



UNIVERSIDAD DE LA REPUBLICA  
FACULTAD DE AGRONOMIA

COMPARACION DE LA CLASIFICACION RACIAL CON UNA CLASIFICACION  
NUMERICA EN LA COLECCIÓN DE MAIZ DEL URUGUAY

por

Lucía GUTIERREZ CHACON

TESIS presentada como uno  
de los requisitos para obtener el  
título de Ingeniero Agrónomo

MONTEVIDEO  
URUGUAY  
2001

Tesis aprobada por:

Director: \_\_\_\_\_  
JORGE FRANCO

\_\_\_\_\_  
TABARE ABADIE

\_\_\_\_\_  
CLARA PRITSCH

Fecha: \_\_\_\_\_

Autor: \_\_\_\_\_  
LUCIA GUTIERREZ

## AGRADECIMIENTOS

Al Dr. Jorge Franco, mi director, que fue mi guía en toda esta etapa, aprendí mucho y me hizo crecer.

Al Dr. Tabaré Abadie, mi orientador, por ayudarme, comprenderme y apoyarme en todo momento.

A la Dr. Clara Pritsch, siempre fue un ejemplo y ha sido una gran experiencia trabajar juntos.

A Juan Carlos y Diego por su apoyo, sobre todo logístico, en los inicios de esta investigación.

A los docentes que me acompañaron durante toda la carrera y con quienes me inicié en la investigación.

A mis amigos y compañeros que en definitiva contribuyeron a que pudiera realizar este trabajo.

A Valentín por su apoyo incondicional y su comprensión.

A mi familia que les debo todo esto.

## TABLA DE CONTENIDO

PAGINA DE APROBACION.....	II
AGRADECIMIENTOS.....	III
LISTA DE CUADROS E ILUSTRACIONES.....	IV
<b>1. INTRODUCCION.....</b>	<b>1</b>
<b>2. REVISION BIBLIOGRAFICA.....</b>	<b>3</b>
2.1 LA CLASIFICACION.....	3
2.1.1 Notación matricial.....	3
2.1.2 Análisis de la parte continua de los datos.....	4
2.1.3 Agrupamiento.....	5
2.2 CLASIFICACION DE LOS RECURSOS GENÉTICOS DE MAÍZ.....	6
2.2.1 Clasificaciones basadas en caracteres morfológicos.....	7
2.2.2 Clasificación numérica.....	7
2.2.3 Cuestionamientos.....	10
<b>3. MATERIALES Y METODOS.....</b>	<b>11</b>
3.1 DATOS EXPERIMENTALES.....	11
3.2 PROCESAMIENTO PRELIMINAR DE LA INFORMACION.....	12
3.3 CLASIFICACION NUMERICA.....	13
3.4 ANALISIS ESTADISTICO.....	14
3.5 CRITERIOS PARA COMPARAR LOS DIFERENTES AGRUPAMIENTOS.....	15
3.5.1 Importancia de las variables continuas y categóricas.....	15
3.5.2 Distancia entre grupos.....	16
3.5.3 Estadísticos multivariados.....	16
3.5.4 Calidad de los agrupamientos.....	17
3.5.5 Representación gráfica de los agrupamientos.....	17
3.5.6 Recuperación de la diversidad fenotípica.....	18
<b>4. RESULTADOS Y DISCUSION.....</b>	<b>19</b>
4.1 DESCRIPCION GENERAL DE LAS VARIABLES.....	19
4.2 ESTRUCTURA DE LAS RAZAS.....	20
4.3 CLASIFICACION NUMERICA.....	20
4.3.1 Estructura de los 7 grupos (Ward-MLM).....	20
4.3.2 Estructura de los 16 grupos (Ward-MLM).....	33
4.3.3 Selección de una clasificación numérica.....	35
4.4 COMPARACION DE LAS CLASIFICACIONES.....	35
4.4.1 Análisis univariado.....	35
4.4.2 Variables que determinan la clasificación.....	35
4.4.3 Distancias entre grupos.....	37
4.4.4 Estadísticos multivariados.....	40
4.4.5 Calidad de los agrupamientos.....	40
4.4.6 Análisis canónico.....	41
4.4.7 Recuperación de la varianza fenotípica.....	44
<b>5. CONCLUSIONES.....</b>	<b>47</b>

<b>6. <u>RESUMEN</u></b> .....	<b>48</b>
<b>7. <u>BIBLIOGRAFIA</u></b> .....	<b>50</b>

## LISTA DE CUADROS E ILUSTRACIONES

	Página
Cuadro N°	
1. Estadísticos simples para las variables utilizadas en el análisis de la Colección de Maíz Uruguaya .....	18
2. Valores medios por raza de todas las variables continuas utilizadas en el análisis de la clasificación de la Colección de Maíz del Uruguay .....	22
3. Frecuencia por raza de todas las variables categóricas utilizadas en el análisis de la clasificación de la Colección de Maíz del Uruguay .....	23
4. Relación entre los individuos pertenecientes a las razas y a los grupos (7 grupos Ward-MLM): porcentaje de observaciones de cada raza en cada grupo (número superior) y de cada grupo en cada raza (número inferior) .....	25
5. Valores medios por grupo (7 grupos Ward-MLM) de todas las variables continuas utilizadas en el análisis de la clasificación de la Colección de Maíz del Uruguay .....	26
6. Frecuencia por grupo (7 grupos Ward-MLM) de todas las variables categóricas utilizadas en el análisis de la clasificación de la Colección de Maíz del Uruguay .....	27
7. Distancias de Mahalanobis para variables continuas entre razas y 7 grupos Ward-MLM para la Colección de Maíz del Uruguay .....	28
8. Comparación de la raza Cateto Sulino (R6), el grupo 3, el grupo 6, los individuos que pertenecen simultáneamente a la raza Cateto Sulino y al grupo 3 y los individuos que pertenecen simultáneamente a la raza Cateto Sulino y al grupo 6, para algunas variables importantes .....	29

9. Comparación del promedio de las variables continuas que diferencian a los grupos 3 y 6 (7 grupos Ward-MLM) en el grupo 3, grupo 6 y todas las observaciones pertenecientes a la raza Canario de Ocho .....	31
10. Relación entre los individuos pertenecientes a las razas y a los grupos (16 grupos Ward-MLM): porcentaje de observaciones de cada raza en cada grupo (número superior) y de cada grupo en cada raza (número inferior) .....	34
11. Análisis Stepwise discriminante para razas y grupos (7 y 16 grupos Ward-MLM) de las variables continuas utilizadas en la Clasificación de la Colección de Maíz del Uruguay .....	36
12. Distancia y similitudes entre razas para todas las variables utilizadas en la clasificación de la Colección de Maíz del Uruguay .....	38
13. Distancia y similitudes entre grupos (7 grupos Ward-MLM) para todas las variables utilizadas en la clasificación de la Colección de Maíz del Uruguay .....	39
14. Estadísticos multivariados para razas y grupos (7 y 16 grupos Ward-MLM) del análisis de la Colección de Maíz del Uruguay .....	41
15. Probabilidad de pertenencia promedio a cada grupo y número de observaciones clasificadas con probabilidad menor a 0.75 y 0.5 en cada grupo de los 7 grupos Ward-MLM .....	41
16. Correlación de las variables originales con las variables canónicas para las razas y grupos (7 y 16 grupos Ward-MLM) de la Colección de Maíz del Uruguay .....	42
17. Promedio de las medias y varianzas de las variables continuas de 100 muestras estratificadas tomando el 20% de cada grupo o raza .....	46
 Figura N°	
1. Todos los individuos pertenecientes a las razas y a los grupos (7 grupos Ward-MLM) en la clasificación de la Colección de Maíz del Uruguay. a) individuos pertenecientes a las razas, b) individuos pertenecientes a los grupos (7 grupos Ward-MLM) .....	21

2. Perfil de verosimilitud desde 1 a 20 grupos, donde se muestran los saltos importantes en la versomilitud que ocurren en los 7 grupos y en los 16 grupos.....	24
3. Primeras variables canónicas para los individuos pertenecientes a la raza Catato Sulino, dentro de los grupos G3 y G6 (a), y según el origen de la colecta (b) .....	30
4. Ubicación de los centroides de las razas respecto de las dos y tres primeras variables canónicas y porcentaje de explicación de cada variable. a) centroides de las razas en las dos primeras variables canónicas, b) centroides de las razas en las tres primeras variables canónicas .....	43
5. Ubicación de los centroides de los grupos (7grupos Ward-MLM) respecto de las dos y tres primeras variables canónicas y porcentaje de explicación de cada variable. a) centroides de los grupos (7gruposWard-MLM) en las dos primeras variables canónicas, b) centroides de los grupos (7gruposWard-MLM) en las tres primeras variables canónicas.....	45



## **1. INTRODUCCION**

La diversidad genética vegetal es el componente principal de cualquier sistema de producción agrícola, incluso de cualquier ecosistema. La misma es clave ya que permite la respuesta evolutiva del sistema a cambios en las condiciones ambientales y bióticas. En el último tiempo ha ocurrido un proceso de erosión genética, se perdieron materiales genéticos, determinándose la extinción de ciertas especies y la pérdida de diversidad de otras (Guarino *et al.*, 1995). Los argumentos manejados para la conservación de los recursos genéticos son desde estrictamente utilitarios, hasta argumentos éticos y morales. Una justificación muy común para la preservación en agricultura es la necesidad de estar preparados para el quiebre de resistencias a enfermedades y plagas, ya que la uniformidad genética aumenta la susceptibilidad del cultivo (Evenson *et al.*, 1998). Mc Neieley *et al.* (1990) considera que todo el material genético tiene un valor potencial ya que hay muy pocas bases para descartar recursos genéticos "utilitarios". Su argumento se basa en que no hay un conocimiento de la tecnología ni de las condiciones ambientales futuras (Evenson *et al.*, 1998). Otros autores consideran a la diversidad de genes de una población como un recurso ya que la variación genética es un recurso para la propia sobrevivencia y evolución futura de la especie; y porque una pequeña fracción de los genes son un recurso potencial para mejorar la productividad de otras poblaciones y especies (Frankel *et al.*, 1995). Existen además justificaciones éticas y morales, propuestas por Shiva *et al.* (1991), Oldfield (1989) y Busch *et al.* (1989), citadas por Evenson *et al.* (1998). Shiva *et al.*, (1991) argumenta que "todas las formas de vida tienen un valor en sí mismo, independiente del valor que el hombre le asigne".

El material genético para la agricultura se encuentra en especies silvestres emparentadas a los cultivos, variedades locales y líneas de los mejoradores (Evenson *et al.*, 1998). La conservación de los recursos genéticos puede realizarse de dos formas: *in situ* y *ex situ*. Siempre que sea posible se debe buscar una conservación *in situ* de los recursos genéticos, ya que permite los procesos evolutivos (Glowka *et al.*, 1998). Hay situaciones en la que esto no es posible, o es complementado mediante la conservación *ex situ*. Esta última es una herramienta muy útil ya que permite la conservación tanto de especies silvestres, como de especies cultivadas. Además permite un acceso más fácil a los recursos genéticos, por lo que es fundamental en programas de mejoramiento y para la caracterización y evaluación de las colecciones. La principal forma de conservación del material genético para la agricultura en todo el mundo, es llevada a cabo en Bancos de Germoplasma, Jardines Botánicos y programas de los mejoradores, es decir: conservación *ex situ* (Evenson *et al.*, 1998). Existen Bancos de Germoplasma en 130 países, con aproximadamente 6.2 millones de accesiones, incluyendo 2 millones de accesiones de cereales (FAO 1996).

El éxito en el uso de los recursos genéticos conservados *ex situ* depende de la accesibilidad y calidad de los datos (Moss y Guarino; 1995). Un buen manejo de los datos es fundamental para determinar las condiciones apropiadas para la regeneración, caracterización y evaluación, inclusión en programas de mejoramiento, planificación de nuevas colectas, así como el manejo para evitar la erosión genética (Crossa *et al.*, 1994; Moss y Guarino; 1995). Dado que el conocimiento de todo el material existente en los bancos de germoplasma es fundamental para realizar una adecuada conservación *ex situ* de los recursos, contar con el material clasificado y ordenado es vital. Es común el uso de análisis de agrupamiento para el estudio de diversidad de las accesiones en los Bancos de Germoplasma (Perry y McIntoch, 1991; Crossa *et al.*, 1994) y formar colecciones núcleo (Brown, 1989). En América Latina se han realizado una cantidad importante de estudios sobre los recursos genéticos de maíz (*Zea mays* L.). Es una de las especies sobre las que más se ha trabajado, por lo que se cuenta con más información. Son de destacar los trabajos sobre clasificaciones de Goodman y Paterniani (1969), Franco (1998), Franco *et al.* (1997a, 1997b, 1998a, 1998b), LAMP (1992), Abadie (1998) y colecciones núcleo de Taba *et al.* (1997), Abadie *et al.* (1999), Malosetti y Abadie (2000).

La primera clasificación subespecífica del maíz fue realizada por Sturtevant (1899), que dividía a los individuos en seis grupos según el tipo de grano y luego los subdividía en base a la forma y color de grano. El criterio se basó en la persistencia del tipo de grano a través de las generaciones (Brandolini, 1970). Esta clasificación fue criticada por Anderson y Cutler (1942), y tomando en cuenta los trabajos de Kulesov (1929) sobre diversidad fenotípica y conocimiento arqueológico y genético, propusieron una clasificación considerada más natural, la clasificación racial. Se definió el concepto de raza y se describieron los caracteres morfológicos que determinan las razas, caracteres de grano y mazorca principalmente (Anderson y Cutler, 1942). Luego se comenzó a clasificar las diferentes colecciones de maíz en razas. A partir de los años 80, con el desarrollo de la computación la taxonomía numérica comienza a cobrar importancia; se comienza a utilizar métodos multivariados para la clasificación de recursos genéticos. Con el objetivo de formar colecciones núcleo se separan subgrupos homogéneos dentro de poblaciones heterogéneas, tomando en cuenta la diversidad fenotípica y utilizando técnicas multivariadas (Brown, 1989; Crossa *et al.*, 1995). En un intento por comparar la clasificación racial con una clasificación numérica, estudiar las características de los grupos y razas y entender en poco mejor la utilidad de los métodos numéricos en las clasificaciones, se plantea este estudio. Los objetivos específicos son: a) realizar clasificaciones numéricas de los Recursos Genéticos de la colección de germoplasma de maíz de Uruguay y b) comparar la clasificación numérica con la clasificación racial.

## 2. REVISION BIBLIOGRAFICA

### 2.1 LA CLASIFICACION

El problema que se enfrenta al querer realizar una clasificación es que se tienen  $n$  individuos (accesiones) a los cuales se les midieron diferentes atributos (variables) y se los quiere agrupar para juntar lo similar y separar lo diferente, o se los supone estructurados naturalmente en  $g$  grupos, que no son conocidos y quieren identificarse. Al utilizarse un número importante de variables, se requiere de análisis de tipo multivariado. La teoría de los métodos multivariados requiere de cierto conocimiento de la notación matricial (Manly, 1994).

Las variables pueden ser continuas o categóricas. Las variables continuas teóricamente pueden tomar un número infinito de valores en la recta de los reales (peso, altura). Las variables categóricas tienen un número finito de valores posibles, pueden ser binarias (presencia o ausencia de una característica), multiestado (color o forma de un órgano) u ordinales (escala visual ordenada).

#### 2.1.1 Notación matricial

La base de datos puede ser escrita como una matriz donde cada una de las hileras corresponde a un individuo (una accesión si se utiliza la nomenclatura de los bancos de germoplasma) y cada una de las columnas corresponde a una variable. De esta forma se tiene una matriz  $n \times (p+q)$ , con  $p$  variables continuas y  $q$  variables categóricas medidas en  $n$  accesiones,  $i=1, 2, \dots, n$  individuos;  $k= 1, 2, \dots, p$  variables continuas y  $l= 1, 2, \dots, q$  variables categóricas.

$$A = \begin{bmatrix} x'_1 \\ x'_2 \\ \cdot \\ \cdot \\ x'_n \end{bmatrix}_{n \times (p+q)} = \begin{bmatrix} y_{11} & y_{12} & \cdot & \cdot & \cdot & y_{1p} & , & Q_{11} & Q_{12} & \cdot & \cdot & \cdot & Q_{1q} \\ y_{21} & y_{22} & \cdot & \cdot & \cdot & y_{2p} & , & Q_{21} & Q_{22} & \cdot & \cdot & \cdot & Q_{2q} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & , & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & , & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ y_{n1} & y_{n2} & \cdot & \cdot & \cdot & y_{np} & , & Q_{n1} & Q_{n2} & \cdot & \cdot & \cdot & Q_{nq} \end{bmatrix}$$

Lawrence y Krzanowski (1996) utilizaron un modelo (Location Model: ML, propuesto por Olkin y Tate, 1961) aplicable al problema de la clasificación utilizando mezcla de variables continuas y categóricas. Reúne toda la información de las variables

categorías en una única variable multinomial  $W$ . La variable  $W$  se obtiene al reemplazar las  $q$  variables categóricas por una única variable, que puede tomar  $m$  valores ( $s = 1, 2, \dots, m$ ), correspondientes a la combinación factorial de cada nivel de cada una de las variables categóricas. Esto significa que si se contara con dos variables,  $A$  y  $B$ , cada una con dos estados posibles, 1 y 2, las combinaciones son:  $A_1B_1, A_1B_2, A_2B_1$  y  $A_2B_2$ . Cada una de estas combinaciones se representa con un valor diferente, por lo que se encuentran  $m=4$  valores para la variable  $W$  ( $s = 1, 2, 3, 4$ ). Se asume que la variable  $W$  se distribuye Multinomial con parámetros  $(p_1, p_2, \dots, p_m)$ . Se genera así una nueva matriz:  $X$ , de  $n$  hileras (accesiones) y  $p+1$  columnas ( $p$  variables continuas + variable  $W$ ).

$$X = \begin{bmatrix} x'_1 \\ x'_2 \\ \vdots \\ x'_n \end{bmatrix}_{p+1} = \begin{bmatrix} y'_1 & s_1 \\ y'_2 & s_2 \\ \vdots & \vdots \\ y'_n & s_n \end{bmatrix} = \begin{bmatrix} Y_{11} & Y_{12} & \dots & Y_{1p} & S_1 \\ Y_{21} & Y_{22} & \dots & Y_{2p} & S_2 \\ \vdots & \vdots & \dots & \vdots & \vdots \\ Y_{n1} & Y_{n2} & \dots & Y_{np} & S_n \end{bmatrix}$$

En la notación matricial utilizada en la literatura cuando se asume que las  $n$  accesiones se encuentran estructuradas en  $g$  grupos, cada valor en la matriz puede ser representado por  $x_{ijks}$ ,  $i = 1, 2, \dots, g$  grupos;  $j = 1, 2, \dots, n_i$  accesiones por grupo;  $k = 1, 2, \dots, p$  variables continuas;  $s = 1, 2, \dots, m$  estados de la variable  $W$ . De esta forma, el vector de la  $j$ -ésima observación en el  $i$ -ésimo grupo es un vector de tamaño  $1 \times (p + 1)$ , conteniendo los valores de cada una de las  $p$  variables continuas y el único valor  $s$  de la variable  $W$ . El vector de medias de las variables continuas del  $i$ -ésimo grupo y media general es  $\bar{y}_i \bar{y}$ , respectivamente (cada uno de tamaño  $p \times 1$ ) (Franco *et al.*, 1997b).

