



UNIVERSIDAD NACIONAL DE ROSARIO
FACULTAD DE CIENCIAS AGRARIAS

VARIABILIDAD Y ESTRUCTURA GENÉTICA EN LÍNEAS ENDOCRIADAS DE MAÍZ

EZEQUIEL ALEJANDRO ROSSI

TESIS PARA OPTAR AL TÍTULO DE MAGISTER EN GENÉTICA VEGETAL

Directora: Ing. Agr. Dra. Natalia Cecilia BONAMICO

Co-Director: Ing. Agr. MSc Miguel Ángel DI RENZO

AÑO 2016

VARIABILIDAD Y ESTRUCTURA GENÉTICA EN LÍNEAS ENDOCRIADAS DE MAÍZ

Ezequiel Alejandro Rossi

Ingeniero Agrónomo - Universidad Nacional de Río Cuarto

Esta Tesis es presentada como parte de los requisitos para optar al grado académico de Magister en Genética Vegetal, de la Universidad Nacional de Rosario y no ha sido previamente presentada para la obtención de otro título en ésta u otra Universidad. La misma contiene los resultados obtenidos en investigaciones realizadas en el campo experimental de la Facultad de Agronomía y Veterinaria de la Universidad Nacional de Río Cuarto, durante el período comprendido entre 2015-2016, bajo la dirección de la Dra. Natalia Cecilia Bonamico.

Nombre y firma del Maestrando

Nombre y firma del Director

Nombre y firma del Co - Director

Defendida:de 2017.

AGRADECIMIENTOS

A mi directora, Ing. Agr. Dra. Natalia Bonamico por su guía y apoyo, por formarme como profesional, por su inmensa colaboración en la recolección de datos y sus aportes e ideas en la escritura de este trabajo.

A mi co-director, Ing. Agr. *Mcs.* Miguel Di Renzo que junto a la Dra. Natalia Bonamico fueron quienes me recomendaron y aconsejaron realizar esta maestría y brindaron todo su apoyo para que la pueda llevar adelante. A Miguel además, por sus consejos, ideas y recomendaciones de libros en el área de la investigación que me han permitido crecer como profesional.

A mi hermano Leonardo por su predisposición para ir a buscarme en aquellos viajes de vuelta desde Zavalla a Río Cuarto en los cuales no conseguía ómnibus.

Al Fondo para la Investigación Científica y Tecnológica (FONCyT) por el financiamiento parcial de mis estudios en esta maestría.

A mis padres y Serenella por su apoyo incondicional en todos mis proyectos.

PUBLICACIONES Y PRESENTACIONES A CONGRESOS

- Análisis de conglomerados para describir la estructura genética en líneas endocriadas de maíz. Rossi, Di Renzo, Bonamico. XVI CONGRESO LATINOAMERICANO DE GENÉTICA, Montevideo, Uruguay. 2016.

ÍNDICE

RESUMEN	10
ABSTRACT	11
INTRODUCCIÓN	12
HIPÓTESIS	17
OBJETIVOS	17
MATERIALES Y MÉTODOS	18
Material vegetal y ensayo de campo	18
Evaluación del fenotipo	18
Evaluación del genotipo	21
Análisis estadístico del fenotipo	21
Análisis estadístico del genotipo	22
<i>Estructura poblacional</i>	22
<i>Diversidad genética</i>	23
<i>Kinship relativo y distancia genética entre líneas</i>	23
<i>Desequilibrio de ligamiento</i>	24
Análisis estadístico fenotipo-genotipo	24
RESULTADOS	27
Evaluación del fenotipo	27
Evaluación del genotipo	36
<i>Estructura poblacional</i>	36
<i>Diversidad genética</i>	38
<i>Kinship relativo y distancia genética entre líneas</i>	39
<i>Desequilibrio de ligamiento</i>	40

Análisis estadístico fenotipo-genotipo	44
DISCUSIÓN	45
CONCLUSIONES	49
BIBLIOGRAFÍA	50

ÍNDICE DE CUADROS

Cuadro 1	Número e identificación de cada una de las 252 líneas de maíz evaluadas en el campo experimental de la FAV-UNRC durante el ciclo agrícola 2015-2016.	25
Cuadro 2	Media, error estándar (EE), valor mínimo (Mín.) y máximo (Máx.) para los siete caracteres agro-morfológicos cuantitativos medidos en las 252 líneas de maíz de CIMMYT evaluadas en el campo experimental de la FAV-UNRC durante el ciclo agrícola 2015-2016.	27
Cuadro 3	Tabla de frecuencia relativa para las categorías de cada uno de los seis caracteres agro-morfológicos categóricos medidos en las 252 líneas de maíz de CIMMYT evaluadas en el campo experimental de la FAV-UNRC durante el ciclo agrícola 2015-2016.	28
Cuadro 4	Análisis molecular de la varianza (AMOVA) entre sub-poblaciones inferidas a partir de las 252 líneas de maíz de CIMMYT.	37
Cuadro 5	Diversidad genética y contenido de información polimórfica (PIC) de las 252 líneas de maíz de CIMMYT y de cada sub-población.	38
Cuadro 6	Desequilibrio de ligamiento de todo el genoma, en la población completa y en los dos grupos con mayor número de genotipos asignados.	40
Cuadro 7	Suma de cuadrados del análisis de procrustes generalizado (APG) combinando información proveniente de caracteres agro-morfológicos y de marcadores moleculares. Los caracteres fueron medidos en las 252 líneas de maíz de CIMMYT evaluadas en el campo experimental de la FAV-UNRC durante el ciclo agrícola 2015-2016.	44

ÍNDICE DE FIGURAS

Figura 1.	Categorías del ángulo de hoja observado en la población de las 252 líneas de maíz de CIMMYT, evaluadas en el campo experimental de la FAV-UNRC durante el ciclo agrícola 2015-2016.	19
Figura 2.	Categorías del tipo de panoja observado en la población de las 252 líneas de maíz de CIMMYT, evaluadas en el campo experimental de la FAV-UNRC durante el ciclo agrícola 2015-2016.	19
Figura 3.	Tipo de grano y color de grano observado en la población de las 252 líneas de maíz de CIMMYT, evaluadas en el campo experimental de la FAV-UNRC durante el ciclo agrícola 2015-2016.	20
Figura 4.	Distribución de la frecuencia relativa de los caracteres días a floración masculina, días a floración femenina e intervalo antesis-estigma en las 252 líneas de maíz de CIMMYT evaluadas en el campo experimental de la FAV-UNRC durante el ciclo agrícola 2015-2016.	29
Figura 5.	Distribución de la frecuencia relativa de los caracteres altura de planta, altura de espiga y la relación altura de planta/altura de espiga en las 252 líneas de maíz de CIMMYT evaluadas en el campo experimental de la FAV-UNRC durante el ciclo agrícola 2015-2016.	30
Figura 6.	Distribución de la frecuencia poblacional para el carácter ángulo de hoja en las 252 líneas de maíz de CIMMYT evaluadas en el campo experimental de la FAV-UNRC durante el ciclo agrícola 2015-2016.	31
Figura 7.	Distribución de la frecuencia relativa del carácter largo de panoja y distribución de la frecuencia poblacional de las tres categorías de tipo de panoja en las 252 líneas de maíz de CIMMYT evaluadas en el campo experimental de la FAV-UNRC durante el ciclo agrícola 2015-2016.	32
Figura 8.	Distribución de la frecuencia poblacional del tipo de grano y del color de grano en las 252 líneas de maíz de CIMMYT evaluadas en el campo experimental de la FAV-UNRC durante el ciclo agrícola 2015-2016.	32

Figura 9.	Distribución de la frecuencia poblacional de los caracteres quebrado y vuelco en las 252 líneas de maíz de CIMMYT evaluadas en el campo experimental de la FAV-UNRC durante el ciclo agrícola 2015-2016.	33
Figura 10.	Gráfico biplot obtenido del análisis de correspondencia múltiple de los caracteres ángulo de hoja, forma de panoja, tipo de grano y color de grano en las 252 líneas de maíz de CIMMYT evaluadas en el campo experimental de la FAV-UNRC durante el ciclo agrícola 2015-2016.	34
Figura 11.	Gráfico biplot del análisis de componentes principales de siete caracteres agro-morfológicos en las 252 líneas de maíz de CIMMYT evaluadas en el campo experimental de la FAV-UNRC durante el ciclo agrícola 2015-2016.	35
Figura 12.	Análisis de la estructura poblacional de 252 líneas de maíz de CIMMYT. Valor del logaritmo de asignación probabilística ($\ln P(D)$) estimado y Δk obtenidos con STRUCTURE. Estructura de la población con k establecido en cuatro sub-poblaciones. Los números indicados debajo de cada sub-población corresponden a los valores del índice de fijación de Wright's (F_{st}).	37
Figura 13.	Gráfico del análisis de coordenadas principales obtenido a partir de 2530 SNP en las 252 líneas de maíz de CIMMYT. No se muestran los individuos clasificados como mixtos.	38
Figura 14.	Distribución de la estimación de parentesco relativo por pares (<i>kinship</i> relativo) y la distancia genética de Roger's para las 252 líneas de maíz de CIMMYT utilizando 2530 SNP. Para simplicidad solo se muestra el porcentaje de los valores de <i>kinship</i> relativo entre 0,05 y 0,50.	39
Figura 15.	Desequilibrio de ligamiento entre 2530 SNP. En el eje X se indica la posición del marcador y en el eje Y el nombre del marcador.	41
Figura 16:	Desequilibrio de ligamiento observado en las 252 líneas de maíz de CIMMYT, con 2530 SNP a través de todo el genoma.	43

RESUMEN

El maíz (*Zea mays* L.) es un excelente modelo vegetal para comprender la variación natural debido a la variabilidad genética disponible. Posee un genoma complejo y un alto nivel de diversidad en comparación con otros cultivos y especies de plantas. En el presente estudio, una población diversa de 252 líneas de maíz desarrolladas en el Centro Internacional de Mejoramiento de Maíz y Trigo (CIMMYT) fue caracterizada por su fenotipo y por su genotipo. La evaluación fenotípica se realizó en el campo experimental de la Facultad de Agronomía y Veterinaria de la Universidad Nacional de Río Cuarto durante el ciclo agrícola 2015/2016. El diseño utilizado fue un diseño completamente al azar con una repetición. La caracterización fenotípica se realizó por medio de 13 caracteres agro-morfológicos. La evaluación genotípica se realizó mediante marcadores moleculares SNP derivados de genotipado por secuenciación. En la evaluación de la diversidad genética, la estructura poblacional y el desequilibrio de ligamiento presentes en la población se utilizaron 2530 SNP. Las líneas de maíz evaluadas en este estudio mostraron amplia variabilidad para los caracteres agro-morfológicos medidos. La descripción de la estructura genética poblacional permitió inferir la presencia de cuatro sub-poblaciones. La diversidad genética y el contenido de información polimórfica medios fueron 0,50 y 0,43, respectivamente. Los valores de los coeficientes *kinship*, en el 85% de las comparaciones por pares de líneas, fueron menores a 0,10 y la distancia genética media fue 0,44. El desequilibrio de ligamiento (DL) en la población total fue 0,20 (r^2). Este valor aumentó al analizar el DL en las sub-poblaciones inferidas. El DL de los marcadores ligados físicamente fue superior al de aquellos marcadores no ligados. EL DL presentó una rápida caída y alcanzó un valor de r^2 menor de 0,10 a una distancia de 300-500 kb. El análisis de componentes principales permitió reducir la dimensión de la matriz de datos fenotípicos y genotípicos, luego mediante el análisis de procrustes generalizado se observó un consenso del 60% entre ambos tipos de ordenamientos. La población de 252 líneas de maíz presenta estratificación en su estructura genética. Esto debe ser considerado al realizar estudios de mapeo de asociación. El ligamiento físico, como principal causa del desequilibrio de ligamiento presente en la población, y la rápida caída del mismo, indican que esta población de líneas de maíz es adecuada para posteriores estudios de mapeo de asociación. La correlación entre la diversidad molecular y fenotípica, observada mediante el moderado valor de consenso, indica desequilibrio de ligamiento entre los marcadores moleculares y los genes que controlan los caracteres agro-morfológicos medidos.

ABSTRACT

Choice of germplasm is critical to the success of association analysis. The germplasm set should encompass as much phenotypic variation as possible. Genetic or phenotypic surveys can be used to identify genotypically diverse subsets of the available germplasm. The present study was conducted to analyze the variability and population structure among 252 CIMMYT maize inbred lines using agro-morphological traits and SNP markers derived from GBS (genotyping by sequencing). Phenotypic evaluation was performed in experimental field of FAV-UNRC during the summer 2015/2016. The experimental design was a randomized complete design with one replication. A total of 2530 informative SNP and was used to estimate the genetic diversity, population structure, and familial relatedness. Structure analysis showed the presence of four sub-populations. The average gene diversity was 0,50 and polymorphic information content was 0,43. The kinship coefficients were minors at 0,10 in 85% of pair of inbred lines. The average genetic distance was 0,44. Linkage disequilibrium in the full population was 0,20 (r^2). This value was higher in the inferred sub-populations. LD in linked markers was higher than in unlinked markers. Linkage disequilibrium showed a rapid decay of 100-300 kb at $r^2=0,10$. Genetic structure in this population should be considered in association mapping studies. Principal component analysis allowed reducing the phenotypic and genotypic data matrix. Generalized procrustes analysis showed a 60% consensus between phenotypic and genotypic ordering. Population is suitable for mapping association studies because it presents a rapid decay of linkage disequilibrium and presence of physical linkage. The correlation between phenotypic and genotypic diversity indicated linkage disequilibrium between molecular markers and the genes controlling these agro-morphological traits.

INTRODUCCIÓN

El maíz (*Zea mays* L.), uno de los cereales cultivados más importante en todo el mundo, fue domesticado hace 9000 años atrás en México a partir del teosinte (*Zea mays* ssp. *parviglumis*) (Prasanna, 2012). La transición del maíz desde una especie silvestre hacia una especie cultivada llevó muchos años de mejora. La presión de selección para cada carácter fue diferente según las preferencias y las características ambientales de los diferentes asentamientos nativos americanos (Hallauer et al., 1988). Durante mucho tiempo, el principal objetivo de los mejoradores de maíz ha sido aumentar el rendimiento en grano. Al aplicar una alta intensidad de selección para un rendimiento más alto y estable, se modificaron otros caracteres por selección indirecta. Así, sin la intención expresa del obtentor y/o selección, se logró obtener panojas más pequeñas, hojas más verticales, y un intervalo antesis-estigma más corto (Duvick, 2004).

