

Caracterización de poblaciones nativas de maíz mediante Análisis de Procrustes Generalizado y Análisis Factorial Múltiple

Maize landraces characterization using Generalized Procrustes Analysis and Multiple Factor Analysis

Paola Zuliani ¹
Andrea Lavalle ¹

Sergio J. Bramardi ^{1,2}
Raquel Defacio ³

Originales: Recepción: 25/04/2011 - Aceptación: 29/12/2011

RESUMEN

El establecimiento de relaciones entre taxones es un paso esencial en el proceso de catalogación y evaluación del material conservado en un Banco de Germoplasma. Existen distintos métodos de evaluación en función del tipo de caracteres estudiados. Cuando el registro de caracteres se repite en el tiempo y en distintos ambientes, se debe separar la variabilidad intrínsecamente genética entre los taxones de aquella que se debe al ambiente, y más aún, de la posible variabilidad debida a la interacción genotipo*ambiente para el posterior establecimiento de relaciones puramente filogenéticas. En el presente trabajo se estudia comparativamente la factibilidad de aplicación de dos estrategias de análisis estadístico para dar solución a este problema. La primera corresponde al análisis tradicional donde se realiza un Análisis de Componentes Principales sobre los caracteres promedios a lo largo de los diferentes ambientes; y la segunda son métodos más complejos en los cuales cada dato es originado por tres modos: individuos, variables y condiciones ambientales, tales como el Análisis Factorial Múltiple y el Análisis de Procrustes Generalizado. Si bien las configuraciones resultantes fueron todas equivalentes, los métodos de tres vías permiten la interpretación de la interacción genotipo*ambiente.

ABSTRACT

The establishment of relationships among taxa is an essential step in the process of cataloging and evaluation of material conserved in a germplasm bank. There are several evaluation methods according to the types of the characters in the study. When the registration of the characters should be repeated in diverse environments and times, it is necessary to separate the genetic variability of the taxa from the variability due to the environment, and from the possible genotype*environment interaction variability. Consequently, pure phylogenetic relationships may be established. In this work, the feasibility of application of two strategies of statistical analysis to give solution to this problem is studied comparatively. The first one is a traditional Principal Component Analysis applied on the average characters. The second one is a set of more complex methods where each datum is originated by three ways: individuals, variables and environmental conditions, as the Multiple Factorial Analysis and the Generalized Procrustes Analysis. While the resulting configurations were all equivalent, three-way methods allow the interpretation of genotype*environment.

- 1 Dpto. de Estadística. Facultad de Economía y Administración. Universidad Nacional del Comahue. Buenos Aires 1400. (8300) Neuquén. Argentina. paola.zuliani@faea.uncoma.edu.ar
- 2 Dpto. de Ciencias Exactas. Facultad de Ciencias Agrarias y Forestales. Universidad Nacional de La Plata. Calle 60 y 119. (1900) La Plata. Argentina
- 3 Estación Experimental Agropecuaria. Instituto Nacional de Tecnología Agropecuaria (INTA). Avenida Frondizi (Ruta 32) km 4,5. (2700) Pergamino. Provincia de Buenos Aires. Argentina.

Palabras clave

datos de tres vías • interacción genotipo*ambiente • Análisis Factorial Múltiple • Análisis de Procrustes Generalizado

Keywords

three-way data • genotype*environment interaction • Multiple Factorial Analysis • Generalized Procrustes Analysis

INTRODUCCIÓN

Con el surgimiento de la agricultura, la domesticación de las especies ha sido fuertemente influenciada por la selección artificial. Esta y el efecto de la selección natural, hicieron que surgieran razas locales (variedades nativas, variedades de los agricultores o poblaciones locales) que mantenían la variabilidad, la adaptación al ambiente y cierta estabilidad en la producción (12).

El desarrollo industrial, la mecanización de la agricultura y las necesidades y exigencias crecientes del mercado, hicieron que los fitomejoradores se vieran en la necesidad de generar cultivos más uniformes, de rápido crecimiento, más productivos y de rendimiento estable a través del tiempo. Esta homogeneidad, que trae como consecuencia erosión génica y pérdida de variabilidad, hace que los cultivos sean más vulnerables a cualquier cambio ambiental y a la aparición de nuevas plagas y enfermedades (10, 14, 15).

La pérdida de variabilidad trajo aparejados muchos problemas en todo el mundo. Las consecuencias del estrechamiento de la base genética quedan demostradas con varios hechos que se han ido sucediendo a lo largo del tiempo. Por ejemplo, en 1970 disminuyó en un 50% la producción de maíz del estado de Iowa debido a una epifitía provocada por tizón (*Helminthosporium maydis*) dada la susceptibilidad del citoplasma estéril T (Texas) a la enfermedad. El citoplasma T se encontraba presente en el 75-90% de los híbridos cultivados. El problema fue resuelto a través del fitomejoramiento, utilizando recursos genéticos obtenidos de otras partes del mundo.

Cuando comenzaron a manifestarse los problemas de la excesiva uniformidad de los cultivos y se observó que las posibles soluciones se encontraban en las variedades locales, se revalorizó a nivel mundial la conservación y uso de la variabilidad genética disponible en los Bancos de Germoplasma.