### 2.1.2 Análisis de la parte continua de los datos

El análisis de la variabilidad de los datos multivariados continuos está dado por las matrices de suma de cuadrados y productos cruzados (de tamaño  $p \times p$ ), correspondiente a la variabilidad total,  $T$ , la variabilidad dentro de grupos,  $W$ , y la variabilidad entre grupos,  $B$  :

$$T = \sum_{i=1}^g \sum_{j=1}^{n_i} (y_{ij} - \bar{y}) (y_{ij} - \bar{y})'$$

$$W = \sum_{i=1}^g \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i) (y_{ij} - \bar{y}_i)'$$

$$B = \sum_{i=1}^g n_i (\bar{y}_i - \bar{y}) (\bar{y}_i - \bar{y})'$$

Las diagonales de  $\mathbf{T}$ ,  $\mathbf{W}$ , y  $\mathbf{B}$  contienen las sumas de cuadrados total, dentro de grupos (SSW), y entre grupos (SSB), respectivamente. La traza de  $\mathbf{W}$ ,  $\text{tr}(\mathbf{W})$  (suma de los términos en la diagonal de  $\mathbf{W}$ ), puede ser escrita como la suma de las distancias Euclidianas cuadradas entre el vector de observaciones,  $\mathbf{y}_{ij}$ , y el vector de medias de cada grupo,  $\bar{\mathbf{y}}_i$  (Kaufman y Rousseeuw, 1990)

$$\text{tr}(\mathbf{W}) = \sum_{i=1}^g \sum_{j=1}^{n_i} \left\| \mathbf{y}_{ij} - \bar{\mathbf{y}}_i \right\|^2$$

De igual forma, la traza de la matriz  $\mathbf{B}$ ,  $\text{tr}(\mathbf{B})$ , expresa la suma de las distancias Euclidianas cuadradas entre el vector de medias de cada grupo,  $\bar{\mathbf{y}}_i$ , (centroide) y el vector de medias general,  $\bar{\mathbf{y}}$  (Kaufman y Rousseeuw, 1990)

$$\text{tr}(\mathbf{B}) = \sum_{i=1}^g n_i \left\| \bar{\mathbf{y}}_i - \bar{\mathbf{y}} \right\|^2$$

El estimador de máxima verosimilitud de la matriz de varianzas-covarianzas dentro de grupos es  $\mathbf{S} = (1/n)\mathbf{W}$ . Se pueden definir dos tipos de varianza a partir de  $\mathbf{S}$  para los datos multivariados (Mardia *et al.*, 1979). Una es la varianza total, definida como la traza de  $\mathbf{S}$ ,  $\text{tr}(\mathbf{S})$ . La otra es la varianza generalizada, definida como el determinante de  $\mathbf{S}$ ,  $\det(\mathbf{S})$ . Altos valores de ambas indican que las observaciones se encuentran muy dispersas alrededor de los vectores de medias. Sin embargo, mientras que la  $\text{tr}(\mathbf{S})$  considera solamente las varianzas, el  $\det(\mathbf{S})$  considera varianzas y covarianzas (Franco *et al.*, 1997b).

### 2.1.3 Agrupamiento

Una clasificación pretende estructurar las observaciones en grupos, de tal forma que los grupos queden con individuos "similares" dentro y que sean más diferentes entre sí. Para esto es necesario encontrar alguna forma de determinar cuán lejos se encuentra una observación de otra. Con este fin se han desarrollado muchas medidas de distancia, tanto para variables continuas como categóricas y para la mezcla de variables.

Definida una medida de distancia, la siguiente etapa consiste en escoger la estrategia a seguir para realizar el agrupamiento. Los métodos de agrupamiento se pueden separar en dos tipos : 1) aquellos que se basan en la configuración geométrica de las observaciones y 2) aquellos que suponen un modelo estadístico y propiedades distribucionales. Estos últimos permiten la incorporación de conceptos probabilísticos en la clasificación.

Las técnicas geométricas pueden dividirse en tres grupos no excluyentes: 1) jerárquicas aglomerativas, 2) de búsqueda de densidades y 3) de partición y optimización de una función objetivo. Las técnicas jerárquicas aglomerativas parten de  $n$  grupos, donde cada grupo corresponde a un individuo. Luego se fusionan los grupos más cercanos cada vez (con menor distancia), hasta que todos los individuos quedan en un solo grupo. Son procesos ordenados y jerárquicos, de forma que los agrupamientos iniciales no son modificados por los posteriores y el resultado puede ser representado en una gráfica de líneas llamada dendrograma. Existen varios métodos de agrupamiento jerárquico que se diferencian en como calculan las distancias entre los grupos, por ejemplo UPGMA, Centroide y Mediana. Las técnicas de búsqueda de densidades tienen como objetivo la búsqueda de regiones con alta densidad de puntos, separadas por otras de baja densidad o nula. Un método de este tipo es el Density propuesto por Wishart (1969b). Las técnicas de partición y optimización de una función objetivo supone que el número de grupos es conocido o se estima al inicio del proceso. Luego asigna cada observación a un grupo y después efectúa reclasificaciones en un proceso iterativo optimizando alguna función objetivo. Se incluyen en este tipo las pruebas tipo  $g$ -medioides (Kaufman y Rosseeuw, 1990) y las que se basan en las propiedades del análisis de varianza multivariado.

El método de Ward (1963) fue la primer estrategia que planteó explícitamente la optimización de una función objetivo (la suma de cuadrados dentro de grupos), pero manteniendo la característica de ser jerárquico. Esto le da dos ventajas muy importantes: a) al ser jerárquica puede estimar el número de grupos en el proceso de clasificación y b) la función objetivo elegida está relacionada con el análisis de varianza. Recientemente se han desarrollado técnicas de agrupamiento que permiten combinar métodos jerárquicos con métodos estadísticos (Franco *et al.*, 1998a).

## 2.2 CLASIFICACION DE LOS RECURSOS GENÉTICOS DE MAÍZ

El maíz (*Zea mays* L.) es el tercer cultivo alimenticio más importante en el mundo (FAO, 1996; Fassio *et al.*, 2000). Las principales zonas productoras de maíz son Estados Unidos, Europa, China, Brasil, México, Sudáfrica, Argentina, India e Indonesia (Goodman, 1976). Se identificaron ancestros de maíz en el sudoeste de Estados Unidos y México que datan de 60.000 a 80.000 años y existe evidencia de domesticación en Perú de 1.000 A.C. Los centros de diversidad más importantes son México y las planicies de América Central, aunque existen otros centros de diversidad de menor importancia en América Latina como ser Perú (Goodman, 1976). El destino principal del maíz a nivel mundial es para raciones balanceadas, aunque una pequeña porción es utilizada como consumo humano directo (Goodman, 1976; Fassio, *et al.*, 2000). A pesar de esto, el maíz es la base del aporte energético de la dieta de muchos de los países de América Latina y Africa, incluidos Bolivia, Brasil, Costa Rica, Cuba, Ecuador, El Salvador, Guatemala,

Honduras, Jamaica, México, Nicaragua, Panamá, Perú, República Dominicana y Venezuela (FAO, 1996).

Se encuentran en la literatura un número importante de clasificaciones de los recursos genéticos de maíz. Algunos trabajos importantes son los trabajos de Kulesov (1929, 1933), Anderson y Cutler (1942), Wellhausen *et al.* (1952), Brieger *et al.* (1958), Goodman (1967, 1968, 1972), Goodman y Paterniani (1969), Paterniani (1970, 1972), Goodman y Bird (1977), Bird y Goodman (1977), Doebley *et al.* (1985), Basford y McLachlan (1985), Brown (1989), Bretting *et al.* (1990), Pineda y Crossa (1994), Crossa *et al.* (1994, 1995), Taba *et al.* (1997), Franco *et al.* (1997a, 1997b, 1998a, 1998b). Se realiza para este estudio un agrupamiento arbitrario de las metodologías, separando aquellas que utilizan algunos caracteres morfológicos y aquellas que utilizan toda la información disponible para realizar los agrupamientos.

### 2.2.1 Clasificaciones basadas en caracteres morfológicos

La clasificación de Kuleshov (1933), toma en cuenta el tipo de endosperma del grano para realizar los grupos. Los ocho grupos resultantes son: "flint", "floury", "dent", "popcorn", "sweet", "starchy-sugar", "waxy" y "pod". Esta clasificación tiene su origen en la hipótesis que los diferentes tipos de maíz se deben a una domesticación independiente por regiones.

Posteriormente a la clasificación de Kulesov, se desarrolló el concepto de raza. Una raza de maíz es definida por Anderson y Cutler (1942), citado por Aguirre Gómez *et al.* (2000), como "un conjunto de plantas de maíz con características en común que le permite ser reconocida como un grupo". Desde el punto de vista genético una raza es "un grupo de individuos con un número significativo de genes en común". La clasificación racial se basa en caracteres de la mazorca (Aguirre Gómez *et al.*, 2000).

### 2.2.2 Clasificación numérica

La técnica propuesta por Ward (1963) fue la primera que planteó explícitamente la optimización de una función objetivo, pero manteniendo la característica de ser jerárquico. La función objetivo que optimiza es la suma de las distancias Euclidianas cuadradas entre cada observación y la media de su grupo ( $\text{tr}(\mathbf{W})$ : suma de cuadrados dentro de grupos). El método minimiza dentro de cada grupo la distancia de cada observación hacia su centroide. Por lo tanto maximiza la distancia entre los centroides de los grupos y los vectores de medias generales ( $\text{tr}(\mathbf{T}) = \text{tr}(\mathbf{W}) + \text{tr}(\mathbf{B})$ ). Posteriormente, Gower (1971) propuso una medida para mezcla de variables continuas y categóricas y demostró que cumple con las propiedades de una medida de distancia métrica Euclidea.

Wishart (1986) desarrolló la técnica de cómputo para la utilización de la distancia de Gower y lo propuso para su uso en métodos de agrupamiento jerárquico como Ward.

El primer modelo estadístico es el Modelo de Localización (LM) propuesto por Olkin y Tate (1961). Lawrence y Krzanowski (1996) lo proponen como un modelo aplicable al problema de clasificaciones utilizando mezcla de variables continuas y discretas. La idea fundamental de este modelo es que reúne toda la información de las variables categóricas en una única variable  $W$ .

Franco *et al.* (1998a) desarrolló una estrategia de clasificación en dos etapas (Ward-MLM). En primera instancia son definidos grupos iniciales por métodos jerárquicos como el Ward, donde se incluyen variables continuas y categóricas. La segunda etapa consiste en la mejora de estos grupos por medio de un método no jerárquico. El uso de toda la información (variables continuas y discretas) produce mejores agrupamientos que aquella con solo parte de la información. Esta estrategia maximiza las distancias entre grupos, obteniéndose grupos mejor diferenciados y más compactos (Franco *et al.*, 1998a).

El propósito del método Ward (1963) es agrupar en cada paso los grupos que producen el mínimo incremento en  $SSW$ . Al construir grupos con mínima varianza, los grupos generados tienden a ser esféricos aunque esta no sea su estructura natural (Franco, 1998). El método MLM mejora los grupos formados por el Ward, asignando cada observación al grupo para el cual presenta una probabilidad de pertenencia máxima. El MLM es una modificación del ML (Lawrence y Krzanowski, 1996), que se diferencia en que considera que la media del vector de variables continuas es independiente del valor multinomial asociado (valor de la variable  $W$ ). El MLM realiza la estimación de la media y varianza dentro de la subpoblación y no por celda, de forma que maximiza la verosimilitud respecto a los centroides o valores promedios de los grupos y no de las celdas. Por lo tanto forma grupos más homogéneos (Franco *et al.*, 1998a). Para la estimación de los parámetros utiliza el algoritmo EM propuesto por Dempster, Laird y Rubin (1977) y un programa escrito en el procedimiento IML del paquete estadístico SAS (1990). El algoritmo Esperanza-Maximización fue propuesto para la estimación por máxima verosimilitud en datos incompletos. Cuando se está clasificando se supone que la matriz de datos es incompleta porque no se conoce a que grupo pertenece cada observación. Se realiza un proceso iterativo que inicia en un punto de arranque, que puede ser un conjunto de valores  $\mu^{(0)}$  y  $\Sigma^{(0)}$  para los vectores de medias y las matrices de varianzas y covarianzas, o algunos valores  $\tau_{ijs}^{(0)}$  para las probabilidades de pertenencia de cada observación a cada grupo. En el siguiente paso se calculan las probabilidades de pertenencia a cada grupo a posteriori (si se inició con valores para las medias y varianzas-covarianzas) o las medias y varianzas-covarianzas (si se inició con valores para las probabilidades). El proceso continúa estimando unos y otros alternadamente, hasta lograr



la convergencia del logaritmo de la función de verosimilitud o de las probabilidades a posteriori (Franco *et al.*, 1998a).

La función a maximizar es el logaritmo de la verosimilitud de las observaciones:

$$l(\Theta; x, z) = \sum_{i=1}^g \sum_{s=1}^m \sum_{j=1}^{n_s} z_{ijs} \left\{ \log(\alpha_i) + \log(p_{is}) - \frac{p}{2} \log(2\Pi) - \frac{1}{2} \log|\Sigma_W| - \frac{1}{2} (y_{sj} - \mu_i)' \Sigma_W^{-1} (y_{sj} - \mu_i) \right\}$$

en donde:

$z_{ijs} = 1$  si  $x_{js}$  pertenece a la  $i$ -ésima población

$= 0$  si  $x_{js}$  no pertenece a la  $i$ -ésima población

$\alpha_i$  = proporción de observaciones de la  $i$ -ésima población

$p_{is}$  = proporción de observaciones de la  $i$ -ésima población con el  $s$ -ésimo valor de la variable  $W$

$\Sigma_W$  = matriz de varianzas y covarianzas "dentro" de poblaciones

$\mu_i$  = vector de promedios de las variables continuas en la  $i$ -ésima población

$y_{sj}$  = parte continua de la observación  $x_{sj}$

la función crece si: a) la proporción de observaciones asignadas a cada población crece, b) la proporción de observaciones con el  $s$ -ésimo valor de la variable  $W$  en cada población crece, c) la matriz de varianzas y covarianzas "dentro" de poblaciones disminuye y d) la diferencia entre cada observación hacia su centroide disminuye.

La estrategia de clasificación propuesta por Franco *et al.* (1998a) tiene una serie de aspectos positivos a recalcar: 1) el proceso obedece a la optimización de dos funciones objetivo relacionadas, en primera etapa la suma de cuadrados dentro de grupo, en segunda etapa la función de verosimilitud de las observaciones; 2) está asociada a una técnica para la definición del número óptimo de grupos; 3) permite el cálculo de una medida de la precisión (calidad) de los agrupamientos; 4) asigna las observaciones a los grupos con base en la probabilidad de pertenencia de cada observación a cada grupo; y 5) utiliza toda la información disponible, tanto de las variables continuas como de las discretas.

### 2.2.3 Cuestionamientos

Tanto la clasificación racial como los métodos estadísticos tradicionales han sido cuestionados. La clasificación racial es cuestionada ya que es satisfactoria para el tipo de grano pero no para diferencias en otros caracteres morfológicos. En algunos casos alcanza con solamente un gen de diferencia para que sean separados en diferentes grupos. Además el tipo de caracteres utilizados presenta gran influencia ambiental (Hallauer y Miranda, 1988).

A su vez, el análisis cluster tiene cuestionamientos en cuanto a cómo utilizar diferentes tipos de variables (continuas, discretas, ordinales, multiestado, binomiales), separar las variables más discriminantes, qué medida de distancia se debe usar, qué tipo de estrategia cluster usar y cuál es el número óptimo de grupos. Los métodos de agrupamiento tradicionales como Ward, UPGMA, centroide, etc., son solamente descriptivos y no calculan la probabilidad de pertenencia al grupo (Franco *et al.*, 1997a). Otro problema adicional es cuando la estrategia de agrupamiento no encuentra grupos naturales o impone a priori una determinada distribución de los datos (Everitt, 1974).

### **3. MATERIALES Y METODOS**

#### **3.1 DATOS EXPERIMENTALES**

Para el presente estudio se utilizaron los datos de la Colección de Maíz del Uruguay, cuya base de datos fue generada por De María *et al.* (1979). Las poblaciones de maíz fueron recolectadas directamente en predios de agricultores durante los meses de marzo a junio de 1978, en el mismo momento o inmediatamente después de la cosecha. Esta colecta estaba a cargo del Ing. Agr. José L. De León y era parte de un proyecto conjunto entre Facultad de Agronomía y el International Board of Plant Genetic Resource (IBPGR). Se recolectaron aquellos materiales que han sido cultivados por los propios agricultores por un período largo de tiempo. Luego de la colecta, las accesiones fueron caracterizadas y evaluadas por De María *et al.* (1979) durante la estación de crecimiento 1978-1979 en la Estación Experimental 'Dr. A. Backhaus' de la Facultad de Agronomía (Universidad de la República), Montevideo, Uruguay (Lat. 34° 50' S, Long. 56° 10' W). Las accesiones se sembraron el 3 de Noviembre de 1978, en parcelas de 6 metros de largo con una distancia entre hileras de 0.70 m, sin repeticiones. Se realizó fertilización a la siembra con 40 Kg. P<sub>2</sub>O<sub>5</sub>.ha<sup>-1</sup> y 46 Kg. N.ha<sup>-1</sup>. Luego de la emergencia, se llevó la densidad de plantas a 47.620 plantas por hectárea (De María *et al.*, 1979). La temperatura normal durante la estación de crecimiento se encuentra entre 18°C y 22°C, y la precipitación mensual alrededor de 90 mm, con ocasionales déficit hídricos en Enero. Las condiciones de la estación de crecimiento 1978-1979 no se diferenciaron significativamente de los promedios históricos del sitio (Malosetti *et al.*, 2000).

Los procedimientos y metodología utilizada en la evaluación de los caracteres se encuentran dentro de la guía desarrollada por el programa 'Colecciones, Conservación, y Evaluación de Germoplasma de Maíz de la Región Oriental de América del Sur' del IBPGR, desarrollado en Sete Lagoas, Brasil en 1977 (De María *et al.*, 1979). Los caracteres evaluados se dividieron en dos grupos: a) caracteres primarios como rendimiento en grano, rendimiento de forraje, vuelco, y b) caracteres accesorios del ciclo, altura de planta y mazorca, número de macollos y mazorcas por planta y caracteres descriptivos de grano y mazorca. La base de datos obtenida fue publicada en el Catálogo de Recursos Genéticos de Maíz de Sudamérica-Uruguay (Fernández *et al.*, 1983). Esta información incluye datos de pasaporte (sitios de colecta y donantes), clasificación racial, y datos de evaluación y caracterización.

Cada accesión de la base de datos generada por De María *et al.* (1979) fue clasificada por el Ing. Agr. José L. De León en una raza de maíz. Las razas presentes en la colección y su tipo de grano son: Dente Branco (dentado-semidentado), Morotí (harinosos), Dente Riograndense (duro-semiduro), Semidentado Riograndense (duro-semiduro), Canario de Ocho (duro-semiduro), Cateto Sulino (duro-semiduro),

Cuarentino (duro-semiduro), Cateto Sulino Grosso (duro-semiduro), Pisingallo (reventador) y Cristal (duro-semiduro).