Debido a la variabilidad genética disponible, el maíz es un excelente modelo para comprender la variación natural (Wallace et al., 2014). Posee un genoma complejo y un alto nivel de diversidad en comparación con otros cultivos y especies de plantas (Romay et al., 2013). El maíz es una especie modelo en estudios sobre los mecanismos genéticos de caracteres cuantitativos (Liu et al., 2014). La disección genética de estos caracteres es importante para mejorar la eficiencia de los programas de mejoramiento y como herramienta para entender los procesos básicos de desarrollo, fisiología y bioquímica de las plantas (Hansey et al., 2011). Las líneas endocriadas de maíz son un recurso fundamental para estudios de genética y mejoramiento vegetal. La utilización de las líneas de maíz para análisis genéticos requiere conocer la relación histórica y entender la partición de la diversidad genética entre ellas (Liu et al., 2003).

La presencia de variabilidad fenotípica en el germoplasma es un requisito fundamental para identificar genotipos con caracteres específicos (Dinesh et al., 2016). La diversidad genética entre líneas endocriadas ha sido estimada tradicionalmente con datos de caracteres morfológicos. Este método tiene muchas limitaciones debido a la interacción que presentan estos caracteres con el ambiente (Mienie and Fourie 2013). Conocer la relación entre las líneas podría ayudar a identificar un conjunto de líneas que tengan la máxima diversidad para luego poder analizar los efectos del fondo genético (Liu et al., 2003). Para estimar las relaciones entre diversas líneas, los marcadores moleculares han sido ampliamente utilizados (Yan et al., 2009). Además según Warburton et al., (2002), el análisis basado sólo en el fenotipo puede perder

mucha variación favorable, por lo cual debe ser complementado con un análisis genotípico mediante marcadores moleculares.

Un marcador molecular es un segmento de ADN con una ubicación específica en un cromosoma. La secuencia puede pertenecer a regiones codificantes o sin función conocida. Las técnicas modernas de biología molecular permitieron generar diversos métodos de detección de polimorfismos a nivel del ADN (Levitus et al., 2010). Los marcadores moleculares son herramientas valiosas para la clasificación de germoplasma y para aplicar selección asistida por marcadores (MAS) (Edwards and McCouch 2007). Los RFLPs ó polimorfismos en la longitud de los fragmentos de restricción fueron los primeros marcadores de ADN. Luego con el descubrimiento de la reacción de la polimerasa en cadena (PCR) se lograron muchos avances en la tecnología de los marcadores moleculares. A partir de aquí surgieron otros tipos de marcadores como son los RAPDs ó ADN polimórfico amplificado aleatoriamente, los AFLP ó polimorfismos en la longitud de los fragmentos amplificados y los SSR ó microsatlélites (Edwards and McCouch 2007). En la actualidad, los marcadores moleculares SNP (polimorfismo de un nucleótido) están siendo ampliamente utilizados en distintos estudios genéticos. Estos marcadores se basan en la detección de polimorfismos resultantes de la alteración de una única base en una secuencia de ADN. Poseen naturaleza bialélica y son muy abundantes en el genoma, con una frecuencia de 1 cada 100-300 pares de bases en genomas de plantas. Debido a esto, los SNP surgen como importantes marcadores para la obtención de mapas genéticos de alta resolución. Estudios en *Arabidopsis* como organismo modelo, muestran que permiten obtener mapas con una resolución 100 veces superior a la obtenida con marcadores convencionales (Levitus et al., 2010). Otra característica importante es que el genotipado mediante SNP no está basado en la medida del tamaño de los alelos, por lo que su distinción puede ser automatizada. El descubrimiento de los SNP permitió a los investigadores explorar nuevas hipótesis. Se considera que muchos de estos SNP están localizados en el interior de secuencias génicas. Esto puede significar una importante reducción en el tiempo y costos para el descubrimiento de genes de interés, comparado con los marcadores utilizados hasta ahora. Asimismo, el desarrollo constante de tecnologías para su aplicación permitirá automatizar el mapeo de alta densidad y reducir los costos de su empleo en estudios moleculares a gran escala (Levitus et al., 2010).

Con la creciente disponibilidad de datos genómicos polimórficos y el potencial de explorar todos los eventos de recombinación que han ocurrido en la historia evolutiva de un germoplasma específico, estudios de mapeo de asociación se han convertido en

una alternativa de gran alcance para la disección de caracteres cuantitativos en plantas (Hao et al., 2015). El mapeo de asociación es un método alternativo para la identificación de secuencias de ADN que contribuyen a caracteres cuantitativos, incluyendo la resistencia a enfermedades (Warburton et al., 2015). Esta estrategia de mapeo utiliza poblaciones con desequilibrio por ligamiento (DL), el cual consiste en asociaciones no al azar entre alelos de diferentes *loci* (Würschum et al., 2012). El DL también se conoce como desequilibrio de fase gamética, desequilibrio gamético, ó asociación alélica. Es la correlación entre los polimorfismos que es causada por su historia compartida de mutación y recombinación. A pesar de que DL y ligamiento están relacionados, son claramente diferentes. Ligamiento se refiere a la herencia correlacionada de *loci* dado por la conexión física en un cromosoma, mientras que DL se refiere a la correlación entre los alelos en una población. El DL puede ser causado por una estructura desconocida de la población y muchas otras fuerzas, incluyendo mutación, deriva génica, selección y específicamente en las plantas, juega un rol importante el nivel de endocría originado según el sistema de apareamiento (Zhao et al., 2014). Cuando el objetivo es aplicar apropiadamente mapeo de asociación en plantas, es un requisito previo caracterizar los niveles y patrones del DL en la población a utilizar. Es importante distinguir entre DL físico y las diferentes fuerzas que pueden crear desequilibrio de ligamiento en las poblaciones para evitar la detección de asociaciones espurias. Cuando existe DL altamente significativo entre los polimorfismos de diferentes cromosomas, puede producir asociaciones entre un marcador y un fenotipo a pesar de que el marcador no está ligado físicamente al *locus* responsable de la variación fenotípica (Flint-García et al., 2003). La clave para el mapeo de asociación es que el DL entre *loci* y marcadores sea debido al ligamiento físico. La caída del DL sobre la distancia física en una población determina la densidad de la cobertura con marcadores para realizar un análisis de asociación. Si el DL decae rápidamente, es necesaria una alta densidad de marcadores para capturar marcadores ligados a *loci* funcionales (Yu and Buckler 2006). El mapeo de asociación explora la variabilidad genética de líneas que no forman parte de un esquema de cruzamiento específico (Zhao et al., 2014). Warburton et al., (2008) sugieren que las líneas de maíz desarrolladas por el Centro Internacional de Mejoramiento de Maíz y Trigo (CIMMYT), son una valiosa fuente para estudios de mapeo de asociación debido a que poseen una gran diversidad alélica. Así es que Dao et al., (2014), Semang et al., (2012) y Suwarno et al., (2015) informaron el uso de líneas de maíz del CIMMYT para un estudio de diversidad genética, caracterización molecular y mapeo de asociación, respectivamente. La identificación de secuencias de ADN, mediante mapeo de asociación, que contribuyen a caracteres cuantitativos pueden ser utilizadas en los

programas de mejoramiento para realizar selección asistida por marcadores (MAS). La selección asistida por marcadores en los programas de mejoramiento implica la selección indirecta de un fenotipo deseado por la presencia o la ausencia de patrones de bandas de marcadores moleculares localizados dentro o cerca de los genes que controlan determinado carácter fenotípico. Se ha observado que la aplicación de MAS en programas de mejoramiento aumenta significativamente la tasa de ganancia genética en comparación con el mejoramiento convencional (Edwards and McCouch 2007).

La elección del germoplasma es fundamental para el éxito de los análisis de asociación. El germoplasma debe abarcar tanta variación fenotípica como sea posible. La caracterización genética puede ser utilizada para identificar diversas sub-poblaciones en el germoplasma (Flint-García et al., 2005). Una restricción para el uso del mapeo de asociación en las plantas es la presencia de estratificación en la estructura genética de la población y la relación genética (Hao et al., 2015, Peña Malavera et al., 2014). El problema surge debido a una distribución desigual de las frecuencias alélicas entre las sub-poblaciones. Las frecuencias alélicas para el fenotipo y los marcadores pueden diferir entre sub-poblaciones. Por lo tanto, valores altos de DL en algunas sub-poblaciones pueden resultar de la estratificación poblacional y no del efecto del ligamiento. Antes de iniciar estudios de mapeo de asociación, resulta necesario calcular empíricamente la magnitud del mismo y describir la estructura poblacional mediante el uso de gran número de marcadores independientes (Pritchard and Rosenberg, 1999). Para la resolución del mapeo de asociación y para detectar estratificación poblacional se han utilizado métodos estadísticos tal como los que utiliza el software STRUCTURE (Pritchard et al., 2000). Este software implementa un método de agrupamiento basado en modelos, que permite inferir la estructura de la población utilizando datos de genotipo basados en marcadores no ligados. Se asume un modelo en el que hay K poblaciones, cada uno de estas K se caracteriza por un conjunto de frecuencias de alelos en cada locus. Los individuos son asignados probabilísticamente a las poblaciones, o conjuntamente a dos o más poblaciones si sus genotipos indican que se mezclan. Los individuos son asignados a las poblaciones de tal manera que dentro de ellas, los *loci* se encuentren en equilibrio Hardy-Weinberg y en equilibrio de ligamiento (Pritchard et al., 2000).

La estadística multivariada es usada para describir y analizar observaciones multidimensionales obtenidas al relevar información sobre varios caracteres para cada uno de los genotipos en estudio (Balzarini et al., 2008). El análisis de componentes principales (ACP) es una técnica generalmente utilizada para reducción de dimensión.

Permite obtener nuevas variables mediante la combinación lineal de las variables originales. Estas componentes principales (CP) son no correlacionadas, por lo que son óptimas para describir la variabilidad entre los genotipos en estudio. Cuando las CP son obtenidas a partir de la combinación de marcadores moleculares de los genotipos de la población permiten describir la variabilidad genética entre los genotipos (Peña Malavera et al., 2014).

La caracterización de materiales de un banco de germoplasma con marcadores moleculares y/o caracteres agro-morfológicos es una práctica común, pero los estudios que usan simultáneamente marcadores moleculares y caracteres agro-morfológicos son menos frecuentes (Bramardi et al., 2005; Demey, 2008). Cuando se han obtenido dos configuraciones sobre la misma muestra de genotipos porque se utilizaron marcadores moleculares y caracteres agro-morfológicos para realizar el ordenamiento, se requiere de una técnica para analizar la congruencia de dichas configuraciones. El análisis de procrustes generalizado propuesto por Gower (1975) armoniza las configuraciones individuales a través de pasos algebraicos iterativos que transforman cada configuración individual. Estos pasos incluyen la traducción, rotación, reflexión y escalamiento de las componentes de sus puntos bajo dos premisas: mantener la distancia relativa entre elementos de las configuraciones individuales y minimizar las sumas de cuadrados entre puntos que corresponden al mismo genotipo bajo diferentes configuraciones.

HIPÓTESIS

Los genotipos de maíz provenientes de colecciones de germoplasmas poseen variabilidad y constituyen un material valioso para estudios de mapeo de asociación. La caracterización de líneas endocriadas de maíz permite analizar la variabilidad y describir la estructura genética presente en una población.

OBJETIVO GENERAL

Analizar la variabilidad y describir la estructura genética existente en una población diversa de líneas endocriadas de maíz mediante caracteres agro-morfológicos y moleculares de ADN.

OBJETIVOS ESPECÍFICOS

- i) Evaluar la variabilidad fenotípica en una población de líneas de maíz.
- ii) Describir la estructura genética en una población de líneas de maíz.
- iii) Evaluar la variabilidad genética en una población de líneas de maíz.
- iv) Analizar el desequilibrio de ligamiento (DL) entre pares de *loci*.
- v) Evaluar el consenso entre ordenamientos fenotípicos y genotípicos.

MATERIALES Y MÉTODOS

Material vegetal y ensayo de campo

Una población diversa de 252 líneas de maíz desarrolladas en el Centro Internacional de Mejoramiento de Maíz y Trigo (CIMMYT) fue caracterizada por su fenotipo y por su genotipo (Cuadro 1). La evaluación de las líneas de maíz se realizó en el campo experimental de la Facultad de Agronomía y Veterinaria de la Universidad Nacional de Río Cuarto (CAMDOCEX, FAV-UNRC; 33° 06' S; 64° 17' O). El diseño utilizado fue un diseño completamente al azar con una repetición. Cada genotipo se estableció en parcelas de un surco de 2,0 m de largo espaciados a 0,52 m. La siembra se realizó durante la primera semana de diciembre de 2015.

Evaluación del fenotipo

La población de 252 líneas de maíz fue caracterizada fenotípicamente mediante los caracteres agro-morfológicos días a floración femenina (DFF), días a floración masculina (DFM), intervalo antesis-estigma (IAE), altura de planta (AP), altura de espiga (AE) y relación AP/AE. Se midió el ángulo de inserción (AH) de las dos primeras hojas por encima de la espiga principal. Para este carácter se establecieron tres categorías: ángulo menor a 30°, ángulo entre 30° y 45° y ángulo mayor a 45° (Figura 1). En la estructura reproductiva masculina los datos obtenidos fueron el largo de panoja (LP) y la forma de panoja (FP). Para esta última se establecieron tres formas: laxa (L), intermedia (I) y compacta (C) (Figura 2). Cada genotipo se clasificó por su tipo de grano (TG) en dentado (D), intermedio (I) y flint (F), y por el color de grano (CG) en anaranjado (A) y blanco (B) (Figura 3). También se evaluó el comportamiento frente a vuelco (VU) y quebrado (QU). Se consideraron plantas volcadas aquellas que se presentaban totalmente caídas desde la raíz a causa de una pudrición en su base. Mientras que se definieron como quebradas todas aquellas plantas que presentaban un quebrado verde del tallo en alguno de sus nudos. Para una mejor interpretación de los datos, los caracteres VU y QU fueron categorizados. Las categorías se definieron teniendo en cuenta el porcentaje de plantas afectadas en cada genotipo de la siguiente manera: 0-25%; 25-50%; 50-75%; 75-100%.

