



PROGRAMA DE DOCTORADO EN CIENCIAS MENCIÓN
INGENIERÍA GENÉTICA VEGETAL UNIVERSIDAD DE TALCA,
INSTITUTO DE CIENCIAS BIOLÓGICAS

**MAPEO ASOCIATIVO DE GENOMA COMPLETO (GWAS) PARA EL
ÁNGULO FOLIAR, ALTURA DE LA PLANTA Y ALTURA DE LA
PRIMERA MAZORCA EN LÍNEAS ENDOGÁMICAS DE MAÍZ (*Zea
mays* L.)**

CARLOS ERNESTO MALDONADO MUÑOZ

Tesis para optar al grado académico de
Doctor en Ciencias mención en
Ingeniería Genética Vegetal

**DIRECTOR DE TESIS
DR. FREDDY MORA POBLETE**

OCTUBRE 2019

CONSTANCIA

La Dirección del Sistema de Bibliotecas a través de su unidad de procesos técnicos certifica que el autor del siguiente trabajo de titulación ha firmado su autorización para la reproducción en forma total o parcial e ilimitada del mismo.



Talca, 2022

**MAPEO ASOCIATIVO DE GENOMA COMPLETO (GWAS)
PARA EL ÁNGULO FOLIAR, ALTURA DE LA PLANTA Y
ALTURA DE LA PRIMERA MAZORCA EN LÍNEAS
ENDOGÁMICAS DE MAÍZ (*Zea mays* L.)**

**GENOME-WIDE ASSOCIATION STUDY (GWAS) FOR THE
LEAF ANGLE, PLANT HEIGHT AND EAR HEIGHT IN
INBRED LINES OF MAIZ (*Zea mays* L.)**

CARLOS ERNESTO MALDONADO MUÑOZ

Fecha de inicio de la Tesis: 1 de abril de 2017

Fecha de término: 15 de octubre de 2019

PROFESOR GUÍA

DR. FREDDY MORA POBLETE

Universidad de Talca

Instituto de Ciencias Biológicas

fmora@utalca.cl

Dr. Fernando Guerra

Ingeniero Forestal, Dr.

Dr. Andrés Schwember

Ingeniero Agrónomo, Dr.

Dr. Nilo Mejía

Ingeniería en Biotecnología Molecular, Dr.

AGRADECIMIENTOS

Quiero dar gracias a Dios por acompañarme siempre y poner en mi camino a las personas con las que compartir este proceso.

Deseo agradecer a mi familia y amigos por todo su apoyo durante estos años. A mi esposa, que me acompañó y entregó todo su apoyo constantemente en esta etapa de mi vida. Agradezco a mis amigos y compañeros, de manera especial a Paulina Ballesta y Osvin Arriagada, quienes fueron compañeros de estudios y siempre tenían palabras de aliento cuando más lo necesité. Agradezco en particular y de manera especial al Dr. Freddy Mora Poblete, por todas las enseñanzas que me entregó, las palabras de aliento, críticas constructivas, enseñanzas de vida y valiosos consejos que me acompañaron durante este proceso y seguirán acompañándome durante mi desarrollo profesional.

Agradezco a CONICYT, por otorgarme la beca Doctorado Nacional año 2017, además del beneficio complementario “Gastos Operacionales”, los cuales me permitieron solventarme durante todo este proceso, además de entregarme recursos para finalizar mi trabajo de Tesis.

ÍNDICE

RESUMEN	iv
ABSTRACT	vi
1. INTRODUCCIÓN	1
1.1 Antecedentes generales:	1
1.2 Selección asistida por marcadores (MAS; <i>Marker-Assisted Selection</i>):	2
1.3 Loci de caracteres cuantitativos (QTL):	3
1.4 Genotipado por secuenciación (GBS; <i>Genotyping By Sequencing</i>):	5
1.5 Importancia del desequilibrio de ligación (LD; <i>Linkage Disequilibrium</i>) en estudios genéticos:	6
1.6 Haplotipos	7
1.7 Estudios de asociación del genoma completo (GWAS; <i>Genome-Wide Association Studies</i>):	8
1.8 Efecto pleiotrópico	9
1.9 Genes candidatos para AF, AP y AM en maíz	11
1.10 Planteamiento del problema y novedad de la propuesta	12
2. HIPOTESIS DE TRABAJO	14
3. OBJETIVO GENERAL	14
3.1 Objetivos Específicos	15
4. CAPÍTULO I: ESTRUCTURA POBLACIONAL, DESEQUILIBRIO DE LIGAMIENTO Y MAPEO ASOCIATIVO EN MAÍZ TROPICAL USANDO UN PANEL DE MARCADORES SNP (GBS) DE ALTA DENSIDAD	16
4.1 Introducción	16
4.2 Materiales y métodos	17
4.2.1 Descripción del material biológico y características fenotípicas evaluadas	17
4.2.2 ADN genómico y descubrimiento de SNP por GBS	17

4.2.3	<i>Estructura poblacional y estimación de matriz kinship</i>	18
4.2.4	<i>Análisis estadístico de los datos fenotípicos</i>	19
4.2.5	<i>Estudio de asociación del genoma completo (GWAS)</i>	20
4.3	Resultados	21
4.3.1	<i>Estructura poblacional</i>	22
4.3.2	<i>Estudio de asociación de genoma amplio con marcadores SNP- individuales</i>	23
4.4	Discusión	29
5.	CAPÍTULO II: MAPEO ASOCIATIVO DE GENOMA AMPLIO BASADO EN HAPLOTIPOS	30
5.1	Introducción	30
5.2	Materiales y métodos	31
5.2.1	<i>Desequilibrio de ligamiento (LD)</i>	31
5.2.2	<i>Bloques de haplotipos</i>	32
5.2.3	<i>Genoma de referencia</i>	32
5.3	Resultados	32
5.3.1	<i>Desequilibrio de ligamiento (LD)</i>	33
5.3.2	<i>Haplotipos asociados con características complejas</i>	33
5.3.3	<i>Anotación de genes basada en SNP y haplotipos individuales</i>	39
5.4	Discusión	40
5.4.1	<i>Haplotipos y regiones genéticas asociadas con características complejas</i>	40
5.4.2	<i>Datos anotados del genoma</i>	44
6.	CAPÍTULO III: EFECTO PLEIOTRÓPICO DE LOCI POLIMÓRFICOS	46
6.1	Introducción	46
6.2	Materiales y métodos	46
6.2.1	<i>Correlaciones genética y pleiotropía</i>	47

6.3 Resultados.....	48
6.4 Discusión	51
7. CONCLUSIONES GENERALES.....	53
8. BIBLIOGRAFÍA	55
9. MATERIAL SUPLEMENTARIO	69
10. ANEXOS.....	70
10.1 ANEXO I. Pseudocódigo para ejecutar paralelamente STRUCTURE	70
10.2 ANEXO II. Scripts para generar marcadores multialélicos a partir de resultados de HAPLOVIEW	71

RESUMEN

Las características relacionadas con la arquitectura de las plantas, incluyendo su capacidad de mantenerse erguidas frente a condiciones de cultivo, son determinantes de la productividad de las plantas en los sistemas de cultivo intensivo de maíz. En este sentido, los estudios de asociación de genoma amplio (GWAS) basados en polimorfismos de nucleótido único (SNP) han proporcionado información valiosa sobre la relación entre la variación de regiones genéticas y fenotipos complejos como los de la arquitectura. Sin embargo, no existen antecedentes, a la fecha, que relacionen caracteres complejos con haplotipos de genoma amplio en maíz. En el presente estudio se caracterizaron aproximadamente 7,800 bloques haplotipos de un conjunto de SNP de alta calidad (~38,000), en un panel de asociación compuesto por líneas endogámicas de maíz tropical. Con ello, se identificaron asociaciones genómicas con el ángulo de la hoja (AF), altura de la planta (AP), altura de la primera mazorca (AM) y la relación AM/AP. La proporción de la variación fenotípica explicada por los SNP individuales varió entre 7%, para el SNP S1_285330124 (ubicado en el cromosoma 9 y asociado con la relación AM/AP), y 22%, para el SNP S1_317085830 (ubicado en el cromosoma 6 y asociado con el ángulo de la hoja). Un total de 40 bloques de haplotipos se asociaron significativamente con las características de interés, explicando hasta el 29% de la variación fenotípica para el ángulo de la hoja, correspondiente al haplotipo hapAF4.04, que se mantuvo estable durante dos temporadas de crecimiento. En general, las asociaciones para AP, AM y la relación AM/AP fueron específicas de la temporada, lo que se confirmó mediante un análisis de comparación de modelos utilizando los criterios de información de Akaike y Schwarz. Además, se identificaron cinco haplotipos estables (83%) y 15 SNP (75%) para el ángulo de la hoja. Otro importante resultado fue que el 29% (4/14) de los loci asociados concomitantemente con AP y AM presentaron una moderada o fuerte evidencia de asociación ($\log_{10}(\text{Factor Bayes}) > 3$ y Probabilidad posteriori de Asociación > 0.5) a estas características en conjunto, proporcionando una clara evidencia de que estos loci tienen un efecto pleiotrópico significativo. Aproximadamente el 62% de las asociaciones con haplotipos (25/40) no contenían SNP detectados en el estudio de asociación utilizando SNP individuales. Este resultado confirma la ventaja de los estudios de asociación de todo el genoma basados en haplotipos para examinar determinantes genéticos que controlan la

arquitectura y el encamado de las plantas de maíz. Adicionalmente, el mapeo asociativo basado en haplotipos proporcionó una reducción de la dimensionalidad basada en la agrupación de SNP a partir del patrón de Desequilibrio de Ligamiento (LD) observado en la población, además aumenta la posibilidad de encontrar segmentos genómicos que controlen la variación de una característica.

ABSTRACT

The traits related to the architecture of the plants, including their ability to remain upright in the face of growing conditions, are determinants of the productivity of the plants in the intensive maize cultivation systems. In this sense, Genome-Wide Association Studies (GWAS) based on Single Nucleotide Polymorphisms (SNPs) have provided valuable information on the relationship between the variation of genetic regions and complex phenotypes such as those of architecture. However, to date, there is no background, that relates complex traits to haplotypes of genome-wide in maize. In the present study, approximately 7,800 haplotype blocks of a set of high-quality SNPs (~38,000) were characterized in an association panel composed of inbred lines of tropical maize. With this, genomic associations were identified with the leaf angle (AF), plant height (AP), ear height (AM) and the AM/AP ratio. The proportion of phenotypic variation explained by individual SNPs varied between 7%, for SNP S1_285330124 (located on chromosome 9 and associated with the AM/AP ratio), and 22%, for SNP S1_317085830 (located on chromosome 6 and associated with the leaf angle). A total of 40 haplotype blocks were significantly associated with the traits of interest, explaining up to 29% of the phenotypic variation for the leaf angle, corresponding to the *hapAF4.04* haplotype, which remained stable for two growing seasons. In general, the associations for AP, AM and the AM/AP ratio were season-specific, which was confirmed by a model comparison analysis using the information criteria of Akaike and Schwarz. Also, five stable haplotypes (83%) and 15 SNPs (75%) were identified for the leaf angle. Another important result was that 29% (4/14) of the loci associated concomitantly with AP and AM presented moderate or strong evidence of association (\log_{10} (Bayes Factor) > 3 and Posterior Probability of Association > 0.5) to these traits in together, providing clear evidence that these loci have a significant pleiotropic effect. Approximately 62% of the associations with haplotypes (25/40) did not contain SNPs detected in the association study using individual SNPs. This result confirms the advantage of haplotype-based genome-wide association studies to examine genetic determinants that control the architecture and lodging of maize plants. Additionally, haplotype-based association mapping provided a reduction of the dimensionality based on the SNP clustering through on the linkage

disequilibrium (LD) pattern observed in the population, also increases the possibility of finding genomic segments that control the variation of traits.

1. INTRODUCCIÓN

1.1 Antecedentes generales:

El maíz (*Zea mays ssp. mays* L), junto con el arroz y el trigo, es uno de los cultivos más plantados a nivel mundial, con una superficie de 226 millones de hectáreas destinadas para la agricultura y la alimentación según la FAO (*Food and Agriculture Organization*). Debido a que el maíz contiene aproximadamente 72% de almidón, 10% de proteínas y 4% de grasas (con una densidad energética de 365 Kcal/100g), es una excelente fuente de alimento (de la Barrera y Martínez 2018). El maíz también es utilizado en la producción de forraje, productos farmacéuticos e industriales (Li et al. 2016; Yang et al. 2011). Los indicadores de la FAO señalan que la producción de maíz para el 2016 fue de aproximadamente 1.300 millones de toneladas, siendo Estados Unidos, China y Brasil los principales productores, cosechando 30%, 18% y 5% de la producción mundial, respectivamente. Adicionalmente, el maíz es un organismo modelo para estudios de genética y biología (Strable y Scanlon 2009). Esta especie tiene muchas ventajas agronómicas, entre las cuales destaca su alta diversidad fenotípica y alta tasa de propagación. Posee inflorescencias separadas (macho y hembra), lo cual permite realizar cruzamientos controlados simples y generar líneas endogámicas (Strable y Scanlon 2009). Diversos estudios de mejoramiento de maíz señalan que esta especie posee la variabilidad genética necesaria para mejorar cualquier característica que se desee (Bedoya et al. 2017; Strable y Scanlon 2009).