### **3.2 PROCESAMIENTO PRELIMINAR DE LA INFORMACION**

Para estudiar el comportamiento de las variables se realizó un análisis de distribución de frecuencias para todas ellas (27), incluyendo categóricas y continuas. Para ello se utilizó el procedimiento PROC UNIVARIATE del paquete estadístico SAS (1996). Luego se seleccionaron 17 variables continuas y 5 variables categóricas. Las variables continuas seleccionadas para la clasificación numérica son:

- 1) floración masculina (FM): número de días transcurridos entre la siembra y el momento en que se ha iniciado la antesis de las flores masculinas en el 50% de las plantas de la parcela;
- 2) floración femenina (FF): número de días transcurridos entre la siembra y el momento en que se ha iniciado la emergencia de los estigmas en el 50% de las plantas de la parcela;
- 3) altura de planta (ALTPLA): promedio de la altura (cm) de 10 plantas sucesivas y competitivas medido desde el nivel del suelo hasta el nudo donde nace la hoja superior y el pedúnculo de la panoja;
- 4) altura de mazorca (ALTMAZ): promedio de la altura (cm) de inserción de la mazorca superior de las mismas 10 plantas medido en el tallo principal desde el nivel del suelo hasta donde nace el nudo de la mazorca superior;
- 5) índice de prolificidad (IPRO): promedio por planta del número de mazorcas producidas por las mismas 10 plantas;
- 6) índice de macollaje (IMAC): promedio por planta del número de tallos contados en las mismas 10 plantas;
- 7) vuelco (VUELCO): número de plantas volcadas desde la raíz contadas en las mismas 10 plantas, multiplicado por 10 para llevarlo a porcentaje;
- 8) relación espiga forraje (ESPFORR): es el cociente entre la materia seca de la espiga y la materia seca total;
- 9) forraje residual (FORRES): es la producción de materia seca total (tt) por hectárea;
- 10) longitud de mazorca (LONGMAZ): promedio de la longitud (cm) de la mazorca superior, medida desde la base del ápice;
- 11) espesor de grano (ESPGRA): promedio del espesor (mm) de 10 granos consecutivos de la parte media de una hilera normal de mazorca superior;
- 12) longitud de grano (LONGRA): promedio de la longitud (mm) de 10 granos escogidos de la parte media de una hilera normal de mazorca superior;
- 13) ancho de grano (ANCHOGRA): promedio del ancho (mm) de los mismos 10 granos;

- 14) peso de 100 granos (P100GRA): promedio del peso (g) de 100 granos secos y sanos tomados al azar de cada una de las 10 mazorcas estudiadas;
- 15) peso de grano por mazorca (PESOGRA): peso total de los granos (g) de las 10 mazorcas estudiadas;
- 16) peso de la mazorca (MAZENT): peso total (g) de las 10 mazorcas evaluadas;
- 17) porcentaje del peso de grano en el total de la mazorca (PORGRA): es el cociente entre el peso de los granos provenientes de las 10 mazorcas evaluadas y el peso total de las mazorcas, expresado como porcentaje.

Las variables categóricas estudiadas fueron: forma de grano (FORGRA), textura de grano (TEXGRA), color de perisperma (COLORPER), color de aleurona (COLORALE) y color de endosperma (COLORENDO).

Las variables diámetro de mazorca y número de hileras por mazorca son variables continuas, que se transformaron a discretas. Diámetro de mazorca (DIAMAZ) se midió como el promedio del diámetro (mm) de la mazorca superior en el punto medio de su longitud y número de hileras por mazorca (NUMHIL) se midió como el promedio del número de hileras de granos de la mazorca superior contando en el centro de la mazorca.

Se eliminaron variables para realizar los análisis, estas fueron: rendimiento en grano, forma de mazorca y color de marlo. Las dos últimas se eliminaron porque presentan un 94 y 98% de sus individuos en una única categoría, respectivamente. Rendimiento en grano se eliminó por ser una característica con heredabilidad baja. Se utilizaron para los análisis 840 de las 852 accesiones existentes. Se eliminaron 12 accesiones por presentar datos faltantes en variables continuas.

### 3.3 CLASIFICACION NUMERICA

La base de datos generado por De María *et al.* (1979) contaba con una clasificación de los datos, la clasificación racial. A esta base se le agregó la clasificación realizada con la estrategia Ward-MLM propuesta por Franco *et al.* (1998a).

El método de Ward es capaz de aportar la estructura en grupos necesaria para el MLM, pero debe definirse cuantos grupos hay en la población. En este estudio el número de grupos se determinó secuencialmente. Se utilizó en primera instancia el Índice de Mojena (1977) y luego se construyó el perfil de verosimilitud asociado a la prueba de razón de verosimilitud (Mardia *et al.*, 1979). Ambos se encuentran entre los 12 mejores criterios de 30 estudiados por Milligan y Cooper (1985).

El método de razón de verosimilitud compara la hipótesis nula de que el número óptimo de grupos es  $g$  contra la hipótesis alterna que es  $g'$  (con  $g < g'$ ). Calcula la razón de verosimilitud como  $\lambda = L_g / L_{g'}$ , siendo  $L_g$  y  $L_{g'}$  la verosimilitud para  $g$  y  $g'$  grupos respectivamente. Wolfe (1971) encontró que bajo esta hipótesis nula el estadístico  $X^2$  se comporta con una distribución chi-cuadrado con  $f$  grados de libertad; esto es,

$$X^2 = - (2/n) (n - 1 - p - 1/2g) \log \lambda \sim \chi^2_f,$$

$f = 2p(g' - g)$ . Por lo tanto, si  $X^2$  es mayor a un valor chi-cuadrado con determinado nivel de significancia, la hipótesis nula es descartada. Esto significa que no conviene aumentar el número de grupos. Binder (1978) mostró que  $X^2$  no se distribuye asintóticamente como chi-cuadrado, por lo tanto el nivel de significancia no es correcto. El perfil de verosimilitud utiliza los criterios de razón de verosimilitud en una interpretación gráfica y fue utilizado en este estudio (Franco *et al.*, 1997a).

El Índice de Mojena (1977) supone que los valores obtenidos luego del proceso de fusión son una muestra de tamaño  $n-1$  de una población que tiene distribución normal, con media y desviación estándar  $\bar{z}$  y  $S_z$ . A partir de este supuesto, se ordenan todos los valores del proceso de fusión y se estandarizan. Se obtiene así el valor  $k_j = (z_j - \bar{z}) / S_z$  para el  $j$ -ésimo valor de fusión. Luego se busca en una tabla el valor de fusión que deja delante de sí, en la distribución supuesta, una probabilidad muy baja ( $\alpha$ ), llamado  $k_\alpha$ . Se selecciona como número más probable de grupos aquel que presenta un valor  $k$  cercano a 1.27 ( $\alpha = 0.10$ ), como recomiendan Milligan *et al.* (1983).

### 3.4 ANALISIS ESTADISTICO

Con el objetivo de estudiar el comportamiento de cada una de las variables continuas e identificar aquellas que presentan diferencias significativas para los grupos o razas se realizó un análisis de varianza univariado. Se probó la hipótesis que los promedios de los grupos o razas son iguales para una variable continua dada, contra la hipótesis de que existe al menos una diferencia estadística entre las medias de los grupos o razas para esa variable. Para ello se utilizó cada variable continua como variable dependiente y a los diferentes grupos o razas como tratamientos. Además se calcularon estadísticos simples para todas las variables continuas: media, desvío, máximo y mínimo.

El tratamiento que se les dio a las variables categóricas con el mismo objetivo fue el cálculo de frecuencias, que luego se estudiaron a través de una prueba de independencia con el estadístico chi-cuadrado. Se colocan los datos en una tabla de doble entrada donde una de las entradas son los grupos o las razas y la otra los estados de la variable discreta probada. Luego se calcula cuanto se aleja cada celda de lo esperado en una distribución independiente, esta medida es el chi-cuadrado. Un valor alto indica que los grupos o razas no son independientes de la variable probada, o lo que es lo mismo que

algunos grupos se asocian claramente con un estado de la variable y otros grupos con otro estado.

### 3.5 CRITERIOS PARA COMPARAR LOS DIFERENTES AGRUPAMIENTOS

Una vez obtenida la clasificación numérica, se utilizaron diferentes metodologías estadísticas para compararla con la clasificación racial existente. Se utilizaron como criterios para comparar las clasificaciones: 1) el número y la importancia de cada variable en cada una de las clasificaciones, 2) la distancia entre los grupos, 3) algunos estadísticos multivariados, 4) el promedio de las probabilidades de pertenencia a cada grupo, 5) análisis canónico y 6) la recuperación de la varianza fenotípica en un proceso de muestreo posterior al agrupamiento.

#### 3.5.1 Importancia de las variables continuas y categóricas

Se quiso determinar cuáles eran las variables que discriminan más a los grupos o razas, es decir cuales son las variables más importantes en diferenciar a los grupos o a las razas. Para ello se utilizó una aproximación “stepwise discriminante” (Klecka, 1980). El procedimiento es iniciado al considerar el siguiente modelo lineal

$$y_{ij} = \mu_i + \varepsilon_{ij},$$

siendo  $y_{ij}$  el valor del  $j$ -ésimo individuo en el  $i$ -ésimo grupo ( $i = 1, 2, \dots, g$  grupos y  $j = 1, 2, \dots, n_i$  observaciones por grupo),  $\mu_i$  la media del  $i$ -ésimo grupo, y  $\varepsilon_{ij}$  el desvío del valor del  $j$ -ésimo individuo sobre la media del  $i$ -ésimo grupo (SAS, 1996).

Se realiza un análisis de varianza utilizando cada variable continua como variable dependiente y se selecciona la variable con el valor de F más alto. En el segundo paso la variable seleccionada ( $X_1$ ) es utilizada como covariable en el modelo

$$y_{ij} = \mu_i + \beta X_{1ij} + \varepsilon_{ij}.$$

Se selecciona una segunda variable y el proceso continúa con dos, tres, etc. covariables hasta que se identifican las variables principales. Se utilizó el procedimiento PROC STEPDISC del paquete estadístico SAS (1996) con el método de selección FORWARD y un nivel de significancia del 15%.

La importancia de las variables categóricas es determinada a través del estadístico  $\chi^2$ , usado en una prueba de independencia entre cada variable categórica y cada agrupamiento. Para ello se genera una tabla de doble entrada, donde se coloca como columna los estados de la variable categórica y como fila cada uno de los grupos o razas. En el cuerpo de la tabla quedan las frecuencias de cada grupo en cada estado de la variable. Para el análisis se utilizó el procedimiento PROC FREQ del paquete estadístico SAS (1996)

### 3.5.2 Distancia entre grupos

La mejor estrategia de agrupamiento debe producir la mayor distancia entre las medias de los grupos, considerando además su dispersión interna; esto es grupos más separados en relación a las variables continuas (Franco *et al.*, 1997a). Se utilizó como criterio para determinar que agrupamiento es mejor, respecto a las variables continuas, la distancia de Mahalanobis (1930). La distancia entre los vectores de medias  $\mu_1$  y  $\mu_2$  de dos poblaciones multinormales, con una matriz de varianza-covarianza  $\Sigma$  es definida como

$\Delta^2 = (\mu_1 - \mu_2)' \Sigma^{-1} (\mu_1 - \mu_2)$ , siendo su estimador muestral

$$\begin{aligned} D^2 &= (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)' S^{-1} (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2) \\ &= n (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)' W^{-1} (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2) \end{aligned}$$

donde  $S$  y  $\bar{\mathbf{x}}_i$  son los estimadores de máxima verosimilitud de  $\Sigma$  y  $\mu_i$ , respectivamente (Franco *et al.*, 1997b).

Krzanowski (1983) presenta medidas de afinidad y distancia entre grupos para el modelo LM. Demuestra que con matrices de varianza-covarianza homogéneas, el coeficiente de afinidad entre dos grupos  $i$  y  $j$ ,  $p_{ij}$ , tiene dos componentes: 1) la afinidad sobre las variables continuas,  $I_{ij}$ , asociada a la distancia  $D^2$  de Mahalanobis (1930),  $I_{ij} = \exp [ -D_{ij}^2/8 ]$ , y 2) la raíz cuadrada del producto de las frecuencias relativas sobre las variables categóricas. El coeficiente es

$$p_{ij} = \sum_s [ (p_{is} p_{js})^{1/2} I_{ij}^{(s)} ],$$

donde  $p_{is}$  y  $p_{js}$  son la proporción de casos del  $s$ -ésimo valor en los grupos  $i$  y  $j$ , respectivamente, y  $I_{ij}^{(s)}$  la afinidad sobre las variables continuas entre los grupos para la celda  $s$ . Esta medida es apropiada para el LM pero no para el MLM, para el cual la no-condicionalidad de las variables continuas y categóricas produce un coeficiente de afinidad  $p_{ij} = \sum [ (p_{is} p_{js})^{1/2} I_{ij} ] = I_{ij} \sum (p_{is} p_{js})^{1/2}$ , donde la afinidad sobre las variables continuas es independiente de la  $s$ -celda, y la medida es el producto de la afinidad sobre las variables continuas por la afinidad sobre las variables categóricas. La medida de distancia adoptada por Kranowski (1983) es la propuesta por Matusita (1959),  $\Delta_{ij} = [2(1 - p_{ij})]^{1/2}$  (Franco *et al.*, 1998a).

### 3.5.3 Estadísticos multivariados

Los estadísticos multivariados usados corrientemente para probar la diferencia entre vectores de medias de  $g$  poblaciones son: lambda de Wilks,  $\Lambda = |W| / |W + B|$ ; traza de Pillai,  $V = \text{tr} [ B(B + W)^{-1} ]$ ; traza de Hotelling-Lawley,  $U = \text{tr}(W^{-1}B)$ ; y máxima raíz característica de Roy,  $\max \{ \lambda_i \} = \lambda_1$  (máxima raíz característica de la matriz  $W^{-1}B$ ). Menores valores de  $\Lambda$  y mayores valores de  $V$ ,  $U$ , y  $\lambda_1$  indican una diferencia mayor entre las medias de los grupos (Franco *et al.*, 1997b).



### 3.5.4 Calidad de los agrupamientos

El modelo MLM permite el cálculo de probabilidades de pertenencia de cada individuo a cada grupo. Se pueden comparar dos estrategias de agrupamiento por el promedio de las probabilidades de pertenencia, que es una medida de la calidad del agrupamiento. El promedio de las probabilidades de pertenencia a cada grupo se calcula como:

$$T = \frac{1}{n} \sum_{j=1}^n \max_{(j)} \left( \hat{\tau}_{ij} \right)$$

siendo  $\tau_{ij}$  la probabilidad de pertenencia a cada grupo. Valores mayores de T indican un mejor agrupamiento.

### 3.5.5 Representación gráfica de los agrupamientos

El análisis canónico del conjunto de datos estructurados en g grupos permite mostrar las diferencias entre grupos en gráficos de dos o tres dimensiones (Mardia *et al.*, 1979). Las n observaciones pueden ser representadas en el espacio de  $t < p$  dimensiones sin grandes distorsiones de las verdaderas relaciones entre y dentro de grupos. Las nuevas coordenadas, llamadas variables canónicas, son una combinación lineal de las variables originales. La relación  $\lambda_i (\sum \lambda_i)^{-1}$  corresponde a la proporción de la variabilidad “entre-dentro” de grupos que es explicada por la i-ésima variable canónica (Franco *et al.*, 1997b).

Este análisis determina variables canónicas que maximizan la relación varianza entre grupos sobre la varianza dentro de grupos, permitiendo una mejor diferenciación de los grupos cuando los datos presentan una estructura. Los coeficientes de correlación entre las variables originales y las variables canónicas proveen una interpretación biológica, permitiendo la caracterización de los grupos y la interpretación de los mismos en el gráfico (Franco *et al.*, 1997a). Agrupamientos cuyas primeras dos o tres variables canónicas explican el 70% o más de la variabilidad “entre-dentro” de grupos pueden ser representadas en gráficos de dos o tres dimensiones sin mucha distorsión o pérdida de información (Franco *et al.*, 1997b).

Se realizó el análisis canónico basado en la clasificación de razas y en los grupos obtenidos con las diferentes estrategias de agrupamiento. El análisis canónico discriminante se realizó utilizando el PROC CANDISC (SAS 1996) (Taba *et al.*, 1997).

### 3.5.6 Recuperación de la diversidad fenotípica

Un criterio para determinar la calidad de un agrupamiento es la recuperación de la varianza fenotípica. Este criterio es muy importante ya que uno de los objetivos de los agrupamientos es realizar luego muestreos de algunas accesiones representativas de la colección. Para determinar cuanto de la varianza fenotípica es recuperada por cada agrupamiento en un muestreo al azar se realizaron 100 muestreos estratificados en cada uno de los dos agrupamientos (razas y 7 grupos). Se muestreó el 20% de cada grupo (o raza) y luego se calculó la media y varianza por muestra, para cada variable. A partir de esto, se calculó el promedio de las varianzas fenotípicas que fue retenida en las 100 muestras, por variable. Un mejor agrupamiento es aquel que retiene mayor proporción de la varianza fenotípica.

## 4. RESULTADOS Y DISCUSION

### 4.1 DESCRIPCION GENERAL DE LAS VARIABLES

Los estadísticos descriptivos de las variables continuas (media, desviación estándar, mínimo y máximo) y el análisis de frecuencia de las variables discretas se presentan en la Tabla 1.

**Tabla 1. Estadísticos simples para las variables utilizadas en el análisis de la Colección de Maíz Uruguaya**

Variable Continua	Mín	Max	Media	Desvío estándar	Variable Categorica <sup>(#)</sup>	Clases	Frec.	%
Floración Masculina	65	102	77.5	5.6	Forma de Grano	1	172	20
Floración Femenina	68	102	82.8	7.4		2	355	42
Altura de Planta	94	204	169.6	14.5		3	290	35
Altura de Mazorca	50	168	107.4	14.4		4	23	3
Índice de Prolificidad	0.04	2.00	0.811	0.255	Textura de Grano	1	87	10
Índice de Macollaje	1.0	3.5	1.450	0.291		2	86	10
Vuelco	7	100	77.0	13.0		3	644	77
Relación espiga-forraje	0.01	0.81	0.242	0.132	Color de Perisperma	4	23	3
Forraje Residual	6.3	40.5	19.7	5.8		0	600	71
Longitud de Mazorca	8	22	15.3	1.941	Color de Aleurona	1	240	29
Espesor de Grano	4	14	6.0	0.957		0	144	17
Longitud de Grano	7	15	10.4	1.344	Color de Endosperma	1	696	83
Ancho de Grano	6	13	8.9	1.196		0	201	24
Peso 100 Granos	6	64	26.1	6.1	Diámetro de Mazorca	1	639	76
Peso de Grano	0.163	1.38	0.748	0.204		0	584	70
Peso Mazorca entera	0.264	1.85	0.988	0.252	Número de Hileras	1	256	30
Porcentaje de Grano	32	88	75.6	5.9		0	432	51
						1	408	49

(#) Forma de Grano. 1: más largo que ancho, 2: redondo aplanado grande, 3: redondo aplanado pequeño, 4: aristado pequeño y grande. Textura de Grano. 1: Harinoso, 2: Dentado y Semidentado, 3: Duro y Semiduro, 4: Reventón. Color de perisperma. 0: Incoloro, 1: Coloro (naranja, amarillo, rojo y rojo variegado). Color de aleurona. 0: Incoloro, 1: Amarillo. Color de endosperma. 0: Blanco, 1: Coloro (amarillo y naranja). Diámetro de mazorca. 0:  $\leq 4$ , 1:  $>4$ . Número de hileras por mazorca. 0:  $\leq 13.4$ , 1:  $>13.4$

## 4.2 ESTRUCTURA DE LAS RAZAS

En la Colección de maíz del Uruguay se encuentran 10 razas. La raza Dente Branco (R1) es la única raza de textura de grano dentado y se caracteriza por presentar mazorcas grandes. La única raza de textura de grano harinosos es Morotí (R2), que también presenta valores altos para las variables continuas. Pisingallo (R9) es una raza de textura reventón y granos aristados, se caracteriza por presentar individuos de plantas pequeñas muy macolladoras. El resto de razas son todas de textura de grano duro-semiduro, en general con valores intermedios para las variables continuas. Resalta la raza Cateto Sulino (R6) ya que es la raza numéricamente más importante de la colección, cuenta con 449 accesiones (53% del total), además es una raza con características intermedias a todas las otras, por lo que se encuentra “mezclada” con los otras razas de textura dura (figura 1). La raza Dente Riograndense (R3) y Semidentado Riograndense (R4) se diferencian a grandes rasgos por presentar individuos mayoritariamente de endosperma blanco. Las razas Cateto Sulino Grosso (R8) y Cuarentino (R7) son de floración más temprana que la raza Cateto Sulino, además Cuarentino tiene altos valores de prolificidad. La raza Cristal (R10) es de floración más tardía y plantas más altas. La raza Canario de Ocho (R5) también tiene características intermedias para las variables continuas (Tablas 2 y 3).