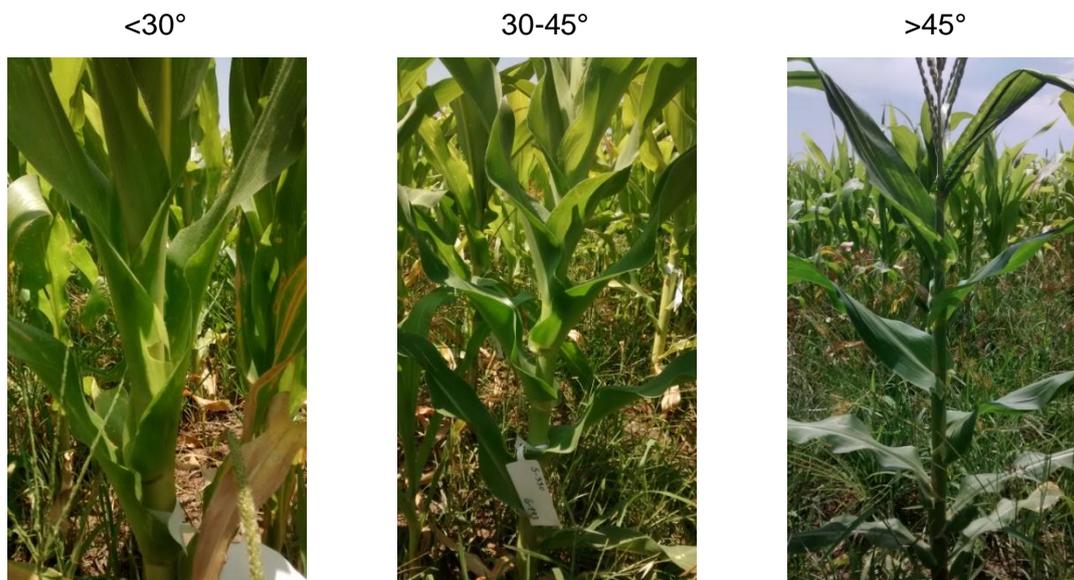


Figura 1. Categorías del ángulo de hoja observado en la población de las 252 líneas de maíz de CIMMYT, evaluadas en el campo experimental de la FAV-UNRC durante el ciclo agrícola 2015-2016.

Laxa

Intermedia

Compacta



Figura 2. Categorías del tipo de panoja observado en la población de las 252 líneas de maíz de CIMMYT, evaluadas en el campo experimental de la FAV-UNRC durante el ciclo agrícola 2015-2016.



Figura 3. Tipo de grano y color de grano observado en la población de las 252 líneas de maíz de CIMMYT, evaluadas en el campo experimental de la FAV-UNRC durante el ciclo agrícola 2015-2016.

D: dentado; I: intermedio; F: flint

A: anaranjado; B: blanco

Evaluación del genotipo

Las líneas de maíz incluidas en este estudio fueron caracterizadas con marcadores moleculares SNP. Estos se obtuvieron mediante genotipado por secuenciación. La información genotípica (SNP) fue provista por CIMMYT. A partir de un total de 53836 SNP polimórficos, un grupo de 2530 SNP se utilizaron para realizar el análisis de diversidad y la descripción de la estructura genética. Los marcadores fueron seleccionados por presentar una tasa de datos faltantes menor al 5% y una frecuencia superior al 0,25 en el alelo menos frecuente. Los marcadores seleccionados corresponden a los 10 cromosomas y tuvieron en promedio un 3% de datos faltantes.

Análisis estadístico del fenotipo

La descripción de los caracteres agro-morfológicos se realizó mediante estadística descriptiva, medidas de resumen para los caracteres cuantitativos y tabla de frecuencia para los caracteres categóricos. Con la finalidad de aproximar la distribución teórica subyacente en cada carácter se realizaron histogramas de frecuencias para los caracteres continuos y gráficos de sectores para los categóricos. Los caracteres AP, AE, AP/AE, DFM, DFF, IAE y LP fueron analizados con un enfoque multivariado como es el análisis de componentes principales (ACP) que permite analizar la interdependencia de variables métricas y encontrar una representación gráfica óptima de la variabilidad presente en la población. Los caracteres TG, FP, CG y AH fueron analizados mediante el análisis de correspondencia múltiple (ACM). Esta es una técnica exploratoria que permite representar gráficamente filas y columnas de una tabla de contingencia. Las observaciones multivariadas se graficaron en un biplot para así poder identificar las asociaciones de mayor peso entre las categorías de cada uno de los tres caracteres cualitativos (Greenacre, 1984). Los análisis se realizaron mediante el software INFOSTAT (Di Rienzo et al., 2016).

Análisis estadístico del genotipo

Estructura poblacional

El modelo bayesiano del software STRUCTURE 2.3.4 (Pritchard et al., 2000) se utilizó para describir la estructura genética de la población diversa de 252 líneas de maíz. Este programa permite realizar inferencias sobre la estructura poblacional a partir de las frecuencias alélicas de distintos *loci*, que puedan compartir un conjunto de genotipos y de esta forma asignarlos probabilísticamente a una sub-población k . El procedimiento utiliza la serie Cadena de Markov Monte Carlo (MCMC), después de un periodo de pre simulación de los procesos iniciales de corrida hasta alcanzar un nivel óptimo de convergencia entre el número de iteraciones y el logaritmo de la asignación probabilística ($\ln P(D)$). Este valor se obtiene calculando primero el logaritmo de verosimilitud de los datos en cada paso de la MCMC. A continuación, se calcula la media de estos valores y se le resta la mitad de su varianza. Los parámetros utilizados por STRUCTURE fueron el modelo ancestral “*admixture*” con un período de 10000 iteraciones seguido de 10000 repeticiones de MCMC. Se realizaron 15 corridas independientes para cada valor de k que varió entre uno y quince. Para determinar el número óptimo de k se utilizó la metodología del número de agrupamientos más probable según Evanno et al., (2005). Estos autores proponen una estadística basada en la tasa de cambio de segundo orden en el logaritmo de probabilidad de los datos entre valores de k sucesivos (Δk). Esta metodología muestra un pico claro en el verdadero número de k . Para aplicar esta metodología se utilizó el sitio web STRUCTURE HARVESTER (Earl and vonHoldt, 2012). Cada genotipo fue asignado a la sub-población, a la cual tenía una probabilidad de pertenencia mayor a 0,70. Un análisis molecular de la varianza (AMOVA) (Excoffier et al., 1992) entre todas las sub-poblaciones se realizó con INFOGEN para investigar diferencias poblacionales entre las sub-poblaciones formadas por STRUCTURE. También se calculó el coeficiente Φ_{iST} que es la correlación de líneas ó genotipos dentro de una misma sub-población, relativa a la correlación entre cualquier par de líneas ó genotipos seleccionado al azar de la población total.

Diversidad genética

La diversidad genética y el contenido de información polimórfica se estimaron mediante el programa INFOGEN (Balzarini y Di Rienzo, 2014). La diversidad genética es una medida de variabilidad apropiada para poblaciones endocriadas donde hay muy pocos individuos heterocigotas pero muchos tipos diferentes de alelos que se encuentran en homocigosis. La diversidad genética es una función de la heterocigosis y se calcula a partir de la suma de cuadrados de frecuencias alélicas. Así, para m loci, la diversidad genética promedio es:

$$D = 1 - \frac{1}{m} \sum_{j=1}^m \sum_{i=1}^l p_{ij}^2$$

donde p_{ij} es la frecuencia del alelo i en el locus j (Balzarini et al., 2010).

El contenido de información polimórfica (PIC) según Botstein et al., (1980) es definido como:

$$PIC = 2 \sum_{i=2}^l \sum_{j=1}^{i-1} [p_i p_j (1 - p_i p_j)]$$

Para el ordenamiento de las líneas basado en la caracterización molecular se calculó la distancia entre ellas, mediante la métrica de Jaccard, combinando la información provista por todos los marcadores. Las distancias de a pares se usaron luego para construir una matriz de distancias a partir de la cual se pueden ordenar los materiales. El ordenamiento (visualización de las distancias en un plano) se realizó mediante análisis de coordenadas principales (ACoP) (Balzarini et al., 2010).

Kinship relativo y distancia genética entre líneas

La matriz de estimaciones de parentesco y la matriz de distancia genética de Roger's, por pares de líneas, se calcularon con el software TASSEL 5.2.30 (Bradbury et al., 2007) y el software Power Marker V3.25 (Liu and Muse, 2005), respectivamente. Estas matrices se realizaron con 2530 marcadores SNP.

Desequilibrio de ligamiento

El software TASSEL 5.2.30 (Bradbury et al., 2007) se utilizó para calcular el desequilibrio por ligamiento entre los 2530 marcadores SNP. Con el objetivo de observar el comportamiento del DL en la población se calculó el valor de r^2 en las 252 líneas de maíz y en las dos sub-poblaciones de mayor tamaño (al menos 20 líneas). Se realizó sólo para estas dos sub-poblaciones porque el software TASSEL sólo calcula los valores de r^2 y p cuando hay al menos 20 individuos. Cada par de marcador se clasificó como no ligado (marcadores localizados en diferentes cromosomas) o ligados (marcadores localizados en el mismo cromosoma). Para tener una visión general del DL presente en todo el genoma y observar bloques de ligamientos se presentan los gráficos de DL por pares de marcadores para cada uno de los diez cromosomas del genoma del maíz. El grado de DL en todo el genoma se evaluó en función de los valores de r^2 (coeficientes de correlación al cuadrado entre pares de marcadores) y la distancia física entre estos SNP (Remington et al. 2001).

Análisis estadístico fenotipo-genotipo

Con el objetivo de analizar el grado de ajuste entre el ordenamiento fenotípico y genotípico se realizó un análisis de procrustes generalizado (Gower 1975). Para el ordenamiento fenotípico se incluyeron los trece caracteres en un nuevo ACP que se realizó mediante el software INFOSTAT (Di Rienzo et al., 2016) En el caso del ordenamiento genotípico se realizó un análisis de componentes principales a partir de los datos de los 2530 marcadores moleculares en las 252 líneas de maíz. Este ACP se realizó mediante el software TASSEL (Bradbury et al., 2007). En ambos casos se seleccionaron las componentes que tenían un auto valor igual ó mayor a uno.

Cuadro 1. Número e identificación de cada una de las 252 líneas de maíz evaluadas en el campo experimental de la FAV-UNRC durante el ciclo agrícola 2015-2016.

Nro.	A	B	Nro.	A	B	Nro.	A	B	Nro.	A	B
1	1	CML1	33	36	CML66	65	74	CML126	97	108	CML276
2	2	CML2	34	37	CML67	66	75	CML127	98	109	CML277
3	3	CML3	35	38	CML70	67	76	CML131	99	110	CML278
4	4	CML4	36	39	CML71	68	77	CML133	100	111	CML279
5	5	CML5	37	40	CML72	69	78	CML135	101	112	CML280
6	6	CML6	38	41	CML76	70	79	CML136	102	113	CML281
7	7	CML7	39	43	CML78	71	80	CML137	103	114	CML282
8	8	CML8	40	44	CML79	72	81	CML177	104	115	CML283
9	9	CML9	41	46	CML81	73	83	CML179	105	116	CML284
10	10	CML10	42	47	CML82	74	85	CML183	106	117	CML285
11	11	CML11	43	48	CML83	75	86	CML184	107	118	CML286
12	12	CML12	44	51	CML92	76	87	CML223	108	119	CML287
13	13	CML13	45	52	CML93	77	88	CML224	109	120	CML288
14	14	CML14	46	53	CML94	78	89	CML225	110	121	CML289
15	16	CML16	47	54	CML95	79	90	CML226	111	122	CML290
16	17	CML17	48	55	CML96	80	91	CML227	112	123	CML291
17	18	CML18	49	56	CML97	81	92	CML228	113	124	CML292
18	19	CML19	50	58	CML99	82	93	CML229	114	125	CML293
19	20	CML20	51	59	CML100	83	94	CML235	115	126	CML294
20	21	CML21	52	60	CML101	84	95	CML247	116	127	CML295
21	22	CML22	53	61	CML102	85	96	CML248	117	128	CML296
22	25	CML34	54	62	CML103	86	97	CML249	118	129	CML297
23	26	CML35	55	63	CML104	87	98	CML266	119	130	CML298
24	27	CML42	56	64	CML105	88	99	CML267	120	131	CML299
25	28	CML43	57	65	CML106	89	100	CML268	121	132	CML300
26	29	CML44	58	67	CML108	90	101	CML269	122	133	CML301
27	30	CML45	59	68	CML109	91	102	CML270	123	134	CML302
28	31	CML46	60	69	CML110	92	103	CML271	124	135	CML303
29	32	CML47	61	70	CML121	93	104	CML272	125	136	CML304
30	33	CML59	62	71	CML123	94	105	CML273	126	140	CML308
31	34	CML60	63	72	CML124	95	106	CML274	127	141	CML311
32	35	CML65	64	73	CML125	96	107	CML275	128	142	CML312

(Continuación Cuadro 1).

Nro.	A	B									
129	143	CML315	161	176	CML420	193	211	CML460	225	246	CML499
130	144	CML316	162	177	CML425	194	212	CML461	226	247	CML500
131	145	CML320	163	178	CML426	195	213	CML462	227	248	CML501
132	146	CML321	164	179	CML427	196	214	CML465	228	249	CML502
133	147	CML358	165	181	CML429	197	215	CML466	229	250	CML503
134	148	CML359	166	182	CML430	198	217	CML468	230	251	CML504
135	149	CML360	167	183	CML431	199	218	CML469	231	252	CML505
136	150	CML361	168	184	CML432	200	220	CML471	232	253	CML506
137	151	CML362	169	185	CML433	201	221	CML472	233	254	CML507
138	152	CML363	170	186	CML434	202	222	CML473	234	255	CML508
139	153	CML364	171	187	CML435	203	223	CML474	235	256	CML509
140	154	CML365	172	188	CML436	204	225	CML476	236	257	CML510
141	155	CML366	173	190	CML438	205	226	CML477	237	258	CML511
142	156	CML379	174	191	CML439	206	227	CML478	238	259	CML515
143	157	CML380	175	192	CML440	207	228	CML479	239	260	CML516
144	158	CML381	176	193	CML441	208	229	CML480	240	261	CML517
145	159	CML382	177	194	CML442	209	230	CML481	241	262	CML520
146	160	CML383	178	195	CML444	210	231	CML482	242	263	CML521
147	161	CML384	179	196	CML445	211	232	CML483	243	264	CML522
148	162	CML385	180	198	CML447	212	233	CML484	244	266	CML530
149	163	CML386	181	199	CML448	213	234	CML485	245	267	CML531
150	164	CML387	182	200	CML449	214	235	CML488	246	268	CML532
151	165	CML388	183	201	CML450	215	236	CML489	247	269	CML533
152	166	CML389	184	202	CML451	216	237	CML490	248	270	CML534
153	167	CML391	185	203	CML452	217	238	CML491	249	271	CML535
154	168	CML392	186	204	CML453	218	239	CML492	250	272	CML536
155	169	CML393	187	205	CML454	219	240	CML493	251	273	CML538
156	170	CML394	188	206	CML455	220	241	CML494	252	274	CML539
157	171	CML395	189	207	CML456	221	242	CML495			
158	173	CML403	190	208	CML457	222	243	CML496			
159	174	CML408	191	209	CML458	223	244	CML497			
160	175	CML409	192	210	CML459	224	245	CML498			

A: identificación utilizada en este estudio.