La domesticación del maíz se realizó hace aproximadamente 10 mil años a partir de teosinte (*Zea mays ssp. parviglumis*) una especie nativa de América central (Piperno y Flannery 2001), y desde entonces el maíz ha sido sujeto de procesos de selección (Schnable et al. 2009). Uno de los puntos más importantes en los programas de mejoramiento, es la selección de plantas que presenten mayor número de granos por mazorca, lo que se asocia a un alto rendimiento en el cultivo. Sin embargo, diversos estudios se han enfocado en el análisis de otras características, como por ejemplo el ángulo foliar (AF), la altura de la planta (AP), la altura de la primera mazorca (AM) o la relación entre AM y AP (relación AM/AP), debido a que determinan indirectamente el rendimiento de grano y la calidad del cultivo (Buescher et al. 2014; Weng et

al. 2011; Kong et al. 2017; Ding et al. 2015; Peiffer et al. 2014; Abdel_Ghani et al. 2016; Cai et al. 2012; Li et al. 2016).

Diversos estudios señalan que la existencia de cambios en la arquitectura de las hojas, como por ejemplo el aumento de tamaño y la verticalidad (hojas más rectas), le permiten a la planta una captura de luz más eficiente (Ding et al. 2015), lo cual incrementa la eficiencia fotosintética (Sinclair y Sheehy 1999). Morinaka et al. (2006) señalan que una mayor verticalidad de las hojas de arroz, no sólo mejora la captura de luz, sino que también ayuda al aumento de la densidad de cultivo y la acumulación de nitrógeno foliar en el proceso de llenado de granos. Por lo tanto, el ángulo foliar es un factor que influye tanto en la eficiencia fotosintética, como en la acumulación de nitrógeno foliar, los cuales son factores que regulan el rendimiento global de un cultivo (Sinclair y Sheehy 1999). Por otro lado, varios autores sugieren que AP y AM son características agronómicas que están significativamente correlacionados entre sí, y a su vez, con la producción de granos y el número de hojas por planta (Abdel-Ghani et al. 2016; Sibov et al. 2003; Li et al. 2016; Cai et al. 2012; Peiffer et al. 2014). Más aún, la AP, AM y la relación entre estas (AM/AP) tienen un efecto importante en el encamado de plantas en sistemas de cultivo intensivo de maíz (Cai et al. 2012).

Usualmente, los programas de mejoramiento de maíz se han centrado en ganancias continuas del rendimiento de granos. Sin embargo, la selección indirecta basada en características relacionadas al rendimiento, tales como AF, AP, AM, y relación AM/AP, han sido parte importante de los programas de mejoramiento. Por lo tanto, una mejor comprensión de la arquitectura genética de estas características permitiría un proceso de selección eficiente, y con ello, el desarrollo de cultivares altamente productivos (líneas endogámicas o híbridos). En este contexto, el uso de diversas herramientas moleculares, permitirá conocer y comprender los mecanismos moleculares que subyacen las características AF, AP, AM, y relación AM/AP, facilitando la selección de cultivares de alto rendimiento (Wang y Li, 2008, Ku et al. 2011).

1.2 Selección asistida por marcadores (MAS; *Marker-Assisted Selection*):

El fitomejoramiento tiene sus inicios con la domesticación de plantas y el comienzo de la agricultura. Tradicionalmente, la selección es realizada mediante la evaluación visual del fenotipo de las plantas madres y su progenie y, posteriormente, los individuos que sean

superiores en características deseables (por ejemplo: altura de las plantas, rendimiento de grano, resistencia a estrés o enfermedades), son utilizados como nuevos progenitores en sucesivos ciclos de mejoramiento.

El avance de las nuevas tecnologías de genotipificación ha permitido aumentar la eficiencia de selección en los programas de mejoramiento tradicionales, gracias al uso de la información genética a nivel molecular (Collard et al. 2005). En este contexto, el uso de marcadores moleculares (o de ADN) permite detectar variaciones alélicas que subyacen una característica de interés. Por lo que, el uso de estos marcadores en programas de mejoramiento (denominado como selección asistida por marcadores moleculares; MAS) permite reducir los ciclos de mejoramiento y seleccionar plantas con determinadas cualidades en cualquier etapa de su desarrollo (Collard et al. 2005).

Los marcadores moleculares han permitido detectar genes o loci de características cuantitativas (QTL; *Quantitative Trait Loci*), asociados con alguna característica de importancia agronómica. Adicionalmente, estos QTL pueden ser posicionados en el genoma de una especie mediante las distancias de recombinación entre marcas moleculares. El proceso de construcción de mapas de ligación e identificación de QTL, se conoce como mapeo de QTL (Collard et al. 2005), el cual es una herramienta importante dentro programas de mejoramiento genético, debido a que permite determinar la arquitectura genética de un carácter complejo (Mora et al. 2008).

1.3 Loci de caracteres cuantitativos (QTL):

Conceptualmente, un QTL se define como la variación alélica de un locus, asociada a la variación observada de un carácter cuantitativo (Mackay et al. 2009; Miles y Wayne 2008; Mora et al. 2008). Muchas de las características de interés agronómico como AF, AP, AM y relación AM/AP en maíz, son características cuantitativas complejas, debido a que están controladas por un gran número de genes, los cuales tienen un efecto aditivo pequeño sobre el fenotipo (Weng et al. 2011). Sin embargo, cabe destacar que en características complejas, una proporción sustancial de la variación fenotípica puede ser explicada por unos pocos loci de gran efecto, y el resto de la variación es atribuible a numerosos loci de pequeño efecto (Miles y Wayne 2008).

Bajo esta idea, el análisis de QTL busca detectar los loci que presenten un mayor efecto en la expresión de una característica cuantitativa (Mora et al. 2008)

Una gran cantidad de QTL asociados a AF, AP, AM y relación AM/AP han sido reportados en maíz mediante el uso de diferentes marcadores moleculares, tales como microsatélites (SSR; *Simple Sequence Repeat*) y polimorfismos de nucleótido único (SNP; *Single Nucleotide Polymorphism*) (Zheng et al. 2013; Li et al 2016; Ku et al. 2011; Tian et al. 2011). Por ejemplo, Ku et al. (2011) identificaron cinco QTL asociados significativamente a AF en 256 familias F_{2:3} de maíz, mediante marcadores SSR. Por otra parte, Tian et al. (2011) mediante un estudio de asociación de genoma amplio (GWAS) (con 1.6 millones de SNP) y un mapeo de asociación anidado (NAM; *Nested Association Mapping*) con 282 líneas endogámicas, detectaron 30 QTL asociados significativamente a AF, los cuales explicaron gran parte de la variación fenotípica total de esta característica (75%).

A diferencia de AF, una mayor cantidad de QTL asociados a AP ha sido reportada en maíz. Por ejemplo, Weng et al. (2011) trabajando con 284 líneas endogámicas, reportaron más de 219 QTL significativamente asociados a AP. Por otro lado, Zhang et al. (2006), Melchinger et al. (1998) y Li et al. (2016) detectaron 9, 31 y 12 QTL en poblaciones F₂, F₃ y F_{2:3}, respectivamente.

AM es una característica significativamente correlacionada con AP (Cai et al. 2012; Li et al. 2016), por lo tanto, es posible observar QTL asociados a una relación entre estas características (AM/AP) (Cai et al. 2012), o QTL que para ambas características se mapean en las mismas regiones genómicas (evidenciando además efectos pleiotrópicos) (Ji-hua et al. 2007; Li et al. 2016). Un ejemplo de esto, es el trabajo de Cai et al. (2012) quienes, usando una población de 218 RIL de maíz, detectaron 22 QTL asociados AP, 15 QTL a AM y 13 QTL a la relación entre ambas AP/AM. En este mismo contexto, Li et al. (2016) empleando una población F_{2:3} con 225 familias de maíz, encontraron QTL para AP y AM en la misma región genómica, como *qPH1-1* (asociado a AP) y *qEH1-1* (asociado a AM) en el bin 1.05 y 1.06 (respectivamente), sugiriendo que los mecanismos genéticos que controlan estas características pueden ser similares.

La identificación de regiones genómicas asociadas con características cuantitativas, controladas por un gran número de genes (Li et al. 2016, Peiffer et al. 2014, Abdel-Ghani et al. 2016, Cai et al. 2012, Yang et al. 2014, Tian et al. 2011, Ku et al. 2016) y altamente susceptibles a la influencia ambiental (Miles y Wayne 2008), ha adquirido una importancia relativa en el desarrollo de híbridos de maíz. Además, en muchas poblaciones de maíz, los eventos aleatorios pueden causar interacciones "genotipo x ambiente" significativas (por ejemplo, la temporada de crecimiento), que pueden ser difíciles de interpretar y cuantificar (Edmeades 2013). Por este motivo, la estabilidad genética de los QTL debe verificarse a lo largo del tiempo (durante diferentes años o temporadas de crecimiento) y/o en diferentes condiciones ambientales (Miles y Wayne 2008, Asaro et al. 2016, García et al. 2009).

1.4 Genotipado por secuenciación (GBS; *Genotyping By Sequencing*):

Genotipado por secuenciación (GBS) es una de metodologías de alto rendimiento para el descubrimiento de millones SNP a un costo razonable. GBS es una técnica robusta de representación parcial del genoma mediante plataformas de secuenciación de próxima generación (NGS; *Next-Generation Sequencing*) (Elshire et al. 2011). Los avances en las tecnologías NGS, han permitido reducir los costos de secuenciación, haciendo factible el genotipado de genomas de alta diversidad y gran tamaño, por ejemplo, maíz y trigo (Poland et al. 2012; Zheng et al. 2013).

GBS involucra la digestión de ADN mediante enzimas de restricción, generando fragmentos de 75 a 400 pb (pares de bases), lo cual reduce rápidamente la complejidad del genoma a secuenciar (Elshire et al. 2011). Cada individuo a ser genotipado contiene una pequeña secuencia indexadora (de 4 a 9 pb), la cual sirve como identificador de todas las secuencias que pertenecen al genotipo. Esta función reduce el tiempo y los costos de esta metodología, gracias al uso de PCR y pasos de purificación (Davey et al. 2011). Uno de los factores claves de GBS es el uso de diversos algoritmos bioinformáticos, ya que éste es probablemente uno de los puntos más críticos y limitantes para explorar plenamente el conjunto de datos (Huang et al. 2014).

Una de las principales aplicaciones del GBS es en el campo del mejoramiento genético, el cual se presenta como una herramienta para el descubrimiento de marcadores moleculares asociados a genes o loci, estudios de diversidad, y selección genómica (He et al. 2014). GBS se

ha optimizado para diversos cultivos, tales como maíz, trigo y cebada (Romay et al. 2013, Gore et al. 2009, Poland y Rife 2012), permitiendo generar una extensa plataforma de SNP. Por ejemplo, el método GBS ha permitido identificar 0.6 millones de SNP en poblaciones endogámicas (Romay et al. 2013) y 1.4 millones en una colección de RIL de maíz (Gore et al. 2009). Los SNP descubiertos por GBS en maíz se han utilizado principalmente para realizar estudios de asociación de genoma amplio con características complejas (Gore et al. 2009; Romay et al. 2013) y en modelos de selección genómica (Crossa et al. 2013).

GBS es una novedosa metodología para el descubrimiento y genotipado de SNP, la cual presenta algunas ventajas frente a otras plataformas que también utilizan NGS, tales como la identificación de una alta densidad de SNP (simultáneo al genotipado) y localización cromosómica de los SNP (cuando un genoma de referencia está disponible) (Elshire et al. 2011; He et al. 2014; Mammadov et al. 2012).

1.5 Importancia del desequilibrio de ligación (LD; *Linkage Disequilibrium*) en estudios genéticos:

El desequilibrio de ligación es definido como la asociación no aleatoria de dos o más alelos de diferentes loci. En otras palabras, se considera que dos loci están en LD cuando la combinación de sus alelos no es aleatoria. Si consideramos marcadores SNP, el LD describiría el grado en el que el alelo de un SNP es asociado o heredado conjuntamente con el alelo de otro SNP dentro de una misma población (Gupta et al. 2005). En este mismo sentido, un haplotipo consiste de un conjunto de alelos de un mismo cromosoma, los cuales se encuentran en un significativo LD (Slatkin 2008).

El LD es un parámetro poblacional, el cual está influenciado por varios factores genéticos y evolutivos, tales como la estructura poblacional, el tamaño de la población, los niveles de endogamia en la población, la presencia de mutaciones, la deriva génica, la selección simultánea de loci y los eventos de recombinación (Slatkin, 2008). Particularmente, la tasa de recombinación altera fuertemente el LD, ya que una mayor frecuencia de eventos de recombinación a través de las generaciones, genera una menor tasa de eventos de LD en una población (Yan et al. 2011).

La importancia de conocer y determinar el patrón de LD en poblaciones de mejoramiento, radica en el hecho de que la detección efectiva de las asociaciones de QTL se encuentra determinado por la extensión del LD con el polimorfismo causal y por el tamaño de la muestra considerada (Zhu et al. 2008). El patrón del LD sobre la distancia física en una población, determina la cantidad de marcadores requeridos y el nivel de resolución que puede ser obtenido en el análisis de asociación (Zhu et al. 2008). Si el LD disminuye rápidamente dentro de una región genómica, un gran número de marcadores puede ser requerido para abarcar la región de interés (mayor densidad de mapeo).