## 4.3 CLASIFICACION NUMERICA

En la determinación del número de grupos, el Índice de Mojena (1977) encontró como posible número de grupos 15. El perfil de verosimilitud (Franco, 1998) que se realizó para los agrupamientos con 1 a 20 grupos presenta mayores incrementos, indicando mayor ganancia en verosimilitud, para 7 y para 16 grupos (figura 2). Se tomó el criterio de estudiar ambos agrupamientos, los 16 grupos por estar más cerca del Índice de Mojena y los 7 grupos por ser el primero donde se produce un incremento importante en la verosimilitud.

### 4.3.1 Estructura de los 7 grupos (Ward-MLM)

Los 7 grupos formados por Ward-MLM mantienen casi totalmente la estructura de las razas más diferenciadas, encontrándose grupos que contienen más del 50% de los individuos pertenecientes a determinadas razas: Dente Branco (R1), Morotí (R2), Dente Riograndense (R3), Semidentado Riograndense (R4), Canario de Ocho (R5), Cuarentino (R7), Pisingallo (R9) y Cristal (R10) (Tabla 4).

El método Ward-MLM con 7 grupos dividió a la raza Cateto Sulino (R6), con 449 observaciones (53% de las observaciones totales), en dos grupos: uno con granos redondos, aplanados, pequeños (G6) y el otro con granos redondos, aplanados, grandes (G3) (Tablas 4, 5 y 6). Presentan pequeña distancia para las variables continuas hacia

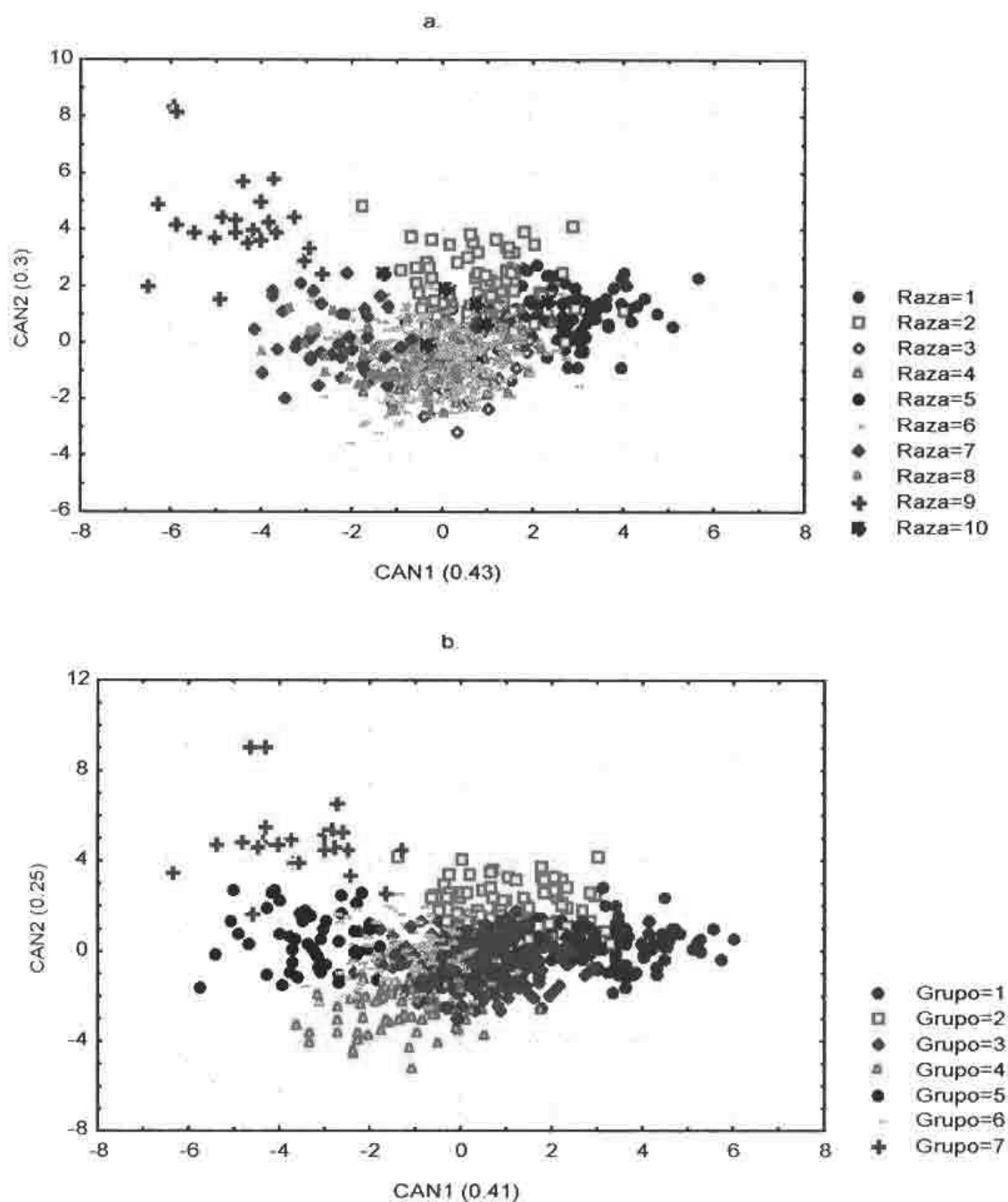


Figura 1. Todos los individuos pertenecientes a las razas y a los grupos (7 grupos Ward-MLM) en la clasificación de la Colección de Maíz del Uruguay. a) Individuos pertenecientes a las razas, b) individuos pertenecientes a los grupos (7 grupos Ward-MLM)

**Tabla 2. Valores medios por raza de todas las variables continuas utilizadas en el análisis de la clasificación de la Colección de Maíz del Uruguay**

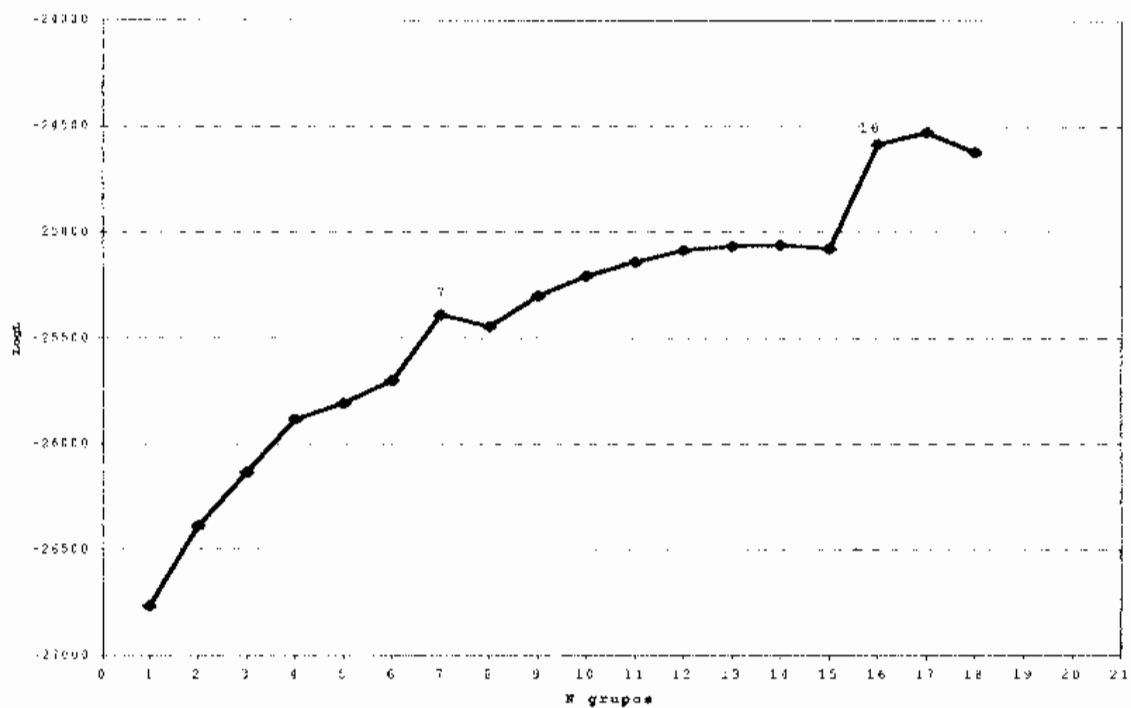
RAZA <sup>(*)</sup>	N <sup>(*)</sup>	PM	FF	ALPTA	ALTMZ	PRO	EMC	VUELCO	ESFORR	FORRES	LONGMAZ	ESFORA	LONGRA	ANCHORR	F100GRA	PESORRA	MAZENT	FORGRA
1	89	81.1	88.1	182.7	125.6	0.681	1.377	73.7	0.200	20.7	14.9	7.45	11.9	10.5	32.3	0.887	1.155	76.7
2	88	82.8	89.7	165.1	113.4	0.645	1.701	68.4	0.135	21.9	15.2	6.13	10.4	9.2	29.9	0.611	0.832	72.4
3	24	75.2	78.9	173.6	101.5	0.815	1.330	72.5	0.348	18.6	14.6	6.38	11.5	10.0	30.2	0.977	1.300	75.3
4	67	75.7	80.1	171.8	107.5	0.781	1.362	75.6	0.282	20.3	14.7	6.03	10.7	9.3	27.2	0.823	1.115	73.8
5	4	71.8	75.5	162.3	87.8	0.895	1.463	75.8	0.225	20.6	15.0	7.00	10.8	9.0	30.3	0.804	1.035	79.8
6	449	76.7	81.8	169.4	105.2	0.820	1.401	80.4	0.248	19.4	16.1	5.76	10.2	8.7	25.8	0.745	0.972	76.7
7	40	75.1	80.0	169.5	103.8	1.006	1.547	67.8	0.253	20.8	14.1	5.85	8.9	7.6	16.5	0.641	0.897	71.0
8	50	73.8	77.3	166.8	101.7	0.930	1.288	77.1	0.369	15.8	13.9	5.86	10.1	8.6	22.0	0.827	1.099	75.1
9	23	78.6	83.9	137.3	85.3	1.258	2.280	79.5	0.176	18.1	11.6	5.52	8.6	7.4	11.8	0.399	0.523	76.1
10	6	85.0	90.0	163.5	108.3	0.753	1.443	78.7	0.203	19.9	14.5	6.33	10.2	9.2	25.0	0.524	0.712	73.3
F <sup>(#)</sup>		25.83	27.19	29.28	37.49	23.04	51.63	12.49	19.00	5.18	31.73	38.87	35.67	48.16	78.98	31.70	34.58	9.66

(#) 1: Dente Branco, 2: Moroti, 3: Dente Riograndense, 4: Semidentado Riograndense, 5: Canario Ocho, 6: Cateño Sulino, 7: Cuarentino, 8: Cateño Sulino Grosso, 9: Pislingallo y 10: Cristal (1) Número de individuos perteneciente a cada raza. FM: Días a Floración Masculina FF: Días a Floración Femenina ALPTA: Altura de la planta ALTMZ: Altura de la mazorca IPRC: Índice de prolificidad IMAC: Índice de macollaje VUELCO: Porcentaje de Vuelco ESPFOR: Relación Espiga Forraje FORRES: Forraje Residual LONGMAZ: Longitud de mazorca ESPGRA: Espesor del Grano LONGRA: Longitud de Grano ANCHOGRA: Ancho de Grano P100GRA: Peso de 100 Granos PESOGRA: Peso de Grano (10 mazorcas) MAZENT: Peso de 10 Mazorcas PORGRA: Porcentaje del Peso de Grano en el total de la mazorca. (\*) Valores significativos (P<0.0001)

**Tabla 3. Frecuencia por raza de todas las variables categóricas utilizadas en el análisis de la clasificación de la Colección de Maíz del Uruguay**

Raza <sup>(1)</sup>	N <sup>(1)</sup>	FORGRA <sup>(1)</sup>			TEXGRA <sup>(2)</sup>			COLORPER <sup>(3)</sup>		COLORALE <sup>(4)</sup>		COLORPER <sup>(5)</sup>		DIAMMAZ <sup>(6)</sup>	NUMHIL <sup>(7)</sup>	
		1	2	3	4	1	2	3	4	0	1	0	1	0	1	
1	89	42	46	1	0	0	78	11	0	89	0	89	0	40	49	
2	88	16	50	22	0	87	0	1	0	88	0	28	60	84	4	
3	24	13	10	1	0	0	7	17	0	23	1	0	24	1	23	
4	67	17	40	10	0	0	1	66	0	64	3	0	67	0	67	
5	4	0	2	2	0	0	0	4	0	3	1	0	4	0	4	
6	449	33	194	222	0	0	0	449	0	240	209	0	449	0	449	
7	40	24	1	15	0	0	0	40	0	28	12	0	40	0	40	
8	50	25	9	16	0	0	0	50	0	41	9	0	50	0	50	
9	23	0	0	0	23	0	0	0	23	18	5	21	2	21	2	
10	6	2	3	1	0	0	0	6	0	6	0	6	0	6	0	
Tot	840	172	355	290	23	87	86	644	23	600	240	144	696	201	639	
$\chi^2/ig1$		17.7														
		43.8														
		26.3														
		71.8														
		96.7														
		9.6														
		12.6														

(#) 1: Dente Branco, 2: Moroti, 3: Dente Riograndense, 4: Semidentado Riograndense, 5: Canario Ocho, 6: Cateto Sulino, 7: Cuarentino, 8: Cateto Sulino Grosso, 9: Pistingallo y 10: Cristal (†) Número de individuos perteneciente a cada raza. (1) Forma de Grano. 1: más largo que ancho, 2: redondo aplanado grande, 3: redondo aplanado pequeño, 4: aristado pequeño y grande. (2) Textura de Grano. 1: Harinoso, 2: Dentado y Semidentado, 3: Duro y Semiduro, 4: Reventón (3) Color de perisperma. 0: Incoloro, 1: Coloro (naranja, amarillo, rojo y rojo variegado) (4) Color de aleurona. 0: Incoloro, 1: Amarillo (5) Color de endosperma. 0: Blanco, 1: Coloro (amarillo y naranja) (6) Diámetro de mazorca. 0: ≤ 4, 1: > 4 (7) Número de hileras por mazorca. 0: ≤ 13,4, 1: > 13,4



**Figura 2. Perfil de verosimilitud desde 1 a 20 grupos, donde se muestran los saltos importantes en la versomilitud que ocurren en los 7 grupos y en los 16 grupos.**



**Tabla 4. Relación entre los individuos pertenecientes a las razas y a los grupos (7 grupos Ward-MLM): porcentaje de observaciones de cada raza en cada grupo (número superior) y de cada grupo en cada raza (número inferior)**

% grupo	RAZA <sup>(#)</sup>										Nº
% raza	1	2	3	4	5	6	7	8	9	10	Total
<b>GRUPO</b>											
1	88.30	4.26	1.06	1.06	0	4.26	0	0	0	1.06	94
	93.26	4.55	4.17	1.49	0	0.89	0	0	0	16.67	
2	4.65	75.58	0	1.16	0	13.95	0	0	0	4.65	86
	4.49	73.86	0	1.49	0	2.67	0	0	0	66.67	
3	0.36	5.76	6.47	13.67	0.72	67.63	0.36	5.04	0	0	278
	1.12	18.18	75.00	56.72	50.00	41.87	2.50	28.00	0	0	
4	0	1.37	5.48	23.29	0	46.58	0	21.92	0	1.37	73
	0	1.14	16.67	25.37	0	7.57	0	32.00	0	16.67	
5	0	0	0	5.66	0	7.55	62.26	24.53	0	0	53
	0	0	0	4.48	0	0.89	82.5	26.00	0	0	
6	0.43	0.43	0.43	3.02	0.86	89.22	2.59	3.02	0	0	232
	1.12	1.14	4.17	10.45	50.00	46.10	15.00	14.00	0	0	
7	0	4.17	0	0	0	0	0	0	95.83	0	24
	0	1.14	0	0	0	0	0	0	100	0	
<b>Nº</b>											
Total	89	88	24	67	4	449	40	50	23	6	840

(#) 1: Dente Branco, 2: Moroti, 3: Dente Riograndense, 4: Semidentado Riograndense, 5: Canario Ocho, 6: Cateto Sulino, 7: Cuarentino, 8: Cateto Sulino Grosso, 9: Pisingailo y 10: Cristal

**Tabla 5. Valores medios por grupo (7 grupos Ward-MLM) de las variables continuas utilizadas en el análisis de la clasificación de la Colección de Maíz del Uruguay**

GRUPO	N <sup>o</sup>	FM	FF	ALTPTA	ALTMZ	IPRO	MAC	VUELCO	ESPFOR	FORRES	LONGMAZ	ESFORA	LONGRA	ANCHOGRA	P100GRA	PESOGRA	MAZENT	PORGRA
1	94	80.6	87.5	182.7	125.2	0.681	1.392	74.6	0.205	20.4	15.0	7.5	12.1	10.6	32.6	0.901	1.170	77.1
2	86	86.0	92.7	165.4	112.5	0.633	1.674	69.3	0.137	21.8	14.8	6.0	10.2	9.0	29.5	0.569	0.798	70.2
3	278	75.8	80.4	170.3	105.6	0.826	1.402	78.6	0.247	19.6	15.8	6.0	11.1	9.6	27.4	0.822	1.068	77.1
4	73	71.5	74.6	168.0	97.1	0.962	1.266	76.5	0.519	13.1	15.0	6.0	10.8	8.9	25.1	0.838	1.098	76.4
5	53	75.0	79.6	164.9	99.8	0.991	1.522	64.5	0.270	20.2	13.6	5.8	9.0	7.6	16.9	0.656	0.928	69.6
6	232	77.4	83.2	169.8	107.7	0.776	1.401	81.6	0.202	20.8	16.0	5.6	9.4	8.0	24.6	0.692	0.916	75.9
7	24	78.8	84.1	139.0	85.8	1.259	2.296	79.0	0.176	18.6	11.6	5.6	8.6	7.4	12.0	0.408	0.534	76.2
F <sup>(*)</sup>		98.12	87.38	43.21	62.38	42.84	78.90	22.44	135.96	23.13	38.86	66.82	167.17	207.83	128.74	64.31	57.33	24.93

(†) Número de individuos perteneciente a cada grupo. FM. Días a Floración Masculina FF. Días a Floración Femenina ALTPTA. Altura de la planta ALTMZ. Altura de la mazorca IPRO. Índice de prolificidad IMAC. Índice de macollaje VUELCO. Porcentaje de Vuelco ESPFOR. Relación Espiga Forraje FORRES. Forraje Residual LONGMAZ. Longitud de Mazorca ESPGRA. Espesor de Grano LONGRA. Longitud de Grano ANCHOGRA. Ancho de Grano P100GRA. Peso de 100 Granos PESOGRA. Peso de Grano (10 mazorcas) MAZENT. Peso de 10 Mazorcas PORGRA. Porcentaje del Peso de Grano en el total de la Mazorca. (\*) Valores significativos (P<0,0001)

Tabla 6. Frecuencia por grupo (7 grupos Ward-MILM) de todas las variables categóricas utilizadas en el análisis de la clasificación de la Colección de Maíz del Uruguay