B: Identificación provista por CIMMYT.

RESULTADOS

Evaluación del fenotipo

Para describir los caracteres agro-morfológicos días a floración masculina y femenina, intervalo antesis–estigma, altura de planta, altura de espiga, relación entre altura de planta y altura de espiga y el largo de panoja se aplicó estadística descriptiva. Las medidas de resumen para los siete caracteres agro-morfológicos cuantitativos se presentan en el Cuadro 2. Se puede observar que la media de los días a floración fue de 75 y 80 días para la floración masculina y femenina, respectivamente. El intervalo antesis-estigma medio fue de 4 días con un amplio rango de variación que muestra la presencia de genotipos con diferente sincronía floral. La altura de planta varió entre 55 y 280 cm con una media de 192 cm. En altura de la espiga se observó una media de 92 cm con un mínimo de 30 y un máximo de 150 cm. En general la AP fue el doble de la AE. El largo de panoja osciló entre 12 y 50 cm con una media de 33 cm.

Cuadro 2. Media, error estándar (EE), valor mínimo (Mín.) y máximo (Máx.) para los siete caracteres agro-morfológicos cuantitativos medidos en las 252 líneas de maíz de CIMMYT evaluadas en el campo experimental de la FAV-UNRC durante el ciclo agrícola 2015-2016.

Carácter	Media	EE	Mín.	Máx.
DFM	75,57	0,34	66,00	91,00
DFF	80,08	0,41	68,00	94,00
IAE	-4,51	0,23	-18,00	3,00
AP	192,22	0,57	55,00	280,00
AE	92,37	0,44	30,00	150,00
AP/AE	2,14	0,01	1,00	4,22
LP	32,70	0,20	12,00	50,00

DFM: días a floración masculina; DFF: días a floración femenina; IAE: intervalo antesis-estigma; AP: altura de planta; AE: altura de espiga; AP/AE: relación AP/AE; LP: largo de panoja

Los caracteres categóricos ángulo de hoja, forma de panoja, tipo de grano, color de grano, vuelco y quebrado se describieron mediante la tabla de frecuencias (Cuadro 3). En el carácter ángulo de hoja, el 71% de los genotipos registró un ángulo menor a 30° mientras que sólo en el 1% de los genotipos el ángulo fue mayor a 45°. El 83% de los genotipos presentó forma de panoja intermedia. En el tipo y color de grano, la distribución de los genotipos en cada categoría fue más equitativa. En la categoría 0-25% de vuelco y de quebrado se ubicaron el 78% y el 97% de los genotipos, respectivamente.

Cuadro 3. Tabla de frecuencia relativa para las categorías de cada uno de los seis caracteres agro-morfológicos categóricos medidos en las 252 líneas de maíz de CIMMYT evaluadas en el campo experimental de la FAV-UNRC durante el ciclo agrícola 2015-2016.

Carácter	Categorías	Frecuencia relativa
Ángulo de hoja	<30°	0,710
	30-45°	0,280
	>45°	0,010
Forma de panoja	Laxa	0,130
	Intermedia	0,830
	Compacta	0,040
Tipo de grano	Dentado	0,290
	Intermedio	0,400
	Flint	0,310
Color de grano	Anaranjado	0,360
	Blanco	0,640
Vuelco	0-25%	0,780
	25-50%	0,150
	50-75%	0,060
	75-100%	0,010
Quebrado	0-25%	0,970
	25-50%	0,020
	50-75%	0,004
	75-100%	0,004

Los histogramas de frecuencias permiten aproximar la distribución teórica subyacente en cada carácter. En la Figura 4 se muestran los gráficos de distribución de frecuencias de los caracteres días a floración masculina (A), días a floración femenina (B) y el intervalo antesis-estigma (C). En el gráfico A se puede observar que la floración masculina, en el 90% de los genotipos, osciló entre 68 y 83 días. Mientras que la floración femenina, ocurrió entre los 75 y 87 días en aproximadamente el 65% de los genotipos (Figura 4 B). El 65% de los genotipos presentaron un intervalo antesis-estigma entre 1 y 7 días, mientras que sólo en el 5% de los genotipos se produjo la aparición de la flor femenina antes que la flor masculina, lo que normalmente es conocido como protoginia (Figura 4 C).

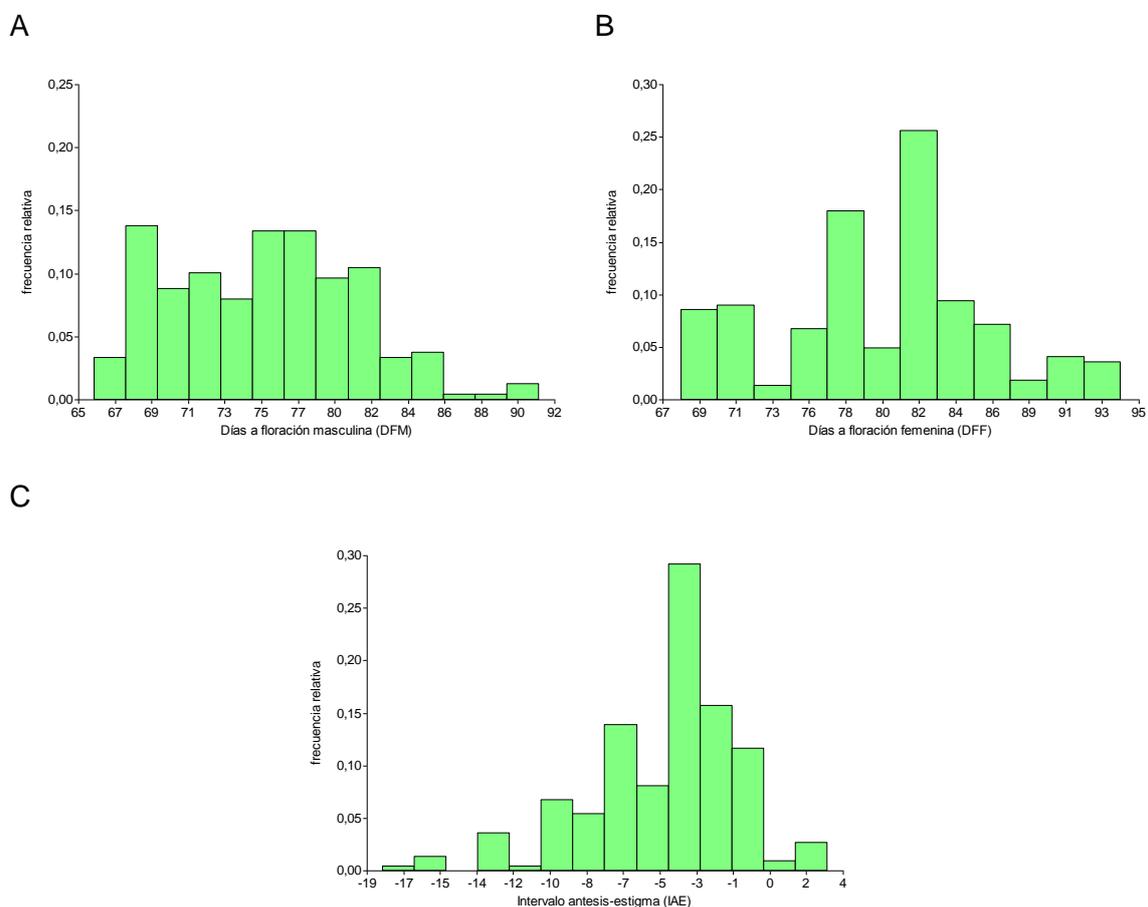


Figura 4. Distribución de la frecuencia relativa de los caracteres días a floración masculina (A), días a floración femenina (B) e intervalo antesis-estigma (C) en las 252 líneas de maíz de CIMMYT evaluadas en el campo experimental de la FAV-UNRC durante el ciclo agrícola 2015-2016.

Las distribuciones para los caracteres altura de planta, altura de espiga y la relación entre altura de planta y altura de espiga se presenta en la Figura 5. La distribución para altura de planta se asemeja a una distribución normal parcialmente sesgada hacia la derecha. El 85% de las líneas de maíz tiene una altura de planta que varía entre 155 y 230 cm (Figura 5 A). El carácter altura de espiga presentó una distribución normal con el 80% de los genotipos entre los valores 65 y 115 cm (Figura 5 B). En la Figura 5 C se observa que en la mayoría de los genotipos la altura de planta duplica a la altura de espiga.

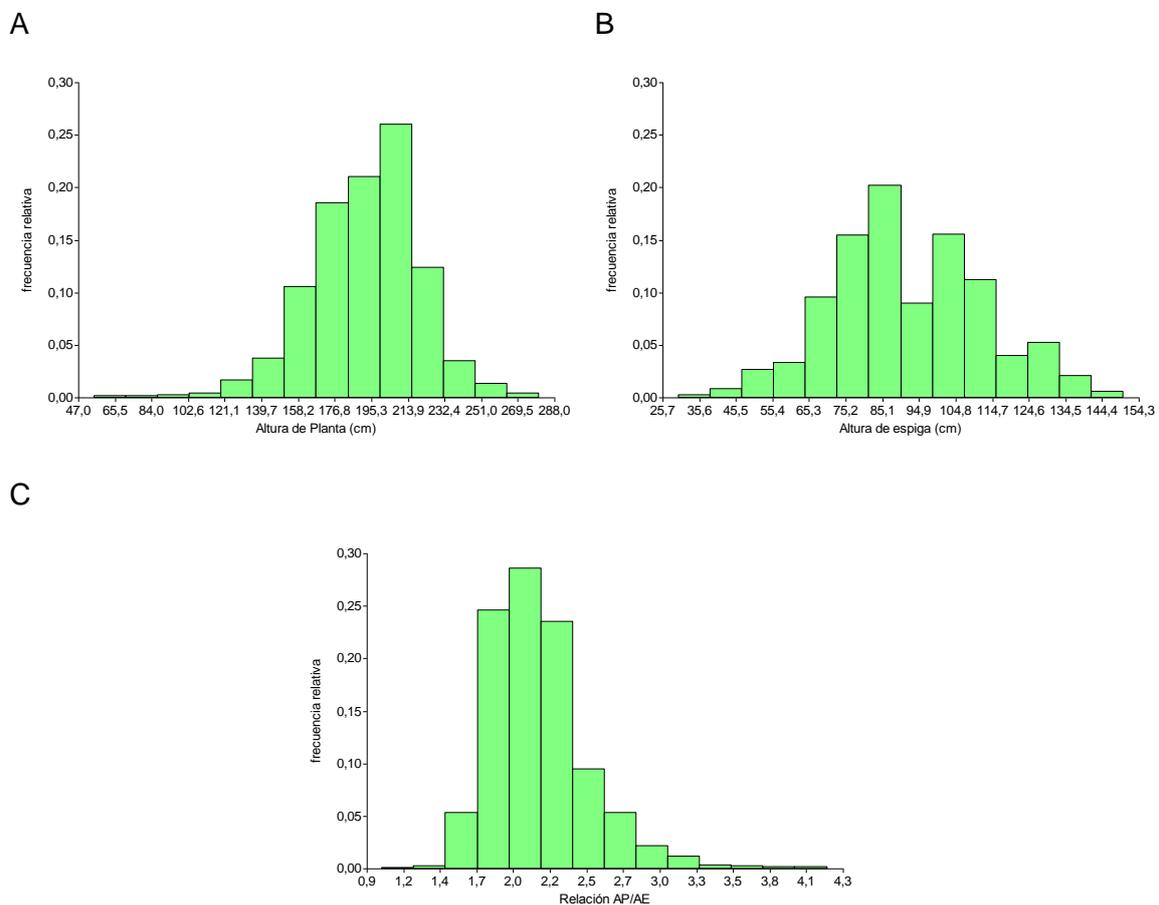


Figura 5. Distribución de la frecuencia relativa de los caracteres altura de planta (A), altura de espiga (B) y la relación altura de planta/altura de espiga (C) en las 252 líneas de maíz de CIMMYT evaluadas en el campo experimental de la FAV-UNRC durante el ciclo agrícola 2015-2016.

El gráfico de sectores es útil para representar contribuciones porcentuales a un total ó la distribución de frecuencias de un carácter categórico. La Figura 6 muestra el gráfico de la distribución poblacional respecto al carácter ángulo de hoja. Se observa que los genotipos se ubicaron en las tres categorías definidas, no obstante se muestra una prevalencia de ángulos de hoja menores a 30° donde se localiza el 70% de las líneas. Mientras que sólo dos genotipos presentaron hojas con ángulo mayor a 45°. Es decir que la mayoría de los genotipos presentan hojas erectas.

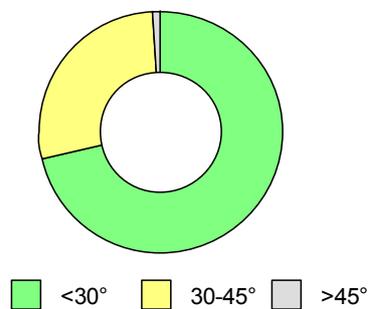


Figura 6. Distribución de la frecuencia poblacional para el carácter ángulo de hoja en las 252 líneas de maíz de CIMMYT evaluadas en el campo experimental de la FAV-UNRC durante el ciclo agrícola 2015-2016.