En el caso del maíz en Argentina, en la década de 1950 empezaron a vislumbrarse las ventajas del uso de híbridos. Con ello, ocurrió el paulatino desplazamiento de las poblaciones locales por cultivares mejorados. Ante esta evidencia, el Grupo de Mejoramiento de Maíz de la Estación Experimental Agropecuaria (EEA) del Instituto Nacional de Tecnología Agropecuaria (INTA) Pergamino, Argentina, comenzó con las colectas de poblaciones locales y, en 1969, se creó el Banco Activo de Germoplasma (BAP) de maíz, el cual actualmente conserva 2569 entradas que representan una amplia variabilidad.

Sin embargo, la misma es poco utilizada. Una de las mayores dificultades para su uso estriba en que, en general, se dispone de un conocimiento bastante limitado del mérito genético de las mismas. Para que el germoplasma pueda ser utilizado en los programas de mejoramiento es necesario que, además de su correcta conservación, sea adecuadamente caracterizado y evaluado. Para ello es preciso implementar y desarrollar análisis estadísticos de manera que se posibilite la tipificación de las poblaciones y el establecimiento de relaciones entre ellas. A partir de esta información, los recursos fitogenéticos debidamente estudiados y catalogados pueden ser fácilmente consultados por los mejoradores en su búsqueda de germoplasma para los programas de mejoramiento (4).

En la evaluación morfológica de poblaciones nativas de maíz disponibles en los Bancos de Germoplasma se registran caracteres que se encuentran muy influenciados por el ambiente. Sin embargo, los mismos permiten evaluar las diferencias entre poblaciones, como también las características potenciales de las mismas para su uso en programas de mejoramiento. Por este motivo, es importante utilizar un diseño experimental adecuado que permita considerar el efecto del ambiente. Además de ello es conveniente evaluar, estimar y eliminar el efecto del ambiente y sus interacciones para así analizar sólo las diferencias debidas al genotipo.

Los datos que se tuvieron en cuenta para este trabajo de investigación, son una muestra de información agro-morfológica perteneciente a 31 poblaciones locales de maíz (*Zea mays* L.) de la raza Cristalino Colorado colectadas en la provincia de Buenos Aires, Argentina, conservadas en el BAP. Para la evaluación del material se llevaron a cabo ensayos en dos localidades: Pergamino (EEA INTA) y Ferré (Escuela Agrotécnica Salesiana).

De esta manera y en términos estadísticos, en cada ambiente se tendrá una tabla de dos modos: individuos por variables; y considerando toda la información se tiene un conjunto de dos tablas de individuos por variables. La información queda resumida en las denominadas Matrices Multivía en las cuales cada dato es originado por tres modos: individuos, variables y condiciones/ambientes.

El análisis tradicional implica la aplicación de técnicas factoriales en las cuales se debe, en primer lugar, organizar los datos en una colección de mediciones sobre p variables diferentes tomadas sobre la misma población. En este caso cada variable es el promedio de los valores correspondientes a la misma población en las distintas condiciones/ambientes.

Sin embargo, en los últimos años se han desarrollado técnicas estadísticas que posibilitan un análisis simultáneo de las tablas en estudio mediante la obtención de una estructura consenso capaz de sintetizar toda la información disponible, y además brindan la posibilidad de analizar el comportamiento de la interacción genotipo*ambiente.

Dentro de los métodos multivariados empleados en el tratamiento de los datos de tres vías, los métodos más usados son: el Análisis Factorial Múltiple (AFM) (9) y el

Análisis de Procrustes Generalizado (APG) (8, 11). Ambos pertenecen a perspectivas de estudio diferentes: el primero es un método de la escuela francesa y el segundo de la escuela anglosajona. Estos métodos tienen como objetivo principal la búsqueda de una estructura común (configuración consenso) para las diferentes condiciones difiriendo en la manera en que dicha estructura se construye. El AFM se basa en el teorema de descomposición de valores singulares (DVS) realizando un Análisis de Componentes Principales (ACP) sobre la tabla X que se obtiene colocando las tablas originales previamente ponderadas una junto a la otra (yuxtaposición), obteniendo la configuración consenso. En cambio, el APG es una técnica enteramente geométrica que se basa en la búsqueda de una configuración consenso óptima, en el sentido de aproximar lo máximo posible las distintas configuraciones asociadas a cada matriz individuos-variables.

Una de las características principales de estas técnicas frente a las tradicionales es que proporcionan resultados de gran interés desde el punto de vista de interpretación. Por un lado, ofrecen representaciones gráficas de la nube de individuos caracterizados por las variables de cada tabla (individuos parciales) y por el conjunto de variables (individuo medio) sobre el mismo plano factorial, pudiéndose observar las proximidades entre los puntos individuos de las diferentes tablas. Asimismo proporcionan una visión global de los datos y una medida de la relación entre las distintas tablas, coeficientes $RV(1)$ y de dimensionalidad de las tablas, coeficiente Ng .

Estas medidas ayudan a tomar decisiones sobre las tablas a eliminar del análisis, y analizarlas por separado, dado su comportamiento específico y distinto del resto. Por otro lado, en el AFM la representación de las variables permite realizar las interpretaciones clásicas de correlaciones y a su vez observar las proximidades entre los vectores variables de las diferentes tablas.

En el caso que se presenta, la configuración consenso establece una estructura común para los dos ambientes en que fueron evaluadas las poblaciones nativas de maíz, y a través de ella es posible caracterizar dichas poblaciones.

En este trabajo se comparan los resultados obtenidos de la aplicación del ACP tradicional sobre los caracteres promedios a lo largo de los diferentes ambientes frente a los métodos AFM y APG con el objetivo de buscar la metodología más adecuada a aplicar.