Varios trabajos han examinado los patrones de LD en diferentes poblaciones de maíz (Jung et al. 2004; Lu et al. 2011; Remington et al. 2001; Yan et al. 2009). Por ejemplo, Jung et al. (2004) evaluaron el LD de un conjunto de 192 líneas endogámicas mediante loci próximos al gen *adh1*, e informaron que en este conjunto de individuos, el LD se extiende por sobre los 500 kpb. Interesantemente, Lu et al. (2011) determinaron que el LD en un germoplasma de maíz tropical disminuye a los 5-10 kb, el cual es dos veces más rápido en comparación con el maíz templado (a los 10-100 kb), indicando que el maíz tropical ha experimentado una tasa de recombinación más alta que el maíz templado.

1.6 Haplotipos

Un haplotipo se define como una combinación de loci polimórficos (variantes alélicas) cercanos (determinado por el LD) a lo largo de un cromosoma, que tienden a heredarse juntos (o en bloque) (International HapMap Consortium 2005). Esto implica que los loci de un haplotipo no son separados por recombinación, por lo que tienden a heredarse en conjunto, formando bloques de haplotipos (Contreras-Soto et al. 2017). La particularidad de un bloque de haplotipo, es que puede establecer diversas combinaciones con las variantes alélicas de los loci, las que pueden estar asociadas con la expresión de un fenotipo (Contreras-Soto et al. 2017; International HapMap Consortium 2005).

El tamaño de los bloques de haplotipos está estrechamente relacionado con el grado de LD presente en la población determinada (Slatkin 2008). Por ejemplo, si el LD disminuye rápidamente, una menor cantidad de loci estarán vinculados, reduciendo el tamaño de los bloques de haplotipos (Slatkin 2008; Zhu et al. 2008). Por otro lado, si el LD disminuye

lentamente se podrán formar bloques de haplotipo de mayor tamaño, permitiendo localizar una mayor cantidad de loci subyacentes al fenotipo (Zhu et al. 2008).

La caracterización de haplotipos en una población de mejoramiento, y su posterior utilización en estudios de asociación marcador-fenotipo, representa una opción a los SNP-individuales (Xiao et al. 2017). La información genética proveniente de bloques de SNPs en LD (bloques de haplotipos) en GWAS podría compensar la limitación bi-alelica de los marcadores SNP, mejorando la eficiencia en la detección de QTL (Chen et al. 2016; Contreras-Soto et al. 2017; Xiao et al. 2017) y proporcionado conocimiento sobre los determinantes genéticos que no pueden ser capturados por el enfoque de marcador independiente (Abdel-Shafy et al. 2014). Ejemplos de lo anterior son los estudios de Chen et al. (2016) y Contreras-Soto et al. (2017), los cuales caracterizaron haplotipos en poblaciones de maíz y soya (respectivamente), y descubrieron loci (bloques de haplotipos) asociados a la resistencia a la pudrición de la mazorca por *Fusarium* y características agronómicas (altura de la planta, rendimiento de semilla y el peso de 100 semillas), los cuales no fueron detectados con GWAS basado en SNP-individuales.

1.7 Estudios de asociación del genoma completo (GWAS; *Genome-Wide Association Studies*):

El mapeo asociativo (o conocido también como mapeo por LD) consiste en la identificación de loci que controlan características cuantitativas u oligogénicas, a partir del mapeo de miles de variantes alélicas en todo el genoma (Yu y Buckler 2006). Uno de los objetivos de los estudios de asociación es explorar los eventos de recombinación histórica a nivel poblacional, buscando las mutaciones asociadas a un carácter, las cuales se han heredado a lo largo del tiempo (Visscher et al. 2012).

La eficiencia de GWAS para identificar una asociación significativa entre un loci y una característica, depende de la varianza fenotípica explicada por el loci dentro de la población (Korte y Farlow 2013). Sin embargo, cuando se tiene un pequeño tamaño poblacional, ciertos patrones de LD en el genoma, y la estructura poblacional, puede reducir la robustez de GWAS en la detección de asociaciones marcador-característica, aumentando las posibilidades de aceptar falsas asociaciones (o falsos positivos) (Visscher et al. 2012).

La presencia de estructura poblacional que no es incluida en los estudios de asociación, ocasiona errores del tipo 1 en la detección de QTL, dado que genera una sobrestimación de loci asociados significativamente a una característica (Korte y Farlow 2013). Una de las formas para evitar el grado de relación genética entre individuos es utilizar una matriz de estructura poblacional y una matriz de parentesco o kinship, como covariables de efecto fijo y de efecto aleatorio, respectivamente, dentro del modelo de asociación (Pritchard et al. 2000, Yu et al. 2006). Esto permite remover las correlaciones genéticas debido al efecto del grupo en el modelo de asociación, y reducir la tasa de falsos positivos (Malosetti et al. 2007).

La estimación de la estructura genética de una población se realiza principalmente mediante métodos de inferencia bayesiana, vía métodos de Monte Carlo con cadenas de Markov (MCMC; *Markov Chain Monte Carlo*). Pritchard et al. (2000), por ejemplo, propuso un procedimiento de agrupamiento Bayesiano, que depende de la existencia de Equilibrio de Hardy-Weinberg (EHW), para inferir de manera iterativa el número más probable de poblaciones (o subpoblaciones) genéticamente diferenciadas, a partir de marcadores no ligados de un grupo de individuos (Pritchard et al. 2000; Evanno et al. 2005). Sin embargo, los resultados de STRUCTURE pueden ser poco apropiados en algunos casos, dado que este método trabaja bajo el supuesto de que las poblaciones se encuentran en EHW y equilibrio de ligación (Pritchard et al. 2000; Evanno et al. 2005). Una alternativa a este procedimiento es el análisis de componentes principales (PCA; *Principal Component Analysis*), el cual es ampliamente utilizado para reducir la dimensionalidad de los datos, y disminuir los tiempos computacionales en comparación a los métodos que emplean modelos MCMC.

Los estudios de asociación de genoma completo pueden ser realizados mediante diferentes supuestos de análisis considerando o no la estructura poblacional, el análisis de componentes principales o la matriz de parentesco. Muchos estudios consideran todos los modelos disponibles, para así obtener el modelo que evitará al máximo la aparición de errores de tipo I (falsos positivos) y errores de tipo II (falsos negativos).

1.8 Efecto pleiotrópico

Mendel (1866) durante el estudio de herencia en plantas de arvejas, describió el primer caso de pleiotropía, sin definirlo formalmente. Mendel, notó que las semillas con cubierta café

siempre tenían las axilas de las hojas manchadas y las flores presentaban coloración violeta. De igual forma, observó que semillas con cubiertas claras o incoloras siempre tenían flores blancas y presentaban ausencia de pigmentación en las axilas de las hojas. Al observar que el color de la cubierta de las semillas siempre se asoció con la coloración de flores y axilas, Mendel señaló que los tres caracteres están correlacionados y podrían estar siendo controlados por un solo factor. A pesar de lo anterior, no fue hasta comienzos del siglo veinte cuando Plate (1910) acuñó por primera vez el término “pleiotropismo”, señalando lo siguiente: “*Yo llamo una unidad de herencia pleiotrópica si varias características dependen de ella; estas características siempre aparecerán juntas y, por lo tanto, pueden parecer correlacionadas*” (Plate 1910). Definición que se ha mantenido válida hasta la fecha.

Etimológicamente, el término pleiotropía se deriva de las palabras griegas *pleios*, que significa ‘*muchos*’, y *tropo*, que significa ‘*cambios*’. La pleiotropía es un fenómeno genético en el cual la expresión de un gen afecta múltiples fenotipos, aparentemente no relacionados (Lemmon y Doebley 2014).

El mapeo de QTL se ha utilizado comúnmente para definir la arquitectura genética que subyace a características de interés en diferentes especies (Lemmon y Doebley 2014). A menudo, los estudios relacionados al mapeo de QTL identifican grupos de loci de gran efecto localizados en un pequeño número de regiones genómicas, lo que sugiere una arquitectura genética común controlando su expresión (posiblemente subyacente de genes pleiotrópicos o altamente ligados) (Miller et al. 2014, Lemmon y Doebley 2014). El efecto de QTLs pleiotrópicos se ha mapeado con éxito en tomate (Frary et al. 2000, Cong et al. 2008), maíz (Lemmon y Doebley 2014, Wang et al. 2005, Studer et al. 2011), arroz (Konishi et al. 2006, Li et al. 2006) y trigo (Simons et al. 2006). Sin embargo, estos genes pleiotrópicos, explican una parte de la variación fenotípica en las especies, dejando una proporción considerable de las diferencias sin explicación.

Lemmon y Doebley (2014) señalan que genes involucrados en la domesticación del maíz están dispersos por todo el genoma, con varias concentraciones de QTLs controlando grandes proporciones de la varianza fenotípica total. Hasta la fecha, se han identificado al menos tres genes pleiotrópicos de gran efecto como actores clave de la domesticación de maíz: *grassy*

tillers1 (*gt1*), *teosinte branched1* (*tb1*) y *teosinte glume architecture1* (*tga1*) (Whipple et al. 2011, Wills et al. 2013, Doebley et al. 1997, Gaudin et al., 2014, Studer et al. 2011, Wang et al. 2005). El gen *gt1*, influye directamente en los macollos y es en gran parte responsable de la concentración de semillas en la mazorca (Whipple et al. 2011, Wills et al. 2013). Por otro lado, *tga1* tiene un impacto en la desarticulación de las mazorcas, la longitud de la rama lateral y la filotaxia de las mazorcas (Wang et al. 2015). Finalmente, el gen *tb1* como un factor de transcripción, se une a muchas regiones del genoma, presentando un impacto en muchas características, tales como: dominancia apical, longitud de las ramas laterales, crecimiento de las hojas en las ramas laterales, desarrollo de las espiguillas pediceladas y arquitectura de la raíz (Gaudin et al., 2014, Studer et al. 2011). Por lo tanto, comprender la pleiotropía puede ayudar a entender cómo surgen las correlaciones entre diferentes características, y arrojar luces sobre genes reguladores del proceso evolutivo en diversas poblaciones.

1.9 Genes candidatos para AF, AP y AM en maíz

El uso de genes candidatos en estudios de asociación está basado en el conocimiento a priori del impacto funcional de uno o varios genes sobre la variación de un fenotipo (Tabor et al. 2002). Esto implica analizar las posibles variantes alélicas dentro de un gen, y determinar si existe una correlación entre esas variantes y la variación fenotípica (Tabor et al. 2002). Sin embargo, otro enfoque está centrado en la identificación de variantes alélicas (SNP) cercanas a un gen, las que podrían estar causando cambios en su expresión (Upadhyaya y Cooper 2013).

Numerosos genes candidatos han sido reportados para las características de AF, AP, AM y la relación AM/AP, colindantes a QTL detectados a través de GWAS (Li et al. 2016). Por ejemplo, Li et al. (2016) reportó 41 SNP asociados significativamente a AP y AM mediante GWAS, de los cuales 29 se encontraban localizados dentro de 19 genes. Por otro lado, Li et al. (2015), Tian et al. (2011) y Ku et al. (2011) encontraron QTL para AF ubicados en el cromosoma 2, los cuales están subyacentes al gen candidato *lg1* (*liguleless1*). El gen *lg1* ha sido reportado como un gen asociado a la rectitud de las hojas y las propiedades de la aurícula en maíz (Moreno et al. 1997). Actualmente, el enfoque de GWAS basado en GBS se ha aplicado con éxito en la identificación a escala masiva (en todo el genoma) de regiones de interés cercanas o dentro de genes candidatos que controlan la variación de características cuantitativas en maíz (Li et al. 2016; Li et al. 2015; Navarro et al. 2017).

1.10 Planteamiento del problema y novedad de la propuesta

El maíz (*Zea mays*) es uno de los cultivos más utilizados alrededor del mundo, debido a que proporciona varios recursos, como por ejemplo, fuente de alimento, forraje y productos agroindustriales (Li et al. 2016; Yang et al. 2011). Cabe destacar que AF, AP, AM y la relación AM/AP, son características agronómicas relacionadas no tan solo al rendimiento de granos, sino también a la calidad del cultivo (Buescher et al. 2014; Weng et al. 2011; Kong et al. 2017; Cai et al. 2012), por lo tanto una mejor comprensión de la base genética de estas características podría ayudar a generar variedades que mejoren el rendimiento de los cultivos (Cai et al. 2012).

Estudios previos en maíz han encontrado que las características AF, AP y AM (Abdel Ghani et al. 2016; Yang et al. 2014; Tian et al. 2011; Ku et al. 2010; Peiffer et al. 2014; Cai et al. 2012; Li et al. 2016), son de moderada a altamente heredables, con valores entre $h^2=0.54-0.92$, $h^2=0.81-0.89$ y $h^2=0.59-0.93$ (respectivamente), indicando que estas características tienen un alto grado de control genético (Li et al. 2016). Por otro lado, Peiffer et al. (2014) y Ku et al. (2010) señalan que AP y AM tienen una correlación genética positiva y significativa ($P<0.01$), con valores de $r=0.77$ y $r=0.55$, respectivamente. Por lo que, estas características podrían estar siendo controladas por factores genéticos comunes y/o presentar regiones con efecto pleiotrópico.