La caracterización de la estructura reproductiva masculina se realizó mediante la evaluación de los caracteres largo de panoja y tipo de panoja. En el largo de panoja, la distribución muestra que los valores oscilaron entre 13 y 50 cm. Sin embargo, el 70% de los genotipos tuvo un largo de panoja entre 27 y 38 cm (Figura 7 A). La Figura 7 B muestra las distribuciones de frecuencias para el carácter tipo de panoja, el gráfico permite observar que existe variabilidad para este carácter, no obstante se puede destacar una prevalencia del tipo de panoja intermedia que reúne aproximadamente al 80% de las líneas evaluadas en este estudio.

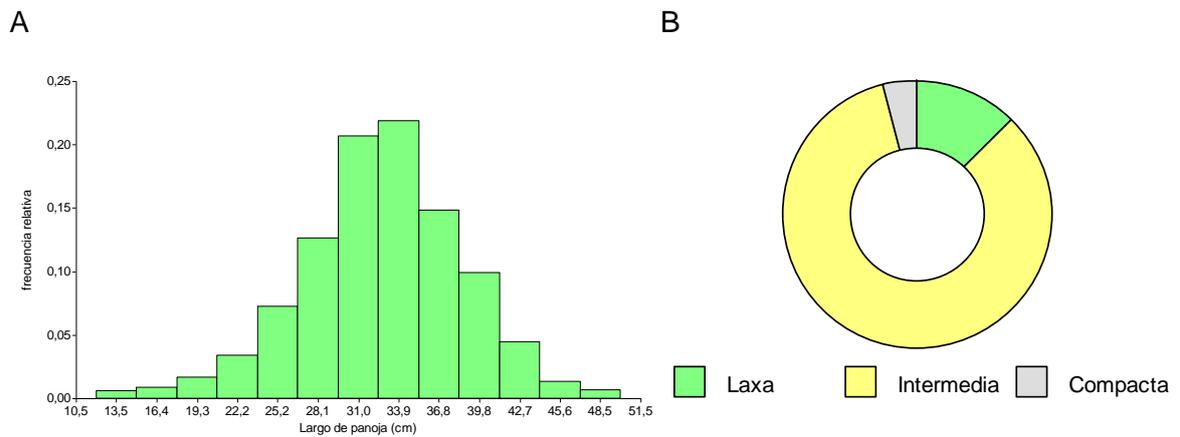


Figura 7. Distribución de la frecuencia relativa de los caracteres largo de panoja (A) y distribución de la frecuencia poblacional de las tres categorías de tipo de panoja (B) en las 252 líneas de maíz de CIMMYT evaluadas en el campo experimental de la FAV-UNRC durante el ciclo agrícola 2015-2016.

Las Figuras 8 A y 8 B muestran la variabilidad presente en la población de líneas de maíz para tipo y color de grano, respectivamente. Amplia variabilidad se observa en el carácter tipo de grano debido a la distribución equitativa de genotipos en cada una de las tres categorías. En la población de líneas de maíz de este estudio hay prevalencia de granos blancos por sobre los anaranjados.

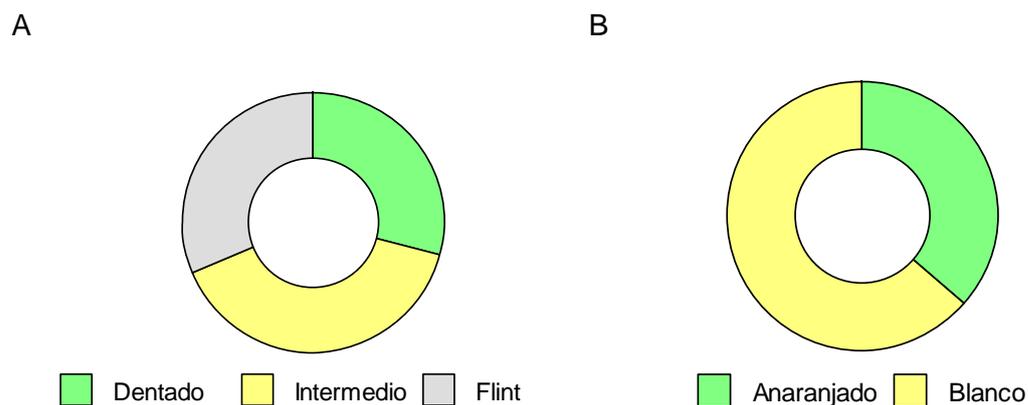


Figura 8. Distribución de la frecuencia poblacional de tipo de grano (A) y color de grano (B) en las 252 líneas de maíz de CIMMYT evaluadas en el campo experimental de la FAV-UNRC durante el ciclo agrícola 2015-2016.

Para una mejor comprensión del comportamiento de la población de líneas de maíz para los caracteres vuelco y quebrado se definieron cuatro categorías. La proporción de las diferentes categorías observadas se muestra en la Figuras 9 A y 9 B. Los gráficos permiten observar un mejor comportamiento de las líneas evaluadas frente a quebrado respecto a vuelco. Ya que 240 genotipos de un total de 252 presentaron menos del 25% de plantas quebradas. Mientras que aproximadamente el 15% y el 6% de las líneas presentaron valores de hasta 50% y 75% de plantas volcadas, respectivamente. En general un buen comportamiento de los genotipos frente a vuelco y quebrado debido a la prevalencia de la categoría que agrupa a los genotipos que presentaron entre 0 y 25% de incidencia de ambos caracteres.

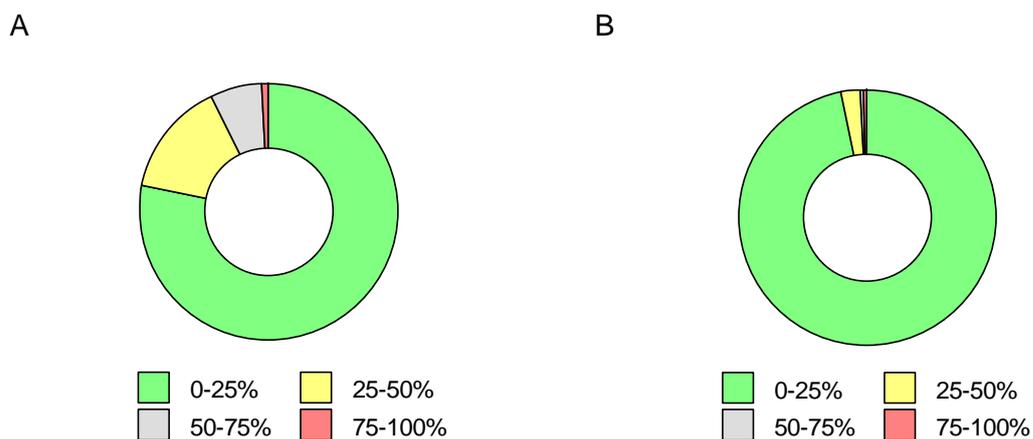


Figura 9. Distribución de la frecuencia poblacional de los caracteres vuelco (A) y quebrado (B) en las 252 líneas de maíz de CIMMYT evaluadas en el campo experimental de la FAV-UNRC durante el ciclo agrícola 2015-2016.

En la Figura 10 se visualizan las dos primeras dimensiones del análisis de correspondencia múltiple de los caracteres tipo de grano, color de grano, ángulo de hoja y forma de panoja. El gráfico sugiere, en su primer eje, que los genotipos con tipo de grano flint son generalmente de color anaranjado, mientras que los genotipos de grano dentado son blancos. Los genotipos con tipo de grano intermedio se asociaron con forma de panoja intermedia y ángulo de hoja erecta. Mientras que los genotipos con ángulo de hoja mayor a 45° y forma de panoja compacta se presentaron aislados en el gráfico.

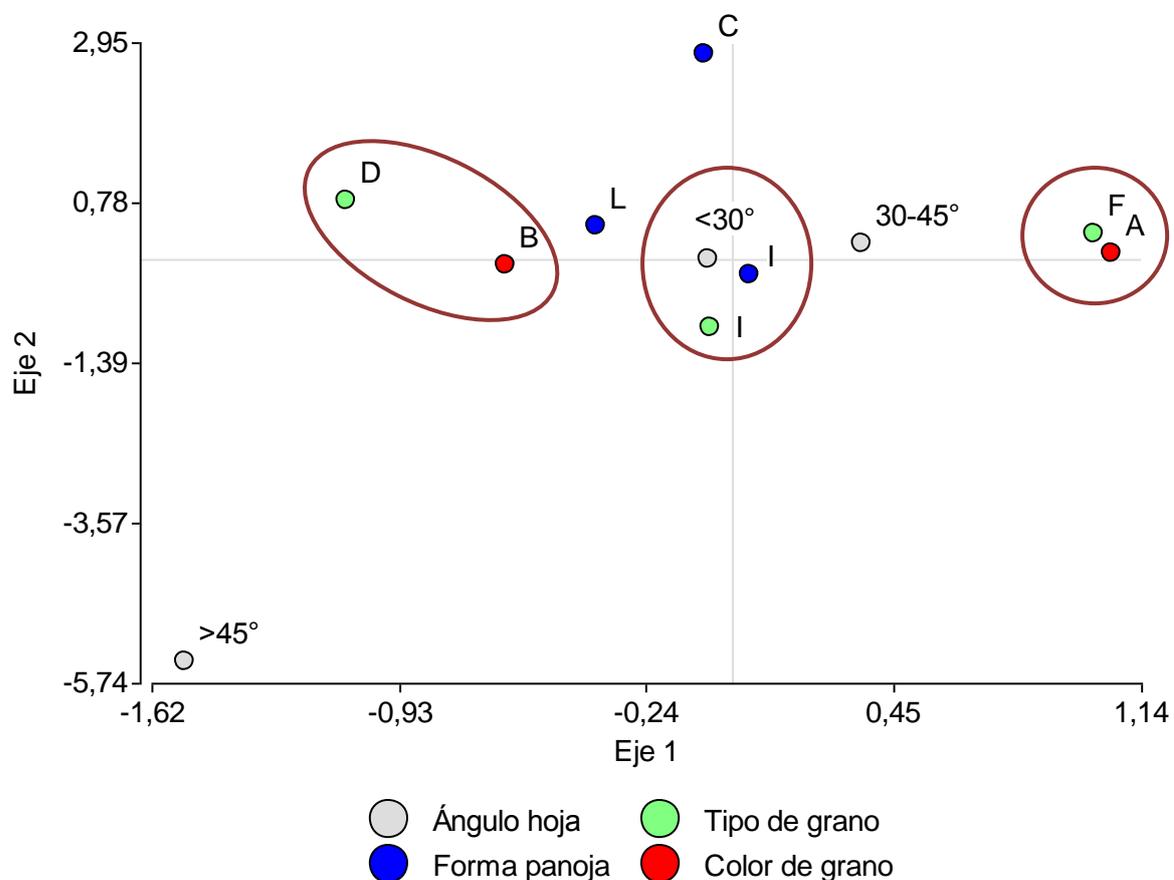


Figura 10. Gráfico biplot obtenido del análisis de correspondencia múltiple de los caracteres ángulo de hoja, forma de panoja, tipo de grano y color de grano en las 252 líneas de maíz de CIMMYT evaluadas en el campo experimental de la FAV-UNRC durante el ciclo agrícola 2015-2016.

<30°: ángulo de hoja menor a 30°, 30-45°: ángulo de hoja entre 30° y 45°, >45°: ángulo de hoja mayor a 45°; C: panoja compacta, I: panoja intermedia, L: panoja laxa; D: grano dentado, I: grano intermedio, F: grano flint; B: grano blanco, A: grano anaranjado.

En la Figura 11 se presenta el gráfico biplot generado a partir de las dos primeras componentes principales (CP) obtenidas del análisis de componentes principales (ACP) con siete caracteres agro-morfológicos. Las dos primeras CP explicaron el 65% de la variabilidad total. En el gráfico se observa que los caracteres que más contribuyen a la CP1 son altura de planta, altura de espiga, días a floración masculina y días a floración femenina, que se encuentran asociados positivamente entre ellos. A su vez se observa asociación negativa entre altura de espiga y la relación AP/AE. La

asociación negativa entre estos caracteres indica que en los genotipos con una altura de espiga mayor, esto no se debía, necesariamente, a una mayor altura de planta, si bien en los gráficos anteriores se observaba que generalmente la altura de planta era el doble de la altura de espiga. Los caracteres que más influyen en la composición de la CP2 son el IAE y el LP que presentan asociación negativa entre ellos. Respecto a la distribución de los genotipos en el gráfico se observa que las líneas 27, 96, 160, 166 y 260 se asociaron con AP, AE, DFM y DFF. Por otro lado, las líneas 48 y 208 se asociaron con el carácter AP/AE. Los genotipos 62, 65 y 222 se presentaron asociados a IAE, mientras que 142, 245 y 256 se asociaron con LP.

