco x Güirua had D4, D5, and D6 values of 4.90, 3.22, and 5.29 s.u., respectively. Both races are from Colombia and have been described by Roberts _et al_. (1957). Cariaco's plants are short and early, with few tillers, and have short to medium length, broad ears, tapering strongly from base to tip. Güirua's plants are tall and early, without tillers, and with predominantly long, slender ears that are slightly tapering from the base to the tip; some, however, are conical. Güirua's origin is unknonw, and it seems to have little influence on the evolution of the other Colombian races of corn. Moreover, it was collected from Indians who spoke no Spanish. The values of the measures used here suggest that the two races are neither distantly nor very closely related.

The last member of Group 2 corresponds to the cross Vandeño x Tepecintle, with D4, D5, and D6 values of 7.21, 2.72 and 4.73 s.u., respectively. Both races are prehistoric mestizos from Mexico and are described by Wellhausen _et al_. (1952). Vandeño, with medium length ears having a slight taper towards the tip, is the most common race on the south Pacific coast of Mexico. Tepecintle has short, broad, cylindrical ears, slightly tapered near the tip. It is from south Mexico but is more common in Guatemala. Tepecintle has been suggested to be an ancestor of Vandeño (Wellhausen _et al_., 1952). The results found here, however, tend to support the findings of Brown and Goodman (1977) and Cervantes _et al_. (1978), who have suggested only a few similarities between the two races.

Group 3. Remotely Related Races. The racial crosses included in this study are: Nal-Tel x Reventador (Nay 15); Pira x Pororo; Andaquí x Guaribero; Amagaceño x Olotón (Gua 686); Chapalote x

Piricinco; Cacahuacintle x Capio; Pollo x Kcello (Ecu 705) and Nal-Tel Nal-Tel x Reventador (Nay 39). All of these crosses have average values of D4, D5, and D6 larger than 5.0 s.u. and less than or equal to 10.00 s.u.; hence the degree of relationship between the pairs of races is less evident than that encompassed in Groups One and Two. The similarities among the crosses Amagaceño x Olotón (Gua 686) and Cacahuacintle x Capio have already been discussed; the remainder will be discussed next.

The first member of Group 3 was Nal-Tel x Reventador (Nay 15), whose D4, D5, and D6 values were 3.59, 2.30, and 10.68 s.u., respectively, suggesting that the two races are probably not closely related. Furthermore, when Reventador collection Nay 39 was used in place of Nay 15, the D4, D5, and D6 values increased to 10.44, 6.06, and 10.14 s.u., respectively. Nal-Tel and Reventador are races from Mexico, described by Wellhausen et al. (1952).
Nal-Tel is an ancient indigeneous race with short, early maturing plants, very few tillers, very short ears having a slight taper at both base and tip, and with small rounded, non-dented kernels. It was probably widely distributed in ancient times on both the east and west coasts of southern Mexico. Reventador, a popcorn with long, narrow ears, is from Northwestern Mexico. The races of corn of Mexico have been extensively restudied by Goodman (1972), Goodman and Bird (1977) and Cervantes et al. (1979), using different multivariate techniques; these studies have not suggested a close relationship between Reventador and Nal-Tel.

The racial cross Pira x Pororo had values D4, D5, and D6 of 3.18, 1.76, and 11.20 s.u., respectively. Pira is a popcorn from Colombia (and Venezuela, but this cross involved Colombian Pira), with plants of medium height, early to medium maturity, and no tillers. The ears are short, very slender and slightly tapered from base to tip or somewhat cigar-shaped (Roberts et al., 1957). Pororo, from Bolivia, is also a popcorn and has been described by Ramírez et al. (1960). It has medium to small, cylindrical to very slightly tapered ears with small, generally rounded kernels. The latter authors have indicated that the two races are similar in some aspects: small plants, slender stalks, small to medium ears, and similar ear and kernel shape. Despite these morphological similarities, Goodman and Bird (1977) classified them in two different popcorn groups. The results found here indicate that the two races are not very closely related

The racial cross Andaquí x Guaribero had D4, D5, and D6 values of 7.60, 1.24, and 9.90 s.u., respectively, indicating that the races are probably not closely related. Andaquí is an early Colombian semi-flint race, with medium to tall plants, no tillers and short to medium length ears that are strongly tapered toward the tip. It is found at low elevations in the southern interior of Colombia (Roberts et al., 1957). Guaribero is from Venezuela, with medium height plants, a zig-zag stalk, and white, round, semi-pop kernels. It is mainly distributed southeast of the city of Caracas (Grant et al., 1960). No relationship among these two races has been reported. These races were included in this study because of their general similarities in ear morphology, but the large D4, D5, and D6 values found here seem to indicate that the two races are not closely related.

The racial cross Chapalote x Piricinco had D4, D5, and D6 values of 5.68, 3.82, and 12.25 s.u., respectively. Piricinco, as indicated earlier, has also been called Coroico by Cutler (1946), Pojoso by Grobman et al. (1961), and Entrelacado by Brieger et al. (1958). This racial complex predominates in the low altitudes of the interior regions east of the central Andes in South America. It characteristically has long, narrow ears with irregular rowing at the base. Chapalote, with medium length, cigar-shaped ears, is from northwestern Mexico. No relationship among the two races has been reported, and the large values of the distances found in this study suggest that the two races are probably unrelated.

The last member of Group 3 corresponds to the cross Pollo x Kcello (Ecu 704) with D4, D5, and D6 values of 13.75, 4.18, and 8.96 s.u., respectively. Pollo is probably the most primitive race of Colombia and Venezuela. The Colombian Pollo was used in this study; it is a popcorn with short, early maturing plants, no tillers, and very short ears, some of which are conical and some of which have a slight taper at both base and tip (Roberts et al., 1957). Kcello is a race from Ecuador and Bolivia (but see Group 4). Kcello Ecuatoriano, involved in this cross, has slender plants, with a prominent zig-zag tendency. It has short, cylindrical and slightly tapered ears (Timothy et al., 1963). Zevallos et al. (1977) have indicated that Kcello Ecuatoriano seems to be related to the archeological kernels found at San Pablo in the southern lowlands of Ecuador; however, Mangeldorf (1977) indicated that Kcello is a distinctly highland race and would not be expected to survive the lowland tropics of coastal Ecuador. No relationship between Kcello and Pollo has been reported, and the results found here seem to indicate that the two races are unrelated.

Group 4. Very Distantly Related Races. The racial crosses for this group are: Kcello (Ecu 704) x Kcello (Bov 948); Clavo x Canilla; Confite Morocho x Enano; Dente Branco x Chapalote; Tabloncillo Perla x Lenha, all of which have average values of D4, D5, and D6 larger than 10.000 s.u.

The first member of this group was the cross Kcello (Ecu 704) x Kcello (Bov 948) whose D4, D5, and D6 values were 12.29, 7.27, and 10.24 s.u.; hence, the two races are probably not related. The races are from Ecuador and Bolivia, typically at high altitudes (2000 to 2600 meters). In general, they share a similar ear type; small-eared with straight and few rows, relatively large, rounded, yellow and flinty kernels (Timothy *et al.*, 1967). Despite these similarities, the results found here seem to indicate that the races are unrelated.

The cross Clavo x Canilla had D4, D5, and D6 values of 15.51, 6.22, and 16.66 s.u., respectively, and the races appear to be unrelated. Clavo from Colombia is probably an ancient introduction from Peru and seems to have had much influence on the evolution of Colombian maize. It has been suggested to be one of the parents of Montaña, Cabuya, Puya and Puya Grande. Its plants have medium height, are early with no tillers, and have long, slender and slightly conical ears (Roberts *et al.*, 1957). Canilla from Venezuela is described by Grant *et al.* (1963) as having tall plants with long, slender ears and deep rounded and hard grains. It is mainly distributed at low altitudes (20 to 450 meters). The results found here contrast with the hypothesis of Grant *et al.* (1963) that Clavo probably originated from a cross between Canilla and an Andean maize with a low row number.

The racial cross Confite Morocho x Enano had D4, D5, and D6 values of 27.04, 7.02, and 16.51 s.u., respectively. Confite Morocho is a small flint or popcorn and is reported to be one of the most ancient races from Peru. It has short plants and short, cylindrical to conical ears. It has been hypothesized to be a primitive ancestor of many races of corn (Grobman et al., 1961). Enano from Bolivia is a small kernelled flour or popcorn with short, early plants, no tillers and very small, slightly conical ears (Ramírez et al., 1961). The results found here contrast with the hypothesis of Goodman and Bird (1977) who included both Confite Morocho and Enano in a round yellow popcorn group, although the inclusion of Confite Morocho seemed doubtful. They also added that, in general, the popcorns have been little studied.

The last two members of Group 5 were Dente Branco x Chapalote and Tabloncillo Perla x Lenha; their use and utility in this study has already discussed.

In general, the results (see Table 3) obtained in this study seem to indicate that there is agreement among classifications based upon the average value of the modified generalized distance, D4, approximate Dempster's distance, D5, Dempster's distance, D6, and the classification based upon the Euclidean distance, D1. The Spearman correlation coefficients among D1 vs. D4, D5, and D6 were $r_{14} = 0.66$, $r_{15} = 0.53$ and $r_{16} = 0.95$ were all positive and highly significant. On the other hand, classifications based upon the Mahalanobis' distance, D2, or the generalized distance, D3, seem not to be in agreement with those classifications based upon the average of D4, D5, and D6. In fact, the distance values obtained (see Table 3) with Mahalanobis distance, D2, and generalized distance, D3, for the crosses: Chococeño x Aragüito, Confite Morocho x Chapalote, Cariaco x Chulpi and Salpor x Cacahuacintle in Group 1, and the crosses Cariaco x Güirua and Tepecintle x Vandeño in Group 2, and most of the crosses in Group 3 are very large compared to the Euclidean distance, D1. Moreover, the Spearman coefficient among Mahalanobis' distance, D2, and approximate Dempster's distance, D5, was $r = 0.23$, not statistically significant from zero. The coefficient among generalized distance, D3, and approximate Dempster's distance was $r = 0.15$, also not statistically significant.

SUMMARY AND CONCLUSIONS

A group of 47 collections of corn from Latin America and 30 F_2 populations, obtained by crossing certain pairs of these collections, were planted under Raleigh, North Carolina, and Homestead, Florida, conditions during a period of six years from 1971 to 1976. Data from these races and F_2 populations were collected for five morphological characters of the ear: length, diameter, row number, kernel length, and kernel thickness.

The purpose of this thesis was to examine the utility of various measures (distances) to provide estimates of racial divergence in races of maize. Six distances were considered for measuring such divergence: Euclidean distance (D1), Mahalanobis' distance (D2), generalized distance (D3), modified generalized distance (D4), approximate Dempster's distance (D5), and Dempster's distance (D6). However, only D4, D5, and D6 were used to infer whether or not two populations showed close relationship.