Grupo	N <sup>(t)</sup>	FORGRA <sup>(1)</sup>				TEXGRA <sup>(2)</sup>				COLORPER <sup>(3)</sup>		COLORALE <sup>(4)</sup>		COLORENDO <sup>(5)</sup>		DIAMMAZ <sup>(6)</sup>		NUMHIL <sup>(7)</sup>	
		1	2	3	4	1	2	3	4	0	1	1	0	0	1	0	1	0	1
1	94	47	47	0	0	4	78	12	0	93	1	8	86	87	7	44	50	54	40
2	86	14	42	30	0	64	1	21	0	85	1	54	32	70	16	68	18	64	22
3	278	49	229	0	0	16	6	256	0	195	83	275	3	18	260	195	83	150	128
4	73	17	37	19	0	1	1	71	0	55	18	72	1	2	71	32	41	15	58
5	53	28	0	25	0	0	0	53	0	41	12	53	0	0	53	42	11	8	46
6	232	16	0	216	0	1	0	231	0	112	120	231	1	2	230	179	53	139	93
7	24	1	0	0	0	1	0	0	0	19	5	3	21	22	2	24	0	2	22
Tot	840	172	355	290	23	87	86	644	23	600	240	696	144	639	639	584	256	432	406
$\chi^2_{19df}$		57.0				102.5				26.7		87.8		100.6		12.0		18.1	

(#) 7 grupos obtenidos a partir del procedimiento Ward-MILM (t) Número de individuos perteneciente a cada raza. (1) Forma de Grano. 1: más largo que ancho, 2: redondo aplanado grande, 3: redondo aplanado pequeño, 4: anisado pequeño y grande. (2) Textura de Grano. 1: Harinoso, 2: Dentado y Semidentado, 3: Duro y Semiduro, 4: Reventón (3) Color de perisperma. 0: Incoloro, 1: Coloro (naranja, amarillo, rojo y rojo variegado) (4) Color de aleurona. 0: Incoloro, 1: Amarillo (5) Color de endosperma. 0: Blanco, 1: Coloro (amarillo y naranja) (6) Diámetro de mazorca. 0:  $\leq 4$ , 1:  $>4$  (7) Número de hileras por mazorca. 0:  $\leq 13$ , 4, 1:  $>13$ , 4

ambos grupos, pero menor para el grupo 6 (Tabla 7). En G3 se agrupó a los individuos que presentan granos de mayor longitud y ancho y madurez más temprana. También se discriminó por índice de prolificidad, espesor de grano, peso de 100 granos, peso de grano de toda la mazorca y peso de la mazorca entera en promedio mayor (G3) de menor (G6) y vuelco promedio mayor (G6) de menor (G3) (tabla 8). Esto se debe tanto a que se separaron a los individuos de la raza Cateto Sulino (R6) con estas características, como al ingreso de individuos de otras razas con las mismas características. En esta división de las accesiones se logró una reducción de la varianza de cada variable continua y del coeficiente de variación (CV), la varianza y el CV es menor en el grupo 3 y el 6 que en la raza 6, además es menor en los individuos pertenecientes a la raza Cateto Sulino que quedaron en los grupos 3 y 6. Presentan medias alrededor de la media de la raza Cateto Sulino (R6), encontrándose uno de los grupos por encima y el otro por debajo (Tabla 8). La división de la raza Cateto Sulino (R6) principalmente en dos grupos no está asociada con la distribución geográfica de las colectas. Los grupos 3 y 6 se separan principalmente por el eje canónico 1 que está correlacionado con variables de grano, la distribución geográfica de los Catetos se correlaciona más con el eje canónico 2, que está correlacionado con variables reproductivas como días a floración (figura 3).

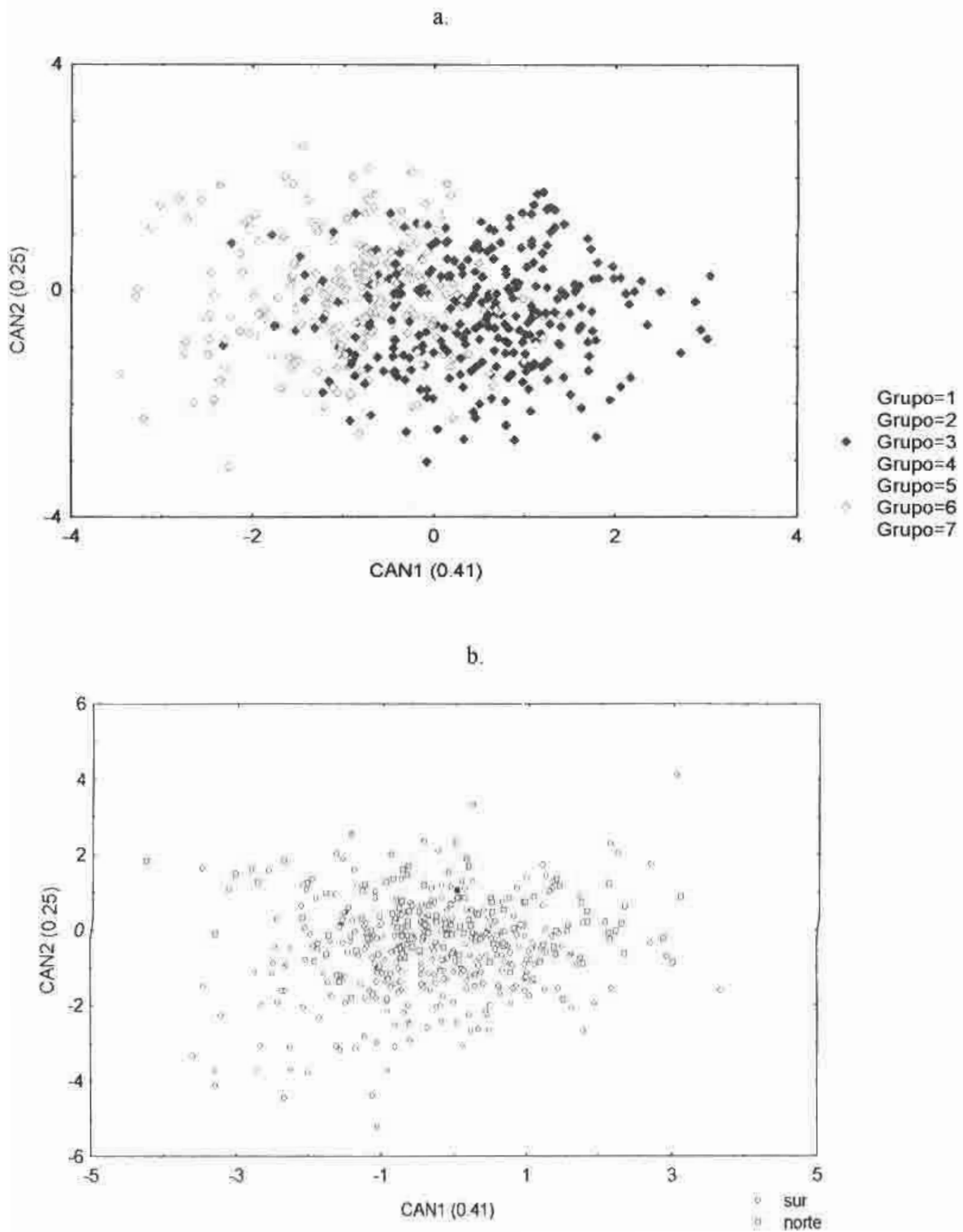
**Tabla 7. Distancias de Mahalanobis para variables continuas entre razas y 7 grupos Ward-MLM para la Colección de Maíz del Uruguay**

		7 GRUPOS						
		G1	G2	G3	G4	G5	G6	G7
RAZAS	R1	0.1	12.7	11.9	29.4	38.4	19.9	72.1
	R2	11.1	1.1	9.1	25.2	24.8	11.8	43.6
	R3	13.9	17.2	4.3	9.2	18.8	10.9	62.6
	R4	14.4	13.7	2.2	8.3	12.6	5.4	52.7
	R5	26.0	25.2	9.4	25.4	19.5	10.8	48.7
	R6	15.9	11.6	1.3	12.0	13.4	1.1	47.8
	R7	36.9	25.9	15.1	19.7	0.5	10.3	30.8
	R8	23.6	21.6	7.0	6.1	7.1	7.2	46.1
	R9	74.5	47.9	51.3	64.9	31.9	48.1	0.0
	R10	14.6	4.3	10.5	21.3	22.7	12.3	44.5

A la raza Canario de Ocho (R5) la clasificación numérica la dividió en dos grupos y colocó a sus individuos con la raza Cateto Sulino (G3 y G6) (Tabla 4). Esta división en G3 y G6 se basa en la forma del grano, como ya se mencionó anteriormente. En el grupo 3 se colocaron los individuos con granos redondos, aplanados, grandes y en el grupo 6 a aquellos con granos redondos, aplanados, pequeños. Las características continuas asociadas con los promedios de los grupos y que discriminaron a los individuos son: vuelco, ancho de grano y peso de grano. El vuelco es menor en los individuos del grupo 3, respecto a los del grupo 6, y el ancho y peso del grano son mayores (Tabla 9). La

**Tabla 8. Comparación de la raza Cateto Sulino (R6), el grupos 3, el grupo 6, los individuos que pertenecen simultáneamente a la raza Cateto Sulino y al grupo 3 y los individuos que pertenecen simultáneamente a la raza Cateto Sulino y al grupo 6; para algunas variables importantes.**

	FM	FF	IPRO	VUELCO	ESPGRA	LONGRA	ANCHOGRA	P100GRA	PESOGRA	MAZENT
<b>R6</b>										
min	65	68	0.050	29	4	7	6	14	0.194	0.358
max	100	103	1.960	100	10	15	12	40	1.332	1.800
media	76.7	81.8	0.820	80.4	5.8	10.2	8.7	25.8	0.745	0.972
varianza	29.26	48.09	0.06	117.77	0.56	1.44	0.99	17.70	0.03	0.04
CV	7.1	8.5	28.7	13.5	13.0	11.7	11.4	16.3	22.5	20.7
<b>G3</b>										
min	66	68	0.150	7	5	10	8	15	0.372	0.502
max	89	99	1.890	100	8	15	13	41	1.270	1.838
media	75.8	80.4	0.826	78.6	6.0	11.1	9.6	27.4	0.822	1.068
varianza	17.33	30.98	0.05	136.94	0.46	0.64	0.50	19.67	0.03	0.05
CV	5.5	6.9	25.9	14.9	11.3	7.2	7.4	16.2	20.7	20.5
<b>G6</b>										
min	68	69	0.040	50	4	7	6	15	0.357	0.532
max	92	102	1.960	100	8	12	9	37	1.097	1.398
media	77.4	83.2	0.776	81.6	5.8	9.4	8.0	24.6	0.692	0.916
varianza	18.70	41.16	0.06	113.23	0.47	0.64	0.47	16.38	0.02	0.03
CV	5.6	7.7	31.9	13.0	12.2	8.5	8.6	16.4	20.2	19.1
<b>G3 catetos</b>										
min	66	68	0.150	50	5	10	8	15	0.384	0.598
max	85	99	1.890	100	8	15	11	40	1.194	1.600
media	75.8	80.5	0.853	79.5	5.9	11.0	9.5	27.0	0.809	1.043
varianza	16.93	31.02	0.05	111.77	0.46	0.67	0.43	17.49	0.03	0.04
CV	5.4	6.9	26.2	13.3	11.5	7.4	6.9	15.5	19.7	18.8
<b>G6 catetos</b>										
min	68	69	0.050	53	4	7	6	15	0.357	0.532
max	88	101	1.960	100	8	12	9	37	1.097	1.398
media	77.4	83.2	0.775	81.5	5.6	9.4	8.0	24.7	0.685	0.903
varianza	16.40	37.70	0.06	108.03	0.47	0.63	0.47	15.91	0.02	0.03
CV	5.2	7.4	31.6	12.8	12.2	8.4	8.5	16.1	20.1	18.4



**Figura 3. Primeras variables canónicas para los individuos pertenecientes a la raza Cateto Sulino, dentro de los grupos G3 y G6 (a), y según el origen de la colecta (b)**

única variable discreta que diferencia a los individuos de esta raza es forma de grano, por lo que el resto de variables no influyeron en la asignación a uno u otro grupo. Canario de Ocho es una raza comercial antigua que se piensa es originaria de la cruce de Morotí por Cateto. La distancia entre la raza y los grupos para las variables continuas no es pequeña, pero la menor distancia es con el grupo 3 (Tabla 7).

**Tabla 9. Comparación del promedio de las variables continuas que diferencian a los grupos 3 y 6 (7 grupos Ward-MLM) en el grupo 3, grupo 6 y todas las observaciones pertenecientes a la raza Canario de Ocho.**

	G	FM	FF	IPRO	VUELCO	ESPGRA	LONGRA	ANCHOGRA	P100GRA	PESOGRA	MAZENT
Promedio G3	3	75.8	80.4	0.826	78.6	6.0	11.1	9.6	27.4	0.822	1.068
Promedio G6	6	77.4	83.2	0.776	81.6	5.6	9.4	8.0	24.6	0.692	0.916
Observación R5											
1064	6	70	73	1.05	85	7	10	9	23	0.703	0.95
3005	6	72	74	0.95	75	7	10	8	33	0.805	1.076
3006	3	72	78	0.95	74	7	12	10	37	0.879	1.112
13057	3	73	77	0.63	69	7	11	9	28	0.829	1.002

La raza Cristal (R10) fue asignada al grupo 2 (Tabla 4), junto a la raza Morotí (R2). El origen de la raza Cristal (raza comercial antigua) es de la cruce de Calchaqui Blanco Duro por Morotí (raza Indígena). La distancia más pequeña con los grupos es con el grupo 2 que recuperó a esta raza (Tabla 7).

La clasificación numérica dividió a la raza Cateto Sulino Grosso (R8) en varias partes, incluyendo algunos individuos en los grupos con los Catetos Sulino (G3 y G6), y a otros en los grupos 4 y 5, siempre con individuos de textura de grano mayoritariamente dura (Tabla 4). La menor distancia para las variables continuas es con el grupo 4 (Tabla 7).

Las razas Dente Riograndense (R3) y Semidentado Riograndense (R4) fueron colocadas con los Cateto Sulino en el grupo 3 (Tabla 4). No varía en forma importante la *media del grupo con el promedio de las razas para las variables continuas* (tablas 2 y 5), la distancia para variables continuas es menor con este grupo, pero no es pequeña (Tabla 7). En cuanto a las variables categóricas, no presenta diferencias salvo que el grupo 3 presenta la mayoría de sus individuos con un número bajo de hileras por mazorca, y la raza Semidentado Riograndense (R4) presenta un mayor número de hileras por mazorca promedio (tablas 3 y 6).

Si se hace la lectura desde el punto de vista de la estructura de los grupos, se puede observar que existen grupos con más de un 50% de sus individuos pertenecientes a una única raza: G1, G2, G3, G5, G6 y G7 (Tabla 4).

El grupo 1 quedó conformado en su mayoría (88%) por individuos pertenecientes a la raza "Dente Branco" (R1), contiene un 93% de los individuos pertenecientes a esta raza (Tabla 4). Presenta características similares a la misma, siendo la distancia para las variables continuas muy pequeña (Tabla 7). Contiene a los individuos con mayor altura de mazorca, espesor de grano, ancho de grano, peso de 100 granos y peso de la mazorca entera de la colección. A su vez presenta índice de macollaje bajo (tabla 5). La textura es mayoritariamente dentado-semidentado, el perisperma y la aleurona son incoloros y el endosperma es blanco (tabla 6).

El grupo 2 se encuentra mayormente conformado por Morotí (R2, 76%), contiene a un 74% de los individuos pertenecientes a la raza Morotí (R2) y un 67% de los individuos pertenecientes a la raza Cristal (R10) (Tabla 4). Presenta características similares a ambas razas, siendo sus individuos más tardíos, de menor índice de prolificidad, forraje residual, espesor de grano, ancho de grano, peso de la mazorca entera y porcentaje de grano que el promedio de las razas Morotí y Cristal. El resto de características continuas se promediaron (Tablas 2 y 5). De todas formas es más similar en cuanto a las variables continuas a las raza 2 que a la raza 10 (Tabla 7). En cuanto a las variables categóricas presenta características de la raza 2 como ser grano mayoritariamente de textura harinosa, aleurona amarilla, diámetro de mazorca pequeño y pocas hileras por mazorca. Comparte la característica de presentar perisperma incoloro (Tablas 3 y 6).

El grupo 3 se encuentra conformado en su mayoría por individuos de la raza Cateto Sulino (R6, 68%). Contiene a un 75% de los individuos pertenecientes a la razas Dente Riograndense (R3), un 57% perteneciente a Semidentado Riograndense (R4) y un 50% perteneciente a la raza Canario de Ocho (R5). Se encuentran además el 42% de los individuos pertenecientes a la raza Cateto Sulino (Tabla 4). En cuanto a sus variables continuas, el grupo se parece más a los Catetos Sulinos que al resto, pero tanto Dente Riograndense, Semidentado Riograndense como Canario de Ocho presentan menores distancias con el grupo 3 que con otros grupos, no así Cateto Sulino (Tabla 7). Es de floración relativamente temprana, altura de mazorca y de planta, índice de macollaje, índice de prolificidad. Los granos son de longitud, ancho y espesor grande (Tablas 5 y 6).

El grupo 4 quedó como una mezcla de las razas Cateto Sulino (R6), Semidentado Riograndense (R4) y Cateto Sulino Grosso (R8) principalmente (Tabla 4), pero se parece más en cuanto a las variables continuas a la los Cateto Sulino Grosso (Tabla 7). Son



individuos de floración masculina temprana, baja altura de mazorca e índice de macollaje. Presentan alta relación espiga forraje y valores intermedios para el resto de características. Son en su mayoría de textura dura, de perisperma incoloro, aleurona amarilla y endosperma blanco. Son de gran espesor de grano (Tablas 5 y 6).

El grupo 5 se encuentra mayoritariamente formado por Cuarentino (R7, 62%), recuperando a un 83% de los individuos pertenecientes a esta raza (Tabla 4). Mantuvo características similares a las de la raza.

El grupo 6 se encuentra en su mayoría conformado por la raza Cateto Sulino (R6, 89%) (tabla 4), siendo esta raza más similar en cuanto a las variables continuas al grupo, que a otros grupos (tabla 7). Este grupo contiene a un 50% de los individuos pertenecientes a la raza Canario de Ocho y un 46% de los Catetos Sulinos (Tabla 4). Son individuos de madurez intermedia, con plantas altas-medias, bajo índice de macollaje e hileras por mazorca. Las mazorcas son largas, con granos de longitud, espesor y ancho pequeño, de textura en su mayor parte dura, aleurona amarilla y endosperma blanco. Las mazorcas son de diámetro pequeño (Tabla 5 y 6).

El grupo 7 abarca la totalidad de la raza Pisingallo (R9) y un individuo Morotí (R2) (Tabla 4). La distancia para las variables continuas es cero (Tabla 7). Se diferencia claramente del resto de grupos por presentar menor altura y longitud de mazorca, número de espiguillas por mazorca, peso de 100 granos y de la mazorca entera y ancho de grano. A su vez es quien presenta mayor índice de macollaje. Para días a floración masculina es intermedio (tabla 5). En cuanto a sus características categóricas presenta granos aristados, pequeños o grandes. La textura del grano es “reventón”, salvo el individuo Morotí. Presenta en su mayoría perisperma y aleurona incoloro y el endosperma blanco. El diámetro de mazorca es pequeño y el número de hileras por mazorca alto (tabla 6).