Estos últimos dos métodos separan la variabilidad intrínsecamente genética entre los taxones a evaluar de aquella que se debe al ambiente, y más aún, de la posible variabilidad debida a la interacción genotipo*ambiente para el posterior establecimiento de relaciones puramente filogenéticas.

También se expone una breve explicación de los métodos, presentando y comparando los resultados obtenidos al final del mismo. Todos los resultados son obtenidos en el entorno de programación y análisis de datos del software R (www.r-project.com).

MATERIALES Y MÉTODOS

Se aplicaron los métodos descriptos a una muestra de datos pertenecientes al Banco de Germoplasma de la EEA INTA Pergamino sobre evaluaciones de 31 poblaciones nativas de maíz de la raza Cristalino Colorado. Sobre la base de la lista de descriptores para maíz del IPGRI (International Plant Genetic Resources Institute) (7), se evaluaron 10 variables cuantitativas de este cultivo de importancia para la región (tabla 1).

Tabla 1. Caracteres agronómicos-morfológicos.

Table 1. Agro-morphological characters.

Código	Carácter
ALTPL ¹	Altura de la planta (cm)
ALTMZ ¹	Altura de la mazorca (cm)
LMZ ²	Longitud de mazorca (cm)
DMZ ²	Diámetro de mazorca (cm)
AGR ³	Ancho de grano (mm)
LGR ³	Largo de grano (mm)
PROL	Prolificidad (espigas/planta)
REND	Rendimiento (g)
GRXM	Granos por metro
PESO	Peso de 1000 granos (g)

(1) Valor promedio de la observación de 10 plantas.

(2) Valor promedio de la observación de 10 mazorcas.

(3) Valor promedio de la observación de 10 granos.

(1) Average value of the observation of 10 plants.

(2) Average value of the observation of 10 ears.

(3) Average value of the observation of 10 grains.

La experiencia se llevó a cabo en dos ambientes de la zona norte de la provincia de Buenos Aires: Pergamino (EEA INTA) y Ferré (Escuela Agrotécnica Salesiana), durante el ciclo agrícola 2003/04, en un diseño en látice.

Análisis de componentes principales sobre la tabla promedio

Dado un conjunto de n individuos y p variables, el Análisis de Componentes Principales (ACP) busca representar adecuadamente esta información en un número menor de variables construidas como combinaciones lineales de las originales, representando óptimamente las observaciones en un espacio de dimensión reducida, lo cual permite identificar variables latentes (13).

Además, al transformar las variables originales, en general correlacionadas, en nuevas variables no correlacionadas, facilita la interpretación de los datos en el sentido de que permite inferir sobre relaciones lineales entre observaciones y variables.

En el caso que se presenta, la tabla a la cual se aplica el ACP se construye a partir del promedio de las observaciones en los dos ambientes o localidades (figura 1).

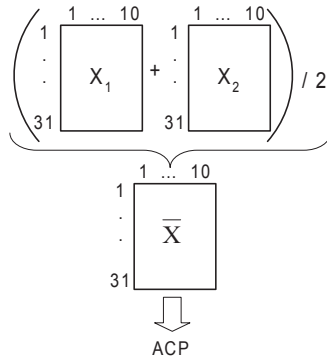


Figura 1. Esquema de la tabla promedio de las observaciones a través de las localidades evaluadas sobre la que se realizó un ACP.

Figure 1. PCA scheme on average of observations across locations table.

Análisis Factorial Múltiple

El Análisis Factorial Múltiple (AFM), desarrollado por Escofier y Pagès (9), en el seno de la Escuela Francesa de Análisis de Datos, es un método factorial adaptado al tratamiento de tablas de datos en las que un mismo conjunto de individuos se describe a través de varios grupos de variables. Los grupos de variables pueden surgir de la utilización conjunta de variables de diferente naturaleza, cuantitativas o cualitativas, o cuando las mismas variables son observadas bajo diferentes condiciones (2). La única restricción es que las variables que integran un grupo sean de la misma naturaleza. Cuando las variables son cuantitativas, como en este caso, el AFM se basa en la metodología de ACP y consta de dos etapas:

1. A cada grupo de variables se asocia una configuración euclídea o nube de individuos denominada nube parcial, que será analizada por separado, obteniendo los factores parciales. Es decir, realiza un ACP de cada una de las tablas X_k , donde X_k recoge la valoración de las p variables del grupo k . Una representación de esta primera etapa del AFM para el caso en estudio en el que $k = 2$, se esquematiza en la figura 2.

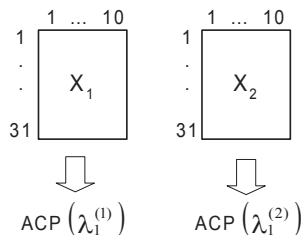


Figura 2. Primera etapa del Análisis Factorial Múltiple.

Figure 2. First step in Multiple Factor Analysis.

- Se realiza un ACP de la tabla global que resulta de yuxtaponer todas las tablas X_k . En este análisis cada tabla es ponderada por el inverso del primer valor propio obtenido en el ACP de dicha tabla $\frac{1}{\lambda_1^k}$.

Esta ponderación mantiene la estructura de cada tabla, ya que todas las variables han recibido la misma ponderación, pero consigue equilibrar la influencia de los grupos, ya que la inercia máxima de cada una de las nubes de individuos, definida por los diferentes grupos, vale 1 en cualquier dirección. Para los datos considerados, una representación de esta segunda etapa del método se muestra en la figura 3.