To construct the D1, D3, D4, D5, and D6 measures, rather than use the within-group variance (and/or covariance), the F_2 variance (and/or covariance) was used. Hence, these are measures of distance between the parents relative to their F_2 generation and appropriate for measuring racial differentiation. A multivariate generalization of Dempster's parental-combination variance was developed, and the relationships among the distance measures were discussed, including their advantages in graphical representation, their facility of computation, and the statistically reasoning for their use.

On the basis of the average of the D4, D5, and D6 measures and upon the fact that the degree of genetic divergence of the cross Zapalote Chico Oax 48 x Zapalote Chico Oax 50 was known to be small and could be used as a standard, the 30 racial crosses were placed in four groups of about the same size each: Group 1, Closely Related Races; Group 2, Somewhat Related Races; Group 3, Remotely Related Races; Group 4, Distantly Related Races.

In general, the degree of relationship between the pairs of races of the racial crosses encompassed in Groups 1, 2, 3, and 4 are in conformity with relationships previously reported on the basis of different multivariate techniques and with those based on the more classical botanical techniques used at the time of the description of the Latin American races of corn. Hence, the measures D4, D5, and D6 indicate that morphological studies of racial F_2 populations are useful as a complementary approach for the study of racial divergence in maize.

Despite the fact that the Euclidean distance, D1, does not consider correlation among characters, and was not used here to imply genetic divergence among pairs of races, the results found in this study seem to indicate that classifications based upon the Euclidean distance, D1 (when used with an F_2 standardization), are in reasonable agreement with those obtained by using the average of D4, D5, and D6.

LIST OF REFERENCES

Anderson, E. 1939. Recombination in species crosses. Genetics 24:668-698.

Anderson, E. 1946. Maize in Mexico. A preliminary survey. Ann. Mo. Bot. Gard. 33:147-157.

Anderson, E. 1947. Field studies of Guatemalan maize. Ann. Mo. Bot. Gard. 34:433-467.

Anderson, E. 1949. Introregressive Hybridization. John Wiley and Sons, Inc., New York.

Anderson, E. and W. L. Brown. 1950. The history of common maize varieties in the United States Corn Belt. Jour. New York Bot. Gard. 1951:252-267.

Anderson, E. and W. L. Brown. 1952. Origin of corn belt maize and its genetic significance. In J. W. Gowen (ed.), Heterosis. Iowa State Coll. Press, Ames, Iowa.

Anderson, E. and H. C. cutler. 1942. Races of Zea Mays. I. Their recognition and classification. Ann. Mo. Bot. Gard. 35:69-88.

Baum, B. R. 1977. Oats: Wild and Cultivated. A Monograph of the Genus Avena L. (Poaceae). Thorn Press, Canada.

Bird, R. M. 1970. Maize and its cultural and natural environment in the Sierra of Huanuco Peru. Ph.D. thesis, University of California, Berkeley.

Bird, R. M. and M. M. Goodman. 1977. The races of maize: V. Grouping races of maize on the basis of ear morphology Econ. Bot. 31:471-481.

Blackith, R. E. and R. A. Reyment. 1971. Multivariate Morphometrics. Academic Press, New York.

Bose, R. C. 1936. On the exact distribution and moment-coefficients of the D^2-statistics. Sankhya 2:143-154.

Brieger, F. G., J. T. A. Gurgel, E Paterniani, A. Blumenschein, and R. M. Alleoni. 1958. Races of maize in Brazil and other eastern South American countries. Nalt. Acad. Sci. Nat. Res. Council Publ. 593. Washington, D. C.

Brown, W. L. 1960. Races of maize in the West Indies. Natl. Acad. Sci. Nat. Res. Council Publ. 792. Washington, D. C.

Brown, W. L. and E. Anderson. 1947. The northern flint corns. Ann. Mo. Bot. Gard. 34:1-29.

Brown, W. L. and E. Anderson. 1948. The southern dent corns. Ann. Mo. Bot. Garden. 35:255-268.

Brown, W. L. and M. M. Goodman. 1977. Races of corn. In G. F. Sprague (ed.). Corn and Corn Improvement. Rev. ed. Amer. Soc. of Agronomy. Madison, Wisconsin.

Chaudhary, B. D. and V. P. Singh. 1975. Genetic divergence in some Indian and exotic barley varieties and their hybrids. Indian J. Genet. Plant Breed. 35:409-413.

Cervantes, T., M. M. Goodman, E. Casas, and J. O. Rawlings. 1978. Use of genetic effects and genotype by environment interactions for the classification of Mexican races of maize. Genetics 90:339-348.

Cutler, H. C. 1946. Races of maize in South America. Bot. Mus. Leaft. 12:257-291. Harvard University.

Davies, O. L. 1961. Statistical Methods in Research and Production. Hafner Publishing Company, New York.

Dempster, E. R. 1949. Effects of linkage on parental-combination and recombination frequencies in F_2. Genetics 34:272-284.

Edwards, R. J. 1966. Comparisons of methods and procedures for intraspecific classification of Zea Mays L. Ph.D. thesis, University of Illinois, Urbana.

Fisher, R. A. 1936. The use of multiple measurements in taxonomic problems. Ann. Eugen. 7:179-188.

Forsyth, A. R. 1930. Geometry of Four Dimensions. Vol. I. Cambridge University Press, Cambridge.

Goodman, M. M. 1967. The races of maize: I. The Use of Mahalanobis' generalized distances to measure morphological similarity. Fitotecnia Latinoamericana 4:1-22.

Goodman, M. M. 1968. The races of maize: II. Use of multivariate analysis of variance to measure morphological similarity. Crop Sci. 8:693-698.

Goodman, M. M. 1969. Measuring evolutionary divergence. Japan J. Genetics 44:310-316.

Goodman, M. M. and E. Paterniani. 1969. The races of maize: III. Choices of appropriate characters for racial classification. Economy Botany 23:265-273.

Goodman, M. M. 1972. Distance analysis in biology. Systematic Zoology 21:174-186.

Goodman, M. M. 1978. A brief survey of races of maize and current attempts to infer racial relationships. In D. B. Walden (ed.), Maize Breeding and Genetics. John Wiley, N. Y.

Goodman, M. M. and R. M. Bird. 1977. The races of maize IV: Tentative grouping of 219 Latin American races. Economic Botany 31:204-221.

Govil, J. H. and B. R. Murty. 1973. Genetic divergence and nature of heterosis in grain sorghum. Indian J. of Genetics and Plant Breed. 34:252-260.

Grant, U. J., W. H. Hatheway, D. H. Timothy, C. Cassalet, and L. M. Roberts. 1963. Races of maize in Venezuela. Natl. Acad. Sci. Nat. Res. Council Publ. 1136. Washington, D. C.

Grobman, A., W. Salhuana and R. Sevilla, with P. C. Mangelsdorf. 1961. Races of maize in Peru. Natl. Acad. Sci. Nat. Res. Council Publ. 915. Washington, D. C.

Hatheway, W. H. 1957. Races of maize in Cuba. Natl. Acad. Sci. Nat. Res. Council Publ. 453. Washington, D. C.

Hernandez, X. E. and G. Alanis. 1970. Estudio morfologico de cinco nuevas razas de maiz de la Sierra Madre Occidental de Mexico: Implicaciones filogeneticas y fitogeograficas. Agrociencia 5:3-30.

Hussaini, S. H., M. M. Goodman and D. H. Timothy. 1977. Multivariate analysis and the geographical distribution of the world collection of finger millet. Crop Sci. 17:257-263.

Jeswani, L. M., B. R. Murty and R. S. Mehra. 1970. Divergence in relation to geographical origin in a world collection of linseed. Indian J. Genetic Plant Breed. 30:11-25.

Jolicoeur, P. and J. E. Mosiman. 1960. Size and shape variation in the painted tuttle. A principal component analysis. Growth 24(4): 339-354.

Kato, T. A. and A. Blumenschein. 1967. Complejos de nudos cromosomicos en los maices de America. Fitotec. Latinoamer. 4(2):13-24.

Kendall, M. G. 1961. A Course in the Geometry of n Dimensions. Hafner Publ. Co., N. Y.

Kendall, M. G. and A. Stuart. 1976. The Advanced Theory of Statistics. Vol. 3, Charles Griffin & Company, Ltd., London.

Longley, A. E. 1938. Chromosomes of maize from North American Indians. J. Agric. Res. 56:177-195.

Longley, A. E. and Kato, Y. 1965. Chromosome morphology of certain races of maize in Latin America. Int. Center for the improvement of maize and wheat. Res. Bull. No. 1. Chapingo, Mexico.

Mahalanobis, P. C. 1936. On the generalized distance in statistics. Proc. Nat. Inst. Sci. India 2:49-55.

Mangelsdorf, P. C. 1974. Corn, Its Origin, Evolution and Improvement. Harvard Univ. Press. Cambridge, Mass.

Mangelsdorf, P. C. 1977. More on the San Pablo corn kernel. (mimeo)

McClintock, B. 1959. Genetic and cytological studies of maize. Carnegie Inst. Wash. Yearbook 58:452-456.

McClintock, B. 1960. Chromosome constitution of Mexican and Guatemalan races of maize. Carnegic Inst. Wash. Yearbook 59: 461-472.

McClintock, B. 1977. Significance of chromosome constitutions in tracing the origin and migration of races of maize. In D. B. Walden (ed.), Maize Breeding and Genetics, John Wiley, N. Y.

Menndirata, P. D., P. S. Phul and N. D. Arora. 1971. Genetic diversity in relation to fodder yield and its components in sorghum. Indian J. Genetics Plant Breed. 31:300-303.

Mochizuki, O. and T. Okuno. 1967. Classification of maize lines collected from Sikoku, Japan, and selection of breeding varieties by application of principal components analysis. Japan J. Breeding 17:284-291.

Morishima, H. and H. Oka. 1960. The pattern of interspecific variation in the genus Oriza: its quantitative representation by statistical methods. Evolution 14:153-165.

Morrison, D. F. 1967. Multivariate Statistical Methods. McGraw-Hill Book Company, N. Y.

Murty, G. S. and M. V. Pavate. 1962. Studies on quantitative inheritance in Nicotiana tabacum L. I. Varietal classification and selection by multivariate analysis. Indian J. of Genetics and Plant Breed. 22:68-77.

Murty, B. R. and M. I. Qadry. 1966. Analysis of divergence in some self-compatible forms of Brassica campestris Var. Brown Sarson. Indian J. of Genetics and Plant Breed. 26:43-57.

Namkoong, G. 1966. Statistical analysis of introgression. Biometrics 22:488-502.

Parker, V. I. and O. Pastorini. 1965. Distribución geografica, clasificación y estudio del maíz (Zea mays) en Chile. Agric. Tecnica 25(2):70-86.

Paterniani, E. and M. M. Goodman. 1977. Races of Maize in Brazil and Adjacent Areas. Cimmyt. Ediciones Las Americas. Mexico.

Ram, J. and D. V. S. Panwar. 1970. Interspecific divergence in rice. Indian J. Genetics and Plant Breed. 30:1-10.