Un GWAS que utiliza diversas accesiones de maíz tropical ofrece el potencial para investigar regiones genómicas o genes que contribuyen a la variabilidad de características complejas, debido a la alta tasa de recombinación que el maíz presenta (Lu et al. 2011). Cabe destacar que esta metodología requiere una alta densidad de marcadores moleculares (principalmente SNP) para cubrir todo el genoma de la especie en estudio. Considerando el bajo costo de la secuenciación, una de las mejores opciones para el descubrimiento de una alta densidad de SNP es la implementación de GBS, la que permite obtener un alto número de polimorfismos (Elshire et al. 2011; He et al. 2014; Mammadov et al. 2012). Por lo tanto, el uso combinado de GBS y GWAS es una poderosa herramienta para la identificación de QTL (a lo largo de todo el genoma) que afectan las características de interés (Yu y Buckler 2006). Adicionalmente, el uso de bloques de haplotipos en el análisis de asociación podría compensar la limitación bi-alélica de los marcadores SNP, proporcionando nuevos conocimientos sobre los

determinantes genéticos que no son capturados por el enfoque de marcador único (SNP) (Abdel-Shafy et al. 2014).

En resumen, considerando que: (I) en el futuro se necesitarán cultivares de maíz que produzcan mayor rendimiento de granos por cultivo y sin pérdidas asociadas a la arquitectura y/o encamado; (II) GWAS proporciona la disección de la base genética de la variabilidad fenotípica presente en una población; y (III) GBS posibilita el descubrimiento de un gran número de loci polimórficos, se plantean las siguientes hipótesis y objetivos:

2. HIPOTESIS DE TRABAJO

A partir de que i) las características AF, AP, AM y la relación AM/AP son controladas por múltiples genes en maíz, ii) GBS permite identificar una alta densidad de loci polimórficos en todo el genoma, y iii) GWAS permite diseccionar la base genética de la variabilidad fenotípica de características complejas, se plantean las siguientes hipótesis:

- i. “El patrón de LD existente en la población de maíz tropical, conformada por líneas endogámicas, posibilita la identificación de bloques de haplotipos que expresen fenotípicamente una respuesta diferencial para las características de AF, AP, AM y la relación AM/AP.”
- ii. “La respuesta fenotípica diferencial de los bloques de haplotipos y loci polimórficos individuales son estables en temporadas agrícolas consecutivas.”
- iii. “SNP individuales y bloques de haplotipos asociados significativamente a más de una característica a la vez, presentan efecto pleiotrópico.”
- iv. “Loci polimórficos asociados significativamente a AF, AP, AM y la relación AM/AP, subyacerán genes candidatos que controlan la expresión de estas características.”

3. OBJETIVO GENERAL

Determinar haplotipos y SNP individuales asociados a AF, AP, AM y la relación AM/AP en dos temporadas agrícolas consecutivas de líneas endogámicas de maíz tropical, validar el efecto pleiotrópico de loci presentes en más de una característica a la vez, e identificar genes candidatos próximos a los polimorfismos de expresión fenotípica diferencial.

3.1 Objetivos Específicos

- i. Determinar la variación inter e intracromosomal del LD y caracterizar bloques de haplotipos en las subpoblaciones endogámicas de maíz tropical (maíz común y popcorn).
- ii. Identificar asociaciones significativas entre los polimorfismos genéticos y las características de AF, AP, AM, y la relación AM/AP, en dos temporadas agrícolas consecutivas.
- iii. Evaluar regiones genéticas con efecto pleiotrópico para las características de interés.
- iv. Identificar genes candidatos subyacentes a los haplotipos y loci polimórficos asociados significativamente a AF, AP, AM y la relación AM/AP.

4. CAPÍTULO I: ESTRUCTURA POBLACIONAL, DESEQUILIBRIO DE LIGAMIENTO Y MAPEO ASOCIATIVO EN MAÍZ TROPICAL USANDO UN PANEL DE MARCADORES SNP (GBS) DE ALTA DENSIDAD

4.1 Introducción

El conocimiento del control genético a nivel molecular como apoyo a los programas de mejoramiento ha mostrado ventajas en la identificación de la base genética de características cuantitativas, permitiendo la identificación de QTLs y genes candidatos a través de las diferencias estadísticas entre loci polimórficos y fenotipos (Paes et al. 2016). Hasta la fecha, gracias al genotipado por secuenciación, se ha desarrollado, para uso directo, un número creciente de marcadores SNP a bajo costo (Elshire et al. 2011). Esto ha permitido explorar la diversidad genética y la estructura poblacional en diversas especies (Siadjeu et al. 2018, Luo et al. 2019, Baral et al. 2018, Dinesh et al. 2016), junto con la construcción de mapas densos de ligación, y el descubrimiento de QTLs asociados con características agronómicas de interés (Song et al. 2015, Su et al. 2017).

La descripción de las relaciones entre la variación genotípica y fenotípica de características complejas, ha sido un elemento clave en la genómica moderna, y ha permitido obtener importantes avances en las ciencias biológicas y el mejoramiento genético (Grattapaglia, 2014). En el maíz (*Zea mays* L.), las características económicamente importantes generalmente se heredan cuantitativamente, poseen una base genética poligénica, con presencia de efectos de interacción entre genes y el medio ambiente (Mafra et al. 2019). En este sentido, el mapeo asociativo de genoma amplio (GWAS) se ha convertido rápidamente en un método estándar en la disección de la base genética de características cuantitativas, el cual permite el descubrimiento de genes candidatos para muchas características en humanos, animales y plantas (Coan et al. 2018, Paes et al. 2016). Gracias a la liberación del genoma de referencia del maíz B73, los estudios de asociación han permitido identificar varias características agronómicamente importantes (Coan et al. 2018), siendo una especie modelo para el uso de mapeo asociativo (Rashid et al. 2018).

A partir de lo anterior, en el presente capítulo un estudio de asociación de genoma amplio, basado en un modelo lineal mixto (MLM), fue aplicado a características relacionadas con el encamado de las plantas (AM, AP, relación AM/AP) y la arquitectura vegetal (AF), en un panel de líneas endogámicas de maíz tropical evaluado en dos temporadas agrícolas consecutivas.

4.2 Materiales y métodos

4.2.1 Descripción del material biológico y características fenotípicas evaluadas

El estudio de asociación de genoma amplio consideró un panel de 183 líneas endogámicas de maíz tropical, el cual se evaluó durante las temporadas agrícolas 2014-2015 y 2015-2016, en la localidad de Iguatemi (latitud: 23S 25'; longitud: 51O 57' y altitud: 550 m), ubicada en los campos experimentales del Departamento de Agronomía de la Universidad Estatal de Maringá (UEM), Estado Federativo de Paraná, Brasil. Las líneas se plantaron acorde a un diseño de bloques completos al azar, con 2 bloques y 28 repeticiones por línea. El manejo del campo se realizó usando irrigación artificial y acorde a las practicas agronómicas sugeridas por Cruz et al. (2015).

Se evaluaron las siguientes características: ángulo foliar (AF), altura de la planta (AP), altura de la primera mazorca (AM) y relación AM/AP. AP se midió desde el suelo hasta la inflorescencia primaria. AM desde el suelo hasta el nodo de la primera mazorca. La relación AM/AP se obtuvo por medio del cociente de AM por AP en cada repetición. AF se midió 10 días después de anthesis en cada repetición y se consideraron 4 hojas por encima de la primera mazorca para estimar visualmente el ángulo de la hoja. La escala considero cuatro niveles: (1) inclinado, (2) erguido, (3) muy erguido y (4) extremadamente erguido. Las características se dividieron en 2 grupos: I) características relacionadas al desplome o encamado (AP, AM y relación AM/AP) y II) arquitectura de la hoja (AF).

4.2.2 ADN genómico y descubrimiento de SNP por GBS

La extracción de ADN se realizó sobre tejido de hoja mediante el método de Bromuro de cetiltrimetilamonio (CTAB) descrito por Helentjaris T. (1995). Se recogieron hojas de 5

plantas aleatoriamente seleccionadas de las 183 líneas endogámicas de maíz. La calidad del ADN se evaluó en un gel de agarosa al 0.8%. Las muestras de ADN de cada línea endogámica se genotiparon con marcadores SNPs via genotipado por secuenciación (GBS), el cual se realizó como se describe en Elshire et al. (2011), y en <https://bitbucket.org/tasseladmin/tassel-5-source/wiki/Tassel5GBSv2Pipeline> (Instituto de Diversidad Genómica, Universidad de Cornell). Brevemente, la técnica de GBS consistió en los siguientes pasos: (I) reducción de la complejidad del genoma mediante digestión con enzimas de restricción sensibles a la metilación de ApeKI, (II) ligación de adaptadores (compatibles con las enzimas de restricción) a los fragmentos de ADN genómico, (III) cebadores de secuenciación Illumina (con un sitio de unión específicos a los adaptadores ligados) que permitieron agrupar, purificar y amplificar el material genómico mediante PCR, (IV) se limpió el producto PCR y se comprobó el tamaño de los fragmentos de las bibliotecas de GBS mediante un analizador de ADN (BioAnalyzer, Agilent Technologies, Inc., USA), y (V) las bibliotecas se cuantificaron y diluyeron para la secuenciación en el Illumina HiSeq 2000 (Illumina Inc., San Diego, CA).

Mediante un pipeline del software TASSEL 5.2 (Bradbury et al. 2007), se alinearon las secuencias con el genoma de referencia *Zea mays* versión AGPV3 (B73 RefGen v3) (Schnable et al. 2009). Después de la alineación, se calculó la heterocigosidad y las variantes de menor frecuencia alélica (MAF; *Minor Allele Frequency*) utilizando VCFtools (versión v0.1.12a). Mediante la puntuación de calidad a escala Phred (calidad del genotipado), se eliminaron las lecturas con una puntuación por debajo de 80 (Chen et al. 2014). Finalmente, se filtró la base de datos considerando un $MAF > 0.01$ y una proporción de datos perdidos por sitio $< 90\%$ (Yu et al. 2017).

4.2.3 Estructura poblacional y estimación de matriz kinship

Se infirió la estructura genética poblacional mediante un modelo de agrupamiento bayesiano implementado en el software STRUCTURE 2.3.4 (Pritchard et al. 2000) y un análisis PCA implementado en TASSEL 5.2 (Bradbury et al. 2007), con un subconjunto de los marcadores aleatorios, cubriendo los 10 cromosomas de la especie. Mediante un script basado en STRUCTURE (ver anexo 1) se estimó el número de subpoblaciones (K) a partir de veinte corridas por cada posible K (que oscila de 2 a 10), considerando un modelo de mezcla o admixture (supuesto que considera que cada individuo ha heredado una fracción de su genoma

de antepasados en la subpoblación K; Pritchard et al. 2000) y frecuencias alélicas correlacionadas (sin considerar información previa de la población), un periodo de descarte (burn-in) de 1×10^5 y un número de repeticiones MCMC después del descarte de 1×10^6 (Evanno et al. 2005). El valor óptimo de K se estimó por medio de la tasa de cambio de segundo orden de la función de probabilidad con respecto a K (ΔK), según lo propuesto por Evanno et al. (2005). Adicionalmente, los resultados de PCA se utilizaron para visualizar y corroborar el número de grupos entregados por STRUCTURE.

Se calculó la matriz de relación inter-líneas o kinship usando los SNPs con un MAF superior al 0.05 y usando muestras con datos faltantes menores a 25%, en el software TASSEL 5.2 (Bradbury et al. 2007).

4.2.4 Análisis estadístico de los datos fenotípicos

El análisis de los datos fenotípicos se realizó usando el siguiente modelo lineal mixto (PROC MIXED) de SAS (versión 9.2):

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}_1\mathbf{l} + \mathbf{Z}_2\mathbf{ls} + \boldsymbol{\varepsilon}$$

donde \mathbf{y} representa los valores fenotípicos (para AF, AP, AM o relación AM/AP); \mathbf{X} , \mathbf{Z}_1 y \mathbf{Z}_2 corresponden a las matrices de incidencia conocidas que relacionan \mathbf{y} con los vectores $\boldsymbol{\beta}$, \mathbf{l} , y \mathbf{ls} , respectivamente; $\boldsymbol{\beta}$ es el vector de efecto fijo de la temporada y el bloque anidado dentro de la temporada; \mathbf{l} corresponden al vector aleatorio del efecto genotípico (líneas); \mathbf{ls} es el vector de efectos aleatorios debido a la interacción línea-temporada; y $\boldsymbol{\varepsilon}$ es el vector de efectos aleatorios residuales. Se utilizó el método de máxima verosimilitud restringida (REML; *REstricted Maximum Likelihood*) para la estimación de los componentes de varianza. Los siguientes criterios de información se utilizaron para determinar estadísticamente la significancia de los efectos aleatorios (comparación de modelos):

$$BIC = -2 \log RL + d \log n$$

$$AIC = -2 \log RL + 2d$$

donde BIC corresponde al criterio de información Bayesiano (o Schwarz); AIC es el criterio de información Akaike; $-2 \log RL$ denota el valor máximo del logaritmo de la verosimilitud (restringida); d corresponde a la dimensión del modelo; y n es el número de variables observadas

por la estimación de RL. Finalmente, se asumió que el mejor modelo (mejor ajuste) fue aquel que minimizó los valores de los criterios de información (Ferrão et al. 2017).