#### 4.3.2 Estructura de los 16 grupos (Ward-MLM)

Con los 16 grupos formados por el método Ward-MLM se recuperó a más del 50% de los individuos pertenecientes a las razas Dente Branco (R1), Morotí (R2), Canario de Ocho (R5), Cuarentino (R7), Pisingallo (R9) y Cristal (R10). A su vez, los grupos Gr1, Gr2, Gr3, Gr6, Gr8, Gr10, Gr11, Gr12, Gr13, Gr14, Gr15 y Gr16 contienen a la mayor parte de sus individuos pertenecientes a una única raza (tabla 10). La mayor parte de los 16 grupos se encuentran asociados a algunas razas, presentando una distancia para las variables continuas pequeña. Los grupos Gr9 y Gr10 no se parecen a ninguna raza para las variables continuas.

La mayor parte de los 16 grupos contiene más de un 75% de sus individuos

**Tabla 10. Relación entre los individuos pertenecientes a las razas y a los grupos (16 grupos Ward-MLM): porcentaje de observaciones de cada raza en cada grupo (número superior) y de cada grupo en cada raza (número inferior)**

Raza <sup>(a)</sup>	16 GRUPOS																Total
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	
1	35.96 80.00	60.67 88.52	0 0	0 0	0 0	0 0	0 0	1.12 0.68	1.12 14.29	0 0	0 0	0 0	0 0	1.12 0.67	0 0	0 0	89
2	4.55 10	4.55 6.56	75.00 91.67	0 0	1.14 2.33	0 0	0 0	10.23 6.08	2.27 28.57	0 0	0 0	1.14 3.13	0 0	0 0	0 0	1.14 4.35	88
3	0 0	0 0	0 0	29.17 8.54	12.5 6.98	8.33 5.71	45.83 34.38	0 0	0 0	0 0	0 0	4.17 3.13	0 0	0 0	0 0	0 0	24
4	1.49 2.5	0 0	0 0	40.30 32.93	11.94 18.6	4.48 8.57	8.96 18.75	17.91 8.11	0 0	0 0	1.49 2.94	1.49 3.13	5.97 10.81	5.97 2.68	0 0	0 0	67
5	0 0	0 0	0 0	25.00 1.22	0 0	0 0	0 0	25.00 0.68	0 0	0 0	0 0	50.00 6.25	0 0	0 0	0 0	0 0	4
6	0.67 7.50	0.45 3.28	0.67 4.17	8.02 43.90	4.23 44.19	5.35 68.57	2.90 40.63	27.39 83.11	0.45 28.57	1.78 80.00	0.22 2.94	5.12 71.88	5.12 62.16	31.63 95.30	6.01 77.14	0 0	449
7	0 0	0 0	0 0	0 0	0 0	10.00 11.43	0 0	2.5 0.68	2.50 14.29	0 0	60.00 70.59	0 0	0 0	5.00 1.34	20.00 22.86	0 0	40
8	0 0	0 0	0 0	22.00 13.41	22.00 25.58	4.00 5.71	4.00 6.25	2.00 0.68	2.00 14.29	0 0	16.00 23.53	8.00 12.5	20.00 27.03	0 0	0 0	0 0	50
9	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0	4.35 10.00	0 0	0 0	0 0	0 0	0 0	95.65 95.65	23
10	0 0	16.67 1.64	50.00 4.17	0 0	16.67 2.33	0 0	0 0	0 0	0 0	16.67 10.00	0 0	0 0	0 0	0 0	0 0	0 0	6
Total	40	61	72	82	43	35	32	148	7	10	34	32	37	149	35	23	840

(#) 1: Dente Branco, 2: Moroti, 3: Dente Riograndense, 4: Semidentado Riograndense, 5: Canario Ocho, 6: Cateto Sulino, 7: Cuarentino, 8: Cateto Sulino Grosso, 9: Pisingallo y 10: Cristal

pertenecientes a uno de los 7 grupos Ward-MLM. Los grupos Gr6, Gr9 y Gr13 están conformados por aproximadamente la mitad de sus individuos pertenecientes a uno de los 7 grupos y la otra mitad a otro grupo. La mayor parte de los 16 grupos se encuentran asociados a los 7 grupos, siendo la distancia entre ellos pequeña. Al igual que con las razas, los grupos Gr9 y Gr10 no se encuentran asociados a ninguno de los 7 grupos Ward-MLM.

#### 4.3.3 Selección de una clasificación numérica

Se realizaron dos clasificaciones numéricas, de las cuales la que resultó más apropiada es la clasificación con 7 grupos. La utilización del agrupamiento con 16 grupos obtenido a partir del procedimiento Ward-MLM no aporta ventajas respecto al de 7 grupos. Esta afirmación se basa en: i) la F es menor en el análisis de varianza tanto univariado como multivariado, ii) el chi-cuadrado es menor, iii) son menos las variables *con valor de F mayor a 20 en el análisis stepdisc*, iv) *el promedio de la distancia entre grupos para las variables continuas, discretas y mezcla de variable es menor*, v) hay más observaciones con probabilidades menores a 0.75 y a 0.5 en los 16 grupos que en los 7 grupos, vi) el porcentaje de explicación de las variables canónicas es menor para los 16 grupos que para los 7 grupos, vii) a los 7 grupos se los dividió para formar los 16 grupos, salvo alguno que es producto de la mezcla de grupos.

### 4.4 COMPARACION DE LAS CLASIFICACIONES

#### 4.4.1 Análisis univariado

El análisis de varianza univariado para la comparación de razas y de grupos y la prueba de chi-cuadrado para la asociación variable discreta - agrupamiento indican que todas las variables, tanto categóricas como continuas, son significativas ( $P < 0.0001$ ) en la clasificación de razas y de 7 grupos. La prueba estadística "F" indica la proporción de la varianza entre grupos, respecto de la varianza dentro de grupos. Para todas las variables continuas se obtiene en los 7 grupos un valor de F mayor que el que se obtiene para las razas. Esto indica que en cada una de las variables, la relación entre la varianza entre grupos y la varianza dentro de grupos es mayor en los 7 grupos (tablas 2 y 5). El chi-cuadrado sobre los grados de libertad también es mayor en los 7 grupos que en las razas, para cada una de las variables categóricas (tablas 3 y 6). Esto indica que los 7 grupos están mejor diferenciados en términos de las variables que las raza, esto es, que los grupos son más homogéneos "dentro" y heterogéneos "entre".

#### 4.4.2 Variables que determinan la clasificación

En el análisis Stepwise discriminante para las razas y los 7 grupos, todas las variables continuas aportan a la diferenciación de grupos o razas ( $P < 0.15$ ), pero los 7

grupos tienen valores más altos de F. Se considera (arbitrariamente) que una variable es importante en la clasificación si la varianza entre grupos es 20 veces mayor que la varianza dentro de los grupos, o lo que es lo mismo, que la F es mayor a 20, puede observarse en la tabla 11, que los 7 grupos Ward-MLM utilizan una mayor cantidad de información (variables) para la diferenciación de los grupos. Las variables más importantes en la clasificación de razas son 6: peso de 100 granos, índice de macollaje, altura de mazorca, longitud de mazorca, peso de mazorca entera y espesor del grano. Las variables que más aportan a los 7 grupos son 12: ancho de grano, relación espiga forraje, índice de macollaje, floración masculina, peso de 100 granos, longitud de mazorca, altura de mazorca, vuelco, índice de prolificidad, forraje residual, espesor del grano y peso de grano. Los 7 grupos y las razas comparten 5 de las 6 variables importantes: peso de 100 granos, índice de macollaje, altura de mazorca, longitud de mazorca y espesor de grano.

**Tabla 11. Análisis Stepwise discriminante para razas y grupos (7 y 16 grupos Ward-MLM) de las variables continuas utilizadas en la Clasificación de la Colección de Maíz del Uruguay**

Paso	RAZAS		7 GRUPOS		16 GRUPOS	
	Variable	F Statistic <sup>(*)</sup>	Variable	F Statistic <sup>(*)</sup>	Variable	F Statistic <sup>(*)</sup>
1	P100GRA	78.880	ANCHOGRA	207.828	ANCHOGRA	98.398
2	IMAC	46.364	ESPFORR	141.451	ESPFORR	81.755
3	ALTMAZ	30.113	IMAC	62.099	FM	48.992
4	LONGMAZ	29.289	FM	55.383	IMAC	35.859
5	MAZENT	28.755	P100GRA	43.808	LONGRA	33.159
6	ESPGRA	22.326	LONGMAZ	34.343	PORGRA	31.774
7	ALTPTA	9.718	ALTMAZ	32.133	LONGMAZ	25.215
8	PORGRA	9.391	VUELCO	24.644	IPRO	22.461
9	FF	9.014	IPRO	24.131	FORRES	19.445
10	FORRES	9.077	FORRES	26.358	ALTMAZ	17.153
11	VUELCO	8.310	ESPGRA	25.000	P100GRA	15.909
12	ANCHOGRA	6.845	PESOGRA	20.493	ALTPTA	13.313
13	IPRO	5.504	MAZENT	15.724	ESPGRA	13.187
14	ESPFORR	6.550	LONGRA	10.295	VUELCO	9.236
15	FM	3.057	ALTPTA	9.025	MAZENT	8.865
16	PESOGRA	1.896	FF	7.006	PESOGRA	28.871
17	LONGRA	1.584	PORGRA	6.538	FF	5.143

FM. Días a Floración Masculina FF. Días a Floración Femenina ALTPTA. Altura de la planta ALTMAZ. Altura de la mazorca IPRO. Índice de prolificidad IMAC. Índice de macollaje VUELCO. Porcentaje de Vuelco ESPFOR. Relación Espiga Forraje FORRES. Forraje Residual LONGMAZ. Longitud de Mazorca ESPGRA. Espesor de Grano LONGRA. Longitud de Grano ANCHOGRA. Ancho de Grano P100GRA. Peso de 100 Granos PESOGRA. Peso de Grano (10 mazorcas) MAZENT. Peso de 10 Mazorcas PORGRA. Porcentaje del Peso de Grano en el total de la Mazorca. (\*) Todas las variables son significativas (P<0.15)

Todas las variables categóricas aportan significativamente a la clasificación de razas y de 7 grupos (tablas 3 y 6), pero para los 7 grupos se obtienen valores de chi-cuadrado de máxima verosimilitud mayores que para razas. Las variables categóricas más importantes que definen la clasificación de razas y los 7 grupos son textura de grano, color de endosperma y color de aleurona.

Este resultado es muy importante ya que si se cuenta con una adecuada selección de las variables, el método de clasificación numérica, utilizando toda la información disponible, encuentra una partición óptima de la población. Es importante además que este comportamiento se muestra tanto en variables continuas como en discretas, señalando la participación de los dos tipos de variables en el agrupamiento.

#### 4.4.3 Distancias entre grupos

Las razas que más se diferencian por las variables continuas son la raza Pisingallo (R9), Dente Branco (R1) y Canario de Ocho (R5). Las razas que más se diferencian para las variables categóricas son la raza Pisingallo (R9), Morotí (R2), Dente Branco (R1) y Cristal (R10). Si se toma la distancia de Krzanowski para mezcla de variables se puede observar que las razas más diferenciadas son la raza Pisingallo (R9), Morotí (R2), Dente Branco (R1) y Cristal (R10) (Tabla 12).

Los grupos que más se diferencian del resto de grupos para las variables continuas son el G7, G1 y G4. Para las variables categóricas, los grupos que más se diferencian del resto son G7, G1 y G2. Al realizar la distancia para mezcla de variables se puede observar que los grupos que se encuentran más diferenciados son los mismos que para discretas (Tabla 16). Se puede observar que hay una asociación muy clara entre los grupos que más se diferencian y su estructura interna (Tabla 4). El grupo G7 es el grupo que contiene a todos los individuos pertenecientes a la raza Pisingallo (R9), que se diferencian claramente por presentar textura de grano reventón así como características de planta muy diferenciadas, con individuos más pequeños y macolladores. El grupo G1 contiene mayoritariamente a los individuos pertenecientes a la raza Dente Branco (R1) que son los únicos dentados de la colección y presentan individuos con valores altos para las variables continuas. El grupo G2 también está bien definido, contiene a los individuos pertenecientes a las razas Morotí (R2) y Cristal (R10), siendo que Cristal aparentemente tienen como ancestro a la raza Morotí. A partir de las medidas de distancia se pone en evidencia la razón por la que la raza Cateto Sulino (R6) fue dividida. Presenta una distancia pequeña hacia la mayoría de razas con textura de grano duro. Además se observa que el grupo G3 se encuentra muy cerca del G6 para las variables continuas, pero se encuentra bien separado para las variables discretas (Tablas 12 y 13).

**Tabla 12. Distancia y similitudes entre razas para todas las variables utilizadas en la clasificación de la Colección de Maíz del Uruguay**

**Distancia de Mahalanobis para variables continuas**

Raza	1	2	3	4	5	6	7	8	9	10	
1	0	9.6	11.8	11.6	23.0	13.4	28.6	18.4	64.6	11.9	
2	9.6	0	13.9	10.2	19.1	8.3	16.9	16.9	40.6	5.6	
3	11.8	13.9	0	1.9	11.8	6.0	5.4	5.4	57.9	13.6	
4	11.6	10.2	1.9	0	11.5	3.0	2.9	2.9	49.2	9.4	
5	23.0	19.1	11.8	11.5	0	9.5	15.4	15.4	46.0	20.2	
6	13.4	8.3	6.0	3.0	9.5	0	5.4	5.4	44.3	8.4	
7	28.6	20.3	15.7	10.4	17.5	10.7	6.2	6.2	28.7	17.4	
8	18.4	16.9	5.4	29.1	15.4	5.4	0	0	42.1	13.1	
9	64.6	40.6	57.9	49.2	46.0	44.3	42.1	42.1	0	40.3	
10	11.9	5.6	13.6	9.4	20.2	8.4	13.1	13.1	40.3	0	
Promedio <sup>(†)</sup>	19.26	14.45	13.79	15.84	17.40	10.89	15.55	12.59	41.36	13.99	17.51

**Distancia (1 – Similitud) para las variables discretas**

Raza	1	2	3	4	5	6	7	8	9	10	
1	0	0,9887	1	1	1	1	1	1	1	0,8168	
2	0,9887	0	1	1	1	1	1	1	1	0,9564	
3	1	1	0	0,363	0,7959	0,5991	0,7064	0,5844	1	1	
4	1	1	0,363	0	0,5906	0,3082	0,5805	0,3218	1	1	
5	1	1	0,7959	0,5906	0	0,4568	0,9209	0,9	1	1	
6	1	1	0,5991	0,3082	0,4568	0	0,4621	0,4033	1	1	
7	1	1	0,7064	0,5805	0,9209	0,4621	0	0,3246	1	1	
8	1	1	0,5844	0,3218	0,9	0,4033	0,3246	0	1	1	
9	1	1	1	1	1	1	1	1	0	1	
10	0,8168	0,9564	1	1	1	1	1	1	1	0	
Promedio <sup>(†)</sup>	0,978	0,994	0,783	0,685	0,852	0,692	0,777	0,726	1	0,975	0,846

**Distancia de Krzanowski para mezcla de variables**

Raza	1	2	3	4	5	6	7	8	9	10	
1	0.0141	1.4118	1.4142	1.4142	1.4142	1.4142	1.4142	1.4142	1.4142	1.3845	
2	1.4118	0	1.4142	1.4142	1.4142	1.4142	1.4142	1.4142	1.4142	1.3988	
3	1.4142	1.4142	0	0.9965	1.3806	1.2724	1.3849	1.2564	1.4142	1.4142	
4	1.4142	1.4142	0.9965	0	1.3440	1.0232	1.3308	1.0283	1.4142	1.4142	
5	1.4142	1.4142	1.3806	1.3440	0	1.2925	1.4079	1.4038	1.4142	1.4142	
6	1.4142	1.4142	1.2724	1.0232	1.2925	0	1.3101	1.1806	1.4142	1.4142	
7	1.4142	1.4142	1.3849	1.3308	1.4079	1.3101	0	1.1733	1.4142	1.4142	
8	1.4142	1.4142	1.2564	1.4016	1.4038	1.1806	1.1733	0	1.4142	1.4142	
9	1.4142	1.4142	1.4142	1.4142	1.4142	1.4142	1.4142	1.4142	0	1.4142	
10	1.3845	1.3988	1.4142	1.4142	1.4142	1.4142	1.4142	1.4142	1.4142	0	
Promedio <sup>(†)</sup>	1.41	1.41	1.33	1.31	1.39	1.30	1.36	1.30	1.41	1.41	1.363

(†) promedio de la distancia entre grupos

**Tabla 13. Distancia y similitudes entre grupos (7 grupos Ward-MLM) para todas las variables utilizadas en la clasificación de la Colección de Maíz del Uruguay**

Distancia de Mahalanobis para variables continuas

	G1	G2	G3	G4	G5	G6	G7	
G1	0	15.1	12.8	35.8	46.2	23.4	76.3	
G2	15.1	0	13.4	30.8	29.1	15.6	45.8	
G3	12.8	13.4	0	17.6	19.7	5.2	51.9	
G4	35.8	30.8	17.6	0	21.6	19.6	69.3	
G5	46.2	29.1	19.7	21.6	0	13.6	29.9	
G6	23.4	15.6	5.2	19.6	13.6	0	47.2	
G7	76.3	45.8	51.9	69.3	29.9	47.2	0	
promedio <sup>(f)</sup>	29.9	21.4	17.2	27.8	22.9	17.8	45.8	26.11

Distancia (1 – Similitud) para las variables discretas

	G1	G2	G3	G4	G5	G6	G7	
G1	0	0,905	0,908	0,967	0,98	0,993	1	
G2	0,905	0	0,807	0,789	0,821	0,799	0,978	
G3	0,908	0,807	0	0,292	0,782	0,945	1	
G4	0,967	0,789	0,292	0	0,528	0,556	1	
G5	0,98	0,821	0,782	0,528	0	0,312	1	
G6	0,993	0,799	0,945	0,556	0,312	0	1	
G7	1	0,978	1	1	1	1	0	
Promedio <sup>(f)</sup>	0,959	0,850	0,789	0,689	0,737	0,768	0,996	0.827

Distancia de Krzanowski para mezcla de variables

	G1	G2	G3	G4	G5	G6	G7	
G1	0.03	1.40	1.40	1.41	1.41	1.41	1.41	
G2	1.40	0.02	1.39	1.41	1.41	1.39	1.41	
G3	1.40	1.39	0	1.36	1.40	1.39	1.41	
G4	1.41	1.41	1.36	0	1.39	1.39	1.41	
G5	1.41	1.41	1.40	1.39	0	1.32	1.41	
G6	1.41	1.39	1.39	1.39	1.32	0	1.41	
G7	1.41	1.41	1.41	1.41	1.41	1.41	0	
Promedio <sup>(f)</sup>	1.410	1.404	1.393	1.396	1.392	1.387	1.414	1.399

(f) promedio de la distancia entre grupos

La distancia de Mahalanobis promedio para los 7 grupos es mayor (49% mayor) que para las razas. Esto indica que los 7 grupos se encuentran mejor separados entre sí, para las variables continuas que las razas. Para las variables discretas se da la situación inversa, la distancia promedio es mayor en las razas que en los 7 grupos (2% mayor). Para mezcla de variables, es mayor la distancia promedio de los grupos que de las razas (2% mayor).