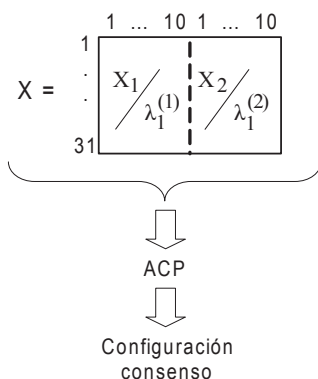


Figura 3. Segunda etapa del Análisis Factorial Múltiple.
Figure 3. Second step of Multiple Factor Analysis.

El objetivo consiste en poner de manifiesto los principales factores de variabilidad de los individuos (2), pero en este caso, estando estos últimos descriptos, de manera equilibrada por los distintos grupos de variables.

Análisis de Procrustes Generalizado

El Análisis de Procrustes Generalizado (APG) propuesto por Gower (11) armoniza las configuraciones individuales provenientes de cada tabla, a través de pasos algebraicos iterativos que incluyen traslación, rotación y escalado de las coordenadas bajo dos premisas: mantener la distancia relativa entre elementos de las configuraciones individuales y minimizar las sumas de cuadrados entre puntos análogos, es decir, puntos que corresponden al mismo elemento bajo configuraciones diferentes.

La configuración de consenso es obtenida como el promedio de todas las configuraciones individuales transformadas.

Cada matriz individual X_k origina una configuración determinada al aplicar un ACP que queda representada por C_k ($k = 1, 2, \dots, K$) con i filas y j columnas donde la i -ésima fila corresponde a las coordenadas del i -ésimo individuo en las j variables (figura 4, pág. 56).

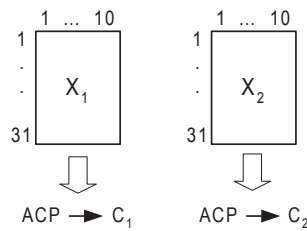


Figura 4. Obtención de las configuraciones iniciales para APG.
Figure 4. Obtaining the initial configurations in GPA.

En términos matriciales, el escalamiento, la rotación y la traslación pueden expresarse por la transformación:

$$Y_k \rightarrow \rho_k C_k H_k + T_k$$

donde: H_k es una matriz ortogonal de rotación, ρ_k el factor de escala y T_k la matriz de traslación.

Estos tres elementos son encontrados minimizando la expresión:

$$SCR = \sum_{i=1}^{31} \sum_{k=1}^2 \Delta^2(P_i^k, G_i)$$

donde $\Delta^2(P_i^k, G_i)$ es la distancia Euclídea entre el punto P_i^k y el centroide de los puntos análogos designado G_i .

Después de la estandarización inicial o la traslación, en cuanto todas las configuraciones han sido transformadas, una iteración es completada. Luego se calcula una configuración de consenso como la media de todas las configuraciones individuales transformadas y una nueva iteración es inicializada. El proceso es repetido hasta que el cambio entre dos pasos consecutivos en las sumas de cuadrados residual (SCR) es menor a un valor en particular (figura 5). Una tolerancia de convergencia de 0,0001 es considerada satisfactoria (5).

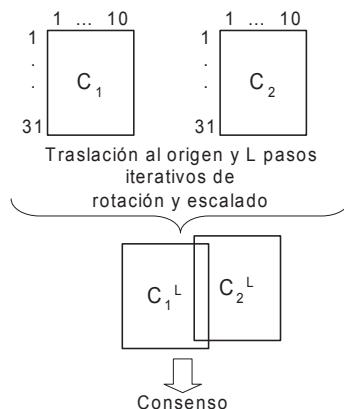


Figura 5. Obtención de la configuración consenso en APG.
Figure 5. Obtaining the consensus configuration in GPA.

Gower (11) propone al finalizar el APG realizar un Análisis de la Varianza que permita identificar la participación relativa de los individuos y las condiciones sobre SCR. Es decir, una vez que el proceso iterativo del método culmina, la variabilidad total de un APG puede ser particionada en forma de tabla de análisis de la varianza.

El APG está siendo utilizado ampliamente en caracterización de material genético cuando la misma se realiza mediante un conjunto de marcadores morfológicos y otro conjunto de marcadores moleculares (3, 6).

En estos casos, las configuraciones iniciales que ingresan al algoritmo iterativo provienen de la aplicación de ACP sobre los marcadores morfológicos y de Análisis de Coordenadas Principales, con una distancia adecuada, sobre los marcadores moleculares.

RESULTADOS Y DISCUSIÓN

Todos los resultados son obtenidos en el entorno de programación y análisis de datos del software R. Se denota con A1 el conjunto de variables de la localidad de Pergamino (EEA INTA) y con A2 el conjunto de variables de la localidad de Ferré (Escuela Agrotécnica Salesiana). Con el fin de identificar el ambiente en el que está medida cada variable, cada etiqueta en la codificación se acompaña con el dígito del mismo.

ACP sobre tabla promedio

La realización de un ACP sobre la tabla obtenida como promedio de las tablas originales correspondientes a los diferentes ambientes permite visualizar en el plano las semejanzas y diferencias medias de las poblaciones en estudio, como así también las correlaciones entre las variables.