Los componentes de varianza se estimaron mediante ASREML 4 (Gilmour et al. 2015). Para el ángulo de la hoja, se utilizó modelo umbral en MTGSAM versión Threshold (Van Tassell y Van Vleck 1995), debido a que soporta variables ordinales. Se calculó la heredabilidad (H^2) de cada carácter a partir de la siguiente expresión:

$$H^2 = \frac{\sigma_g^2}{\sigma_g^2 + (\sigma_{gs}^2/n) + (\sigma_\varepsilon^2/n \cdot r)}$$

donde σ_g^2 es la varianza genotípica, σ_{gs}^2 representa la varianza de la interacción entre los genotipos y el ambiente, σ_ε^2 corresponde a la varianza residual, n es el número de ambientes (temporadas agrícolas), y r indica el número de repeticiones.

4.2.5 Estudio de asociación del genoma completo (GWAS)

El análisis de asociación SNP-característica, se realizó usando el enfoque de Modelos Lineales Mixtos (MLM; *Mixed Linear Model*) implementado en TASSEL 5.2. El modelo incluyó la matriz de parentesco o kinship (K) y la estructura poblacional, usando el coeficiente de STRUCTURE (Q) (Yu et al. 2006). El modelo estadístico puede ser descrito como:

$$\mathbf{y} = \mathbf{S}\boldsymbol{\alpha} + \mathbf{Q}\mathbf{v} + \mathbf{Z}\boldsymbol{\mu} + \boldsymbol{\varepsilon}$$

donde \mathbf{y} es el vector de observaciones fenotípicas ajustadas por el diseño experimental, $\boldsymbol{\alpha}$ es el vector de efectos fijos de los marcadores moleculares SNP, \mathbf{v} es el vector del efecto fijo de la estructura poblacional, $\boldsymbol{\mu}$ es el vector del efecto aleatorio del parentesco, y $\boldsymbol{\varepsilon}$ es el vector de efectos residuales aleatorios. \mathbf{S} , \mathbf{Q} y \mathbf{Z} son las matrices de incidencia relativa a \mathbf{y} con $\boldsymbol{\alpha}$, \mathbf{v} y $\boldsymbol{\mu}$ (respectivamente).

Los marcadores SNP con un MAF inferior al 0.05 y con datos faltantes mayor a 25% se descartaron del análisis de asociación para reducir la tasa de falsos positivos (o error tipo I) (Myles et al. 2009). Adicionalmente, se realizó un test de corrección de la tasa de detección de falsos positivos (FDR; *False Discovery Rate*) en el software R (R Development Core Team

2008) con la función *p.adjust()*, para conocer la proporción de falsos positivos entre todos los descubrimientos.

4.3 Resultados

Acorde a los criterios de información, el modelo que mejor se ajustó a las características de despome o encamado fue el modelo completo (M3), el cual incluye el efecto de la interacción de las líneas con el ambiente (Tabla 1). Por otro lado, para AF el modelo sin interacción genotipo ambiente (M2) fue el de mejor ajuste (Tabla 1).

Tabla 1. Resultados de la selección de modelos basado en los criterios de información Bayesiano (BIC) y Akaike (AIC) para AF, AM, AP y relación AM/AP en líneas endogámicas de maíz tropical.

Característica	CI	Modelo		
		<i>M1</i>	<i>M2</i>	<i>M3</i>
Arquitectura				
	-2LRL	1025.0	324.0*	324.2
<i>Ángulo Foliar (AF)</i>	AIC	1027.0	328.0*	328.2
	BIC	1031.4	334.0*	334.2
Encamado				
	-2LRL	-1109.0	-2941.0	-3247.0*
<i>Altura primera Mazorca (AM)</i>	AIC	-1107.0	-2937.0	-3241.0*
	BIC	-1101.0	-2931.0	-3232.0*
	-2LRL	530.0	-1607.0	-1953.0*
<i>Altura Planta (AP)</i>	AIC	532.0	-1603.0	-1947.0*
	BIC	532.0	-1597.0	-1938.0*
	-2LRL	-5116.0	-5808.0	-6578.0*
<i>Relación AM/AP</i>	AIC	-5114.0	-5804.0	-6570.0*
	BIC	-5108.0	-5798.0	-6558.0*

CI: criterio de información; *modelo de mejor ajuste; -2LRL indica el valor máximo de la probabilidad logarítmica (restringida); *M1* corresponde al modelo que no incluye el efecto de la línea ni la interacción línea-ambiente; *M2* considera el efecto debido a la línea; *M3* corresponde al modelo completo (considerando el efecto de la línea y la interacción línea-ambiente).

El moderado coeficiente de variación (CV) encontrado en las cuatro características (19% en AF, 10% en AP, 16% en AM y 11% en la relación AM/AP; Tabla suplementaria 1) indica un adecuado nivel de precisión experimental (Mora y Arriagada 2016). Los valores promedio variaron en cada temporada para AP (1.52 y 1.27, respectivamente) y AM (0.81 y 0.58, respectivamente), con valores más altos durante la temporada 2014-2015, indicando que el ambiente tuvo un impacto en la variación de estas características. Por otro lado, el AF no mostró diferencias significativas entre las temporadas (1.287 y 1.291, para 2014-2015 y 2015-2016, respectivamente; Tabla suplementaria 1). Los altos valores de heredabilidad en las características de arquitectura de la hoja y encamado de la planta ($H^2 = 0.74, 0.95, 0.94$ y 0.83 , para AF, AP, AM y relación AM/AP, respectivamente; Tabla suplementaria 1) fueron similares a lo reportado en estudios previos (Li et al. 2016, Weng et al. 2011, Tian et al. 2011, Zhang et al. 2010). El fuerte control genético observado en este panel de maíz sustenta la hipótesis de encontrar regiones genómicas asociadas con las características de interés (Brachi et al. 2011).

4.3.1 Estructura poblacional

De acuerdo con el método propuesto por Evanno et al. (2005), vinculado a los resultados de STRUCTURE, los 183 genotipos de maíz se dividieron en dos grupos genéticamente diferenciados: maíz común (98 genotipos) y popcorn (85 genotipos) (Fig. 1A). Estos resultados fueron similares a lo que se visualizó con los dos primeros componentes principales del método PCA (Fig. 1B) y el dendrograma neighbor-joining (Fig. 1A). Los ejes PC1 y PC2 explicaron el 11.6% y el 3.4% de la variación en los datos genotípicos, respectivamente, y fueron los primeros ejes que separaron la mayoría de las líneas endogámicas (Fig. 1B).

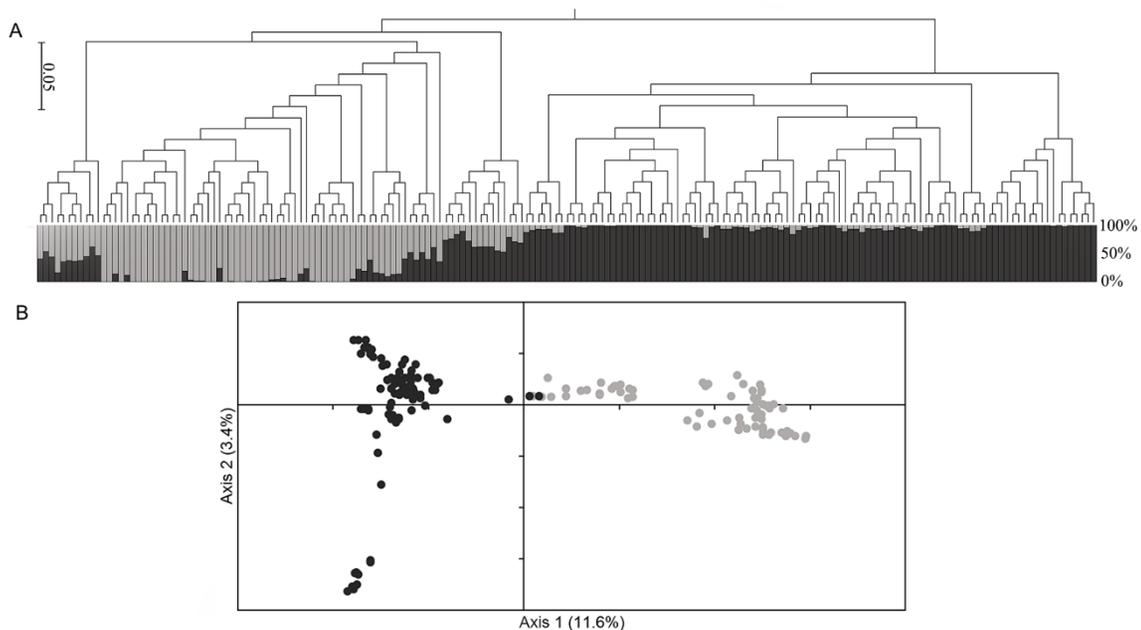


Figura 1. Distribución espacial de las 183 líneas de maíz endogámico en las dos subpoblaciones (maíz común y popcorn) utilizando un subgrupo de 5000 marcadores SNP distribuidos en todo el genoma de maíz. A) Estructura genética inferida por un modelo de agrupación bayesiana usando STRUCTURE y un dendrograma obtenido usando neighbor-joining basado en las distancias genéticas de Nei. La proporción de segmentos de color gris claro y gris oscuro indican la proporción del genoma extraído de los dos subgrupos (maíz común y popcorn, respectivamente). B) Análisis de componentes principales (PCA) que muestra las 183 líneas endogámicas distribuidas espacialmente en relación con los dos primeros componentes principales. Los valores entre paréntesis en los ejes, indican el porcentaje de variación explicado por cada componente principal.

4.3.2 Estudio de asociación de genoma amplio con marcadores SNP- individuales

El modelo lineal utilizado en el GWAS consideró la información de las matrices Q (estructura genética) y K (parentesco), para reducir la tasa de falsos positivos y corregir las asociaciones espurias en el análisis (Yu et al. 2017). Según el análisis de modelos mixtos, se identificó un total de 163 SNP asociados con las cuatro características de interés, 15 de los cuales se asociaron en ambas temporadas agrícolas y 11 se compartieron entre múltiples características. Veinticinco de los SNP detectados se asociaron con AF, 37 presentaron asociaciones significativas con AM, 24 se asociaron con AP y 67 SNP se asociaron con la relación AM/AP. La información sobre los SNP detectados en este estudio se resume en la Tabla 2.

Tabla 2. Resumen de las asociaciones detectadas por GWAS, basado en SNP-individuales, para el ángulo foliar (AF), altura de la primera mazorca (AM), altura de la planta (AP) y la relación AM/AP, evaluadas en líneas de maíz tropical sobre dos temporadas agrícolas.

Característica	2014-2015			2015-2016		
	N ^o _s	Cromosoma (N ^o _s)	PV (%)	N ^o _s	Cromosoma (N ^o _s)	PV (%)
Arquitectura						
AF	20	1(2), 2(3), 4(5), 5(6), 6(1), 7(1) y 8(2)	10.2-22.2	15	1(2), 2(2), 4(4), 5(5), 6(1) y 8(1)	11.1-21
Encamado						
AP	9	1(3), 2(1), 3(1), 4(1), 7(1) y 8(2)	7.9-13.6	15	1(3), 2(1), 3(1), 5(2), 6(1), 7(1), 8(1), 9(3) y 10(2)	7.8-15.3
AM	17	1(1), 3(1), 6(1), 7(5), 8(7) y 10(2)	7.1-12.0	20	1(1), 2(1), 3(1), 4(4), 5(2), 6(2), 7(5), 8(1), 9(1) y 19(2)	8.3-15.2
AM/AP	46	1(3), 2(4), 4(9), 5(4), 6(3), 7(5), 8(5), 9(7) y 10(6)	7.0-11.4	21	1(5), 2(1), 4(2), 5(1), 6(1), 7(2), 8(6), 9(1) y 10(2)	8.2-12.0

PV (%): porcentaje de variación fenotípica explicada por los SNP; N^o_s: número de asociaciones SNP-individuales significativas.