Los valores de la distancia propuesta por Matusita (1959) se encuentran entre 0 a 1.41, esto hace que la interpretación sea más compleja. Al utilizar el vector  $W$  para el cálculo, pondera en forma importante a las variables categóricas ya que se generan muchas celdas vacías. Estas celdas vacías determinan que la similitud sea cero, por lo tanto, al introducirlo en la fórmula de Krzanowski no se toma en cuenta a la distancia para las variables continuas y se le coloca la máxima distancia para mezcla de variables. Por ejemplo, la raza 1 se diferencia de las razas 3 a la 9 en diferente forma para las variables continuas, pero al presentar similitud cero para las variables discretas, quedan igual de distanciadas para la mezcla de variables. El hecho de que existan celdas cero para la variable  $W$  entre dos grupos no es razón para afirmar que se encuentran distanciadas lo máximo posible, ya que pueden coincidir para algunas variables categóricas, pero para otras no. Siguiendo el ejemplo de las razas 3 y 9, ambas presentan perisperma mayoritariamente incoloro. Aún queda mucho que estudiar para encontrar mejores medidas de distancia para las variables discretas y para las mezclas de variables.

#### 4.4.4 Estadísticos multivariados

La relación variabilidad dentro de grupos respecto a la variabilidad total ( $\lambda$  de Wilks) es mayor para las razas que para los 7 grupos (tabla 14). La suma de varianzas entre grupos referida al total (traza de Pillai) y referida a la varianzas dentro de grupos (traza Hotelling-Lawley), así como la mayor relación de la varianzas entre y dentro de grupos (mayor raíz de Roy) es mayor para los 7 grupos que para las razas. Por lo tanto, en todos los estadísticos multivariados la aproximación  $F$  es mayor en los 7 grupos que en las razas. Todos estos estadísticos tienen que ver con el análisis de varianzas, por lo que este era un resultado esperable en la medida en que todos los valores de la prueba  $F$  tanto para el análisis univariado como para el multivariado eran mayores para los 7 grupos que para las razas.

#### 4.4.5 Calidad de los agrupamientos

En cuanto a las probabilidades de pertenencia los 7 grupos están bien definidos. Las observaciones presentan en promedio una probabilidad de pertenencia muy alta: 0.966 (tabla 15). El número de observaciones clasificadas que presentan probabilidad de pertenencia menor al 0.75 son solamente el 5%; y hay solamente una observación



clasificada con una probabilidad menor a 0.5. Los grupos G3, G4 y G6 son los que presentan menores probabilidades de pertenencia y más observaciones con probabilidad menor a 0.75. En este sentido, el grupo G7 y el G1 son los que están mejor definidos.

**Tabla 14. Estadísticos multivariados para razas y grupos (7 y 16 grupos Ward-MLM) del análisis de la Colección de Maíz del Uruguay**

Estadístico	Razas		7 Grupos		16 Grupos	
	(S=9 Valor	M=3,5 N=406 F <sup>(*)</sup>	(S=6 Valor	M=5 N=407,5 F <sup>(*)</sup>	(s=15 Valor	M=0.5 N=403 F <sup>(*)</sup>
Wilks' Lambda	0,059	18,05	0,015	49,53	0,001	29,93
Pillai's Trace	2,039	14,16	2,765	41,32	4,615	21,49
Hotelling-Lawley Trace	4,227	22,44	7,141	57,08	11,620	36,73
Roy's Greatest Root	1,799	86,98	2,922	141,28	3,491	168,81

(\*) valores significativos (P=0.0001)

**Tabla 15. Probabilidad de pertenencia promedio a cada grupo y número de observaciones clasificadas con probabilidad menor a 0.75 y 0.5 en cada grupo de los 7 grupos Ward-MLM**

GRUPO	PROMEDIO	N (P< 0.75)	N (P< 0.5)
1	0.983	2	0
2	0.975	3	0
3	0.963	17	1
4	0.920	10	0
5	0.965	2	0
6	0.970	10	0
7	1.000	0	0
TOTAL	0.966	44	1

#### 4.4.6 Análisis canónico

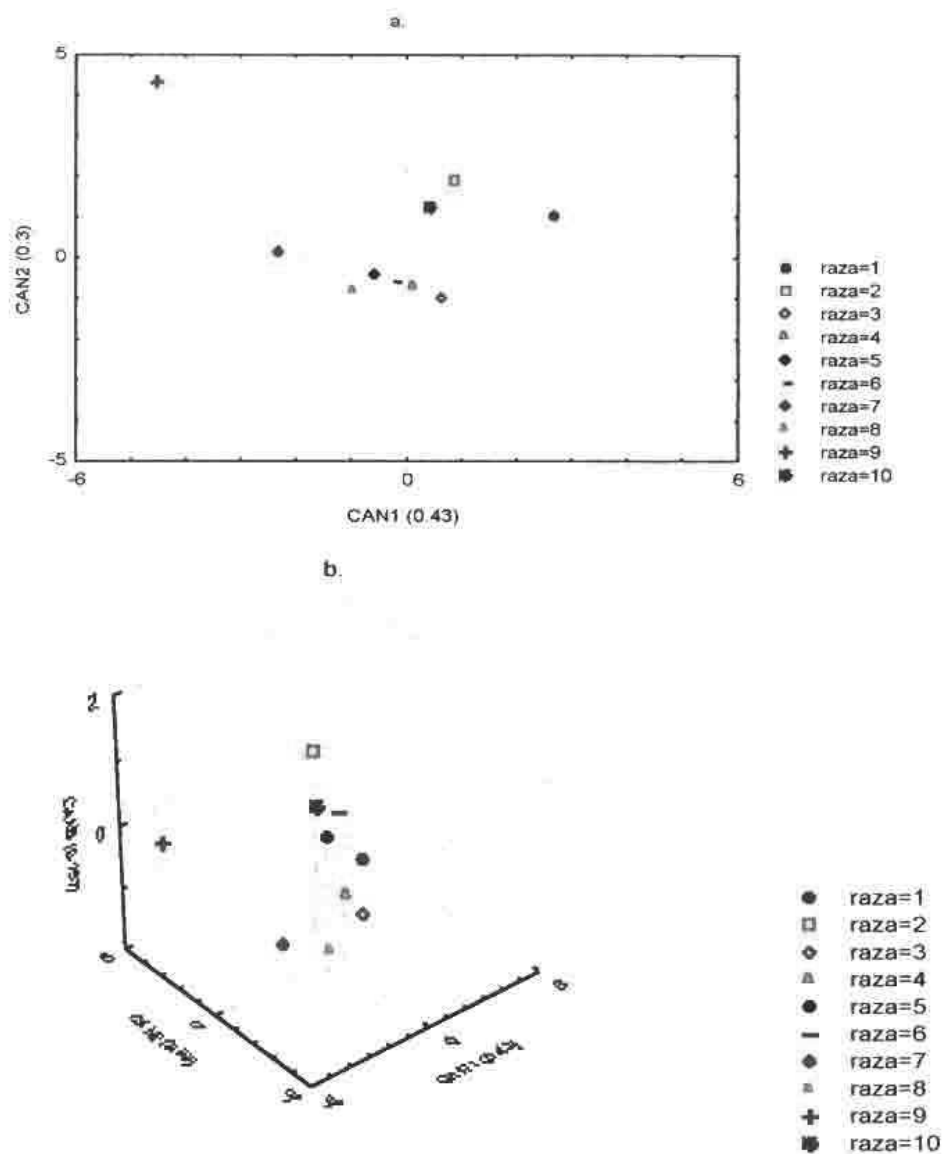
Los coeficientes de correlación entre las variables originales y las tres primeras variables canónicas se muestran en la tabla 16. La variable canónica que más explica la variación entre razas (43%) se encuentra muy correlacionada con variables de grano (peso de 100 granos, ancho de grano y longitud de grano). Peso de 100 granos es la variable continua que explica más a las razas (Stepwise discriminante). Ancho y longitud de grano no se encuentran entre las primeras del stepwise discriminante, tienen una correlación de 0.5 con peso de 100 granos y de 0.8 entre ellas. Otra variable correlacionada con la primer variable canónica en las razas es altura de mazorca, en el

Stepwise discriminante altura de mazorca entró 3°. La segunda variable canónica (30%) está muy correlacionada con índice de macollaje y variables reproductivas (floración femenina y floración masculina). Índice de macollaje es una de las variables continuas que más aporta a la diferenciación en razas. Los días a floración no son tan importantes en el stepwise discriminante, están muy correlacionadas entre ellas (0.9) y poco con índice de macollaje (0.2). La tercera variable canónica (15%) está muy correlacionada con variables de mazorca (longitud de mazorca y peso de la mazorca entera) y espesor de grano. Longitud de mazorca es una de las primeras variables continuas en el stepwise discriminante, le sigue peso de mazorca entera y espesor de grano (tabla 16).

**Tabla 16. Correlación de las variables originales con las variables canónicas para las razas y grupos (7 y 16 grupos Ward-MLM) de la Colección de Maíz del Uruguay**

Variable	RAZAS			7 GRUPOS			16 GRUPOS		
	CAN1	CAN2	CAN3	CAN1	CAN2	CAN3	CAN1	CAN2	CAN3
Floración Masculina	0.220	0.352 **	0.174	0.234	0.465 **	-0.012	0.409 *	-0.362	-0.053
Floración Femenina	0.234	0.355 **	0.184	0.228	0.434 **	0.036	0.368	-0.349	-0.038
Altura de Planta	0.338	-0.210	-0.187	0.244	-0.200	0.109	0.187	0.175	0.268
Altura de Mazorca	0.434 *	0.105	-0.103	0.339	0.058	0.109	0.381	-0.031	0.204
Índice de Prolificidad	-0.340	0.001	-0.171	-0.245	-0.022	-0.147	-0.255	-0.049	-0.009
Índice de Macollaje	-0.239	0.576 **	0.175	-0.112	0.517 **	-0.148	0.004	-0.434	-0.266
Vuelco	-0.069	-0.171	0.173	-0.008	-0.097	0.228	-0.037	-0.005	0.224
Relación espiga-forraje	-0.097	-0.305 *	-0.273	-0.158	-0.561 **	-0.499 **	-0.340	0.578 **	-0.160
Forraje Residual	0.081	0.087	0.083	0.061	0.191	0.233	0.137	-0.258	0.220
Longitud de Mazorca	0.148	-0.303 *	0.515 **	0.103	-0.226	0.317 *	0.082	0.037	0.407 **
Espesor de Grano	0.378	0.186	-0.367 *	0.332	0.012	-0.185	0.259	0.210	-0.146
Longitud de Grano	0.436 **	-0.030	-0.141	0.541 **	-0.239	-0.235	0.416 *	0.500 **	-0.125
Ancho de Grano	0.507 **	0.043	-0.192	0.638 **	-0.163	-0.219	0.536 **	0.435 **	-0.233
Peso 100 Granos	0.661 **	-0.049	0.198	0.536 **	-0.150	0.025	0.429 *	0.292	0.042
Peso de Grano	0.262	-0.310 *	-0.323 *	0.222	-0.363 *	0.010	0.084	0.396 *	0.145
Peso Mazorca entera	0.267	-0.316 *	-0.389 *	0.203	-0.349 *	-0.011	0.077	0.385 *	0.126
Porcentaje de Grano	0.032	-0.103	0.108	0.070	-0.145	0.129	0.040	0.153	0.326 *
%									
Explicación	0.43	0.30	0.15	0.41	0.25	0.17	0.30	0.22	0.12

Los centroides de las razas Dente Branco (R1), Morotí (R2), Pisingallo (R9) y Cristal (R10) se encuentran separados por la primera variable canónica (Figura 4). En la gráfica con los primeros tres ejes canónicos se puede observar como por ejemplo las razas Canario de Ocho (R5) y Cateto Sulino Grosso (R8) se encuentran separadas por la tercer variable canónica. Esto es, por características de la mazorca principalmente. La correlación entre las variables originales y las variables canónicas es similar a la obtenida



**Figura 4. Ubicación de los centroides de las razas respecto de las dos y tres primeras variables canónicas y porcentaje de explicación de cada variable. a) centroides de las razas en las dos primeras variables canónicas, b) centroides de las razas en las tres primeras variables canónicas**

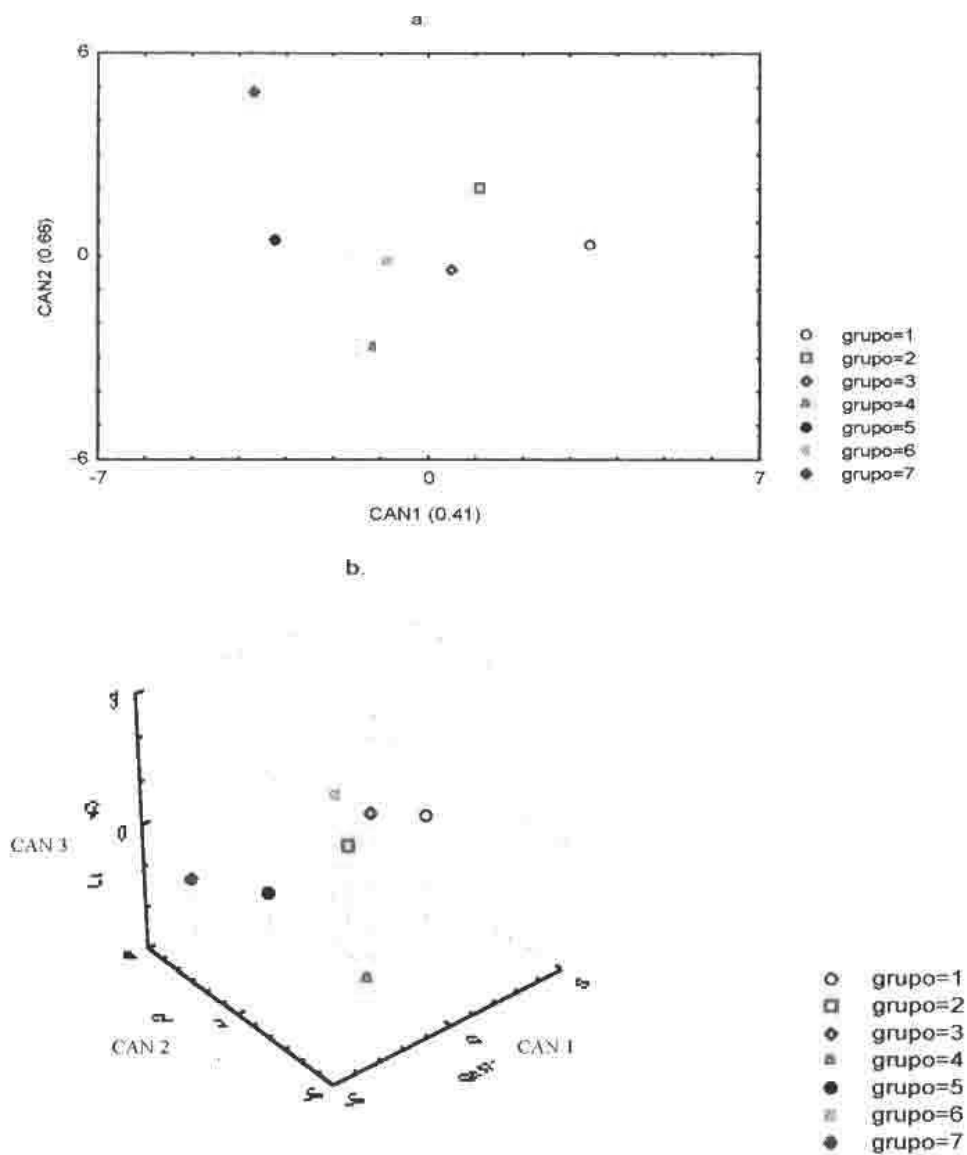
por Malosetti y Abadie (2000) respecto a los componentes principales (PCA). Dente Branco (R1), Morotí (R2) y Pisingallo (R9) se separaron bien por los PCA. El resto de razas se concentró en un único grupo.

Los 7 grupos formados por Ward-MLM presentan correlaciones similares a las razas entre las variables originales y las dos primeras variables canónicas. Las variables correlacionadas con la variable canónica 1 son ancho de grano, peso de 100 granos y longitud de grano. Ancho de grano es la primer variable del análisis stepwise discriminante, peso de 100 granos está en el quinto lugar y longitud de grano en 14° lugar, aunque esta última está muy correlacionada con ancho de grano. La segunda variable canónica se correlaciona con la relación espiga forraje y con índice de macollaje, floración masculina y floración femenina. La relación espiga forraje es la segunda variable del stepwise discriminante, índice de macollaje la tercera y floración masculina la cuarta. Floración femenina está en el 16° lugar, pero se encuentra muy correlacionada con floración masculina. La tercera variable canónica está correlacionada en menor medida con las variables de mazorca que en la clasificación de razas, y se correlaciona además con la relación espiga forraje. El porcentaje de explicación de la relación varianza entre-varianza dentro de grupos es 41%, 25% y 17% para las primeras tres canónicas respectivamente. Los centroides de los 7 grupos Ward-MLM se encuentran bien separados tanto por las dos primeras como por las tres primeras variables canónicas (Figura 5).

#### 4.4.7 Recuperación de la varianza fenotípica

Se calculó el promedio por variable, de la varianza fenotípica de las 100 muestras tomadas de cada agrupamiento (7 grupos y razas). Los 7 grupos y las razas recuperaron aproximadamente la misma varianza fenotípica. No hay grandes diferencias entre las razas y los 7 grupos (2.5 veces como máximo en todas las variables salvo en la relación espiga forraje que es un 6% mayor en los grupos que las razas). De todas formas para los 7 grupos es un poco mejor la recuperación de la varianza fenotípica en 10 de las 17 variables. Los rangos de las varianzas se mantienen aproximadamente igual que las medias.

El método de muestreo estratificado con asignación proporcional parece no ser el mejor método de muestreo. Se han hecho estudios sobre el método de muestreo más adecuado para seleccionar individuos en una colección núcleo, pero los estudios utilizan un número reducido de muestreos. Se podría realizar un número importante de muestreos utilizando diferentes técnicas y luego comparar qué método es el más adecuado. Esto se podría realizar sobre la colección del Uruguay con las dos clasificaciones.



**Figura 5.** Ubicación de los centroides de los grupos (7grupos Ward-MLM) respecto de las dos y tres primeras variables canónicas y porcentaje de explicación de cada variable. a) centroides de los grupos(7gruposWard-MLM) en las dos primeras variables canónicas, b) centroides de los grupos(7gruposWard-MLM) en las tres primeras variables canónicas

**Tabla 17. Promedio de las medias y varianzas de las variables continuas de 100 muestras estratificadas tomando el 20% de cada grupo (raza).**

Variable	Media		Varianza	
	razas	7 grupos	razas	7 grupos
Floración Masculina	77.5	77.4	31.090	30.553
Floración Femenina	82.8	82.7	54.202	53.308
Altura de Planta	169.5	169.6	211.299	216.142
Altura de Mazorca	107.4	107.4	210.055	211.848
Índice de Prolificidad	0.812	0.815	0.066	0.065
Índice de Macollaje	1.452	1.448	0.086	0.085
Vuelco	76.9	76.9	170.413	171.250
Relación espiga-forraje	0.241	0.244	0.017	0.018
Forraje Residual	19.7	19.6	33.114	33.430
Longitud de Mazorca	15.3	15.3	3.827	3.769
Espesor de Grano	6.0	6.0	0.901	0.926
Longitud de Grano	10.4	10.4	1.798	1.802
Ancho de Grano	8.9	8.9	1.433	1.438
Peso 100 Granos	26.2	26.1	38.361	38.593
Peso de Grano	0.747	0.751	0.042	0.042
Peso Mazorca entera	0.966	0.991	0.064	0.065
Porcentaje de Grano	75.5	75.5	40.499	40.208

## **5. CONCLUSIONES**

Es conveniente recordar los objetivos del estudio que son comparar la clasificación racial con una clasificación numérica, estudiar las características de los grupos y razas y entender en poco mejor la utilidad de los métodos numéricos en las clasificaciones. Los objetivos específicos son: realizar clasificaciones numéricas de los Recursos Genéticos de la colección de germoplasma de maíz de Uruguay y comparar la clasificación numérica con la clasificación racial.