Rao, C. R. 1960. Multivariate analysis: An indispensable statistical aid in applied research. Sankya 22:317-338.

Ramírez, E. R., D. H. Timothy, E. Diaz and U. J. Grant, with G. E. Nicholson Calle, E. Anderson and W. L. Brown. 1960. Races of maize in Bolivia. Natl. Acad. Sci. Nat. Res. Council Publ. 747. Washington, D. C.

Roberts, L. M., U. J. Grant, R. Ramírez, W. H. Hatheway and D. L. Smith, with P. C. Mangelsdorf. 1957. Races of maize in Colombia. Natl. Acad. Sci. Nat. Res. Council Publ. 510. Washington, D. C.

Rodríquez A., M. Moreno, J. Quiroga, and G. Avila with A. Brandolini. Maices Bolivianos, FAO. Rome, Italy.

Rummel, J. R. 1970. Applied Factor Analysis. Northwestern University Press, Evanston, Ill.

Salhuana, W. S. 1969. Study of the genetic relationships and heterosis in inter and intra race of Peruvian maize. Ph. D. Thesis, University of Minnesota, St. Paul, Minn.

Searle, S. R. 1971. Linear Models. John Wiley & Sons, Inc., N. Y.

Smouse, P. E. 1972. The canonical analysis of multiple species hybridization. Biometrics 28:361-371.

Sneath, P. H. and R. R. Sokal. 1973. Numerical Taxonomy. W. H. Freeman and Co. San Francisco, Calif.

Sneath, P. H. 1976. Some applications of numerical taxonomy to plant breeding. Z. Pltanzenüchtg. 76:19-46.

Sokal, R. R. and P. H. Sneath. 1963. Principles of Numerical Taxonomy. W. H. Freeman and Co. San Francisco, Calif.

Somayajulu, P. L. M., A. B. Joshi and B. R. Murty. 1970. Genetic divergence in wheat. Indian J. of Genetics and Plant Breed. 30:47-58.

Smith, H. H. 1950. Developmental restrictions on recombination in Nicotiana. Evolution 4:202-211.

Sturtevant, E. L. 1899. Varieties of corn. USDA Office of Exp. Stn. Bull. 57. Washington, D. C.

Timothy, D. H., B. Peña and R. Ramírez, with W. L. Brown and E. Anderson. 1961. Races of maize in Chile. Natl. Acad. Sci. Nat. Res. Council Publ. 847. Washington, D. C.

Timothy, D. H., W. H. Hatheway, U. J. Grant, M. Torregroza, D. Sarria and D. Varela. 1963. Races of maize in Ecuador. Natl. Acad. Sci. Nat. Res. Council Publ. 975. Washington, D. C.

Timothy, D. H. and M. M. Goodman. 1979. Germplasm preservation: The basis of future feast or famine-genetic resources of maize an example. In Rubenstein I., R. L. Phillips, C. E. Green and B. G. Gengelbech (eds.), The Plant Seed: Development, Preservation and Germination. Academic Press, N. Y.

Wellhausen, E. J., L. M. Roberts, and E. Hernandez, with P. C. Mangelsdorf. P. C. Mangelsdorf. 1952. Races of maize in Mexico. The Bussey Institute, Harvard University. Cambridge, Mass.

Wellhausen, E. J., A. Fuentes, and E. Hernandez,with P. C. Mangelsdorf. 1957. Races of maize in Central America. Natl. Acad. Sci. Nat. Res. Council Publ. 511. Washington, D. C.

Zevallos, C. M., W. C. Galinat, D. W. Lathop, E. R. Leng, J. G. Marcos and K. M. Klumpp. 1977. The San Pablo corn and its friends. Science. 196:385-389.

APPENDIX A.1

Let $\underline{x}_1$ and $\underline{x}_2$ be two independent random vectors with $E(\underline{x}_i) = \underline{\mu}_i$, i=1,2 and $Cov(\underline{x}_i) = \Sigma = \{\sigma_{k\ell}\}$, $k,\ell = 1,\ldots,p$. Then a formal definition of the Mahalanobis generalized distance is

$$D^2(\underline{\mu}_1,\underline{\mu}_2,\Sigma) = (\underline{\mu}_1 - \underline{\mu}_2)' \, \Sigma^{-1} \, (\underline{\mu}_1 - \underline{\mu}_2)$$

FACT: The Mahalanobis distance is invariant under the following linear transformation:

$$\underline{y}_i = A\underline{x}_i - \underline{c}$$

where

$\underline{c}$ = vector of known constants

$A = \text{Diagonal } \{\sigma_{11},\sigma_{22},\ldots,\sigma_{pp}\}$

then, $E(\underline{y}_i) = A\underline{\mu}_i - \underline{c} = \underline{\mu}_i^*$

$$\begin{aligned} Var(\underline{y}_i) &= Var\ (A\underline{x}_i - \underline{c}) \\ &= Var\ (A\underline{x}_i) \\ &= A\Sigma A \\ &= \mathbb{R} \end{aligned}$$

hence,

$$\begin{aligned} D^2(\underline{\mu}_1^*, \underline{\mu}_2^*, \mathbb{R}) &= (A(\underline{\mu}_1 - \underline{\mu}_2))' \, \mathbb{R}^{-1} \, (A(\underline{\mu}_1 - \underline{\mu}_2)) \\ &= (\underline{\mu}_1 - \underline{\mu}_2)' \, A(A\Sigma A)^{-1} \, A(\underline{\mu}_1 - \underline{\mu}_2) \\ &= (\underline{\mu}_1 - \underline{\mu}_2)' \, \Sigma^{-1} \, (\underline{\mu}_1 - \underline{\mu}_2) \\ &= D^2(\underline{\mu}_1,\underline{\mu}_2,\Sigma) \end{aligned}$$

APPENDIX A.2

RELATIONSHIP BETWEEN PRINCIPAL COMPONENTS AND MAHALANOBIS' DISTANCE

Principal components are formally defined as follows:

Let $x_1, x_2, \ldots, x_p$ be the elements of the random vector $\underline{x}$ with mean $\underline{\mu}$ and covariance matrix Σ positive semidefinite. Let $\lambda_1 \geq \lambda_2 \geq, \ldots, \geq \lambda_p$ be the eigenvalues of Σ. From matrix theory, we can find an orthonormal matrix P of order pxp whose columns are the associate orthonormal characteristic vectors of Σ such that

$$\Sigma P = P\Lambda \text{ and } \Sigma = P\Lambda P'$$

where

$$\Lambda = \text{Diagonal } \{\lambda_1, \lambda_2, \ldots, \lambda_p\}$$

Consider the orthogonal transformation

$$\underline{y} = P'\underline{x}$$

then, $y_1, y_2, \ldots, y_p$, the elements of $\underline{y}$, are called the principal components of $\underline{x}$. The first principal component of $\underline{x}$; which correspond to the maximum eigen-value λ_1 is y_1, y_2 is the second, etc.

Note,

$$E(\underline{y}) = P'\underline{\mu} \text{ and } Var(\underline{y}) = P'\Sigma P = \Lambda$$

Now, consider $\underline{x}_1, \underline{x}_2$ be two random vectors with $E(\underline{x}_i) = \underline{\mu}_i$ and $Var(\underline{x}_i) = \Sigma$, $i = 1,2$.

Let P be defined as above, then the transformation

$$\underline{y}_i = P'\underline{x}_i$$

implies, $E(\underline{y}_i) = P'\underline{\mu}_i = \underline{\xi}_i$, Var $(\underline{y}_i) = \Lambda$. From Appendix A1,

$$D^2(\underline{\xi}_i, \underline{\xi}_2, \Lambda) = (P'\underline{\mu}_1 - P'\underline{\mu}_2)' \Lambda^{-1} (P'\underline{\mu}_1 - P'\underline{\mu}_2)$$

$$= (\underline{\xi}_1 - \underline{\xi}_2)' \Lambda^{-1} (\underline{\xi}_1 - \underline{\xi}_2)$$

$$= \sum_{k=1}^{p} [(\xi_{1k} - \xi_{2k})^2/\lambda_k]$$

i.e., Mahalanobis' distance in the new axes is a standardized Euclidean distance.

APPENDIX A.3

Kendall (1961) represented a line in p-dimensional space as

$$\frac{X_1-c_1}{m_1} = \frac{X_2-c_2}{m_2} = \ldots\ldots = \frac{X_p-c_p}{m_p} \tag{1}$$

where,

c_i, $i=1,\ldots\ldots,p$ is any fixed point on the line, and m_i, $i=1,\ldots\ldots,p$ are the direction cosines of the line. The direction cosines have the property that $\Sigma m_i^2 = 1$. Consider $Y_1, Y_2,\ldots\ldots,Y_p$ as any external point relative to the straight line (1). Then a generalization of the methodology given by Forsyth (1930) allows us to obtain the coordinates of the intersection of the line with the perpendicular (the orthogonal projection) from the external point to the straight line (1). Let $v_1,v_2,\ldots\ldots,v_p$ be any point on the line (1) at a distance g from $c_1,c_2,\ldots\ldots,c_p$, then $v_1,v_2,\ldots\ldots,v_p$ can be expressed as

$$v_1 = c_1 + m_1 g \;;\; v_2 = c_2 + m_2 g \;;\; \ldots\ldots;\; v_p = c_p + m_p g \tag{2}$$

The Euclidean distance D^2 from $Y_1,Y_2,\ldots\ldots,Y_p$ to $v_1,v_2,\ldots\ldots,v_p$ is given by

$$D^2 = (y_1-v_1)^2 + (y_2-v_2)^2 + \ldots\ldots + (y_p-v_p)^2$$

Now we want to obtain the coordinates of the intersection of the perpendicular from $y_1,y_2,\ldots\ldots,y_p$ to the line (1). To do this we only need that value g, which makes D^2 a minimum and then substitute this value in (2).

Hence we have

$$D^2 = \sum_{i=1}^{p} (y_i - v_i)^2$$

$$= \sum_{i=1}^{p} (y_i - c_i - m_i g)^2$$

$$= \sum_{i=1}^{p} (y_i - c_i)^2 - 2g \sum_{i=1}^{p} m_i (y_i - c_i) + g^2 \sum_{i=1}^{p} m_i^2$$

$$\frac{d(D^2)}{dg} = -2 \sum_{i=1}^{p} m_i (y_i - c_i) + 2g = 0$$

Hence $g = \sum_{i=1}^{p} m_i (y_i - c_i)$

Then the coordinates Z_i at the point of intersection are:

$$Z_i = c_i + m_i \sum_{i=1}^{p} m_i (y_i - c_i) \; ; \; i = 1, \ldots\ldots, p$$

Now we can use these Z_i to obtain an estimator of Dempster's parental-combination variance. For an F_2 population consider a random sample of n observations each with p-characters, and denote a single observation by y_{ji}, $j=1,\ldots\ldots,n$, $i=1,\ldots\ldots,p$.