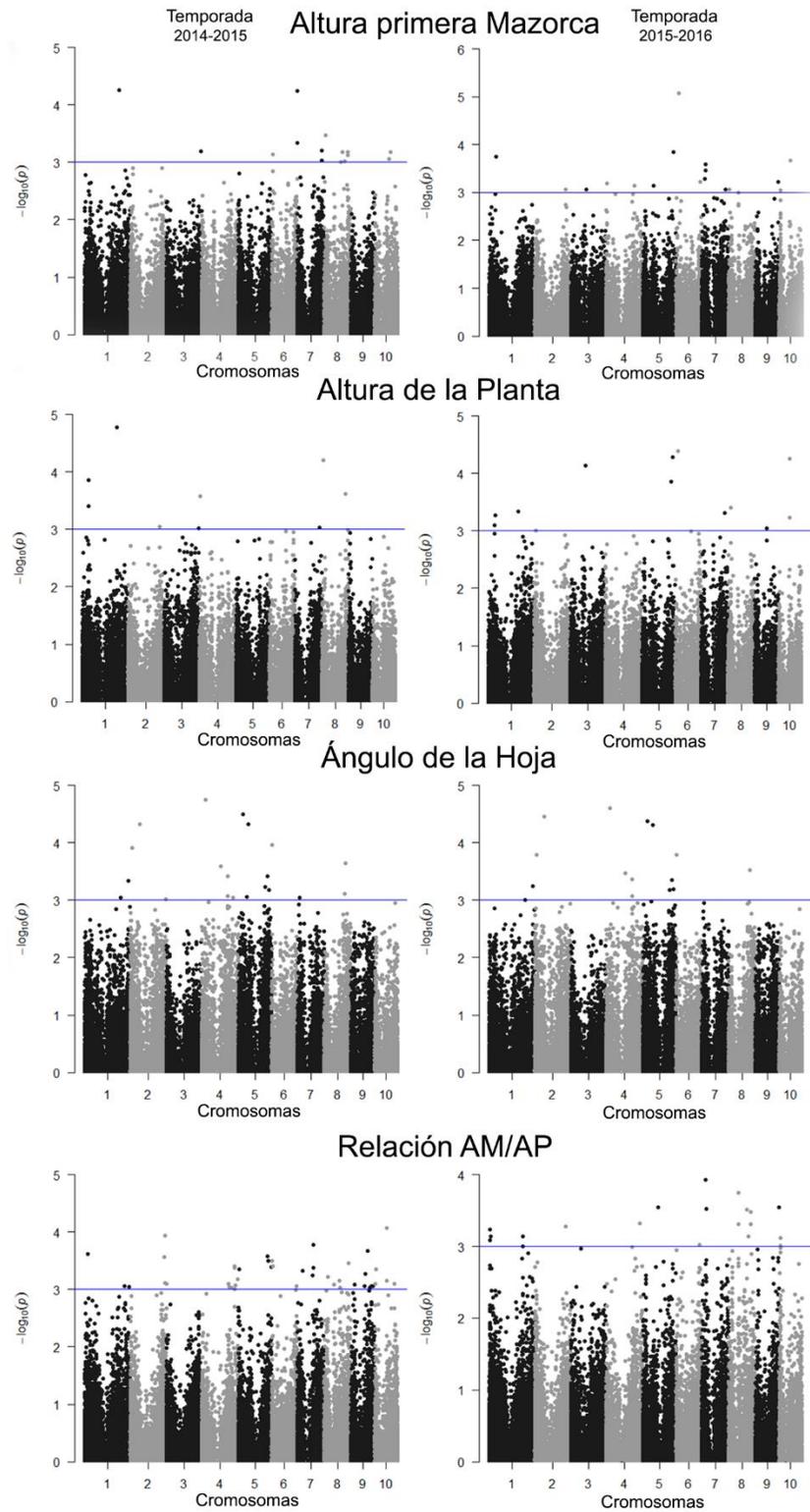


Figura 2. Gráficos de Manhattan de la significancia estadística de los SNP asociados con las características de altura de la primera mazorca (AM), altura de la planta (AP), ángulo de la hoja

(AF) y relación AM/AP, evaluados en un panel de asociación de maíz tropical durante las temporadas de crecimiento 2014-2015 y 2015-2016. Los $-\log_{10}$ (p-values) de los SNP se representaron para todos los cromosomas en su conjunto. Los SNP significativamente asociados con las características están por encima de la línea de umbral ($p\text{-value} < 1.0 \times 10^{-3}$).

La Figura 2 muestra un Manhattan con los resultados de GWAS. Un total de 17 SNP asociados con AM estuvieron presentes en la temporada de crecimiento 2014-2015, y se identificaron 20 SNP diferentes en la siguiente temporada. Del mismo modo, para AF, se asociaron 20 y 15 SNP en las temporadas de crecimiento 2014-2015 y 2015-2016, respectivamente. Para el AP, 9 y 15 SNP se asociaron en las temporadas de crecimiento 2014-2015 y 2015-2016, respectivamente. Para la relación AM/AP, 46 SNP se asociaron en la temporada de crecimiento 2014-2015, y 21 SNP se asociaron en la temporada 2015-2016.

En el ángulo foliar, la proporción de la variación fenotípica explicada por los SNP varió entre 10.2 y 22.2%, siendo el SNP S1_317085830, ubicado en el cromosoma 6, el loci que presentó el mayor efecto durante la temporada agrícola 2014-2015 (Tabla 3). En general, los valores de r^2 fueron moderados a relativamente altos, variando entre el 7% (SNP S1_285330124, asociado con la relación AM/AP durante la temporada 2014-2015, ubicada en el cromosoma 9) y el 22.2% (SNP S1_317085830, asociado con AF durante la temporada 2014-2015) (Tabla 3). Los valores de r^2 en este estudio fueron relativamente más altos que los encontrados en la literatura para estas características (Xiao et al. 2017). Dos de los loci con efectos relativamente importantes relacionados con el encamado de las plantas, los SNPs S1_333289434 (ubicados en el cromosoma 6) y S1_81113660 (en el cromosoma 10), se asociaron concomitantemente con AM y AP en la temporada agrícola 2015-2016, explicando entre 9.5-15.2% y 10.6-15.3% de las variaciones fenotípicas, respectivamente. Además, otras asociaciones de marcador-característica se identificaron conjuntamente para AP y AM, como los SNPs S2_47326011 (bin 1.03), S2_230948259 (bin 1.08), S1_1158043959 (bin 3.04), S1_1280061812 (bin 3.09), S1_1037703810 (bin 5.06), S1_821562289 (bin 7.04), S1_494063499 (bin 8.01) y S1_639521797 (bin 8.06). Del mismo modo, el SNP S1_16427856 (bin 10.02) se asoció conjuntamente con las características AM y la relación AM/AP. Estos resultados sugieren un posible efecto pleiotrópico para las características relacionados con el encamado de las plantas (Li et al. 2016).

En este estudio, se encontraron más de cuatro asociaciones en cada uno de los siguientes Bin cromosómicos 1.03, 1.08, 4.07, 4.08, 4.09, 5.03, 5.05, 6.00, 7.02, 7.04, 8.03, 8.04, 8.05, 8.06, 9.03 y 10.03. El Bin 1.03 contenía una gran cantidad de SNP (9) asociados con las características de interés, seis de los cuales están asociados con AP, dos con AM y uno con AF. En particular, esta región genómica tuvo asociaciones para las tres características, lo que sugiere una posible ligación entre ellas.

Todas las asociaciones detectadas en AP, AM y la relación AM/AP fueron específicas de cada temporada. Estos resultados están en concordancia con los valores de los criterios de información AIC y BIC, que apuntan a una interacción genotipo-ambiente (línea x temporada) significativa para AP, AM y la relación AM/AP (Tabla 1). Por otro lado, para AF, el 75% de las asociaciones estuvieron presentes en ambos años, lo que indica que la mayoría de los QTLs para AF fueron estables entre temporadas agrícolas. El modelo que no incluía la interacción genotipo-ambiente (M2) se ajustó mejor a los datos de AF, según los criterios AIC y BIC (Tabla 1), lo que indica que la interacción genotipo-ambiente no proporciona información al modelo completo (Burnham y Anderson 2004). Por lo tanto, las líneas endogámicas no mostraron interacción con las temporadas, mostrando estabilidad en AF y, en consecuencia, en las asociaciones identificadas a través de GWAS.

Tabla 3. Información detallada de las asociaciones con efectos relativamente importantes (>15%), detectados en un GWAS basado en SNP individuales para las características de ángulo de la hoja (AF), altura de la primera mazorca (AM), altura de la planta (AP) y relación AM/AP, medidas en líneas endogámicas de maíz tropical durante dos temporadas (2014-2015 y 2015-2016).

Temporada	Característica	Marcador	Cr	Posición (pb)	Bin	p-value	PV
2014-2015	Arquitectura						
	AF	S1_1348935692	2	62476194	2.04	4.7E-06	16.5%
	AF	S1_901406458	5	65152212	5.03	4.7E-06	16.9%
	AF	S1_1296983734	2	10524236	2.02	1.2E-05	17.3%
	AF	S1_867050714	5	30796468	5.03	3.1E-06	17.7%
	AF	S1_1648381508	4	124004442	4.05	2.6E-05	17.8%
	AF	S1_1550828517	4	26451451	4.04	1.8E-06	18.6%
	AF	S1_317085830	6	2443523	6.00	1.1E-05	22.2%
2015-2016	Arquitectura						
	AF	S1_901406458	5	65152212	5.03	4.9E-06	16.8%
	AF	S1_1296983734	2	10524236	2.02	1.6E-05	16.8%
	AF	S1_1348935692	2	62476194	2.04	3.5E-06	17.0%
	AF	S1_867050714	5	30796468	5.03	4.3E-06	17.1%
	AF	S1_1648381508	4	124004442	4.05	3.4E-05	17.3%
	AF	S1_1550828517	4	26451451	4.04	2.5E-06	17.9%
	AF	S1_317085830	6	2443523	6.00	1.6E-05	20.9%
	Encamado						
	AP	S1_81113660	10	73141785	10.03	5.5E-05	15.3%
	AM	S1_333289434	6	18647127	6.01	8.3E-06	15.2%

PV: porcentaje de variación fenotípica explicada por un SNP; Cr: cromosoma al cual pertenece el loci; Bin: o “contenedor”, indica una posición en el mapa genético con un patrón de segregación único, y que está separado de los contenedores adyacentes por un único evento de recombinación (Van et al. 2006).

4.4 Discusión

Para AF, se identificaron 15 SNP (75%) significativos, presentes en ambas temporadas de crecimiento. La selección del modelo que no considera la interacción genotipo-ambiente (de acuerdo con los criterios AIC y BIC) sustenta la hipótesis de estabilidad en las asociaciones detectadas en el ángulo foliar. Sin embargo, para las características de encamado (AP, AM y la relación AM/AP), el modelo de mejor ajuste, según AIC y BIC, incluye los efectos de las interacciones genotipo-ambiente (modelo completo), lo que indica que la respuesta de las líneas endogámicas depende de la temporada agrícola. Este resultado nos permite explicar la ausencia de SNP estables durante ambas temporadas. Nuestros resultados indican que las interacciones genotipo-ambiente fueron importantes para el estudio de las características relacionadas con el encamado de las plantas. Además, este resultado podría verse influenciado por las diferencias entre las dos temporadas, ya que tanto AP como AM mostraron valores promedio más altos en la primera temporada, lo que indica mejores condiciones de cultivo. Asaro et al. (2016) mencionó que las diferencias locales del suelo pueden favorecer la identificación diferencial de QTL entre dos temporadas agrícolas, lo que sugiere que las asociaciones identificadas pueden estar respondiendo a componentes como la precipitación, la temperatura u otros factores climáticos. La falta de estabilidad entre los QTL detectados para las características de encamado se puede atribuir a las variaciones ambientales que ocurren durante las temporadas, lo que demuestra la importancia de realizar pruebas longitudinales en el tiempo y/o en diferentes ambientes (García et al. 2009). Según Edmeades (2013), en muchas poblaciones de maíz existen eventos aleatorios que pueden causar interacciones genotipo-año que son difíciles de interpretar. Por esta razón, para obtener asociaciones entre la estabilidad genética y las altas variaciones fenotípicas explicadas, el mismo material genético debe evaluarse durante diferentes estaciones o en diferentes ambientes.

Las temporadas agrícolas en las que se evaluaron las líneas de maíz endogámicas no afectaron la estabilidad de las asociaciones encontradas para LA, lo cual está de acuerdo con estudios relacionados (Chen et al. 2015; Liu et al. 2016). El maíz puede cultivarse en una amplia gama de condiciones, pero las variables de estrés abiótico, como la temperatura, la fertilidad del suelo y los factores climáticos, afectan su producción (Ramirez-Cabral et al. 2017). Por lo tanto, la detección de loci estables durante diferentes temporadas de crecimiento, puede ser muy útil

para los programas de mejoramiento que buscan identificar las mejores variedades capaces de adaptarse a los cambios ambientales.

5. CAPÍTULO II: MAPEO ASOCIATIVO DE GENOMA AMPLIO BASADO EN HAPLOTIPOS

5.1 Introducción

Actualmente, el desarrollo de métodos de genotipado que utilizan matrices de alta densidad, como el genotipado por secuenciación, ha facilitado la identificación de QTL para diferentes características de interés en varios cultivos, incluido el maíz (Li et al. 2016, Abdel et al. 2016, Coan et al. 2018), y el estudio de genes adyacentes, utilizando estudios de asociación de genoma amplio (Contreras-Soto et al. 2017, Elshire et al. 2011, Yu et al. 2006). En este sentido, el uso de polimorfismos de un solo nucleótido ha demostrado ser una herramienta útil para los programas de mejoramiento de diferentes cultivos agrícolas (Yu et al. 2014, Mora et al. 2015, Coan et al. 2018, Li et al. 2016). En este sentido, el uso de haplotipos, definidos a partir de arreglos de SNP con alto desequilibrio de ligamiento, ha surgido como una variante metodológica en la identificación de regiones genómicas y/o en la prospección de genes candidatos de GWAS. Este enfoque de análisis considera la dependencia natural que existe entre los SNP (Clark 2004), que se vuelve más relevante cuando se consideran arreglos de ADN de alta densidad.