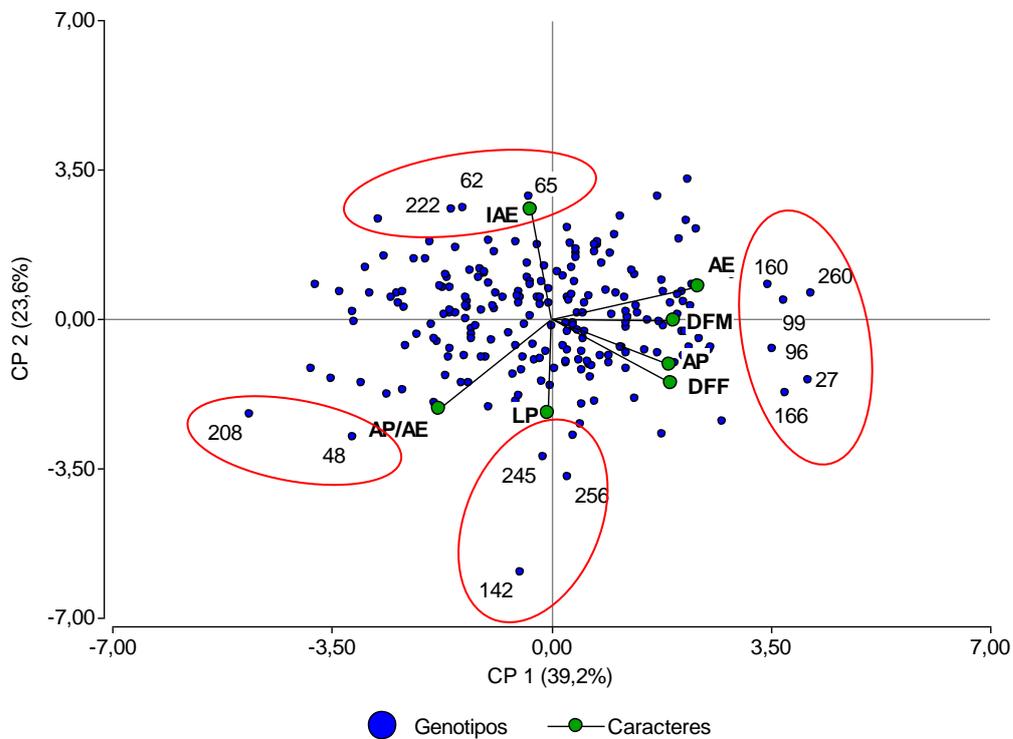


Figura 11. Gráfico biplot del análisis de componentes principales de siete caracteres agro-morfológicos en las 252 líneas de maíz de CIMMYT evaluadas en el campo experimental de la FAV-UNRC durante el ciclo agrícola 2015-2016. AP: altura de planta; AE: altura de espiga; AP/AE: relación altura de planta y altura de espiga; DFM: días a floración masculina; DFF: días a floración femenina; IAE: intervalo anthesis-estigma; LP: largo de panoja

Evaluación del genotipo

Estructura poblacional

Para describir la estructura genética presente en la población de 252 líneas de maíz se utilizaron 2530 marcadores moleculares SNP. El valor del logaritmo de asignación probabilística ($\ln P(D)$) para cada hipotético k aumentó con el incremento de k desde uno hasta diez sin mostrar ningún pico. Entre k diez y quince se observan los mayores desvíos para cada uno de los k inferidos (Figura 12 A). Sin embargo el Δk mostró un claro pico para $k=4$ y un rápido descenso en $k=5$ (Figura 12 B) indicando que las 252 líneas de maíz se pueden agrupar en cuatro sub-poblaciones. Cada línea está representada por una barra vertical, dividida en segmentos de colores donde la longitud de cada segmento indica la proporción del genoma de cada genotipo que pertenece a $k= 1, 2, 3$ ó 4 (Figura 12 C). Las líneas fueron asignadas a la sub-población para la cual tenían una probabilidad de pertenencia mayor a 0,70. Los grupos uno y tres fueron los de menor tamaño con seis y cinco genotipos, respectivamente. Los genotipos que no fueron asignados a ningún grupo, en este caso 79, se definieron como mixtos. Mientras que 104 líneas fueron asignadas al grupo dos y las restantes 58 al grupo cuatro.

La presencia de estratificación en la estructura genética de la población fue observada también al realizar el análisis molecular de la varianza, el cual indica que la variación genética total se repartió en un 4% entre sub-poblaciones y un 96% dentro de las sub-poblaciones (Cuadro 4). El valor del coeficiente Phi_{ST} fue de 0,04 y estadísticamente significativo ($P<0,0001$), lo cual sugiere que a pesar de la variabilidad observada dentro de sub-poblaciones, el factor población produce una estructuración genética. El índice de fijación de Wright's (F_{st}) presentó valores altos para las sub-poblaciones de menor tamaño y valores pequeños para aquellas de mayor tamaño.

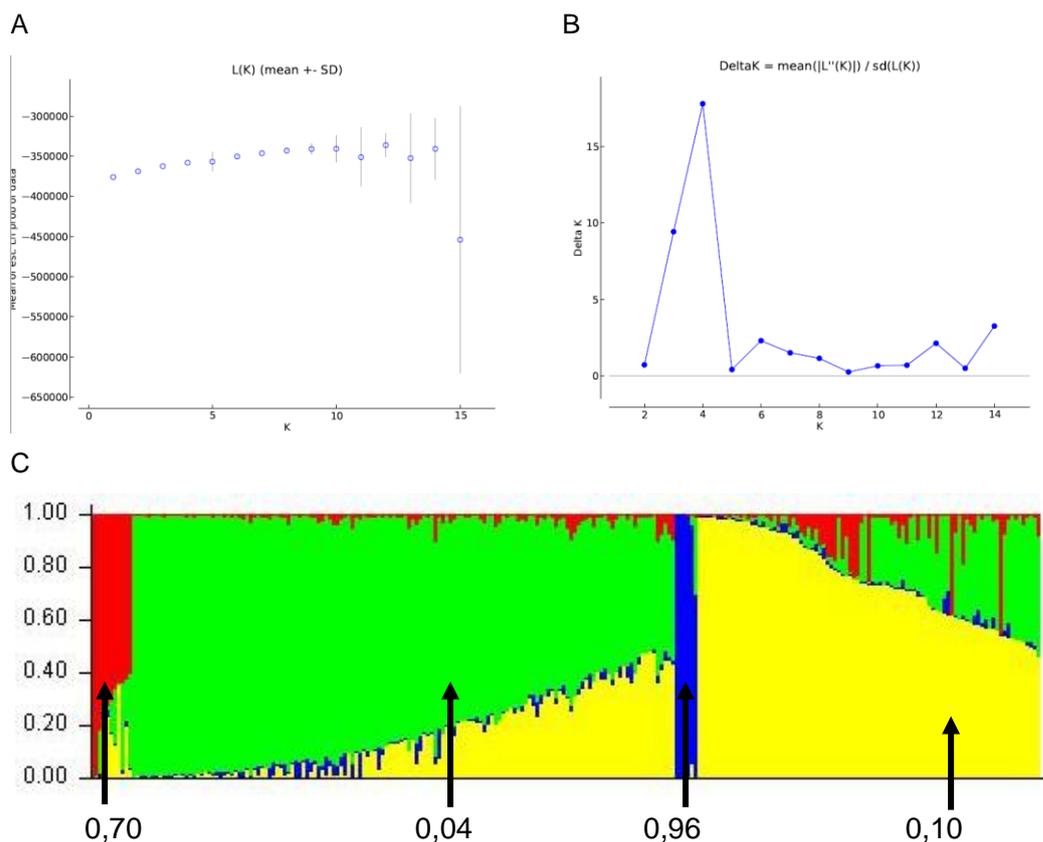


Figura 12. Análisis de la estructura poblacional de 252 líneas de maíz de CIMMYT. Valor del logaritmo de asignación probabilística ($\ln P(D)$) estimado (A) y Δk (B) obtenidos con STRUCTURE. Estructura de la población con k establecido en cuatro sub-poblaciones (C). Los números indicados debajo de cada sub-población corresponden a los valores del índice de fijación de Wright's (F_{st}).

Cuadro 4. Análisis molecular de la varianza (AMOVA) entre sub-poblaciones inferidas a partir de las 252 líneas de maíz de CIMMYT.

Fuente de variación	SC	GL	CM	P-valor	Porcentaje
Entre	3,76	4	0,94	<0,0001	4,08
Dentro	82,49	247	0,33	<0,0001	95,92
Total	86,25	251	0,34		100,00

SC: Suma de cuadrados; GL: Grados de libertad; CM: Cuadrado medio

Diversidad genética

La diversidad genética y el contenido de información polimórfica (PIC) fueron calculados para estimar la variabilidad genética de las 252 líneas de maíz de CIMMYT y para cada uno de las sub-poblaciones inferidas. La diversidad genética y el PIC totales fueron 0,50 y 0,43. Para cada sub-población los valores de diversidad genética variaron entre 0,04 y 0,49 y los valores de PIC entre 0,03 y 0,42 (Cuadro 5). La Figura 13 muestra el gráfico obtenido a partir del análisis de coordenadas principales (ACoP). Los distintos colores indican los individuos que pertenecen a cada una de las cuatro sub-poblaciones inferidas mediante STRUCTURE. Se puede observar que en general los individuos pertenecientes a un misma sub-población presentan menor distancia genética entre ellos.

Cuadro 5. Diversidad genética y contenido de información polimórfica (PIC) de las 252 líneas de maíz de CIMMYT y de cada sub-población.

Variabilidad	Total	Sub-población 1	Sub-población 2	Sub-población 3	Sub-población 4	Mixtos
Diversidad genética	0,50	0,24	0,49	0,04	0,48	0,50
PIC	0,43	0,20	0,42	0,03	0,41	0,42

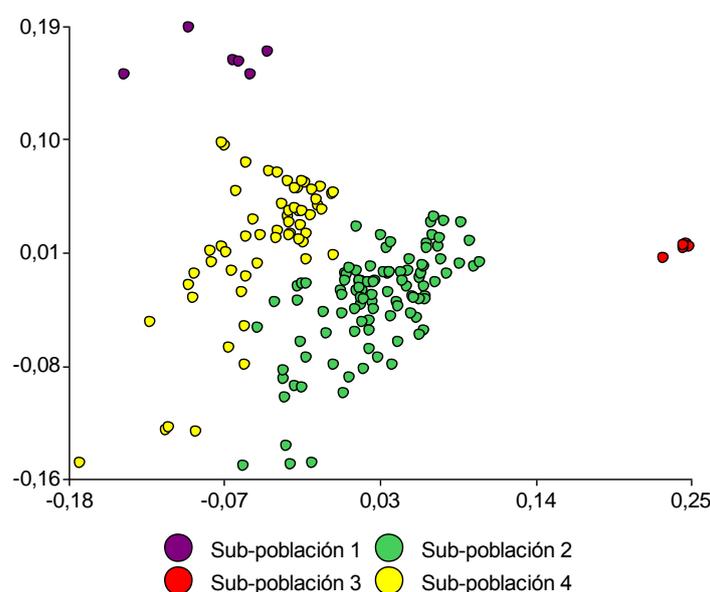


Figura 13. Gráfico del análisis de coordenadas principales obtenido a partir de 2530 SNP en las 252 líneas de maíz de CIMMYT. No se muestran los individuos clasificados como mixtos.

Kinship relativo y distancia genética entre líneas

Las estimaciones de parentesco relativo (*kinship* relativo) basados en los datos de SNP indican que el 85% de las estimaciones de parentesco entre pares de líneas registraron un valor menor a 0,10 (Figura 14 A). No obstante, en algunos pares de líneas, se observaron valores de las estimaciones de parentesco superiores a 0,30. La distancia genética de Roger's media entre las comparaciones por pares de las 252 líneas fue 0,44, dado que el 97% de las comparaciones registró valores entre 0,40 y 0,50 (Figura 14 B).

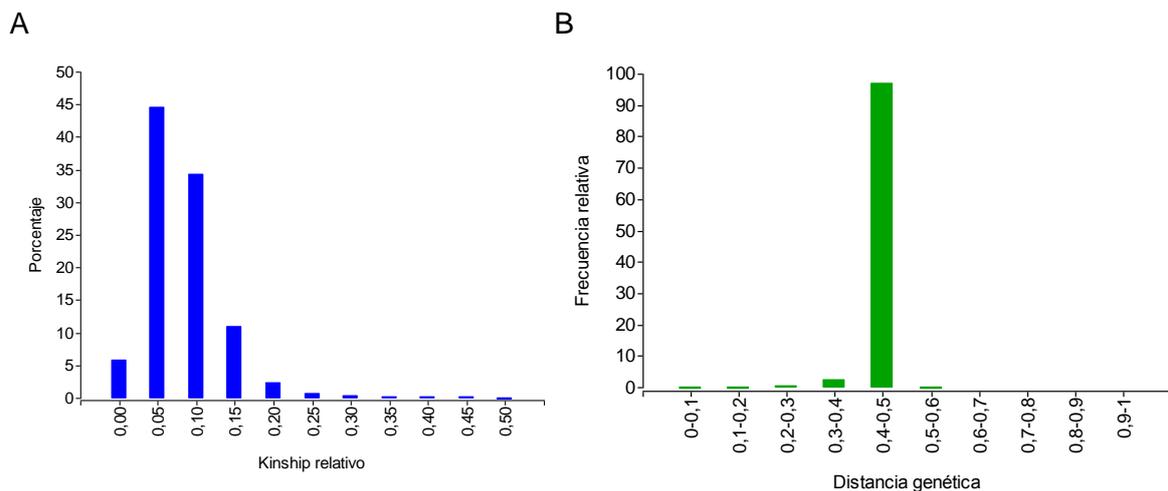


Figura 14. Distribución de la estimación de parentesco relativo por pares (*kinship* relativo) (A) y la distancia genética de Roger's (B) para las 252 líneas de maíz de CIMMYT utilizando 2530 SNP. Para simplicidad solo se muestra el porcentaje de los valores de *kinship* relativo entre 0,05 y 0,50.

Desequilibrio de ligamiento

Las estimaciones del desequilibrio de ligamiento por pares de marcadores se realizaron para las 252 líneas de maíz y para las dos sub-poblaciones que agrupaban al menos 20 individuos en cada una de ellas (Cuadro 6). Todos los valores de r^2 presentados en el Cuadro 6 corresponden al DL entre pares de marcadores que resultaron estadísticamente significativos ($p < 0,001$). En la población total, el r^2 medio de todos los marcadores fue 0,20. En las sub-poblaciones 2 y 4 el valor de r^2 fue 0,36 y 0,30, respectivamente. Al comparar el DL entre los marcadores ligados físicamente y aquellos que no poseen ligamiento físico se puede observar que el valor de r^2 es mayor para los marcadores ligados, en la población total y en cada una de las sub-poblaciones consideradas.

Cuadro 6. Desequilibrio de ligamiento de todo el genoma, en la población completa y los dos grupos con mayor número de genotipos asignados.

Población	Marcadores totales	Marcadores no ligados	Marcadores ligados
	r^2	r^2	r^2
Población total	0,20	0,07	0,22
Sub-población 2	0,36	0,17	0,36
Sub-población 4	0,30	0,18	0,86

Sub-población: definida según los resultados del análisis de estructura de las 252 líneas de maíz.

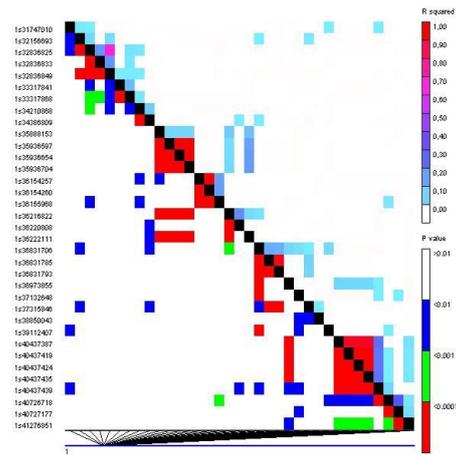
r^2 : coeficientes de correlación al cuadrado entre pares de marcadores SNP.