Without loss of generality assume that each character is standardized to mean zero, variance one. Denote the estimator S^2 defined by

$$s^2 = \frac{1}{n-1} \sum_{j=1}^{n} \sum_{i=1}^{p} (Z_{ji} - \bar{Z}_{.i})^2$$

$$= \frac{1}{n-1} \sum_{j=1}^{n} \sum_{i=1}^{p} [c_i + m_i \sum_{i=1}^{p} m_i (y_{ji} - c_i) - c_i - m_i \sum_{i=1}^{p} m_i (\bar{y}_{.i} - c_i)]^2$$

$$= \frac{1}{n-1} \sum_{j=1}^{n} \sum_{i=1}^{p} (m_i \sum_{i=1}^{p} m_i y_{ji})^2$$

$$= \frac{1}{n-1} \left[\sum_{j=1}^{n} \sum_{i=1}^{p} \left(m_i^2 \sum_{i=1}^{p} m_i^2 y_{ji}^2 + 2m_i^2 \sum_{k<k'}^{p} m_k m_{k'} y_{jk} y_{jk'} \right) \right]$$

$$s^2 = \frac{1}{n-1} \left[(n-1) \sum_{i=1}^{p} \left[m_i^2 \sum_{i=1}^{p} m_i^2 \sum_{j=1}^{n} \frac{y_{ji}^2}{n-1} + 2\, m_i^2 \sum_{k<k'}^{p} m_k m_{k'} \sum_{j=1}^{n} \frac{y_{jk} y_{jk'}}{n-1} \right] \right]$$

$$= 1 + 2 \sum_{k<k'}^{p} m_k m_{k'}\, r_{kk'},$$

where $r_{kk'}$ is the correlation coefficient between y_k and $y_{k'}$, and the m_k correspond to the direction cosines. Since we are interested in the variation through the direction of the parents, then Dempster's parental combination variance would be

$$\hat{s}^2 = 1 + 2 \sum_{k<k'}^{p} m_k^* m_{k'}^* r_{kk'},$$

where the

$$m_k^* = \frac{T_{ik} - T_{jk}}{\left[\sum_{k=1}^{p} (T_{ik} - T_{jk})^2 \right]^{\frac{1}{2}}}$$

are the direction cosines of the line between the parents, $r_{kk'}$ is the simple correlation coefficient between characters k and k' in the F_2 population and

$$T_{ik} = \frac{\bar{x}_{ik}}{s_k}, \quad k = 1,\ldots,p;$$

$\bar{x}_{ik}$ is the mean of parent i for k^{th} character, and s_k is the standard deviation of the F_2 for the k^{th} character.

APPENDIX A.4

a. Standard error for D1.

Searle (1971) gives the following theorem: if $\underline{x}$ is $N_p(\mu,\Sigma)$ and A is a real symmetric matrix, pxp, then

$$\mathrm{Var}(\underline{x}'A\underline{x}) = 2 \text{ trace } (A\Sigma)^2 + 4\mu' A\Sigma A\mu$$

The estimated squared Euclidean distance (D1) between race 1 and race 2 can also be expressed in matrix form as:

$$D1 = (\bar{\underline{z}}_1 - \bar{\underline{z}}_2)' I \, (\bar{\underline{z}}_1 - \bar{\underline{z}}_2)$$

where

I = identity matrix of order p and

$\bar{\underline{z}}_i$ = the standardized (with respect to the F_2) vector of means for race i, i = 1,2;

i.e., $\bar{\underline{z}}_i = B\bar{\underline{x}}_i$

B = is a diagonal matrix whose elements are the reciprocal of the F_2 standard deviations

and $\bar{\underline{x}}_i$ = the vector of means for race i.

Now, if $\underline{x}_i$ is $N_p(\mu_i,\Sigma)$ then

$$\mu_i^* = E(\bar{\underline{z}}_i) = E(B\bar{\underline{x}}_i) \doteq B\mu_i$$

$$\Sigma^* = \mathrm{Var}(\bar{\underline{z}}_i) = \mathrm{Var}(B\bar{\underline{x}}_i) \doteq \frac{1}{n_i} B\Sigma B'$$

and n_i is the number of observations for race i. Note,

$$\mathrm{Var}(\bar{\underline{z}}_1 - \bar{\underline{z}}_2) = \frac{1}{n_1} B\Sigma B' + \frac{1}{n_2} B\Sigma B'$$

$$= (\frac{1}{n_1} + \frac{1}{n_2}) B\Sigma B'$$

Assuming $\underline{\bar{z}}_i$ is $N_p(\mu_i^*, \Sigma^*)$ and using Searle's theorem, the variance for the squared Euclidean distance (D1) is

$$\text{Var(D1)} = \text{Var } (\underline{\bar{z}}_1 - \bar{z}_2)'I\ (\bar{z}_1 - \bar{z}_2)$$

$$= 2tr[I(\frac{1}{n_1} + \frac{1}{n_2})B\Sigma B']^2 + 4(\mu_1^*-\mu_2^*)'I(\frac{1}{n_1} + \frac{1}{n_2})B\Sigma B'I(\mu_1^*-\mu_2^*)$$

$$= 2(\frac{1}{n_1} + \frac{1}{n_2})^2 tr(B\Sigma B')^2 + 4(\frac{1}{n_1} + \frac{1}{n_2})(\mu_1^* - \mu_2^*)'B\Sigma B'(\mu_1^* - \mu_2^*)$$

Then an estimator of the variance for D1 is

$$\text{Var(D1)} = 2(\frac{1}{n_1} + \frac{1}{n_2})^2 tr(BSB')^2 + 4(\frac{1}{n_1} + \frac{1}{n_2})(\underline{\bar{z}}_1 - \underline{\bar{z}}_2)'BSB'(\underline{\bar{z}}_1 - \underline{\bar{z}}_2)$$

where S is the pooled variance covariance matrix of races 1 and 2.

b. Standard error for D2.

The squared Mahalanobis distance D2 between race 1 and race 2 was estimated by

$$D2 = (\underline{\bar{x}}_1 - \underline{\bar{x}}_2)'S^{-1}(\underline{\bar{x}}_1 - \underline{\bar{x}}_2),$$

where

$\underline{\bar{x}}_i$ = vector of means for race i,i=1,2; and

S = the pooled variance covariance matrix of races 1 and 2.

Assuming that the samples sizes are large enough that $S^{-1} = \Sigma^{-1}$ and $\underline{x}_i$, is $N_p(\mu_i, \Sigma)$,i=1,2; then an approximate variance for the squared D2 distance can be obtained by using Searle's result.

$$\text{Var(D2)} \doteq \text{Var}((\underline{\bar{x}}_1 - \underline{\bar{x}}_2)'\Sigma^{-1}(\underline{\bar{x}}_1 - \underline{\bar{x}}_2))$$

$$= 2tr(\Sigma^{-1}(\frac{\Sigma}{n_1} - \frac{\Sigma}{n_2}))^2 + 4(\mu_1 - \mu_2)'\Sigma^{-1}(\frac{\Sigma}{n_1} + \frac{\Sigma}{n_2})\Sigma^{-1}(\mu_1 - \mu_2)$$

$$= 2(\frac{1}{n_1} + \frac{1}{n_2})^2 tr(I_p)^2 + 4(\frac{1}{n_1} + \frac{1}{n_2})(\mu_1 - \mu_2)'\Sigma^{-1}(\mu_1 - \mu_2)$$

$$= 2p(\frac{1}{n_1} + \frac{1}{n_2})^2 + 4(\frac{1}{n_1} + \frac{1}{n_2})D^2$$

where D is the population Mahalanobis' distance. This approximate variance agrees with the exact variance of the studentized-D^2 statistics obtained by Bose (1936). The studentized-D^2 statistics is a linear transformation of the squared Mahalanobis' distance D2. From above, an estimator for the variance of D2 is

$$\hat{Var}(D2) = 2p(\frac{1}{n_1} + \frac{1}{n_2})^2 + 4(\frac{1}{n_1} + \frac{1}{n_2})D2.$$

c. <u>Standard error for D3.</u>

The squared generalized distance (D3) between race 1 and race 2 can also be expressed in matrix form as

$$D3 = (\underline{\bar{x}}_1 - \underline{\bar{x}}_2)' S_2^{-1} (\underline{\bar{x}}_1 - \underline{\bar{x}}_2),$$

where

S_2 = is the estimated covariance matrix of the F_2, and

$\underline{\bar{x}}_i$ = is the vector of means for race i, i=1,2.

Assuming that the samples sizes are large enough that $S_2^{-1} = \Sigma_2^{-1}$, where Σ_2 is the covariance matrix of the F_2 and $\underline{x}_i$ is $N_p(\mu_i, \Sigma)$, i=1,2; then an approximate variance for D3 can be obtained by using Searle's result.

$$Var(D3) \doteq Var((\underline{\bar{x}}_1 - \underline{\bar{x}}_2)' \Sigma_2^{-1} (\underline{\bar{x}}_1 - \underline{\bar{x}}_2))$$

$$= 2\, tr(\Sigma_2^{-1}(\frac{\Sigma}{n_1} + \frac{\Sigma}{n_2}))^2 + 4(\mu_1 - \mu_2)'\Sigma_2^{-1}(\frac{\Sigma}{n_1} + \frac{\Sigma}{n_2})\Sigma_2^{-1}(\mu_1 - \mu_2)$$

$$= 2(\frac{1}{n_1} + \frac{1}{n_2})^2 tr(\Sigma_2^{-1}\Sigma)^2 + 4(\frac{1}{n_1} + \frac{1}{n_2})(\mu_1 - \mu_2)'\Sigma_2^{-1}\Sigma\Sigma_2^{-1}(\mu_1 - \mu_2)$$

$$= 2(\frac{1}{n_1} + \frac{1}{n_2})^2 tr(\Sigma_2^{-1}\Sigma)^2 + 4(\frac{1}{n_1} + \frac{1}{n_2})(\mu_1 - \mu_2)'\Sigma_2^{-1}\Sigma\Sigma_2^{-1}(\mu_1 - \mu_2)$$

Then an estimator for the variance of D3 is

$$\hat{Var}(D3) = 2(\frac{1}{n_1} + \frac{1}{n_2})^2 tr(\underset{\sim}{S}_2^{-1}S)^2 + 4(\frac{1}{n_1} + \frac{1}{n_2})(\underline{\bar{x}}_1 - \underline{\bar{x}}_2)'S_2^{-1}SS_2^{-1}(\underline{\bar{x}}_1 - \underline{\bar{x}}_2)$$

where

S = is the pooled covariance matrix of races 1 and 2.

d. <u>Standard error for D4.</u>

The modified squared generalized distance (D4) between race 1 and race 2 was estimated by

$$D4 = \sum_{k=1}^{q} [(y_{1k} - y_{2k})^2/d_k]; d_k \geq 1 \text{ , } 1 \leq q \leq 5;$$

where

$y_{ik} = \underline{a}_k'\underline{\bar{z}}_i, i = 1,2;$

$\underline{a}_k'$ = the k^{th} characteristic vector of R, the correlation matrix of the F_2;

$\underline{\bar{z}}_i$ = the standardized (with respect to the F_2) vector of means for race i; and

d_k = the k^{th} characteristic root of R.