Un haplotipo se define como un conjunto de SNP cercanos (polimórficos), con un fuerte desequilibrio de ligamiento entre ellos (Contreras-Soto et al. 2017). Desde este punto de vista, es importante tener en cuenta que algunos SNP tienen un fuerte desequilibrio de ligamiento (Clark 2004) y, por lo tanto, el desarrollo de GWAS basado en bloques de haplotipos puede representar una alternativa más eficiente que el uso de marcadores individuales (Xiao et al. 2017). Además, el uso de haplotipos puede compensar la limitación bi-alélica de los SNP, mejorando la eficacia de la identificación de QTLs (Contreras-Soto et al. 2017) y proporcionando conocimientos sobre los determinantes genéticos que no pueden ser capturados

por enfoques de marcadores individuales o independientes (Abdel-Shafy et al. 2014). Por ejemplo, Chen et al. (2016) y Contreras-Soto et al. (2017) detectaron haplotipos asociados con la resistencia a la pudrición de las mazorcas de maíz (Chen et al. 2016) y diferentes características agronómicas en la soja (altura de la planta, rendimiento de las semillas y peso de 100 semillas) (Contreras-Soto et al. 2017), las cuales no fueron capturadas por marcadores SNP individuales. Por lo tanto, con la intención de compensar las limitaciones de los análisis de asociación basados en marcadores de SNP individuales (Contreras-Soto et al. 2017), en el presente capítulo se propusieron los siguientes objetivos: (i) determinar y caracterizar bloques de haplotipos basados en un panel de SNP de alta densidad, (ii) identificar las asociaciones haplotipo-característica utilizando GWAS para características relacionadas a la arquitectura de la planta (AF) y el encamado (AP, AM y la relación AM/AP); (iii) comparar los resultados de GWAS basados en haplotipos con su equivalente GWAS basado en marcadores de SNP individuales; (iv) determinar la estabilidad de las asociaciones durante dos temporadas agrícolas consecutivas; y (v) examinar las regiones genómicas que flanquean las asociaciones genéticas (haplotipos y SNP) que fueron determinadas por GWAS.

5.2 Materiales y métodos

5.2.1 Desequilibrio de ligamiento (LD)

El análisis de LD se realizó en TASSEL 5.2 (Bradbury et al. 2007), utilizando SNP que tenían menos del 25% de datos faltantes y un MAF > 0.05. El LD entre pares de marcadores se estimó utilizando los coeficientes de correlación de las frecuencias alélicas (r^2) considerando todas las combinaciones posibles de los alelos. El nivel de significancia (p-value) se calculó utilizando 10.000 permutaciones.

El valor crítico de r^2 se estimó de acuerdo con Breseghello y Sorrells (2006) y Laidò et al. (2014), utilizando la transformación de la raíz cuadrada de los valores de r^2 y tomando el percentil 95 de estos datos como un umbral en el cual es probable que el LD sea causado por una ligación física real. Finalmente, los patrones de LD se evaluaron mediante una regresión no lineal entre los valores de r^2 y la distancia física de los SNP (en pares de bases) (Marroni et al. 2011).

5.2.2 Bloques de haplotipos

Se construyeron bloques de haplotipos para cada cromosoma, considerando los SNP con un MAF mayor a 0.05 y con menos del 25% de datos faltantes. La formación de los bloques se realizó utilizando el método de intervalos de confianza, descrito por Gabriel et al. (Gabriel et al. 2002) e implementado en HAPLOVIEW (Barrett et al. 2004). Este método considera los intervalos de confianza del 95% de los valores de D' , clasificando el LD como "LD fuerte", "no concluyente" o "recombinación fuerte". El D' entre dos loci, fue definido como sigue:

$$D'_{AB} = D/D_{MAX}$$

donde D y D_{MAX} son calculados como

$$D = P_{A_1B_1}P_{A_2B_2} - P_{A_1B_2}P_{A_2B_1}$$
$$D_{MAX} = \begin{cases} -\min\{P_{A_1}P_{B_1}, P_{A_2}P_{B_2}\}, & \text{cuando } D > 0 \\ \min\{P_{A_1}P_{B_2}, P_{A_2}P_{B_1}\}, & \text{cuando } D \leq 0 \end{cases}$$

Finalmente, se construyeron bloques si el 95% de las comparaciones eran del tipo "LD fuerte". Los bloques de haplotipos se transformaron posteriormente en marcadores multialélicos (con un scripts desarrollado en Perl, ver anexo II), para los análisis de asociación haplotipo-característica posteriores, con respecto a las combinaciones alélicas dentro de cada bloque como alelos independientes.

5.2.3 Genoma de referencia

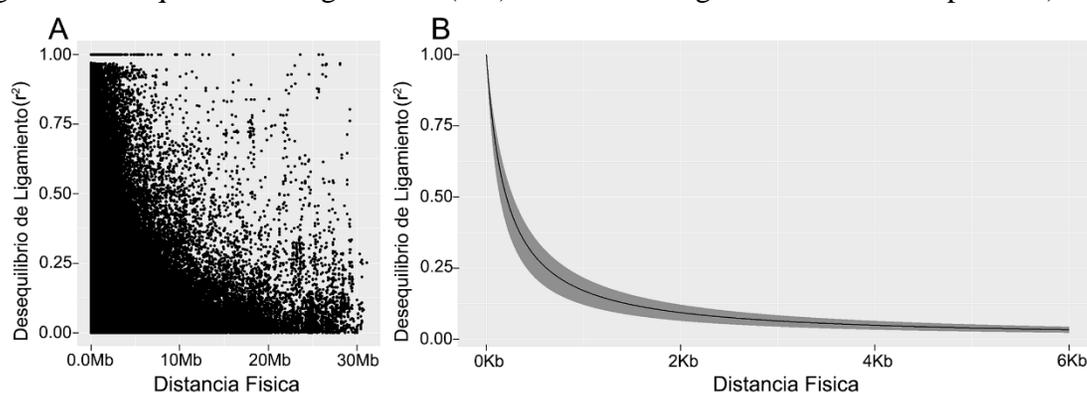
Los genes y QTLs registrados en el genoma de maíz AGPV3 (B73 RefGen_v3), disponible en la base de datos MaizeGDB (<http://www.maizegdb.org/>) (Harper et al. 2011), se utilizaron como referencias. Para esto, se estableció una ventana (o umbral) del doble de la distancia indicada por el LD (0.9 kb, definido en este estudio, más el tamaño del loci), colocando el SNP o haplotipo en el centro de la ventana.

5.3 Resultados

5.3.1 Desequilibrio de ligamiento (LD)

El patrón de LD se estimó a nivel del genoma completo y para cada cromosoma, considerando una alta densidad de marcadores SNP (38K). El LD presentó una rápida disminución (Fig. 3), mostrando variaciones entre los diferentes cromosomas, siendo el cromosoma 7 el que disminuye más rápidamente, y el cromosoma 5 más lentamente. El valor umbral (r^2) en el cual el LD es probable que sea causado por una ligación física real (Brescaglio y Sorrells 2006) fue de 0.19 para todos los cromosomas. El LD varió de 0.6 a 0.7 kb en los cromosomas 7 y 10, de 0.8 a 0.82 kb en los cromosomas 1 y 2, de 0.84 a 0.92 kb en los cromosomas 3 y 9, y de 1.01 a 1.2 kb en los cromosomas 4, 5, 6 y 8. El LD del genoma completo fue de 0.90 kb, el cual coincide con el promedio de todos los cromosomas (0.92 kb).

Figura 3. Desequilibrio de ligamiento (LD) en líneas endogámicas de maíz tropical. A) LD entre



todos los pares de SNP (38K) (r^2) y su distancia (Mega bases). B) Patrón de LD en los 10 cromosomas, con los cromosomas 7 y 5 teniendo los niveles de decaimiento más rápida y lenta (margen inferior y superior, respectivamente).

5.3.2 Haplotipos asociados con características complejas

Se identificó un total de 7.831 haplotipos (con 2 a 20 SNP cada uno) en los 10 cromosomas de maíz. De estos haplotipos, el 55% contenía dos SNP, el 23% contenía tres SNP, el 10% contenía cuatro SNP y el 12% contenía de cinco a 20 SNP. A su vez, el análisis de asociación del genoma amplio basado en haplotipos identificó 40 bloques que se asociaron significativamente con las características de interés, de los cuales 6 se asociaron con AM, 11 con AF, 15 con AP y 8 con la relación AM/AP (Tabla 4) (Fig. 4). Tres haplotipos (*hapAM5.07*,

hapAM7.02 y *hapAM8.06*) se asociaron concomitantemente con AM y AP (en el mismo año de medición), lo que sugiere un posible efecto pleiotrópico. Además, el 83% (5/6) de los haplotipos detectados para AF estuvieron presentes en ambas estaciones de crecimiento, lo que sugiere la posible estabilidad de estas asociaciones.

Tabla 4. Resumen de los haplotipos asociados con las características de ángulo de la hoja (AF), altura de la primera mazorca (AM), altura de la planta (AP) y la relación AM/AP, en líneas endogámicas de maíz tropical durante dos temporadas de crecimiento (2014- 2015 y 2015-2016).

Característica	2014-2015			2015-2016		
	N ^o _h	Cromosoma (N ^o _h)	PV (%)	N ^o _h	Cromosoma (N ^o _h)	PV (%)
Arquitectura						
AF	6	1(1), 4(3), 5(1) y 7(1)	16-29	5	4(3), 5(1) y 7(1)	18-28
Encamado						
AP	7	1(3), 5(1), 6(1), 8(1) y 9(1)	8-17	8	5(2), 6(1), 7(2), 8(2) y 10(1)	8-17
AM	5	4(1), 5(1), 8(2) y 10(1)	9-18	1	7(1)	15
AM/AP	4	1(1), 4(1), 7(1) y 10(1)	8-16	4	1(1), 3(1), 7(1) y 8(1)	10-16

PV (%): porcentaje de la variación fenotípica explicada por los bloques de haplotipos; N^o_h: número de asociaciones significativas basadas en haplotipos.

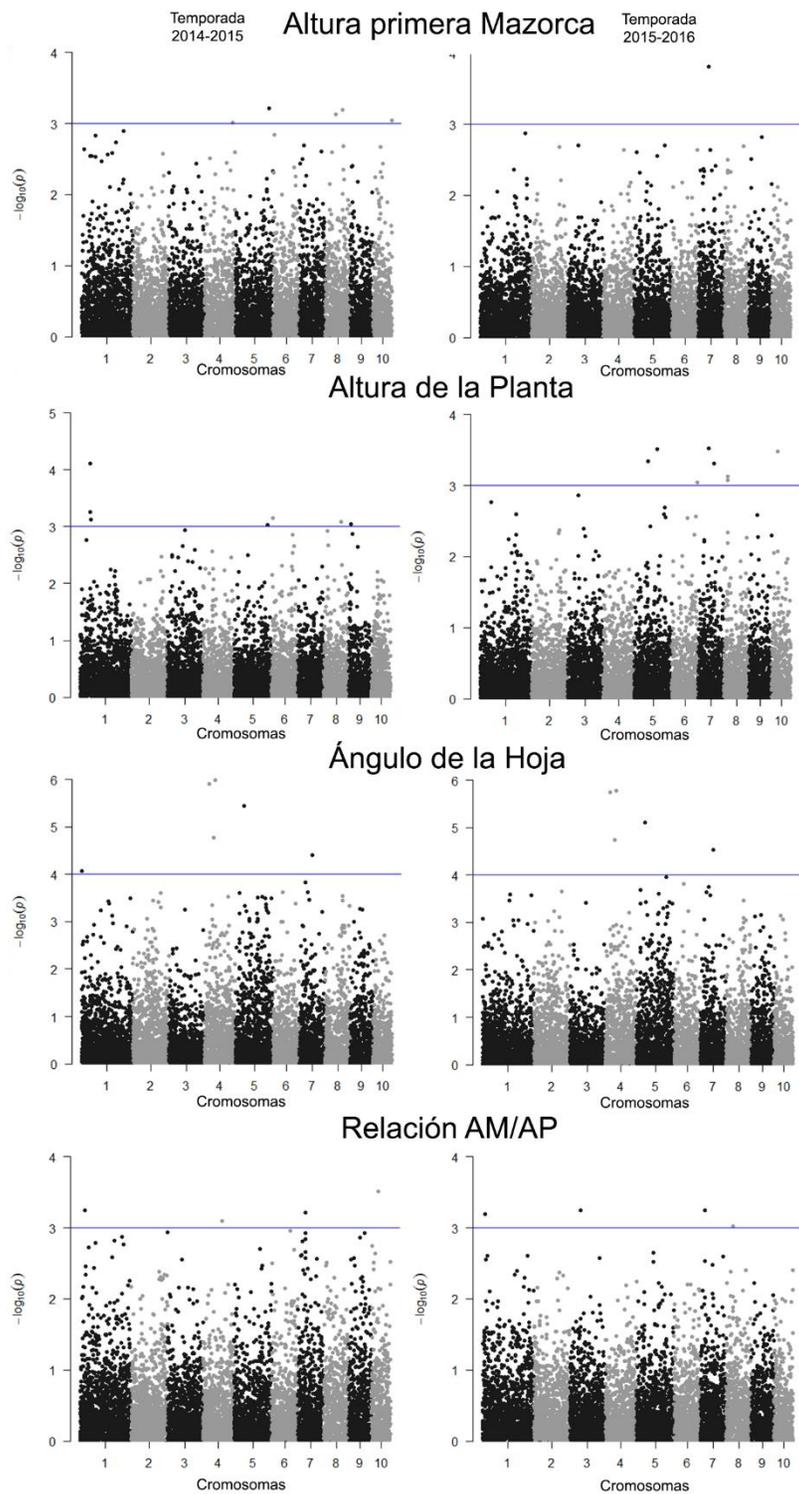


Figura 4. Gráficos de Manhattan de la significancia estadística de los haplotipos asociados con las características: ángulo de la hoja (AF), altura de la primera mazorca (AM), altura de la planta (AP) y relación AM/AP, evaluados en un panel de asociación de maíz tropical durante las

temporadas 2014-2015 y 2015-2016. El $-\log_{10}$ (p-values) de los haplotipos se trazaron para todos los cromosomas en su conjunto. Los haplotipos significativamente asociados con las características están por encima de la línea de umbral (p-value < 0.001).