Se observa un primer factor o componente principal claramente preponderante, que recoge un 44,68% de la variabilidad total de los datos, proyectándose en el plano principal (plano factorial óptimo compuesto por las dos primeras componentes) más de la mitad de la inercia, en concreto un 59,98% de la misma ($\lambda_1 = 4,46$; $\lambda_2 = 1,53$).

Del análisis de la representación de los vectores-variable (figura 6, pág. 58) merecen ser destacados los siguientes resultados:

- Las variables DMZ, GRXM, PESO y REND son las que más contribuyen a la formación del primer factor. Las demás lo hacen de una manera moderada y equilibrada, a excepción de AGR que aporta débilmente a su formación. Del gráfico también se deduce que la totalidad de las variables tienen coordenada de igual signo, lo que podría traducirse en un efecto talla.
- En lo que se refiere al segundo factor hay que señalar que este pone de manifiesto una segunda caracterización de las poblaciones de maíz, distinta de la reflejada por el factor anterior. Las variables que más contribuyen a su formación son: ALTMZ, ALTPL enfrentadas en el plano a AGR. Las demás variables aportan débilmente a su formación.

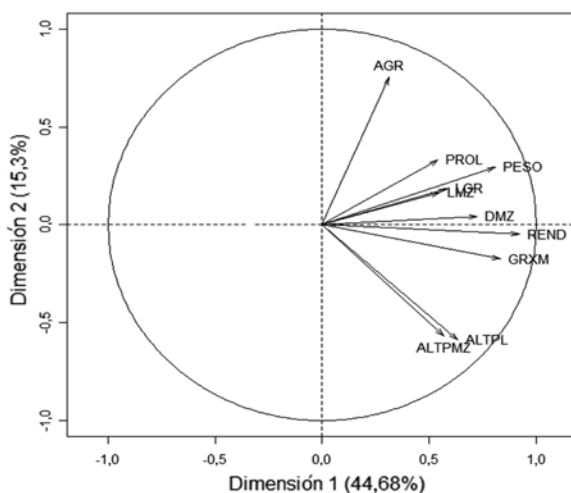


Figura 6. ACP-Proyección de las variables sobre el plano principal.

Figure 6. PCA-Projection of variables on the principal plane.

Del análisis de la nube de puntos-individuo (figura 7) se pueden caracterizar las poblaciones representando el comportamiento medio en ambos ambientes. Se destacan los siguientes resultados:

- La población 14, ocupa a lo largo del eje horizontal, la primera posición (mayor coordenada en valor absoluto). En menor medida, le siguen las poblaciones 10, 12, 7 y 19, estas dos últimas en posición opuesta a las tres anteriores.
- La población 14 tiene comportamiento específico que merece ser interpretado con minuciosidad.
- En el segundo factor se destacan las poblaciones 15 y 6, que se oponen a las poblaciones 1, 18, 14 y 30.

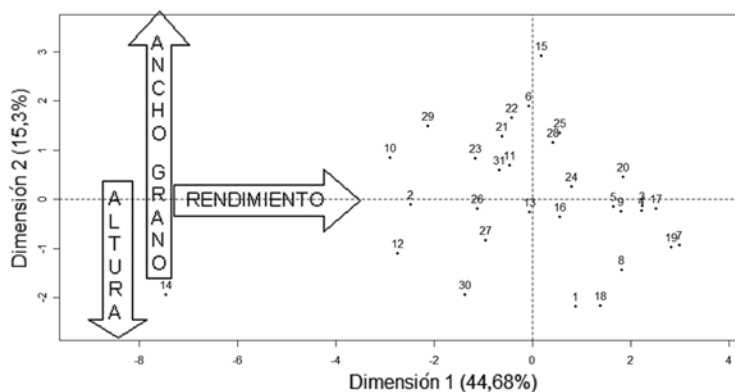


Figura 7. Configuración de los individuos (poblaciones locales de maíz) sobre el plano principal resultante del ACP.

Figure 7. Individuals configuration (native maize populations) on the PCA principal factorial plane.

De lo anterior y a modo de ejemplo, se destacan las siguientes observaciones:

La población 14 es la de menor rendimiento, presenta plantas de gran tamaño, con inserción de espiga elevada y granos pequeños.

El grupo de poblaciones 10, 12, 2, 29, 23, 26, 27 y 30 presenta rendimientos por debajo de la media, diferenciándose notablemente las poblaciones 10, 29 y 23 de la 30 por ser las primeras plantas pequeñas en contraposición a la gran altura que presenta la última.

El grupo constituido por las poblaciones 7, 19, 17, 3, 4, 20, 5, 9, 8 y 18 son las de mayor rendimiento. Entre estas la población 20 presenta altura por debajo de la media y la 18 y la 8 son de importante altura.

Dentro de las restantes poblaciones que presentan rendimientos medios se destacan la 15, 6, 22, 25 y 28 por ser plantas bajas y de considerable ancho de grano, en contraposición a la 1 que corresponde a plantas altas de grano pequeño.

Es importante destacar que esta caracterización representa un comportamiento medio de ambos ambientes lo cual hace que si existe interacción de las poblaciones con los ambientes no pueda visualizarse en este análisis. Una alternativa menos tradicional en este enfoque de ACP que no se presenta en este trabajo, sería realizar una combinación individuo/ambiente constituyendo así una tabla de 62 filas.

Análisis Factorial Múltiple

Primera etapa

Se realiza un ACP en ambas tablas X_1 y X_2 reteniendo el primer autovalor para cada tabla: $\lambda_1^{(1)} = 4,67$ y $\lambda_1^{(2)} = 3,62$.