A standard error for D4 is difficult to attain because the distributions of y_{1k}, y_{2k}, and d_k obtained from sample correlation matrices have not been subjected to thorough study and are complicated to obtain. For this reason and because Var(D3) $\geq$ Var(D4), as shown below, an upper bound of the standard error of D4 was considered to be the standard error of D3.

Note that

$$\text{Var (D3)} = \text{Var} \left[\sum_{k=1}^{5} [(y_{1k} - y_{2k})^2/d_k] \right]$$

$$= \sum_{k=1}^{5} [\text{Var}[(y_{1k}-y_{2k})^2/d_k]],$$ because principal components are independent.

Now, $\text{Var}[(y_{1k}-y_{2k})^2/d_k] \geq 0$, for any k, k=1,2,...5.

Then for any q, $1 \leq q \leq 5$.

$$\text{Var(D3)} \geq \left[\sum_{k=1}^{q} [\text{Var } (y_{1k}-y_{2k})^2/d_k]\right] = \text{Var}\left[\sum_{k=1}^{q} [(y_{1k}-y_{2k})^2/d_k]\right] = \text{Var(D4)}$$

e. Standard error for D5.

The squared approximate Dempster's distance (D5) between race 1 and race 2 was estimated from

$$D5 = (y_{11} - y_{21})^2/d_1,$$

where

$y_{i1} = \underline{a}_1' \bar{\underline{z}}_i,$

$\underline{a}_1'$ = the 1st characteristic vector of R, the correlation matrix of the F_2,

$\bar{\underline{z}}_i$ = the standardized (with respect to the F_2) vector of means for race i; i=1,2; and

d_1 = the 1^{st} characteristic root of R.

As in the case of D4, a standard error for D5 is also difficult to obtain. However, an upper bound of the standard error of D5 was obtained by considering the squared Euclidean distance (D1) as follows

$$D1 = (\bar{\underline{z}}_1 - \bar{\underline{z}}_2)'I(\bar{\underline{z}}_1 - \bar{\underline{z}}_2)$$

$$= (\bar{\underline{z}}_1 - \bar{\underline{z}}_2)'PP'(\bar{\underline{z}}_1 - \bar{\underline{z}}_2)$$

where

P = is a orthonormal matrix, whose elements are the orthonormal characteristic vectors of R.

Let $\underline{y}_i = P'\bar{\underline{z}}_i$,

then

$$D1 = (P'\bar{\underline{z}}_1 - P'\bar{\underline{z}}_2)'I(P'\bar{\underline{z}}_1 - P'\bar{\underline{z}}_2) = (\underline{y}_1 - \underline{y}_2)'I(\underline{y}_1 - \underline{y}_2)$$

$$= \sum_{k=1}^{p} (y_{1k} - y_{2k})^2;$$

i.e., the Euclidean distance is invariant under orthonormal transformations.

Define, $D5^* = D1/d_1$

Claim: $Var(D5^*) \gtrless Var(D5)$

Indeed,

$$Var(D5^*) = Var\left[\frac{1}{d_1} \sum_{k=1}^{p} (y_{1k} - y_{2k})^2\right]$$

$$= \sum_{k=1}^{p} Var\left[\frac{(y_{1k}-y_{2k})^2}{d_1}\right] + \sum_{k=1}^{p} \sum_{\substack{k'=1 \\ k'\neq k}}^{p} Cov\left[\frac{(y_{1k}-y_{2k})^2}{d_1}, \frac{(y_{1k'}-y_{2k'})^2}{d_1}\right]$$

$$= \sum_{k=1}^{p} Var\left[\frac{(y_{1k}-y_{2k})^2}{d_1}\right] ,$$

where y_{ik}, as defined above, is the k^{th} principal component for race i. The last statement is valid because a characteristic root is independent of the elements of its associated characteristic vector and because principal components are themselves uncorrelated (Morrison 1967).

Hence,

$$Var(d5*) \gtrsim Var\ [\frac{(y_{11}-y_{21})^2}{d_1}] = Var\ (D5).$$

Now, we have to obtain the variance of D5*.

Let $f = D5* = \frac{D1}{d_1} = \frac{D}{w}$, where d_1 is independent of D1 (Morrison, 1967).

Using the approximate variance for function of random variables (Davies, 1961):

$$Var\ (D5*) = Var(f) \doteq [\frac{df}{dD}]^2_{D=D1} Var(D) + [\frac{df}{dw}]^2_{w=d1} Var(w)$$

$$= \frac{1}{w^2} [Var(D) + \frac{D^2}{w^2} Var(w)].$$

$$Var\ (w) \doteq \frac{2}{n} [w^2 + \sum_{\substack{k=1 \\ k \neq k'}}^{p} \sum_{k'=1}^{p} l_k^2 l_{k'}^2 r_{kk'}^2 - 2w \sum_{k=1}^{p} l_k^4]$$, Kendall and Stuart (1976);

where $r_{kk'}$ is the sample correlation coefficient among character k and k' of the F_2 generation, the l_k are the elements of the characteristic vector associated with w, the first characteristic root, and n is the number of observations in the F_2.

f. <u>Standard error for D6.</u>

The squared Dempster's distance (D6), between race 1 and race 2 was estimated from

$$D6 = \sum_{k=1}^{p} (\bar{z}_{1k} - \bar{z}_{2k})^2/s^2$$

where

$s^2 = 1 + 2 \sum_{k<k'}^{p} m_k m_{k'} r_{kk'}$, the Dempster's parental-combination variance,

$$m_k = \frac{\bar{z}_{1k} - \bar{z}_{2k}}{\left(\sum_{k=1}^{p} (z_{1k}-z_{2k})^2\right)^{1/2}} ,$$

$$\bar{z}_{1k} = \frac{\bar{x}_{1k}}{s_k} ,$$

$\bar{x}_{1k}$ = the mean of parent 1 for the k^{th} character, and

s_k = the standard deviation of the k^{th} character in the F_2.

Then,

$D6 = D1/s^2$.

Define $D6 = D1/s^2 = D/w = f$.

Hence the procedure to obtain Var(D6) is analogous to that used for Var(D5*),

$$\mathrm{Var}(D6) = \frac{1}{w^2} [\mathrm{Var}(D) + \frac{D^2}{w^2} \mathrm{Var}(w)] ,$$

and

$$\mathrm{Var}(w) = \frac{2w^2}{(n-1)+2} .$$

Appendix Table B. 1 . Planting dates, number of plots, number of plants per plot and locations for the F_2 populations and and the parents in study.

No.	Populations	Date	No. plot	Plants/ plot	Location
1	Tabloncillo Perla x Harinoso de Ocho	4-25-71	41	1	N.C.
2	Tabloncillo Perla	4-28-75	3	30	N.C.
3	Harinoso de Ocho	4-28-75	3	30	N.C.
4	Negrito x Moroti Precoce	4-25-71	36	1	N.C.
5	Negrito	4-28-75	3	30	N.C.
6	Moroti Precoce	4-28-75	3	30	N.C.
7	Chococeño x Aragüito	11- 1-75	5	30	Fla.
8	Chococeño	11- 1-75	2	30	Fla.
9	Aragüito	11- 1-75	2	30	Fla.
10	Confite Morocho x Chapalote	11- 7-72	2	30	Fla.
11	Confite Morocho	11-15-74	2	30	Fla.
12	Chapalote	11- 4-74	1	30	Fla.
13	Cateto Assis x Cateto Grande	4-25-71	57	1	N.C.
14	Cateto Assis	4-28-75	3	30	N.C.
15	Cateto Grande	4-28-75	3	30	N.C.
16	Cariaco x Chulpi	11- 4-74	3	30	Fla.
17	Cariaco	11- 4-74	1	30	Fla.
18	Chulpi	11- 4-74	1	30	Fla.
19	Capio x Salpor	10-17-75	6	30	Fla.
20	Capio	11- 4-74	2	30	Fla.
21	Salpor	10-17-75	2	30	Fla.
22	Salpor x Cacahuacintle	11- 1-75	6	30	Fla.
23	Salpor	11- 1-75	2	30	Fla.
24	Cacahuacintle	11- 1-75	2	30	Fla.
25	Zapalote Chico Oax 48 x Zapalote Chico Oax 50	4-25-71	33	1	N.C.
26	Zapalote Chico Oax 48	4-28-75	3	30	N.C.
27	Zapalote Chico Oax 50	4-28-75	3	30	N.C.
28	Mochero x Chuncho	11- 4-74	3	30	Fla.
29	Mochero	11- 4-74	1	30	Fla.
30	Chuncho	11- 4-74	1	30	Fla.
31	Chapalote x Reventador Nay 15	4-25-71	50	1	N.C.
32	Chapalote	4-28-75	2	30	N.C.
33	Reventador Nay 15	4-28-75	2	30	N.C.
34	Piricinco x Morado	11- 1-75	120	1	N.C.
35	Piricinco	11- 1-75	2	30	N.C.
36	Morado	11- 1-75	2	30	N.C.
37	Amagaceño x Olotón Gua 639 (2)	10-30-71	15	5	Fla.
38	Amagaceño	11- 1-75	2	30	Fla.
39	Olotón Gua 639	11- 1-75	2	30	Fla.
40	Amagaceño x Olotón Gua 639 (1)	10-30-71	15	5	Fla.

Appendix Table B. 1 (continued).