Marcadores no ligados: Pares de marcadores de diferentes cromosomas.

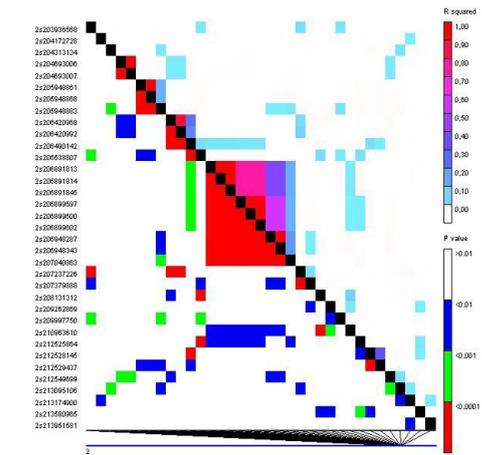
Marcadores ligados: Pares de marcadores de un mismo cromosoma.

Para tener una visión general del desequilibrio de ligamiento presente en todo el genoma y observar bloques de ligamientos se presentan los gráficos de desequilibrio de ligamiento por pares de marcadores para cada uno de los diez cromosomas del genoma del maíz. Para los 2530 SNP el DL se observó principalmente entre SNP con ligamiento físico. En los cromosomas 2, 8 y 9 se observaron grandes bloques de ligamiento. Mientras que en los cromosomas 5, 6 y 10 los bloques de ligamiento son de tamaño intermedio. En el resto de los cromosomas se observaron bloques de ligamiento de menor tamaño (Figura 15).

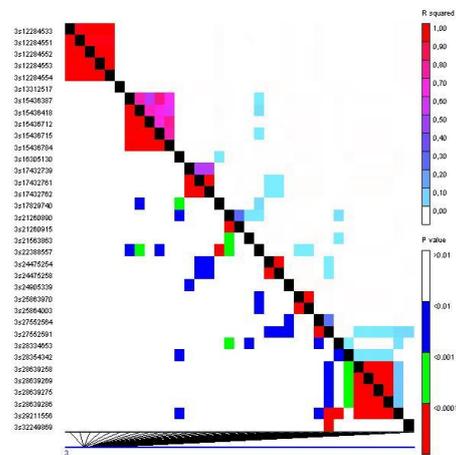
Cromosoma 1



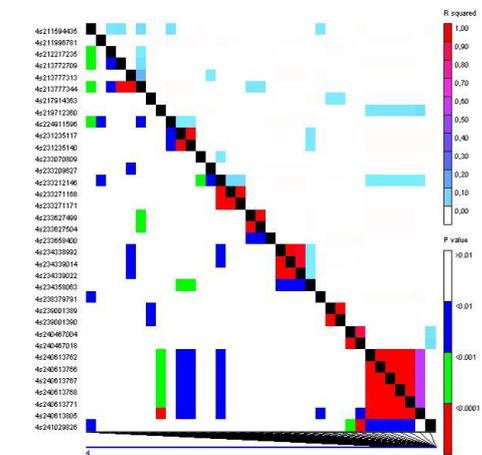
Cromosoma 2



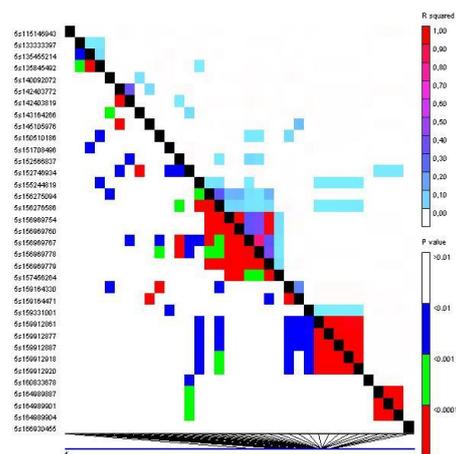
Cromosoma 3



Cromosoma 4



Cromosoma 5



Cromosoma 6

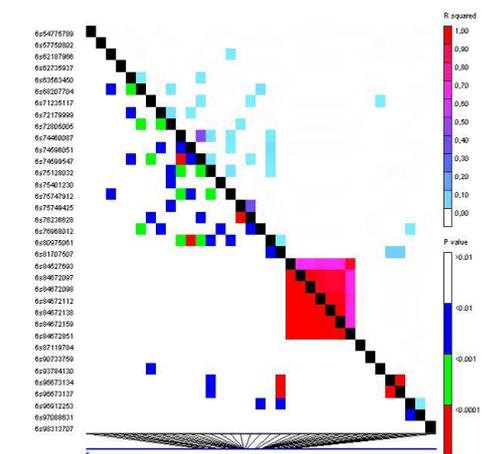


Figura 15. Desequilibrio de ligamiento entre 2530 SNP. En el eje X se indica la posición del marcador y en el eje Y el nombre de los marcadores.

(Fig. 15 continúa en pág. 42)

Los valores de coeficientes de correlación al cuadrado entre pares de marcadores (r^2) en función de la distancia física entre estos SNP se evaluó para observar la caída del desequilibrio de ligamiento en todo el genoma. La Figura 16 muestra que a medida que aumenta la distancia entre dos marcadores, el valor de DL disminuye rápidamente indicando que este se da entre SNP adyacentes. En esta población de líneas de maíz, el DL a través de todo el genoma decayó por debajo del valor $r^2=0,10$ a partir de una distancia de 100-300 kb.

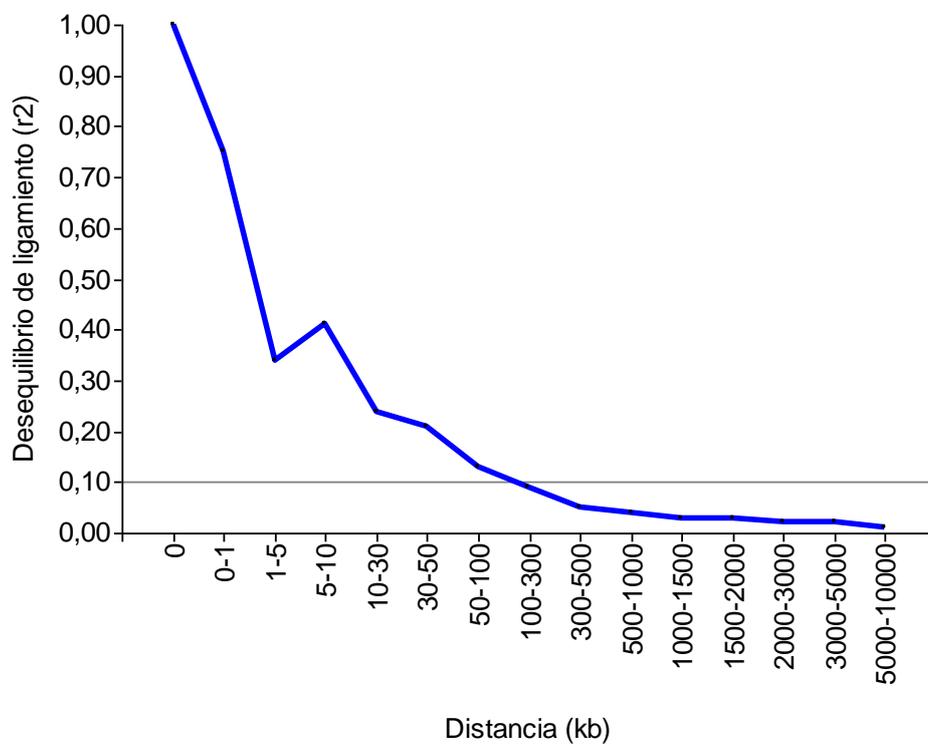


Figura 16. Desequilibrio de ligamiento observado en las 252 líneas de maíz de CIMMYT, medido con 2530 SNP a través de todo el genoma.

Análisis estadístico fenotipo-genotipo

El análisis de componentes principales permitió reducir la dimensión de la matriz de datos de los 13 caracteres agro-morfológicos medidos a sólo seis componentes principales, que explicaban el 75% de la variabilidad. En tanto que, la matriz de datos de marcadores moleculares se redujo de 2530 marcadores SNP a 147 componentes principales que explicaban el 79% de la variabilidad presente. En el Cuadro 7 se presenta la proporción de consenso obtenida a partir del análisis de procrustes generalizado realizado con 147 CP, obtenidas del análisis de componentes principales a partir de los 2530 marcadores SNP, y 6 CP obtenidas del análisis de componentes principales a partir de los 13 caracteres agro-morfológicos.

Cuadro 7. Suma de cuadrados del análisis de procrustes generalizado (APG) al combinar información proveniente de caracteres agro-morfológicos y de marcadores moleculares. Los caracteres fueron medidos en las 252 líneas de maíz de CIMMYT evaluadas en el campo experimental de la FAV-UNRC durante el ciclo agrícola 2015-2016.

Ordenamiento	Consenso	Residuo	Total	Proporción consenso
Fenotípico	0,59	0,41	1,00	0,59
Genotípico	0,59	0,41	1,00	0,59
Total	1,18	0,81	2,00	0,59

DISCUSIÓN

El CIMMYT es uno de los centros más importantes de recolección, conservación y utilización de germoplasma de maíz (Yan et al., 2009). La población de líneas de maíz evaluadas en este estudio mostró estar sometidas a varios años de mejoramiento ya que en general presentan panojas pequeñas, hojas erectas e intervalos antesis-estigma breves. Estos caracteres contribuyen en la eficiencia de transformación en granos de la luz solar, los nutrientes del suelo y el dióxido de carbono. Las hojas verticales mejoran la capacidad de la planta de capturar la luz solar para la fotosíntesis cuando la densidad de plantas es alta. Las panojas más pequeñas reducen los requerimientos energéticos de la planta para mantenimiento general y deja disponible más energía para la producción de granos. En los programas de mejoramiento no se aplicó selección específica para estos caracteres. Sin embargo, al seleccionar para mayor rendimiento durante un largo período de tiempo, se realizó selección indirecta para todos los caracteres que contribuyen al rendimiento (Duvick, 2004). Por otra parte, en el presente estudio se observó que las líneas evaluadas presentaron buen comportamiento frente a vuelco y quebrado, ya que la mayoría registró valores menores al 25% de incidencia en ambos caracteres.

Los marcadores moleculares proveen importante información biológica. Las técnicas biotecnológicas que involucran el uso de marcadores moleculares han facilitado la identificación y la cuantificación de la variabilidad genética para el uso y conservación de especies de interés. El conocimiento a nivel molecular de la diversidad genética de líneas *elite* y de poblaciones vegetales cultivadas y silvestres es fundamental para el uso, manejo y conservación de germoplasma (Edwards and McCouch 2007). El grado de variabilidad genética en una población se evalúa mediante el contenido de información polimórfica (PIC), la diversidad genética, los valores de los coeficientes *kinship* y la distancia genética entre individuos (Dinesh et al., 2016). En el presente estudio, la variabilidad genética observada en la población de las 252 líneas de maíz de CIMMYT fue caracterizada con 2530 marcadores moleculares SNP. Los valores de diversidad genética y PIC obtenidos concuerdan con los informados por Dinesh et al., (2016) y Semang et al., (2012) quienes también evaluaron mediante este tipo de marcadores moleculares a 64 y 450 líneas de maíz de CIMMYT, respectivamente. Todos estos resultados muestran la amplia variabilidad genética presente en las líneas de maíz de CIMMYT.

En general, las líneas endocriadas de maíz tienen una historia compleja, han sido derivadas de variedades de polinización abierta y luego cruzadas entre ellas. Esta

historia dificulta la localización de las líneas dentro de grupos que reflejen el grado de similitud genética. La información genealógica puede ser una importante guía, dado que líneas derivadas de antecesores comunes pueden compartir genes y por lo tanto tienen mayor probabilidad de ser agrupadas respecto a aquellas líneas que provienen de antecesores no comunes. Sin embargo, la selección y la deriva genética durante la endocria pueden causar discrepancias entre la genealogía y la constitución genética (Liu et al., 2003). La descripción de la estructura genética del maíz es de gran importancia para la selección de líneas parentales en el mejoramiento y en la reducción de falsas asociaciones en estudios de mapeo de asociación (Liu et al., 2014). En el presente estudio se clasificaron las líneas mediante el método de agrupamiento basado en modelos bayesianos, implementados en el software STRUCTURE (Pritchard et al., 2000). Este análisis permitió clasificar las 252 líneas de maíz de CIMMYT en cuatro sub-poblaciones. Estudios previos (Semang et al., 2012; Suwarno et al., 2015) también encontraron estratificación en la estructura genética poblacional de líneas de maíz de CIMMYT. Semang et al., (2012) identificaron que los patrones de agrupamiento no están basados en el fenotipo, es decir no se correlaciona con la adaptación ambiental, el color o tipo de grano ó con los grupos de madurez. En el presente estudio también fue observada esa falta de correlación entre los patrones de agrupamiento de las líneas con el fenotipo.

Los marcadores moleculares pueden ser usados para calcular el parentesco relativo entre pares de individuos en un estudio. El parentesco relativo refleja la identidad aproximada entre dos Individuos sobre la probabilidad media de identidad entre dos individuos aleatorios (Yan et al., 2009). En este estudio, los valores de los coeficientes del análisis de parentesco fueron en su mayoría menores a 0,50. Además, sólo un 0,16% de las distancias genéticas fueron menores a 0,10 indicando que cada línea de maíz aporta cierta variabilidad genética que puede ser de interés en un programa de mejoramiento. Los valores observados en este estudio concuerdan con los informados por Dinesh et al., (2016) quienes evaluaron la diversidad y la estructura presente en una población de 64 líneas de maíz de CIMMYT que representaban el germoplasma del programa tropical. Mientras que Semang et al., (2012) también informan valores similares en su estudio de caracterización molecular de 450 líneas de maíz de CIMMYT pertenecientes a los programas de mejoramiento de África. Por otro lado, la distancia genética del presente estudio es levemente superior a la informada por Dao et al., (2014) quienes mediante 1057 SNP evaluaron la diversidad genética presente en una población de 100 líneas de maíz que representaban el germoplasma del instituto de investigaciones ambientales y agrícolas (INERA), del instituto

internacional de agricultura tropical (IITA) y del centro internacional de mejoramiento de maíz y trigo (CIMMYT).