El coeficiente de correlación vectorial RV, con un valor igual a 0,56, no sugiere que ambos ambientes estén estrechamente relacionados pero tampoco que presenten grandes diferencias. En definitiva, las configuraciones proporcionadas por cada ambiente son medianamente similares, ya que toma un valor próximo a 0,5.

El coeficiente de dimensionalidad Ng indica el número de factores con destacable importancia en el seno de cada tabla. Los valores obtenidos: $Ng_1 = 1,126$ y $Ng_2 = 1,3$ indican el fuerte predominio de la componente 1 sobre la 2 en ambos ambientes.

Análisis global

Los valores propios obtenidos en esta etapa confirman que el primer factor de dirección global de mayor inercia es común a los dos ambientes (su valor está próximo al número de grupos considerados). Es un primer eje global preponderante, en el sentido de que recoge un importante porcentaje de inercia, prácticamente un 37%, frente al 14,58% de variabilidad que se proyecta en el segundo factor global (Autovalores: 1,83 frente a 0,72). La tercera componente recoge una inercia de 10,92%, pero al igual que en las secciones anteriores, se tuvo en cuenta para el análisis el plano factorial generado por las dos primeras componentes, ya que en ellas se observan las principales características de estas poblaciones de maíz, las diferencias y semejanzas más relevantes.

Análisis de los vectores-variable (figura 8). Se observa que la mayoría de los vectores que representan la misma variable en ambas condiciones están bastante próximos.

Las variables ALTPL, ALTMZ para los dos ambientes son las que presentan un ángulo mayor que el resto de las variables analizadas, lo cual indica que tienen un comportamiento menos estable que las demás. Esta es una conclusión importante, ya que serían las variables más afectadas por el ambiente.

El primer factor pone de manifiesto una caracterización global de estas poblaciones de maíz, siendo REND, PESO, GRXM y DMZ las que más contribuyen a su formación. En cambio, en lo que se refiere al segundo factor, las variables que más contribuyen son AGR con carga negativa y ALTPL y ALTMZ correspondientes al ambiente 2 con cargas positivas.

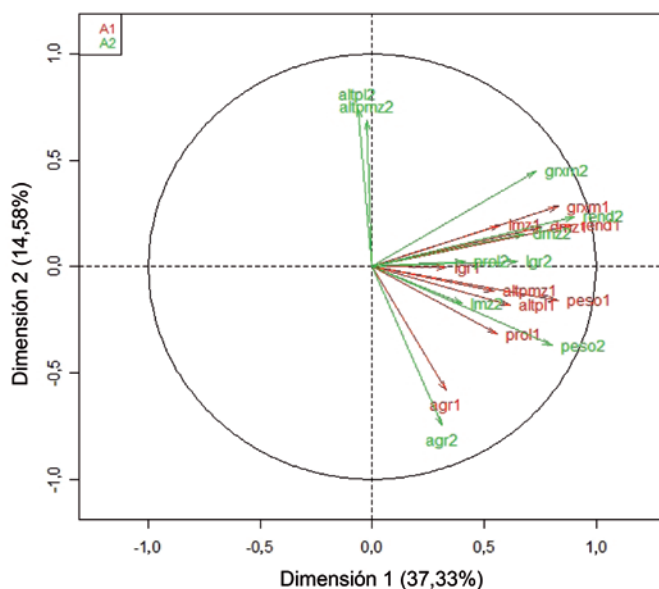


Figura 8. AFM-Proyección de las variables sobre el plano principal.
Figure 8. MFA-Projection of variables on the principal plane.

En la nube de puntos-individuo (figura 9, pág. 61) se muestran tanto los individuos medios (posición media que ocupa la población de maíz entre los dos ambientes considerados), como los individuos parciales (diferentes posiciones que ocupan las poblaciones de maíz en cada uno de los ambientes) proyectados sobre los dos primeros ejes factoriales.

Esto es: cada individuo está representado por tres puntos, los cuales se unen con línea sostenida para el caso del ambiente 1 y línea de puntos para el ambiente 2.

En general, se observa disparidad en la proximidad entre los puntos que representan la misma población de maíz (puntos análogos) en ambos ambientes. Merecen ser destacadas, dado su comportamiento menos homogéneo, las siguientes poblaciones: 6, 10, 12, 18, 23, 27 y 30 entre otras, lo que indicaría que en estas poblaciones es en las cuales se presenta el mayor efecto ambiente.

Además, las mayores diferencias en las primeras seis poblaciones se deben a la altura de la planta y la altura de inserción de la mazorca, mientras que para la población 30 las mayores diferencias entre los ambientes se deben a rendimiento y peso. Esta situación contrastada con las poblaciones de comportamiento homogéneo en los dos ambientes pone en evidencia la presencia de interacción genotipo*ambiente. Al igual que en el ACP tradicional la población 14 se diferencia de todas las otras, principalmente por tener las peores características agronómicas, bajo rendimiento, bajo peso y bajo diámetro. Sin embargo, es una población que se comportó de manera estable en ambas localidades.