No.	Populations	Date	No. plot	Plants/ plot	Location
41	Amagaceño	11- 1-75	2	30	Fla.
42	Olotón Gua 639	11- 1-75	2	30	Fla.
43	Chaparreño x Camba	10-19-73	4	30	Fla.
44	Chaparreño	10-19-73	2	30	Fla.
45	Camba	10-19-73	2	30	Fla.
46	Cariaco x Güirua	11- 1-73	3	30	Fla.
47	Cariaco	11- 1-75	2	30	Fla.
48	Güirua	11- 1-75	2	30	Fla.
49	Tepecintle x Vandeño	4-25-71	53	1	N.C.
50	Tepecintle	4-28-75	3	30	N.C.
51	Vandeño	4-28-75	3	30	N.C.
52	Nal-Tel x Reventador Nay 15	4-25-75	50	1	N.C.
53	Nal-Tel	4-28-75	3	30	N.C.
54	Reventador Nay 15	4-28-75	3	30	N.C.
55	Pira x Pororo	10-23-72	3	30	Fla.
56	Pira	11- 1-75	3	30	Fla.
57	Pororo	11- 1-75	2	30	Fla.
58	Andaquí x Guaribero	11- 1-75	3	30	Fla.
59	Andaquí	11- 1-75	2	30	Fla.
60	Guaribero	11- 1-75	2	30	Fla.
61	Amagaceño x Olotón Gua 686	11- 1-75	3	30	Fla.
62	Amagaceño	11- 1-75	3	30	Fla.
63	Olotón Gua 686	11- 1-75	2	30	Fla.
64	Chapalote x Piricinco	11- 4-74	3	30	Fla.
65	Chapalote	11- 4-74	1	30	Fla.
66	Piricinco	11- 4-74	1	30	Fla.
67	Cacahuacintle x Capio	11- 4-74	3	30	Fla.
68	Cacahuacintle	11- 4-74	1	30	Fla.
69	Capio	11- 4-74	1	30	Fla.
70	Pollo x Kcello Ecu 704	9-29-76	3	30	Fla.
71	Pollo	9-29-76	2	30	Fla.
72	Kcello Ecu 704	9-29-76	2	30	Fla.
73	Nal-Tel x Reventador Nay 39	4-25-71	33	1	N.C.
74	Nal-Tel	4-28-75	3	30	N.C.
75	Reventador Nay 39	4-28-75	3	30	N.C.
76	Kcello Ecu 704 x Kcello Bov 948	9-29-76	3	30	Fla.
77	Kcello Ecu 704	9-29-76	2	30	Fla.
78	Kcello Bov 948	9-29-76	2	30	Fla.
79	Clavo x Canilla	10-17-75	6	30	Fla.
80	Clavo	10-17-75	2	30	Fla.
81	Canilla	10-17-75	2	30	Fla.
82	Confite Morocho x Enano	11-15-74	4	30	Fla.
83	Confite Morocho	11-15-74	2	30	Fla.
84	Enano	11-15-74	2	30	Fla.
85	Tabloncillo Perla x Lenha	4-25-71	53	1	N.C.

Appendix Table B. 1 (continued).

No.	Populations	Date	No. plot	Plants/ plot	Location
86	Tabloncillo Perla	4-28-75	3	30	N.C.
87	Lenha	4-28-75	3	30	N.C.
88	Dente Branco x Chapalote	4-25-71	48	1	N.C.
89	Dente Branco	4-28-75	3	30	N.C.
90	Chapalote	11- 4-74	1	30	N.C.

Appendix B. Table 2. Means and variances for the five characters and for the Latin American races of corn.

Race Name	Ear length $\bar{x}$	Ear length s^2	Ear diameter $\bar{x}$	Ear diameter s^2	No. of rows $\bar{x}$	No. of rows s^2	Kernel length $\bar{x}$	Kernel length s^2	Kernel thickness $\bar{x}$	Kernel thickness s^2
Amagaceño	13.36	4.20	3.50	0.10	10.96	3.04	8.86	0.80	5.14	0.17
Andaquí	10.44	3.33	3.63	0.11	11.36	3.15	9.36	0.45	4.74	0.17
Araguito	7.53	0.83	2.68	0.05	12.55	2.90	7.50	0.43	3.69	0.09
Cacahuacintle (Fla. 74)	6.48	1.24	3.96	0.07	11.00	1.33	11.80	0.35	5.55	0.81
Cacahuacintle (Fla. 75)	10.37	8.87	4.38	0.16	11.48	1.97	11.54	1.27	5.24	0.34
Canilla	15.09	5.45	2.94	0.15	13.07	2.35	9.21	1.37	3.91	0.33
Camba	14.21	7.64	4.26	0.13	15.09	5.11	11.89	0.81	4.40	0.24
Capio (Composite)	11.41	6.71	3.96	0.13	13.25	3.13	8.87	0.68	5.34	0.29
Capio (Ant. 318)	15.48	6.15	4.78	0.14	12.22	1.83	10.07	1.04	6.70	0.27
Cariaco (Cor. 338)	8.50	3.27	4.22	0.18	11.25	4.80	9.75	0.68	4.36	0.08
Cariaco (Composite)	10.20	2.80	4.35	0.15	15.09	3.50	10.56	0.76	4.33	0.17
Cateto Assis Brasil	20.10	14.64	3.86	0.13	12.82	2.91	10.23	1.11	4.36	0.17
Cateto Grande	20.32	9.12	3.59	0.09	10.87	3.03	9.12	0.65	4.66	0.22
Chapalote (Fla. 74)	10.52	3.90	2.86	0.02	9.91	1.99	7.90	0.42	3.90	0.13
Chapalote (N.C. 75)	14.01	11.75	2.86	0.08	11.72	2.78	8.01	0.67	4.01	0.14
Chaparreño	10.37	2.86	4.00	0.10	13.41	1.25	11.25	1.28	4.51	0.24
Chococeño	7.96	2.16	2.89	0.10	15.42	4.32	6.33	0.87	3.73	0.13
Chulpi	8.60	4.93	4.35	0.09	18.47	6.26	10.95	0.37	4.74	0.36
Chuncho	11.82	11.24	4.18	0.15	13.09	5.09	10.93	0.96	4.59	0.24
Clavo	15.04	8.37	3.09	0.18	10.35	1.11	8.53	1.16	5.40	0.34
Confite Morocho	7.68	5.39	2.57	0.11	11.75	2.20	8.85	0.69	4.28	0.17
Dente Branco	18.84	6.81	5.02	0.29	13.87	3.47	14.51	1.35	3.60	0.16
Enano	4.09	0.74	2.14	0.06	14.43	4.85	5.15	0.30	3.41	0.07
Guaribero	9.25	2.70	3.27	0.16	14.24	5.44	9.13	0.43	3.88	0.15
Guirua	14.19	6.83	3.43	0.18	13.85	5.14	9.18	0.47	4.51	0.11
H. de Ocho Occidentales	21.14	16.71	3.77	0.17	10.06	1.93	10.22	1.29	4.44	0.31

Appendix B. Table 2 (continued).

Race Name	Ear length $\bar{x}$	Ear length s^2	Ear diameter $\bar{x}$	Ear diameter s^2	No. of rows $\bar{x}$	No. of rows s^2	Kernel length $\bar{x}$	Kernel length s^2	Kernel thickness $\bar{x}$	Kernel thickness s^2
Kcello (Bov 948)	10.82	2.50	4.04	0.11	11.69	2.36	12.95	1.23	4.19	0.19
Kcello (Ecu 704)[1]	11.37	3.53	3.25	0.11	10.38	0.89	10.16	1.26	4.83	0.16
Kcello (Ecu 704)[2]	10.61	3.92	3.32	0.09	10.29	0.91	9.82	0.95	4.74	0.13
Lenha	16.80	12.41	5.04	0.25	21.31	8.16	10.14	0.91	4.20	0.13
Mochero	7.76	2.27	3.92	0.08	13.00	2.19	10.20	1.00	4.43	0.23
Morado	12.03	6.96	3.11	0.10	11.49	1.73	8.24	0.48	4.43	0.25
Moroti Precoce	17.40	11.87	3.73	0.14	13.00	4.36	9.32	0.91	5.04	0.16
Nal-Tel	8.81	3.22	2.79	0.23	10.80	10.17	8.34	1.75	4.34	0.26
Negrito	16.61	4.14	3.83	0.10	12.18	1.87	10.44	0.57	4.93	0.13
Olotón (Gua 639)	15.55	7.46	3.82	0.05	12.35	3.11	9.65	0.44	4.47	0.22
Oloton (Gua 686)	16.78	10.24	4.21	0.09	12.62	2.31	9.98	0.70	4.83	0.11
Pira	10.24	3.95	2.61	0.07	10.43	1.28	9.30	0.44	3.59	0.12
Piricinco (Fla. 74)	15.56	8.41	3.17	0.05	11.92	7.16	8.58	0.30	5.32	0.25
Piricinco (Fla. 75)	15.23	10.93	2.97	0.06	11.93	5.29	8.02	0.26	5.10	0.39
Pollo	8.80	1.47	2.84	0.05	10.52	1.17	8.13	0.31	4.31	0.13
Pororo	11.56	5.84	2.67	0.07	14.32	2.43	7.84	0.29	3.07	0.05
Reventador (Nay 15)	18.11	11.13	2.93	0.09	13.35	4.05	8.25	0.55	3.43	0.12
Reventador (Nay 39)	19.38	17.92	3.10	0.10	13.20	2.00	8.42	0.98	4.05	0.20
Salpor[3]	17.07	9.56	4.64	0.14	13.10	1.29	9.50	0.54	5.87	0.16
Salpor[4]	14.52	20.20	4.21	0.37	12.27	1.64	9.38	0.75	5.72	0.23
Tabloncillo Perla	19.23	11.50	3.61	0.13	9.11	1.64	9.62	1.42	4.67	0.15
Tepecintle	14.18	4.13	4.43	0.13	13.25	2.86	10.14	0.75	3.82	0.15
Vandeño	19.02	5.66	3.66	0.11	12.74	1.82	9.35	0.59	4.07	0.13
Zapalote Chico Oax 48	12.24	2.62	3.90	0.12	10.82	1.48	10.67	1.33	4.09	0.18
Zapalote Chico Oax 50	12.50	4.79	4.32	0.15	12.80	3.35	10.64	1.03	4.01	0.34

[1] Used in the cross Kcello Ecu 704 x Pollo.

[2] Used in the cross Kcello Ecu 704 x Kcello Bov 948.

[3] Used in the cross Salpor x Capio.

[4] Used in the cross Salpor x Cacahuacintle.

Appendix Table B. 3. Means, variances and correlations for the F_2 populations: Harinoso de Ocho x Tabloncillo Perla (1) and Negrito x Moroti Precoce (2).

F_2 populations		Character				
		Ear length (1)	Ear diameter (2)	No. of rows (3)	Kernel length (4)	Kernel thickness (5)
(1)	$\bar{x}$	20.86	3.72	9.80	9.95	4.69
(1)	s^2	23.16	0.08	2.76	0.80	0.36
(2)	$\bar{x}$	15.09	3.82	12.56	9.34	4.30
(2)	s^2	13.03	0.16	3.11	1.03	0.24

Correlations: Harinoso de Ocho x Tabloncillo Perla/Negrito x Moroti Precoce

Character	Character (1)	(2)	(3)	(4)	(5)
1	1.0	0.4219	0.3254	0.4860	-0.1390
2	0.3812	1.0	0.4626	0.6122	-0.3428
3	0.0782	0.4221	1.0	0.4944	0.2153
4	0.0403	0.5404	0.0713	1.0	0.0241
5	-0.0980	-0.2213	-0.1679	-0.0962	1.0

Appendix Table B. 4. Means, variances and correlations for the F_2 populations: Chococeño x Aragüito (1) and Confite Morocho x Chapalote (2).