La medida del desequilibrio de ligamiento en cualquier población utilizada para asociación tiene influencia en los resultados de mapeo debido a que la existencia de grandes bloques de ligamiento y la presencia de una baja tasa de disminución del DL disminuyen la resolución del mapeo (Suwarno et al., 2015). En las líneas de maíz de este estudio el DL entre marcadores ligados fue mayor al DL presente en los marcadores no ligados, indicando que el ligamiento es predominante en la determinación del DL en comparación con otras posibles causas. Los resultados concuerdan con lo informado por Zhao et al., (2014) quienes al analizar el DL presente en una población de 329 accesiones de algodón con 212 marcadores microsatelites (SSR) también encontraron que el ligamiento físico era la principal causa del DL observado. Para el total de marcadores, el DL presente en las dos sub-poblaciones analizadas en este estudio fue mayor al DL presente la población completa de líneas de maíz, sugiriendo que el DL aumentó cuando la población se estratificó en sub-poblaciones.

La estadística multivariada es usada para describir y analizar observaciones multidimensionales obtenidas al relevar información sobre varios caracteres para cada uno de los genotipos en estudio (Balzarini et al., 2008). El análisis de componentes principales permitió reducir la dimensión de la matriz de datos genotípicos de 2530 SNP a 147 componentes principales que explicaron el 79% de la variabilidad presente. Por otro lado la matriz de datos agro-morfológicos se redujo de 13 caracteres agro-morfológicos a seis CP que explicaron el 75% de la variabilidad total. El consenso entre caracteres fenotípicos (agro-morfológicos) y genotípicos (marcadores moleculares SNP) permitió ver la asociación de aquellos respecto al genoma en estudio. Esto puede considerarse como una herramienta inicial que permite la identificación de asociaciones subyacentes a los datos y la ubicación de caracteres cuantitativos (QTL), con la ventaja adicional que pueden realizarse con más de un carácter simultáneamente (Demey, 2008). El análisis conjunto de los datos fenotípicos y genotípicos mediante el análisis de procrustes generalizado permitió observar un 60% de consenso entre el ordenamiento de los genotipos mediante marcadores moleculares y caracteres agro-morfológicos. Este valor fue levemente inferior al observado por Hernández et al., (2010) quienes informaron un 67% de consenso entre 14 caracteres fenotípicos y 40 marcadores moleculares, conformados por 20 RAPD y 20 SSR, en 14 líneas de maíz en Venezuela. Según Baranger et al., (2004) la correlación entre la diversidad molecular y agro-morfológica indica desequilibrio de

ligamiento entre los marcadores moleculares y los genes que controlan los caracteres agro-morfológicos evaluados.

Los conocimientos del presente trabajo muestran la presencia de variabilidad en la población de 252 líneas de maíz de CIMMYT para los 13 caracteres agro-morfológicos medidos. Esta variabilidad puede ser utilizada para la identificación de QTL mediante la realización de estudios de mapeo de asociación. A su vez, la identificación de la estructura genética poblacional y los análisis de relación entre las líneas pueden ser de gran valor para la implementación de estudios de mapeo de asociación con esta población. Esta información, en conjunto con datos fenotípicos de caracteres de interés, obtenidos en múltiples ambientes de evaluación, permitiría identificar QTL para realizar selección asistida por marcadores (MAS) en programas de mejoramiento de maíz. Lo cual le brindaría numerosas ventajas a los mejoradores genéticos de instituciones tanto públicas como privadas. La selección asistida por marcadores moleculares permite evaluar una población de plantas en estadios juveniles aun cuando se trate de caracteres que se expresan al final de la vida del organismo. Tiene la capacidad de permitir seleccionar caracteres de interés agronómico que son difíciles de evaluar fenotípicamente e insumen un alto costo y tiempo. Tal es el caso de la enfermedad viral Mal de Río Cuarto, así como también otras enfermedades virales del maíz donde mediante la MAS es posible obtener genotipos con resistencia genética, lo cual es eficiente y ambientalmente sustentable. Permite distinguir individuos homocigotos de los heterocigotos en una sola generación, sin necesidad de realizar pruebas de progenie.

CONCLUSIONES

La población de líneas de maíz evaluada en este estudio presenta amplia variabilidad fenotípica para los caracteres agro-morfológicos evaluados.

La población de líneas de maíz posee una amplia variabilidad genética.

La población de 252 líneas de maíz presenta estratificación en su estructura genética.

El ligamiento físico y la rápida caída del mismo, indican que la población de líneas de maíz, es adecuada para posteriores estudios de mapeo de asociación.

La correlación entre la diversidad molecular y fenotípica, observada mediante el moderado valor de consenso, indica desequilibrio de ligamiento entre los marcadores moleculares y los genes que controlan los caracteres agro-morfológicos medidos.

BIBLIOGRAFÍA

- Balzarini M, L Gonzalez, M Tablada, F Casanoves, J Di Rienzo, C Robledo. 2008. Infostat. Manual del Usuario, Editorial Brujas, Córdoba, Argentina.
- Balzarini M, C Bruno, A Peña, I Teich, J Di Rienzo. 2010. Estadística en Biotecnología. Aplicaciones en Info-Gen Ed. Grupo de Editores. Córdoba, Argentina.
- Balzarini M, J Di Rienzo. InfoGen versión 2014. FCA, Universidad Nacional de Córdoba, Argentina. URL <http://www.info-gen.com.ar>
- Baranger, A, G Aubert, G Arnau, A Lainé, G Deniot, J Potier, C Weinachter, I Lejeune-Hénaut, J Lallemand, J Burstin, 2004 Genetic diversity within *Pisum sativum* using protein and PCR based markers. *Theor. Appl. Genet.* 108: 1309-1321.
- Botstein D, R White, M Skolnick, R Davis. 1980. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am. J. Hum. Genet.* 32: 314-331.
- Bradbury P, Z Zhang, D Kroon, T Casstevens, Y Ramdoss, E Buckler. 2007. TASSEL: Software for association mapping of complex traits in diverse samples. *Bioinformatics* 23: 2633-2635.
- Bramardi S, G Bernet, M Asins, E Carbonell. 2005. Simultaneous agronomic and molecular characterization of genotypes via the generalised procrustes analysis: An application to Cucumber. *Crop Sci.* 45: 1603-1609. Disponible en: <http://doi:10.2135/cropsci2004.0633>
- Dao A, J Sanou, S Mitchell, V Gracen, E Danquah. 2014. Genetic diversity among INERA maize inbred lines with single nucleotide polymorphism (SNP) markers and their relationship with CIMMYT, IITA, and temperate lines. *BMC Genetics.* 15: 127. Disponible en: <http://doi:10.1186/s12863-014-0127-2>
- Dinesh A, A Patil, P Zaidi, P Kuchanur, M Vinayan, K Seetharam. 2016. Genetic diversity, linkage disequilibrium and population structure among CIMMYT maize inbred lines, selected for heat tolerance study. *Maydica.* 61.3 – M 29.
- Di Rienzo J, F Casanoves, M Balzarini, L Gonzalez, M Tablada, C Robledo. InfoStat versión 2016. Grupo InfoStat, FCA, Universidad Nacional de Córdoba, Argentina. URL <http://www.infostat.com.ar>

- Duvick, D. 2004. Long-term selection in a commercial hybrid maize breeding program. *Plant Breed. Rev.* 24, Part 2. 109-151 In J. Janick (eds.) John Wiley & Sons, Inc., Hoboken, NJ.
- Earl D, B vonHoldt. 2012. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genet. Resour.* 4: 359-361. Disponible en: <http://doi:10.1007/s12686-011-9548-7>
- Edwards J, S McCouch. 2007. Molecular markers for use in plant molecular breeding and germplasm evaluation. In *MARKER-ASSISTED SELECTION: Current status and future perspectives in crops, livestock, forestry and fish* (pp 29-50) FAO.
- Evanno G, S Regnaut, J Goudet. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol. Ecol.* 14: 2611-2620
- Excoffier L, P Smouse, J Quattro. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* 131: 479-491
- Flint-García S, J Thornsberry, E Buckler. 2003. Structure of linkage disequilibrium in plants. *Ann. Rev. Plant Biol.* 54: 357-374.
- Flint-Garcia S, A Thuillet, J Yu, G Pressoir, S Romero, S Mitchell, J Doebley, S Kresovich, M Goodman, E Buckler. 2005. Maize association population: a high-resolution platform for quantitative trait *locus* dissection. *The Plant J.* 44: 1054-1064.
- Gower J. 1975. Generalized procrustes analysis. *Psychometrika* 40: 33-51.
- Greenacre, M. (1984). *Theory and applications of correspondence analysis*. London: Academic Press.
- Hallauer A, J Miranda. 1988. *Quantitative genetics in maize breeding*. 2nd edition, Iowa State University Press, Ames, IA.
- Hansey C, J Johnson, R Sekhon, S Kaeppler, N de Leon. 2011. Genetic diversity of a maize association population with restricted phenology. *Crop Sci.* 51: 704-715.
- Hao D, Cheng Y, Chen G, H Lu, M Shi, Z Zhang, X Huang, Y Mao, L Xue. 2015. Identification of significant single nucleotide polymorphisms for resistance to maize rough dwarf disease in elite maize (*Zea mays* L.) inbred lines. *Euphytica.* 203: 109-20.

- Hernández A, F San Vicente, R Figueroa-Ruiz. 2010. Evaluación y caracterización de líneas parentales de híbridos de maíz (*Zea mays* L.) en tres ambientes de Venezuela. *Interciencia* 35: 290-298.
- Levitus G, V Echenique, C Rubinstein, E Hopp, L Mroginski. 2010. Biotecnología y Mejoramiento Genético Vegetal II. 2^{da} Ed. INTA, Argentina. 650 p.
- Liu K, M Goodman, S Muse, J Smith, E Buckler, J Doebley. 2003. Genetic structure and diversity among maize inbred lines as inferred from DNA microsatellites. *Genetics* 165: 2117-2128.
- Liu K, S Muse. 2005. Power Marker: an integrated analysis environment for genetic marker analysis. *Bioinformatics* 21: 2128-2129.
- Liu Y, J Guo, D Zhang, Y Zhao, L Zhu, Y Huang, J Chen. 2014. Genetic diversity and linkage disequilibrium estimation among the maize breeding germplasm for association mapping. *Int. J. Agric. Biol.*, 16: 851-861.
- Mienie, C, A Fourie. 2013. Genetic diversity in South African maize (*Zea mays* L.) genotypes as determined with microsatellite markers. *Afr. J. Biotechnol.* 12: 123-133.
- Peña Malavera A, L Gutierrez, M Balzarini. 2014. Principal components in associative mapping componentes principales en mapeo asociativo. *J. of Basic & Appl. Genet.* 25
- Prasanna B. 2012. Diversity in global maize germplasm: Characterization and utilization. *J. Biosci.* 37: 843-855.
- Pritchard J, N Rosenberg. 1999. Use of unlinked genetic markers to detect population stratification in association studies. *Am. J. Hum. Genet.* 65: 220-228.
- Pritchard J, M Stephens, P Donnelly. 2000. Inference of population structure using multilocus genotype data. *Genetics* 155: 945-959.
- Remington D, J Thornsberry, Y Matsuoka, L Wilson, S Whitt, J Doebley, S Kresovich, M Goodman. 2001. Structure of linkage disequilibrium and phenotypic associations in the maize genome. *lv ESB* 98: 11479-11484
- Romay M, M Millard, J Glaubitz, J Peiffer, K Swarts, T Casstevens, R Elshire, C Acharya, S Mitchell, S Flint-Garcia, M McMullen, J Holland, E Buckler, C Gardner. 2013. Comprehensive genotyping of the USA national maize inbred seed bank. *Genome Biol.* 14: R55. Disponible en: <http://doi:10.1186/gb-2013-14-6-r55>

- Semang K, C Magorokosho, B Vivek, D Makumbi, Y Beyene, S Mugo, B Prasanna, M Warburton. 2012. Molecular characterization of diverse CIMMYT maize inbred lines from eastern and southern Africa using single nucleotide polymorphic markers. *BMC Genomics*. 13: 113.
- Suwarno W, K Pixley, N Palacios-Rojas, S Kaeppler, R Babu. 2015. Genome-wide association analysis reveals new targets for carotenoid biofortification in maize. *Theor. Appl. Genet.* 128: 851-864
- Wallace J, P Bradbury, N Zhang, Y Gibon, M Stitt, E Buckler. 2014. Association mapping across numerous traits reveals patterns of functional variation in maize. Disponible en: <http://dx.doi.org/10.1101/010207>
- Warburton M, X Xia, J Crossa, J Franco, A Melchinger, M Frisch, M Bohn, D Hoisington. 2002. Genetic characterization of CIMMYT maize inbred lines and open pollinated populations using large scale fingerprinting methods. *Crop Sci.* 42: 1832-1840.
- Warburton M, J Reif, M Frisch, M Bohn, C Bedoya, X Xia, J Crossa, J Franco, D Hoisington, K Pixley, S Taba, A Melchinger. 2008. Genetic diversity in CIMMYT nontemperate maize germplasm: landraces, open pollinated varieties, and Inbred lines. *Crop. Sci.* 48: 617-624
- Warburton M, J Tang, G Windham, L Hawkins, S Murray, W Xu, D Boykin, A Perkins, W Williams. 2015. Genome-Wide Association Mapping of *Aspergillus flavus* and Aflatoxin Accumulation Resistance in Maize. *Crop. Sci.* 55: 1-11.
- Würschum T. 2012. Mapping QTL for agronomic traits in breeding populations. *Theor. Appl. Genet.* 125: 201-210.
- Yan J, T Shah, M Warburton, E Buckler, M McMullen. 2009. Genetic characterization and linkage disequilibrium estimation of a global maize collection using SNP markers. *PLoS ONE* 4(12): e8451. Disponible en: <http://doi:10.1371/journal.pone.0008451>
- Yu J, E Buckler. 2006. Genetic association mapping and genome organization of maize. *Curr. Opin. Biotechnol.* 17: 155-160
- Zhao Y, H Wang, W Chen, Y Li. 2014. Genetic structure, linkage disequilibrium and association mapping of *Verticillium* wilt resistance in elite cotton (*Gossypium hirsutum* L.) germplasm population. *PLOS ONE*. Disponible en: <http://doi:10.1371/journal.pone.0086308>