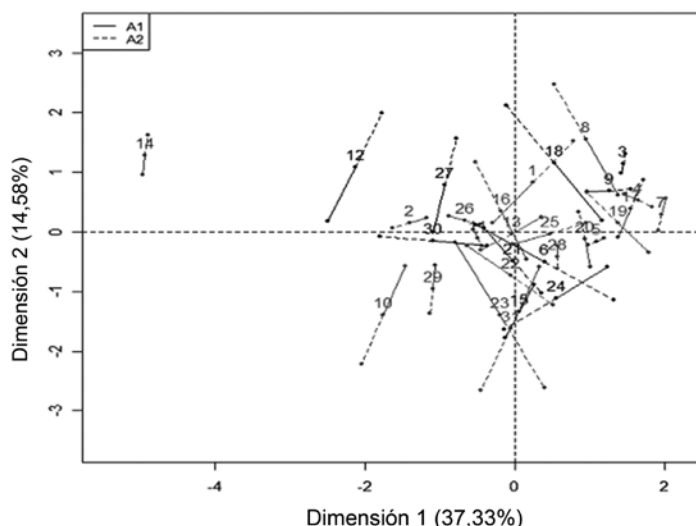


Figura 9. AFM-Proyección de los individuos medios y parciales.
Figure 9. MFA-Projection of average and partial individuals.

Análisis de Procrustes Generalizado

Como se menciona en la sección "Análisis de Procrustes Generalizado" (pág. 55), en el APG la discrepancia entre las tablas es expresada en forma de análisis de la varianza. La variabilidad total es el equivalente a la suma de cuadrados total y puede ser particionada como:

$$V_{total} = V_{consenso} + V_{residual}$$

donde:

la suma de cuadrados debida a la componente "dentro" puede descomponerse por tablas o por individuos.

A partir de la proyección de los individuos (medios y parciales) sobre el primer plano factorial (figura 10, pág. 62) y de la suma de cuadrados residual (SCR) para las

poblaciones (tabla 2, pág. 63) se destacan algunas diferencias parciales. Por ejemplo: las poblaciones 1, 6, 15, 18, 19, 23 y 27 son las que manifestaron mayor discrepancia entre los ambientes por presentar mayores valores de SCR. Es decir, los puntos que representan esas poblaciones en los dos ambientes (individuos parciales) están más distanciados. También es notorio el comportamiento que tiene la población 14 diferente del resto que evidencia la máxima suma de cuadrados total.

Sin embargo, se observa en el mismo gráfico, que en esta población no hay diferencias importantes entre los dos ambientes ya que los puntos que representan esas poblaciones en los dos ambientes (individuos parciales) están próximos del punto medio (individuo medio).

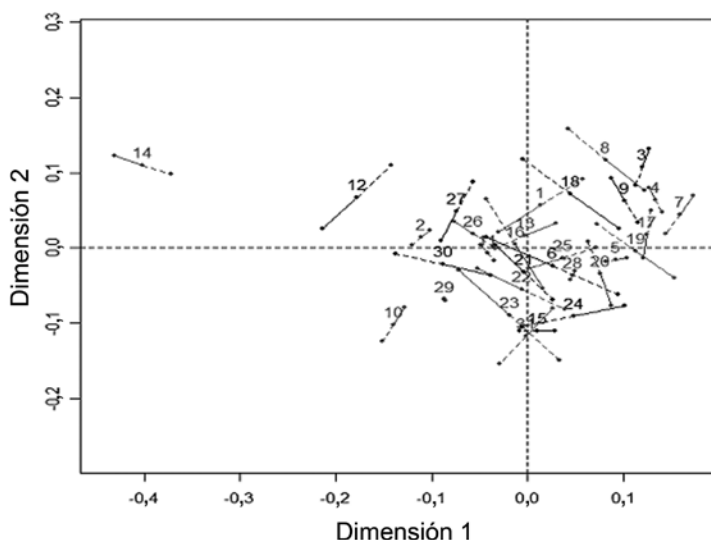


Figura 10. APG-Proyección de los individuos medios y parciales.
Figure 10. GPA-Projection of average and partial individuals.

Comparación de técnicas

Para poder comparar estos métodos se obtuvo la matriz de distancias entre los individuos en las tres configuraciones (individuos medios en AFM y APG, e individuos del ACP a partir del promedio de variables para ambos ambientes) y a partir de ellas, se calculó el coeficiente de correlación lineal de Pearson.

Se observa una alta concordancia entre las caracterizaciones obtenidas con los tres métodos (estadísticamente altamente significativas).

La correlación entre AFM y APG con el método tradicional fue de 0,90 y 0,89 respectivamente, mientras que la correlación entre las dos alternativas de tres vías resultó igual a 0,97, lo que indicaría que estas dos metodologías presentan configuraciones consenso prácticamente equivalentes.

Tabla 2. Cuadro de Análisis de la Varianza (APG).

Table 2. Analysis of variance (GPA).