F_2 populations		Character: Ear length (1)	Ear diameter (2)	No. of rows (3)	Kernel length (4)	Kernel thickness (5)
(1)	$\bar{x}$	8.56	2.88	15.21	7.36	3.85
(1)	s^2	1.51	0.10	4.23	0.60	0.13
(2)	$\bar{x}$	11.60	3.15	12.30	9.29	3.98
(2)	s^2	3.50	0.08	2.44	0.70	0.19

Correlations: Chococeño x Aragüito/Confite Morocho x Chapalote

Character	Character: (1)	(2)	(3)	(4)	(5)
1	1.0	0.3428	0.3610	0.2937	-0.1144
2	0.2089	1.0	0.6063	0.7517	-0.0689
3	0.2111	0.4842	1.0	0.3265	-0.2085
4	0.1757	0.6500	0.3540	1.0	-0.0899
5	-0.0480	-0.1573	-0.1763	-0.1161	1.0

Appendix Table B. 5. Means, variances and correlations for the F_2 populations: Cateto Assis Brasil x Cateto Grande (1) and Cariaco x Chulpi (2).

F_2 populations		Character				
		Ear length (1)	Ear diameter (2)	No. of rows (3)	Kernel length (4)	Kernel thickness (5)
(1)	$\bar{x}$	19.82	3.68	11.96	9.16	4.93
(1)	s^2	18.94	0.10	2.93	0.79	0.29
(2)	$\bar{x}$	9.93	4.93	17.37	11.48	4.55
(2)	s^2	2.78	0.20	7.85	1.45	0.16

Correlations: Cateto Assis Brasil x Cateto Grande/Cariaco x Chulpi

Character	Character (1)	(2)	(3)	(4)	(5)
1	1.0	0.3154	0.1967	0.3013	-0.0674
2	0.0983	1.0	0.5433	0.6450	0.1124
3	-0.2061	0.5004	1.0	0.5905	0.2822
4	0.1726	0.7029	0.3323	1.0	-0.1284
5	-0.1619	-0.1189	-0.1418	-0.1089	1.0

Appendix Table B. 6. Means, variances and correlations for the F_2 populations: Capio x Salpor (1) and Salpor x Cacahuacintle (2).

F_2 populations		Character				
		Ear length (1)	Ear diameter (2)	No. of rows (3)	Kernel length (4)	Kernel thickness (5)
(1)	$\bar{x}$	18.25	4.73	12.35	10.27	6.13
(1)	s^2	7.60	0.17	2.05	0.74	0.25
(2)	$\bar{x}$	11.77	4.42	12.72	10.60	5.32
(2)	s^2	9.51	0.19	2.75	1.45	0.26

Correlations: Capio x Salpor/Salpor x Cacahuacintle.

Character	Character (1)	(2)	(3)	(4)	(5)
1	1.0	0.0696	-0.0009	0.1818	-0.0287
2	0.2920	1.0	0.3640	0.4738	-0.1789
3	-0.0697	0.4775	1.0	0.1884	-0.0842
4	0.3974	0.6654	0.2432	1.0	-0.1985
5	-0.3266	-0.2072	-0.0338	-0.3575	1.0

Appendix Table B. 7. Means, variances and correlations for the F_2 populations: Zapalote Chico Oax 48 x Zapalote Chico Oax 50 (1) and Mochero x Chuncho (2).

F_2 populations		Character				
		Ear length (1)	Ear diameter (2)	No. of rows (3)	Kernel length (4)	Kernel thickness (5)
(1)	$\bar{x}$	12.26	4.23	11.52	10.95	4.19
(1)	s^2	4.16	0.13	2.76	1.34	0.29
(2)	$\bar{x}$	11.51	4.28	13.83	11.39	4.37
(2)	s^2	7.12	0.13	4.09	0.99	0.12

Correlations: Zapalote Chico Oax 48 x Zapalote Chico Oax 50/ Mochero x Chuncho.

Character	Character (1)	(2)	(3)	(4)	(5)
1	1.0	0.1391	0.1196	0.0791	-0.0132
2	0.2502	1.0	0.3763	0.8505	-0.4155
3	0.1473	0.6067	1.0	0.3659	-0.0459
4	0.3740	0.7106	0.3070	1.0	-0.5080
5	-0.2566	-0.2083	-0.1054	-0.2335	1.0

Appendix Table B. 8. Means, variances and correlations for the F_2 populations: Chapalote x Reventador (Nay 15) (1) and Piricinco x Morado (2).

F_2 populations		Character: Ear length (1)	Ear diameter (2)	No. of rows (3)	Kernel length (4)	Kernel thickness (5)
(1)	$\bar{x}$	15.63	2.97	11.60	8.22	4.10
(1)	s^2	17.38	0.07	2.31	0.58	0.30
(2)	$\bar{x}$	13.96	3.13	13.33	8.37	5.34
(2)	s^2	7.83	0.06	3.65	0.41	0.31

Correlations: Chapalote x Reventador (Nay 15)/Piricinco x Morado

Character	Character: (1)	(2)	(3)	(4)	(5)
1	1.0	0.4073	0.0322	0.4563	-0.1904
2	0.2155	1.0	0.3003	0.7949	-0.1338
3	0.3287	0.4770	1.0	0.1098	-0.0061
4	0.2324	0.7058	0.2327	1.0	-0.0579
5	-0.0735	-0.3334	-0.2522	-0.3182	1.0

Appendix Table B. 9. Means, variances and correlations for the F_2 populations: Amagaceño x Olotón Gua 639 Cross 1 (1) and Amagaceño x Olotón Gua 639 Cross 2 (2).

F_2 populations		Character: Ear length (1)	Ear diameter (2)	No. of rows (3)	Kernel length (4)	Kernel thickness (5)
(1)	$\bar{x}$	16.06	3.38	10.55	8.03	5.17
(1)	s^2	9.27	0.07	3.04	0.40	0.47
(2)	$\bar{x}$	16.91	3.70	11.28	8.74	4.85
(2)	s^2	8.07	0.10	2.00	0.63	0.43

Correlations: Amagaceño x Olotón Gua 639 Cross 1/Amagaceño x Olotón Gua 639 Cross 2.

Character	Character: (1)	(2)	(3)	(4)	(5)
1	1.0	0.2054	-0.1087	0.3381	-0.2618
2	0.4983	1.0	0.3060	0.6855	-0.0412
3	0.2247	0.4400	1.0	-0.0046	0.2923
4	0.5485	0.6438	0.2212	1.0	-0.2963
5	0.0093	-0.1889	-0.0537	-0.2795	1.0

Appendix Table B.10. Means, variances and correlations for the F_2 populations: Chaparreño x Camba (1) and Cariaco x Güirua (2).

F_2 populations		Ear length (1)	Ear diameter (2)	No. of rows (3)	Kernel length (4)	Kernel thickness (5)
(1)	$\bar{x}$	13.76	4.42	14.90	12.43	4.45
(1)	s^2	4.59	0.11	3.56	1.07	0.19
(2)	$\bar{x}$	12.37	3.20	10.78	8.86	4.17
(2)	s^2	3.25	0.14	2.89	0.62	0.14

Correlations: Chaparreño x Camba/Cariaco x Güirua.

Character	(1)	(2)	(3)	(4)	(5)
1	1.0	0.0582	-0.1797	0.2735	-0.0565
2	-0.0060	1.0	0.5520	0.5721	-0.1908
3	-0.0713	0.6808	1.0	0.3095	-0.1208
4	0.5782	0.4329	0.2318	1.0	-0.1733
5	-0.1566	0.0824	0.0080	-0.2342	1.0

Appendix Table B.11. Means, variances and correlations for the F_2 populations: Tepecintle x Vandeño (1) and Nal-Tel x Reventador (May 15) (2).

F_2 populations		Character: Ear length (1)	Ear diameter (2)	No. of rows (3)	Kernel length (4)	Kernel thickness (5)
(1)	$\bar{x}$	15.09	3.82	12.56	9.34	4.30
(1)	s^2	13.03	0.16	3.11	1.03	0.24
(2)	$\bar{x}$	14.53	3.02	13.44	8.25	4.02
(2)	s^2	8.86	0.13	4.58	0.87	0.37

Correlations: Tepecintle x Vandeño/Nal-Tel x Reventador (May 15)

Character	Character: (1)	(2)	(3)	(4)	(5)
1	1.0	0.3812	0.0782	0.0403	-0.0980
2	0.1354	1.0	0.4221	0.5404	-0.2213
3	0.1396	0.6777	1.0	0.0713	-0.1679
4	0.0909	0.7651	0.5040	1.0	-0.0962
5	0.3108	-0.1033	0.0356	-0.1094	1.0

Appendix Table B.12. Means, variances and correlations for the F_2 populations: Pira x Pororo (1) and Andaquí x Guaribero (2).

F_2 populations		Character: Ear length (1)	Ear diameter (2)	No. of rows (3)	Kernel length (4)	Kernel thickness (5)
(1)	$\bar{x}$	15.16	2.86	12.30	8.94	3.35
(1)	s^2	3.95	0.09	3.01	0.58	0.09
(2)	$\bar{x}$	11.13	3.41	12.09	8.77	4.38
(2)	s^2	2.46	0.11	3.45	0.59	0.13

Correlations: Pira x Pororo/Andaquí x Guaribero.

Character	Character: (1)	(2)	(3)	(4)	(5)
1	1.0	-0.0296	-0.0617	0.0611	0.4849
2	0.1746	1.0	0.3819	0.4303	-0.0856
3	0.0792	0.6984	1.0	0.3962	-0.2677
4	0.3059	0.5648	0.5006	1.0	-0.0002
5	0.1061	-0.2839	-0.2158	-0.3275	1.0

Appendix Table B.13. Means, variances and correlations for the F_2 populations: Amagaceño x Olotón (Gua 686) (1) and Chapalote x Piricinco (2).

F_2 populations		Ear length (1)	Ear diameter (2)	No. of rows (3)	Kernel length (4)	Kernel thickness (5)
		Character				
(1)	$\bar{x}$	16.15	4.02	12.03	9.48	5.27
(1)	s^2	8.57	0.13	3.57	0.69	0.28
(2)	$\bar{x}$	15.49	3.06	12.71	8.43	4.91
(2)	s^2	11.78	0.07	4.16	0.54	0.19

Correlations: Amagaceño x Olotón (Gua 686)/Chapalote x Piricinco

Character	Character (1)	(2)	(3)	(4)	(5)
1	1.0	-0.1104	-0.2520	0.0366	0.0279
2	0.0987	1.0	0.5544	0.5612	-0.1549
3	0.1292	0.5499	1.0	0.2838	0.0532
4	0.1181	0.7393	0.2576	1.0	-0.1113
5	0.4371	-0.0618	-0.0281	-0.1021	1.0

Appendix Table B.14. Means, variances and correlations for the F_2 populations: Cacahuacintle x Capio (1) and Pollo x Kcello (Ecu 704) (2).

F_2 populations		Character: Ear length (1)	Ear diameter (2)	No. of rows (3)	Kernel length (4)	Kernel thickness (5)
(1)	$\bar{x}$	12.82	4.34	12.71	11.40	4.97
(1)	s^2	6.36	0.13	2.77	1.33	0.18
(2)	$\bar{x}$	10.51	3.16	9.97	9.42	4.42
(2)	s^2	2.14	0.13	1.77	0.80	0.12

Correlations: Cacahuacintle x Capio/Pollo x Kcello (Ecu 704).