En cuanto al objetivo de estudiar las características de los grupos y las razas, se observa que el agrupamiento con 7 grupos mantiene una estructura similar a la de las razas, conservando prácticamente la estructura interna de las razas más diferenciadas. Estas razas son Dente Branco (R1), Morotí (R2), Cuarentino (R7), Pisingallo (R9) y Cristal (R10). En cambio a la raza Cateto Sulino (R6) que es la más numerosa de la colección y se encuentra muy cerca de las otras razas, la divide principalmente en dos grupos más homogéneos, con menor coeficiente de variación y varianza. También divide a la raza Cateto Sulino Grosso (R8), Dente Riograndense (R3) y Semidentado Riograndense (R4), todas ellas muy similares a la raza Cateto Sulino.

Respecto a la comparación de la clasificación numérica con la clasificación racial se observa: 1) los 7 grupos Ward-MLM respecto a las variables estudiadas son más homogéneos dentro y heterogéneos entre que las razas, 2) los 7 grupos se construyeron utilizando más información y utilizándola de mejor forma que las razas, 3) el agrupamiento en 7 grupos utiliza tanto información de variables continuas como de variables discretas, 4) no hubo ganancia importante en la varianza fenotípica retenida en el muestreo proporcional de los 7 grupos respecto de las razas, 5) la raza Cateto Sulino que presenta distancias pequeñas hacia las otras razas fue dividida en dos grupos más separados de los otros. Por lo tanto, dado un conjunto de observaciones y variables, la clasificación numérica produce una partición con muy buenas características numéricas, mejor en término de los criterios utilizados que la clasificación racial.

## 6. RESUMEN

La diversidad genética vegetal es el componente principal de cualquier ecosistema, siendo fundamental una buena clasificación de los recursos genéticos para lograr un mejor manejo de los materiales. En maíz existen muchas clasificaciones de los recursos genéticos, entre ellas la clasificación racial que ha sido muy utilizada, pero que ha tenido serios cuestionamientos. A partir de los años 80, con el desarrollo de la computación la taxonomía numérica comienza a cobrar importancia; se comienza a utilizar métodos multivariados para la clasificación de recursos genéticos. Este estudio tiene como objetivos comparar la clasificación racial con una clasificación numérica, estudiar las características de los grupos y razas y entender en poco mejor la utilidad de los métodos numéricos en las clasificaciones. Los objetivos específicos son: a) realizar clasificaciones numéricas de los Recursos Genéticos de la colección de germoplasma de maíz de Uruguay y b) comparar la clasificación numérica con la clasificación racial.

Se utilizó la base de datos generada por De María *et al.* (1979), publicada en el catálogo de Recursos Genéticos de Maíz de Sudamérica – Uruguay (Fernández *et al.*, 1983). Las colectas fueron evaluadas durante el año agrícola 1978/79 en la Estación Experimental Dr. A. Backhaus de la Facultad de Agronomía. La evaluación fue realizada en parcelas de un surco de 6 metros a una densidad de 47.620 plantas por hectárea. Se cuenta con 852 accesiones y 27 variables, incluyendo variables continuas y categóricas. La Colección de Germoplasma de Maíz de Uruguay fue clasificada por el Ing. Agr. L. De León en 10 razas: Dente Branco, Morotí, Dente Riograndense, Semidentado Riograndense, Canario de Ocho, Cateto Sulino, Cuarentino, Cateto Sulino Grosso, Pisingallo y Cristal.

Se realizaron clasificaciones numéricas de la Colección de Germoplasma de maíz de Uruguay. Para la clasificación numérica se utilizó el método desarrollado por Franco *et al.*, (1998a), que consiste en emplear una estrategia de clasificación en dos etapas. En la primera etapa se utiliza un método jerárquico con una función objetivo (Ward, 1963) para la formación de grupos iniciales. La segunda etapa consiste en la mejora de esos grupos por medio del modelo de localización modificado (MLM), que se acerca a un máximo, local o global, de la verosimilitud.

Luego de obtenida la mejor clasificación numérica a partir de la estrategia Ward-MLM, se la comparó contra la clasificación racial. Los criterios utilizados para la comparación de la clasificación racial con la clasificación numérica son los siguientes: 1) el número y la importancia de cada variable en cada una de las clasificaciones, 2) la distancia entre los grupos, 3) algunos estadísticos multivariados, 4) el promedio de las probabilidades de pertenencia a cada grupo, 5) análisis canónico y 6) la recuperación de la varianza fenotípica en un proceso de muestreo posterior al agrupamiento.



En la comparación de la clasificación numérica con la clasificación racial se observa: a) el agrupamiento en 7 grupos mantiene una estructura similar a la de las razas, mantiene la estructura de las razas más diferenciadas y divide a las otras, b) los 7 grupos Ward-MLM tienen una relación varianza entre grupos – varianza dentro de grupos mayor que las razas para el análisis univariado, c) la asociación entre variables discretas y grupos o razas es mayor para los 7 grupos que para las razas, d) las variables continuas más importantes son 12 en los 7 grupos y 6 en las razas, de las cuales comparten 5, e) el promedio de la distancia de mahalanobis para las variables continuas es un 49% mayor en los 7 grupos que en las razas, para las variables discretas es un 2% mayor en las razas y para la mezcla de variables también es un 2% mayor en las razas, f) en todos los estadísticos multivariados la F es mayor para los 7 grupos que para las razas, g) las variables que definen la representación gráfica son las mismas en razas que en 7 grupos, h) en el muestreo estratificado se recupera aproximadamente la misma varianza fenotípica en las razas que en los 7 grupos, i) la partición de la raza Cateto Sulino en dos grupos produce grupos más homogéneos, con menor coeficiente de variación y varianza.

Respecto a la comparación de las clasificaciones se concluye que: 1) los 7 grupos Ward-MLM son más homogéneos dentro y heterogéneos entre sí que las razas respecto a las variables estudiadas, 2) los 7 grupos se construyeron utilizando más información y utilizándola de mejor forma que las razas, 3) el agrupamiento en 7 grupos utiliza tanto información de variables continuas como de variables discretas, 4) no hubo ganancia importante en la varianza fenotípica retenida en el muestreo proporcional de los 7 grupos respecto de las razas, 5) la raza Cateto Sulino que presenta distancias pequeñas hacia las otras razas fue dividida en dos grupos más separados de los otros. Por lo tanto, dado un conjunto de observaciones y variables, la clasificación numérica produce una partición con muy buenas características numéricas, mejor en término de los criterios utilizados que la clasificación racial.

## **7. BIBLIOGRAFIA**

- ABADIE, T; MAGALHAES, JR; CORDEIRO, CT; PARENTONI, S; DE ANDRADE, R. 1998. A clasificación for Brazilian maize landraces. *Plant Genetic Resources Newsletter* 114: 43-44.
- \_\_\_\_\_. MAGALHAES, JR; PARENTONI, S; CORDEIRO, CT; DE ANDRADE, R. 1999. The Core Collection of Maize Germplasm of Brazil. *Plant Genetic Resources Newsletter* 117: 55-56.
- AGUIRRE GOMEZ, JA; BELLON, MR; SMALE, M. 2000. A regional analysis of maize biological diversity in southestern Guanajuato, México. *Economic Botany* 54 (1): 60-72.
- ANDERSON, E; CUTTLER, H. 1942. Races of *Zea mays*: I. Their recognition and classification. *Annals of the Missouri Botanical Garden* 21: 69-88
- BASFORD, KE; McLACHLAN, GJ. 1985. The mixture method of clustering applied to three way data. *J. Classif.* 2:109-125.
- BINDER, DA. 1978. Bayesian cluster analysis. *Biometrika*. 65: 31-38
- BIRD, RMCK; GOODMAN, M.M. 1977. The races of maize. V. Grouping maize races on the bases of ear morphology. *Econ. Bot.* 23: 471-481.
- BRANDOLINI, A. 1970. Maize. In Frankel, OH; Bennett, E. (eds). *Genetic Resources in Plant: their exploration and conservation* Bell and Bain Ltd, Glasgow and The Kemp Hall Bindery, Oxford
- BRETTING, PK; GOODMAN, MM. 1989. Karyotypic variation in mesoamerican races of maize and its systematic significance. *Economic Botany* 43: 107-124.
- \_\_\_\_\_. GOODMAN, MM; STUBER, CW. 1990. Isozymatic variation in Guatemalan races of maize. *Am. J. Bot.* 77:211-225.
- BRIEGER, FG; GURGEL, JTA; PATERNIANI, E; BLUMENSCHNEIN, A; ALLEONI, MR. 1958. Races of maize in Brazil and other eastern South American countries. *Natl. Acad. Sci. – Natl. Res. Council Publ.* 593. Washington DC. 283 p.
- BROWN, AHD. 1989. Core Collections: A practical approach to genetic resources management. *Genome*. 31: 818-824
- BUSCH, L; LACY, WB; BUKHARDT, J. 1989. Ethical and policy issues. In Knutson, L; Stener, AK (eds) *Biotic Diversity and Germplasm Global Imperatives. Invited papers presented at symposium held 9-11 may 1988, Beltsville Agricultural Research Center, Beltsville, Maryland.* Klewer Academic Publishers, Netherlands. P 43-62.
- CROSSA, J; TABA, S; EBENHART, SA; BRETTING, P; VENCOVSKY, R. 1994. Practical considerations for maintaining germplasm in maize. *Theor. Appl. Genet.* 89:89-95.
- \_\_\_\_\_. BASFORD, K; TABA, S; DeLACY, I; SILVA, E. 1995. Three-mode analyses of maize using morphological and agronomic attributes measured in multilocation trials. *Crop Sci.* 35:1483-1491.

- DE MARÍA, F; FERNÁNDEZ, G; ZOPPOLO, G. 1979. Caracterización agronómica y clasificación racial de las muestras de maíz colectadas en Uruguay bajo el Proyecto IBPGR y Facultad de Agronomía. Tesis Ing. Agr. Universidad de la República, Uruguay.
- DEMPSTER, AP; LAIRD, NM; RUBIN, DB. 1977. Maximum likelihood for incomplete data via the EM algorithm (with discussion). *Journal of the Royal Statistics Society B.* 39: 1-38
- DOEBLEY, JF; GOODMAN, MM; STUBER, CW. 1983. Isozyme variation in maize from the southwestern United States: Taxonomic and Anthropological implications. *Maydica* 28: 97-120
- \_\_\_\_\_. GOODMAN, MM; STUBER, CW. 1985. Isozyme variation in races of maize from México. *Am. J. Bot.* 72:629-639
- EVENSON, RE; GOLLIN, D; SANTANIELLO, V. 1998. *Agricultural Values of Plant Genetic Resources.* CABI Publishing, Wallingford, UK. 285 p.
- FAO. 1996. Cumbre Mundial Sobre la Alimentación, 13-17 de noviembre de 1996. Roma, Italia. Organización de las Naciones Unidas para la Agricultura y la Alimentación.
- FERNANDEZ, G; FRUTOS, E; MAIOLA, C. 1983. Catálogo de Recursos Genéticos de Maíz de Sudamérica-Uruguay. EERA-Pergamino INTA CIRF. Pergamino, Argentina.
- FRANCO, J; CROSSA, J; VILLASEÑOR, J; TABA, S; EBERHART, SA. 1997a. Classifying Mexican maize accessions using hierarchical and density search methods. *Crop Sci.* 37 (3): 972-980.
- \_\_\_\_\_. CROSSA, J; DÍAZ, J; TABA, S; VILLASEÑOR, J; EBERHART, SA. 1997b. A sequential clustering strategy for classifying gene bank accessions. *Crop Sci.* 37 (5): 1656-1662.
- \_\_\_\_\_. 1998. Clasificación de observaciones utilizando variables discretas y continuas simultáneamente. Tesis PhD. Montecillo, México, Institución de Enseñanza e Investigación en Ciencias Agrícolas. 170p.
- \_\_\_\_\_. CROSSA, J; VILLASEÑOR, J; TABA, S; EBERHART, SA. 1998a. A Method for classifying genetic resources using categorical and continuous variables. *Crop Sci.*
- \_\_\_\_\_. CROSSA, J; VILLASEÑOR, J; TABA, S; EBERHART, SA. 1998b. A two stage three-way method for classifying genetic resources using categorical and continuous variables. *Crop Sci.*
- FRANKEL, O; BROWN, A; BURDON, J. 1995. *The Conservation of Plant Biodiversity.* Cambridge University Press. Cambridge. UK. 299p.
- GOODMAN, MM. 1967. The race of maize: I. The use of Mahalanobi's generalized distances to measure morphological similarity. *Fitotec. Latinoamer.* 4(1): 1-22.
- \_\_\_\_\_. 1968. The race of maize II: Use of multivariate analysis of variance to measure morphological similarity. *Crop. Sci.* 8:693-698.
- \_\_\_\_\_. 1972. Distance analysis in biology. *Syst. Zool.* 21:174-186.

- \_\_\_\_\_. 1976. Maize In: Simmonds, N.W. (Ed.) Evolution of Crop Plants. Longman Inc, New York. pp 128-136
- \_\_\_\_\_. PATERNIANI, EE. 1969. The races of maize. III. Choices of appropriate characters for racial classification. *Econ. Bot.* 23: 65-273.
- \_\_\_\_\_. BIRD, RMCK. 1977. The races of maize. IV. Tentative grouping of 219 Latin American races. *Econ. Bot.* 31:204-221.
- GOWER, JC. 1971. A general coefficient of similarity and some of its properties. *Biometrics* 27:857-874.
- GUARINO, L; RAMANATHA RAO, V; REID, R. 1995. Collecting Plant Genetic Diversity. Technical Guidelines. Cambridge University Press. Cambridge. UK.
- HALLAUER, AR; MIRANDA, JB. 1988. Quantitative Genetics in Maize Breeding. Iowa State University Press. United States of America.
- KAUFMAN, L; ROUSSEUW, PJ. 1990. Finding groups in data. Jhon Wiley & Sons, New York
- KLECKA, WR. 1980. Discriminant analysis. Sage University Paper Series on Quantitative Applications in the Social Science, series n° 7-19. Beverly Hills, Sage Publications, Thousand Oaks, CA.
- KRZANOWSKI, WJ. 1983. Distance between populations using mixed continuous and categorical variables. *Biometrika*. 70: 235-243
- KULESOV, NN. 1929, The geographical distribution of the varietal diversity of maize in the word. *Trudy po Prikladnoi Botanike, Genetike, i Seleksii (Bull. Appl. Bot., Genet. And Plant Breeding, Lenin Acad. Agric. Sci., URSS)* 20:506-510
- KULESOV, NN. 1933. Word's diversity of phenotypes of maize. *J. Am. Soc. Agron.* 25: 688-700.
- LAMP. 1992. Data of the Latin American Maize Proyect. CD-ROM
- LAWRENCE, CJ; KRZANOWSKI, WJ. 1996. Mixture separation for mixed-mode data. *Statistics and Computing*. 6: 85-92
- MAHALANOBIS, PC. 1930. On tests and measures of group divergences. *J. Proc. ASiat. Soc. Bengal* 26: 541-588
- MALOSETTI, M; ABADIE, T. 2000. Sampling strategy to develop a core collection of Uruguayan maize landraces based on morphological traits. *Plant Genetic Resources Newsletter*
- MANLY, BFJ. 1994. *Multivariate Statistical Methods: A primer*. 2ª. Edición. UK. Chapman & Hall. 215p.
- MARDIA, KV; KENT, JT; BIBBY, JM. 1979. *Multivariate Analysis*. Academic Press, London.
- MATUSITA, K. 1956. Decision rules based on the distance for problems of fit, two samples, and estimation. *Annals of Mathematical Statistics*. 8: 67-77

- McNEELY, JA; MILLER, KR; REID, WV; MITTERNEIER, RA; WERNER, TB. 1990. *Conserving the World's Biological Diversity*. IUCN, WRI, CIWLF – US, World Bank. Gland Switzerland and Washington D.C.
- MILLIGAN, GW; SOON, SC; SOKOL, LM. 1983. The effect of cluster size, dimensionality, and the number of clusters on recovery of true cluster structure. *IEEE Trans. Pattern Anal. Machine Intelligence* 5 (1): 40-47
- \_\_\_\_\_. COOPER, M. 1985. An examination of procedures for determining the number of clusters in data set. *Psychometrika*. 50-2: 159-179
- MOJENA, R. 1977. Hierarchical grouping methods and stopping rules: An evaluation. *The Computer Journal* 20: 359-363
- OLDFIELDS, ML. 1989. *The value of conserving Genetic Resources*. Sinauer Associates, Sunderland, Massachusetts.
- OLKIN, I; TATE, RF. 1961 Multivariate correlation models with mixed discrete and continuous variables. *Annals of Mathematical Statistics* 32:448-465
- PATERNIANI, E. 1970. Heterose em cruzamentos intervarietais de milho. Universidade de Sao Paulo, Escola Superior de Agricultura "Luiz de Queiroz," departamento e Instituto de genética Relatório Científico 1970:95-100. Piracicaba, S.P., Brazil, 158 p.
- \_\_\_\_\_. Capacidade de combinacao de linhagens exóticos em cruzamento intervarietal. Universidade de Sao Paulo, Escola Superior de Agricultura "Luiz de Queiroz," departamento e Instituto de genética Relatório Científico 1970:95-100. Piracicaba, S.P., Brazil, 128 p.
- PERRY, MC; Mc INTOSH, MS. 1991. Geographical patterns of variation in the USDA Soybean germplasm collection: I morphological traits. *Crop Sci.* 31: 1350-1355
- PINEDA; CROSSA, J. 1994. Forming core subsets from Tuxpeño race complex. In: S. Taba (ed). *The CIMMYT maize germplasm bank: genetic resource preservation, regeneration, maintenance, and use*. CIMMYT Maize Program Special Report. México DF, México.
- PFLÜGER, LA; SCHLATTER, AR. 1996. Isozyme variation in some races of maize from Argentina. *Genetic Resources and Crop Evolution* 43: 357-362
- SANCHEZ, JJ; GOODMAN, MM. 1992. Relationships among mexican races of maize. *Economic Botany* 46: 72-85
- SAS Institute Inc. 1990. *SAS/STAT® User's Guide, versión 6*. Fourth Ed., N.C.: SAS Institute Inc.
- SAS Institute Inc. 1996. *SAS/STAT® Software Changes and Enhancements trough Release 6.11*. Cary, NC. 1104 pp.
- SHIVA (eds) 1991. *Biodiversity: social and Ecological Perspectives*. Zed Books Ltd. London and New Jersey, with World Rainforest movement, Penang, Malaysia.

- STUREVANT, EL. 1899. Varieties of corn. U.S.D.A. Off. Expt. Sta. Bull. 57.
- TABA, S; DÍAZ, J; FRANCO, J; CROSSA, J. 1997. Evaluation of caribbean maize accessions to develop a core subset. *Crop Sci*
- WARD, J. 1963. Hierarchical grouping to optimize an objective function. *J. Amer. Stat. Assoc.* 58: 236-244.
- WELLHAUSEN, EJ; ROBERTS, E; HERNANDEZ, XE; in collaboration with Mangelsdorf PC. 1952. Races of maize in México: their origin, characteristics and distribution. The Bussey Institution of Harvard University. Cambridge, MA.
- WISHART, D. 1986. Hierarchical cluster analysis with messy data. In *Classification as a Tool of Research* (W. Gaul and M. Schader eds.). Elsevier Science Publishers B.V., Amsterdam, Holland p. 453-460
- WOLFE, JH. 1971. A Monte Carlo study of the sampling distribution of the likelihood ratio for mixture of multinormal distributions. *Naval Personnel and Training Res. Lab. Tech. Bull. STB 72-2*