Individuo	Sumas de cuadrados dentro por individuo			
	Consenso	Residual	Total	Prop Cons
1	0,027	0,017	0,043	0,618
2	0,035	0,007	0,042	0,835
3	0,060	0,004	0,065	0,936
4	0,058	0,004	0,063	0,932
5	0,027	0,004	0,031	0,863
6	0,029	0,023	0,051	0,559
7	0,073	0,009	0,082	0,895
8	0,048	0,013	0,061	0,793
9	0,047	0,006	0,053	0,888
10	0,080	0,008	0,088	0,913
11	0,093	0,003	0,096	0,969
12	0,097	0,009	0,106	0,914
13	0,018	0,007	0,024	0,724
14	0,358	0,009	0,366	0,977
15	0,076	0,019	0,094	0,800
16	0,011	0,015	0,026	0,410
17	0,037	0,005	0,043	0,871
18	0,047	0,021	0,068	0,688
19	0,033	0,018	0,051	0,648
20	0,026	0,009	0,035	0,744
21	0,013	0,015	0,028	0,455
22	0,013	0,007	0,021	0,642
23	0,035	0,026	0,062	0,577
24	0,037	0,009	0,046	0,806
25	0,014	0,008	0,022	0,649
26	0,022	0,005	0,026	0,829
27	0,048	0,022	0,070	0,684
28	0,052	0,008	0,060	0,866
29	0,054	0,016	0,070	0,768
30	0,033	0,016	0,049	0,676
31	0,050	0,008	0,059	0,857
Total	1,651	0,349	2,000	0,826

CONCLUSIONES

Si el objetivo del estudio es encontrar similitudes entre poblaciones de maíz evaluadas en dos ambientes, o condiciones agronómicas diferentes, es posible utilizar cualquiera de las técnicas ya que existe una alta concordancia entre las configuraciones consenso.

Sin embargo, los resultados que se obtienen con los métodos AFM y APG frente a la técnica tradicional, poseen información mucho más rica en lo que se refiere a la interpretación del efecto ambiente y sus interacciones.

Por ejemplo, del análisis con los tres métodos se observa que la población 14 tiene un comportamiento diferente a la del resto. No obstante, con la técnica tradicional no se podría haber concluido que en esta población no se observan diferencias importantes entre las dos localidades (menor efecto ambiente). Por otro lado, del análisis con las tres técnicas se concluye la misma posición relativa de la población 23 respecto de las demás, aunque a partir de las representaciones gráficas que surgen de aplicar AFM o APG se observa también que manifiesta la mayor discrepancia entre los dos ambientes. Estas dos situaciones ponen en evidencia la existencia de una importante interacción genotipo*ambiente.

La técnica AFM además permite estudiar qué variables resultan más sensibles a los cambios ambientales. En este caso particular, fue la situación presentada por los caracteres ALTPL y ALTMZ.

Para poder generalizar una conclusión se prevé en breve realizar simulaciones con distintas situaciones de efecto ambiente e interacciones genotipo*ambiente.

BIBLIOGRAFÍA

1. Abdi, H. 2007. RV coefficient and congruence coefficient. In: N. J. Salkind (Ed.). 2007. Encyclopedia of Measurement and Statistics. Thousand Oaks (CA): Sage. p. 849-853.
2. Abdi, H.; Valentin, D. 2007. Multiple factor analysis. In: N. J. Salkind (Ed.). 2007. Encyclopedia of Measurement and Statistics. Thousand Oaks (CA): Sage. p. 657-663.
3. Balzarini, M.; Teich, I.; Bruno, C.; Peña, A. 2011. Making genetic biodiversity measurable: a review of statistical multivariate methods to study variability at gene level. Revista de la Facultad de Ciencias Agrarias. Universidad Nacional de Cuyo. 43(1): 261-275.
4. Bramardi, S. J. 2000. Estrategias para el análisis de datos en la caracterización de recursos genéticos. Tesis doctoral. Departamento de Estadística e Investigación Operativa. Universidad Politécnica de Valencia, España. 390 p.
5. Bramardi, S. J.; Bernet, G.; Asíns, M.J.; Carbonell, E. 2005. Simultaneous Agronomic and Molecular Characterization of Genotypes via the Generalised Procrustes Analysis: An Application to Cucumber. Crop Sci. 45: 1603-1609.
6. Bruno, C.; Balzarini, M. 2010. Ordenaciones de material genético a partir de información multidimensional. Revista de la Facultad de Ciencias Agrarias. Universidad Nacional de Cuyo. 42(2): 183-200.
7. Centro Internacional de Mejoramiento de Maíz y Trigo (CIMMYT/IBPGR). 1991. Descriptores de maíz. México-Roma. 88 p.
8. Dijksterhuis, G. B.; Gower, J. C. 1991. The interpretation of Generalized Procrustes Analysis and allied Methods. Food Quality and Preference. p. 67-87.
9. Escofier, B.; Pagès, J. 1990. Análisis Factoriales Simples y Múltiples: Objetivos, Métodos e Interpretación. Universidad del País Vasco. Bilbao. 285 p.
10. Esquinas Alcázar, J. 2005. Protecting crop genetic diversity for food security: political, ethical and technical challenges. Nature Publishing Group. Volumen 6: 946-953.
11. Gower, J. C. 1975. Generalized Procrustes Analysis. Psychometrika, 40: 33-51.
12. Lucchin, M.; Barcaccia, G.; Parrini, P. 2003. Characterization of a flint maize (*Zea mays* L. convar. Mays) Italian landrace: I. Morpho-phenological and agronomic traits. Genetic Resource and Crop Evolution 50: 315-327.
13. Peña, D. 2002. Análisis de datos multivariantes. Mc Graw Hill. Madrid. 539 p.
14. Salhuana, W.; Pollak, L. M.; Ferrer, M. E.; Paratori, O.; Vivo, G. 1998. Breeding Potential of Maize Accessions from Argentina, Chile, USA, and Uruguay. Crop Sci. 38: 866-872.
15. Troyer, A. F. 1988. Measurement of Genetic Diversity Among Popular Commercial Corn Hybrid. Crop Sci. 28: 481-485.