Character	Character: (1)	(2)	(3)	(4)	(5)
1	1.0	0.0787	0.0157	0.2441	-0.2410
2	0.4318	1.0	0.3202	0.3356	0.2380
3	0.0559	0.4999	1.0	-0.0814	-0.0699
4	0.3116	0.7172	0.4868	1.0	-0.3712
5	-0.0992	-0.1804	-0.1248	-0.2325	1.0

Appendix Table B.15. Means, variances and correlations for the F_2 populations: Nal-Tel x Reventador (Nay 39) (1) and Kcello (Ecu 704) x Kcello (Bov 948) (2).

F_2 populations		Character: Ear length (1)	Ear diameter (2)	No. of rows (3)	Kernel length (4)	Kernel thickness (5)
(1)	$\bar{x}$	15.00	3.09	13.03	8.50	4.36
(1)	s^2	11.20	0.12	2.28	0.78	0.24
(2)	$\bar{x}$	10.99	3.94	11.36	12.33	4.34
(2)	s^2	2.94	0.11	2.35	1.74	0.24

Correlations: Nal-Tel x Reventador (Nay 39)/Kcello (Ecu 704) x Kcello (Bov 948).

Character	Character: (1)	(2)	(3)	(4)	(5)
1	1.0	0.4331	0.3803	0.1881	-0.2024
2	-0.0427	1.0	0.4449	0.7131	-0.3576
3	0.0332	0.3684	1.0	0.2087	0.0204
4	0.2687	0.2389	0.1176	1.0	-0.2972
5	-0.3930	0.0051	0.2378	-0.3355	1.0

Appendix Table B.16. Means, variances and correlations for the F_2 populations: Clavo x Canilla (1) and Confite Morocho x Enano (2).

F_2 populations		Character				
		Ear length (1)	Ear diameter (2)	No. of rows (3)	Kernel length (4)	Kernel thickness (5)
(1)	$\bar{x}$	15.59	3.08	11.58	8.56	4.60
(1)	s^2	7.19	0.15	2.14	0.46	0.28
(2)	$\bar{x}$	6.82	2.47	12.49	7.59	3.65
(2)	s^2	4.64	0.12	4.25	0.91	0.16

Correlations: Clavo x Canilla/Confite Morocho x Enano.

Character	Character				
	(1)	(2)	(3)	(4)	(5)
1	1.0	0.0142	-0.0257	0.2906	0.0414
2	0.5250	1.0	0.3726	0.2763	-0.0019
3	0.2960	0.4393	1.0	0.0468	-0.1633
4	0.5378	0.7437	0.1526	1.0	-0.2670
5	0.0130	-0.0516	-0.2566	-0.0495	1.0

Appendix Table B.17. Means, variances and correlations for F_2 populations: Dente Branco x Chapalote (1) and Tabloncillo Perla x Lenha (2).

F_2 populations		Character				
		Ear length (1)	Ear diameter (2)	No. of rows (3)	Kernel length (4)	Kernel thickness (5)
(1)	$\bar{x}$	15.83	3.90	13.54	10.25	4.37
(1)	s^2	14.99	0.12	5.32	1.06	0.35
(2)	$\bar{x}$	20.44	4.36	13.74	9.75	5.22
(2)	s^2	17.97	0.24	3.54	0.93	0.52

Correlations: Dente Branco x Chapalote/Tabloncillo Perla x Lenha.

Character	Character (1)	(2)	(3)	(4)	(5)
1	1.0	0.0741	-0.0498	0.2907	-0.3344
2	0.0880	1.0	0.3404	0.8242	-0.4852
3	0.2593	0.4824	1.0	0.1122	-0.1136
4	0.2153	0.6286	0.2370	1.0	0.5373
5	0.0901	-0.2854	0.1689	-0.2478	1.0

Appendix Table B.18. Euclidean distance (D1), Mahalanobis' distance (D2), generalized distance (D3), modified generalized distance (D4), approximate Dempster's distance (D5), and Dempster's distance (D6) for the racial crosses in study.

No.	Pedigree	Squared Distances					
		D1	D2	D3	D4	D5	D6
1	Harinoso de Ocho x Tabloncillo Perla	1.37	0.96	0.77	0.55	0.5233	0.63
2	Negrito x Moroti Precoce	1.58	3.28	2.69	0.44	0.4345	1.21
3	Chococeño x Aragüito	4.72	11.16	9.71	0.18	0.0676	2.42
4	Confite Morocho x Chapalote	6.66	18.42	13.34	0.05	0.0001	3.97
5	Cateto Assis Brasil x Cateto Grande	3.87	2.28	2.55	1.33	1.2122	1.81
6	Cariaco x Chulpi	3.25	10.87	4.23	1.67	0.6007	2.55
7	Capio x Salpor	3.83	4.06	4.91	1.54	0.0534	3.83
8	Salpor x Cacahuacintle	6.09	10.58	10.43	1.66	0.4153	3.04
9	Zapalote Chico Oax 48 x Zapalote Chico Oax 50	3.86	3.58	4.73	1.97	1.4351	2.55
	Group 2						
10	Mochero x Chuncho	3.49	4.66	3.47	2.38	0.6159	2.44
11	Chapalote x Reventador (Nay 15)	3.35	6.52	3.31	2.19	0.6616	3.04
12	Piricinco x Morado	3.17	4.53	3.26	2.84	0.0901	3.15
13	Amagaceño x Olotón Gua 639 (2)	4.63	5.17	2.36	2.01	1.7172	2.64
14	Amagaceño x Olotón Gua 639 (1)	5.14	5.17	3.24	2.36	2.2560	3.37
15	Chaparreño x Camba	5.04	4.11	4.88	2.47	1.1014	3.84
16	Cariaco x Güirua	8.07	13.82	11.37	2.41	1.3019	4.66
17	Tepecintle x Vandeño	6.39	17.55	10.27	3.80	1.0101	4.14

Appendix Table B.18. (continued)

No.	Pedigree	Squared Distances					
		D1	D2	D3	D4	D5	D6
	Group 3						
18	Nal-Tel x Reventador (Nay 15)	13.41	16.39	18.90	1.62	0.7645	9.73
19	Pira x Pororo	12.56	26.39	18.47	1.37	0.4448	10.09
20	Andaquí x Guaribero	9.37	12.71	19.69	4.04	0.1416	8.99
21	Amagaceño x Oloton (Gua 686)	8.21	6.61	5.96	5.00	2.5609	5.42
22	Chapalote x Piricinco	16.22	18.08	13.27	2.88	1.6519	11.19
23	Cacahuacintle x Capio	12.56	33.16	16.94	5.58	0.0356	11.74
24	Pollo x Kcello (Ecu 704)	11.84	10.68	13.17	7.76	1.8587	8.11
25	Nal-Tel x Reventador (Nay 39)	13.67	15.64	10.77	5.51	2.9643	9.21
	Group 4						
26	Kcello Ecu 704 x Kcello (Bov 948)	12.77	11.93	9.24	6.87	3.6868	10.28
27	Clavo x Canilla	12.56	11.56	11.32	8.82	3.0525	15.34
28	Confite Morocho x Enano	25.78	46.32	26.64	15.79	3.5237	15.20
29	Dente Branco x Chapalote	86.75	41.71	60.08	29.66	28.6783	46.02
30	Tabloncillo Perla x Lenha	51.68	37.68	52.13	14.46	11.8625	35.04

Appendix Table B.19. Standard errors of squared Euclidean distance (D1), squared Mahalanobis distance (D2), generalized squared distance (D3), modified generalized squared distance (D4), approximate squared Dempster's distance (D5), and squared Dempster's distance (D6).

No.	Racial Cross	Standard Error					
		D1*	D2	D3*	D4*	D5*	D6**
	Group 1						
1	Harinoso de Ocho x Tabloncillo Perla	1.13	0.51	0.68	0.68	0.48	0.54
2	Negrito x Moroti Precoce	0.71	0.95	0.87	0.87	0.35	0.61
3	Chococeño x Aragüito	1.35	1.55	2.71	2.71	0.61	0.75
4	Confite Morocho x Chapalote	1.75	2.40	4.12	4.12	0.95	1.23
5	Cateto Assis Brasil x Cateto Grande	1.36	0.59	0.85	0.85	0.65	0.72
6	Cariaco x Chulpi	1.51	2.11	1.63	1.64	0.76	1.56
7	Capio x Salpor	1.71	1.23	2.13	2.13	0.98	1.78
8	Salpor x Cacahuacintle	2.63	2.15	4.23	4.23	1.13	2.34
9	Zapalote Chico Oax 48 x Zapalote Chico Oax 50	1.02	0.71	1.79	1.79	0.49	0.86
	Group 2						
10	Mochero x Chuncho	1.98	1.65	1.89	1.89	0.87	1.44
11	Chapalote x Revent. (Nay 15)	1.31	1.31	1.18	1.18	0.70	1.33
12	Piricinco x Morado	1.14	1.12	1.17	1.17	0.52	1.20
13	Amagaceño x Olotón Gua 639 (2)	2.28	1.46	1.19	1.19	1.05	1.36
14	Amagaceño x Olotón Gua 639 (1)	2.58	1.46	1.62	1.62	0.93	1.85
15	Chaparreño x Camba	1.77	0.97	1.64	1.64	0.95	1.49
16	Cariaco x Güirua	2.09	1.63	2.85	2.85	1.22	1.42
17	Tepecintle x Vandeño	1.76	1.87	2.56	2.56	1.18	1.39
	Group 3						
18	Nal-Tel x Revent. (Nay 15)	4.78	2.53	6.40	6.40	2.34	3.96
19	Pira x Pororo	3.20	2.44	4.75	4.75	2.17	3.14
20	Andaquí x Guaribero	2.50	1.53	5.03	5.03	1.24	2.75
21	Amagaceño x Olotón (Gua 686)	2.96	1.43	2.13	2.13	1.63	2.17
22	Chapalote x Piricinco	5.29	1.43	4.44	4.44	2.82	4.03
23	Cacahuacintle x Capio	6.45	2.47	9.29	9.29	4.22	6.52
24	Pollo x Kcello (Ecu 704)	3.46	6.68	3.93	3.93	1.66	2.74
25	Nal-Tel x Revent. (Nay 39)	5.49	1.62	4.13	4.13	2.72	4.32

Appendix Table B.19 (continued).

No.	Racial Cross	Standard Errors D1*	D2	D3*	D4*	D5*	D6**
	Group 4						
26	Kcello Ecu 704 x Kcello Bov 948	4.56	2.61	3.23	3.23	3.04	4.15
27	Clavo x Canilla	4.20	1.97	3.58	3.58	2.83	3.31
28	Confite Morocho x Enano	2.64	3.79	7.95	7.95	3.67	5.22
29	Tabloncillo Perla x Lenha	14.53	3.22	14.53	14.53	22.34	11.97
30	Dente Branco x Chapalote	19.43	3.04	13.28	13.28	10.83	13.88

* Upper bound of the standard error.

** Approximate standard error.