Los haplotipos *hapAM10.07* y *hapAM4.09* mostraron diferencias superiores al 40% entre sus variantes alélicas para la característica AM. En efecto, los individuos que presentaron el alelo "b" del haplotipo *hapAM10.07* tuvieron una altura de la primera mazorca 62% y 93% más alta que la de los individuos con los alelos "a" y "c", respectivamente. Por otro lado, los individuos con el alelo "d" del haplotipo *hapAM4.09* presentaron AM 43%, 88% y 48% más altos que aquellos de los individuos con los alelos "a", "b" y "c", respectivamente. Estos resultados muestran la capacidad de estos haplotipos para diferenciar a los individuos con mayor altura de la primera mazorca. Para los loci con posibles efectos pleiotrópicos, los individuos que presentaron los alelos "a" y "c" del haplotipo *hapAM7.02-hapAP7.02* mostraron diferencias significativas en comparación con los individuos con el alelo "b", para ambas características (AM y AP) en la temporada 2015-2016. De hecho, las plantas que presentaron los alelos "a" y "c" (haplotipo *hapAP7.02*) eran 19% y 23% más altas, respectivamente, que las de los individuos con el alelo "b", y tuvieron 29% y 42% más AM (*hapAM7.02*), respectivamente, que los individuos con el alelo "b". Además, los bloques *hapAM5.07-hapAP5.07* y *hapAM8.06-hapAP8.06* también presentaron diferencias entre sus variantes alélicas para ambas características (AM y AP) durante la temporada de crecimiento 2014-2015. Las plantas que presentaron el alelo "b" en el haplotipo *hapAP5.07* fueron 15% y 22% más altas que los individuos con las variantes alélicas "a" y "c", mientras que para AM, las plantas que presentaron el alelo "b" (*hapAM5.07*) tuvieron 22% y 33% más AM que los individuos con los alelos "a" y "c". Del mismo modo, los individuos con los alelos "c" y "d" del haplotipo *hapAP8.06* eran 10% a 14% más altos y tuvieron una AM de 9% a 21% más alto que los de los individuos con los alelos "a" y "b".

Un total del 38% de los haplotipos significativamente asociados con una determinada característica contenía uno o más SNP que se detectaron previamente durante el análisis de asociación utilizando marcadores SNP individuales. Este resultado muestra que el análisis de

haplotipos permitió la identificación de regiones genómicas que no fueron detectadas por GWAS con el uso de marcadores individuales.

La variación fenotípica explicada por los haplotipos varió entre 16% y 29% para AF, entre 8% y 17% para AP, entre 9% y 18% para AM, y entre 8% y 16% para la relación AM/AP (Tabla 4). Al igual que en el análisis con SNP individuales, la característica AF presentó los valores más altos de r^2 (Tabla 5), en el cual dos haplotipos pueden explicar más del 23% de la variación fenotípica (haplotipos *hapAF4.04* y *hapAF4.05B*, ubicados en el cromosoma 4; Tabla 5). Estos valores son ligeramente más altos que los observados en el análisis de SNP individuales. Aunque los haplotipos abarcan una región genética más grande (debido a la agrupación de SNP), el 45% de éstos no aumentó la variación fenotípica explicada en comparación con los SNP que los componen; por ejemplo, el haplotipo *hapAP9.02* explicó el 8% de la variación fenotípica, mientras que la suma de sus SNP explicó el 14%. Sin embargo, el 55% de los haplotipos explican una variación fenotípica mayor que la suma de los SNP que los componen; por ejemplo, el haplotipo *hapAF4.05B* explica el 23% de la variación, mientras que la suma de sus 3 SNP solo explica el 6%. Este último ejemplo muestra el potencial presentado de los estudios GWAS basado en bloques de haplotipos sobre GWAS basado en SNP individuales, aumentando la fracción de la variación fenotípica explicada por un marcador.

Tabla 5. Asociaciones basadas en haplotipos con efectos relativamente importantes (>15%) para las características de ángulo de la hoja (AF), altura de la mazorca (AM), altura de la planta (AP)

y la relación AM/AP, medida en líneas endogámicas de tropical maíz durante dos temporadas agrícolas.

Temporada	Característica	Marcador	Chr	Posición (pb)	Bin	p-value	PV(%)
2014-2015	Arquitectura						
	AF	<i>hapAF1.01</i>	1	2993589-2993592	1.01	8.5E-05	16.3%
	AF	<i>hapAF7.03</i>	7	132216972-132217768	7.03	4.0E-05	17.0%
	AF	<i>hapAF4.05A</i>	4	74484290-74484291	4.05	1.7E-05	18.3%
	AF	<i>hapAF5.03</i>	5	30796468-30796505	5.03	3.5E-06	19.9%
	AF	<i>hapAF4.05B</i>	4	118922762-118922809	4.05	1.0E-06	23.3%
	AF	<i>hapAF4.04</i>	4	26450981-26451472	4.04	1.2E-06	28.8%
	Encamado						
	AP	<i>hapAP6.00</i>	6	8145260-8148648	6.00	7.0E-04	15.0%
	AP	<i>hapAP8.06</i>	8	155471554-155472264	8.06	8.1E-04	17.4%
	AM	<i>hapAM8.06</i>	8	155471554-155472264	8.06	6.5E-04	18.0%
	AM	<i>hapAM4.09</i>	4	235925519-235927228	4.09	9.8E-04	18.4%
	AM/AP	<i>hapAM/AP10.03</i>	10	79150221-79159436	10.03	3.1E-04	15.9%
2015-2016	Arquitectura						
	AF	<i>hapAF7.03</i>	7	132216972-132217768	7.03	2.9E-05	17.8%
	AF	<i>hapAF4.05A</i>	4	74484290-74484291	4.05	1.8E-05	18.1%
	AF	<i>hapAF5.03</i>	5	30796468-30796505	5.03	7.9E-06	18.4%
	AF	<i>hapAF4.05B</i>	4	118922762-118922809	4.05	1.6E-06	22.8%
	AF	<i>hapAF4.04</i>	4	26450981-26451472	4.04	1.8E-06	28.3%
	Encamado						
	AP	<i>hapAP7.02</i>	7	127179852-127179870	7.02	3.0E-04	15.2%
	AP	<i>hapAP10.03</i>	10	73141503-73141785	10.03	3.4E-04	16.0%
	AP	<i>hapAP6.07</i>	6	166163113-166163458	6.07	9.0E-04	17.0%

AM	<i>hapAM7.02</i>	7	127179852- 127179870	7.02	1.5E- 04	15.1%
AM/AP	<i>hapAM/AP1.01</i>	1	9391163-9391230	1.01	6.4E- 04	16.2%

PV(%): porcentaje de la variación fenotípica explicada por bloques de haplotipos; Cr: cromosoma al cual pertenece el loci; Bin: o “contenedor”, indica una posición en el mapa genético con un patrón de segregación único, y que está separado de los contenedores adyacentes por un único evento de recombinación (Van et al. 2006).

5.3.3 Anotación de genes basada en SNP y haplotipos individuales

Según la posición física (genoma de referencia B73) de los SNP y haplotipos significativamente asociados, según GWAS, 122 SNP y 31 haplotipos están cerca de genes candidatos. Veinte de estos SNP y ocho haplotipos se asociaron con más de una característica o una temporada a la vez. Además, algunos SNP y haplotipos están cerca de los mismos genes candidatos. Por lo tanto, solo se encontraron 95 genes candidatos únicos en el presente análisis. Seis de estos genes candidatos son de particular interés debido a su probada participación en la expresión de las características de interés (*GRMZM2G141386*, *GRMZM2G130675*, *GRMZM2G104262*, *GRMZM5G845755*, *AC205122.4_FG003* y *GRMZM2G042429*).

Para AF, el gen *GRMZM2G104262* tiene un ortólogo en *Arabidopsis thaliana* que codifica el cryptochrome 1 (CRY1), el cual, en longitudes de onda bajas de la luz azul, establece la inducción rápida del crecimiento hiponástico del pecíolo (Millenaar et al. 2009), independiente del etileno (Millenaar et al. 2009, Kozuka et al. 2005). Además, Wu y Yang (2010) demostraron que la sobreexpresión de *CRY1* regula positivamente la resistencia adquirida sistémica (SAR; *Systemic Acquired Resistance*) y la expresión de la proteína R, que media la resistencia a *Pseudomonas syringae*, limitando el crecimiento de bacterias en *Arabidopsis*.

Para AM, el gen *AC205122.4_FG003* codifica para una proteína no caracterizada en el maíz; sin embargo, su homólogo en *Arabidopsis* codifica la proteína C2-domain ABA-related 5 (CAR5), la cual está relacionada con la sensibilidad al ácido abscísico (ABA) (Rodriguez et al. 2014). De hecho, los mutantes CAR5 mostraron una sensibilidad reducida a ABA y un rápido crecimiento (Rodriguez et al. 2014). Además, el gen *GRMZM5G845755* codifica la casein

kinase II subunit alpha-4 (CKA4), un componente de la proteína quinasa CK2 (un complejo tetramero estructural), que tiene propiedades muy similares a las descritas en otras plantas (Riera et al. 2001). Lee y col. (1999) demostraron que esta proteína parece estar involucrada en el crecimiento de las plantas. Además, Wang et al. (2016) reportaron que un mutante knockout de *CKA4* exhibe defectos tanto en el desarrollo como en el alargamiento del hipocotilo, retrasando el crecimiento general de *Arabidopsis*.

Para las asociaciones relacionadas a AP, el gen *GRMZM2G141386* codifica una putativa RNA-binding protein (ARP1). Makabe et al. (2017) caracterizaron esta proteína como un activador de resistencia a enfermedades y un represor del crecimiento de plantas en *Nicotiana tabaccum*. Los genes *GRMZM2G042429* (asociaciones cercanas de loci-AP) y *GRMZM2G130675* (relación AM/AP) expresan una small auxin-up RNA 52 (SAUR52) y SAUR1, respectivamente, dos miembros de la familia SAUR, la cual está compuesta de genes con respuestas tempranas a auxina y son efectores clave de las señales hormonales y ambientales que regulan el crecimiento y desarrollo de las plantas (Ren y Gray 2015). En la soja y *Arabidopsis thaliana*, estas proteínas pueden promover el alargamiento del hipocótilo cuando son estimuladas por las auxinas (Hagen y Guilfoyle 2002, Chae et al. 2012).

5.4 Discusión

5.4.1 Haplotipos y regiones genéticas asociadas con características complejas.

La rápida disminución del LD es característico de los germoplasmas de maíz tropical (Coan et al. 2018, Romay et al. 2013). Remington et al. (2001) y Yan et al. (2009) reportaron que el LD disminuye dentro del rango de 0.1 a 10 kb, dependiendo de la población y la región genética estudiada, y siempre disminuye más rápido en maíz tropical (Romay et al. 2013). En este sentido, el tamaño de los haplotipos está estrechamente relacionado con el grado de LD presente en la población estudiada (Slatkin 2008), en el cual una rápida disminución produce haplotipos más pequeños porque se vincularán un menor número de loci (Slatkin 2008, Zhu et al. 2008). La mayoría de los bloques de haplotipos se formaron con dos o tres SNP (78%), debido a la rápida disminución del LD (0.9 kb). Lorenz et al. (2010) y Contreras-Soto et al.

(2017) demostraron que GWAS basado en haplotipo (H-GWAS) puede ser beneficioso al identificar asociaciones marcador-fenotipo y puede ofrecer ventajas para la disección genética de los loci subyacentes a características complejas. Además, H-GWAS reduce el número de comparaciones múltiples (o pruebas múltiples), en comparación con el análisis de asociación basado en SNP-individuales, ya que los haplotipos pueden agrupar SNP a partir del patrón LD observado en los datos. La reducción de la dimensionalidad de las pruebas múltiples, desde una perspectiva biológica (haplotipos) se torna más relevante cuando se consideran las nuevas plataformas de datos genómicos (datos genómicos de alta densidad), lo que aumenta la posibilidad de encontrar regiones genómicas que controlen la variación de una característica.

Un total de 38% de los haplotipos significativamente asociados contuvo uno o más SNP que se detectaron previamente en el GWAS, lo que indica una eficiencia similar a GWAS basado en SNP-individuales en la detección de regiones asociadas a las características complejas. Más aún, 25 nuevas regiones (62%) fueron identificadas mediante H-GWAS, confirmando la ventaja que tienen los haplotipos en la detección de múltiples variantes de ADN (Contreras-Soto et al. 2017). Los umbrales seleccionados de SNP-GWAS y H-GWAS se basaron en el método de corrección de Bonferroni, estos umbrales fueron equivalentes a los p-valores de 2.6×10^{-6} y 1.3×10^{-5} , respectivamente. El método de corrección basado en FDR indicó que los loci presentaban un valor de probabilidad (p-valor) corregido con valor de 1. Aunque los loci identificados no superaron los umbrales de Bonferroni y de FDR, los p-valores en ambos análisis fueron significativos (p-valores $< 1 \times 10^{-3}$). Vale la pena señalar que el 16% de las variantes de ADN se detectaron consistentemente en ambos métodos, confirmando la alta confiabilidad de los loci detectados. En este estudio, los haplotipos asociados explican una mayor proporción de la variabilidad fenotípica de una característica determinada en comparación con los SNP que lo componen. Por ejemplo, los haplotipos *hapAF4.05B* y *hapAM4.09* explicaron el 23% y el 18% de la variación fenotípica para AF y AM, respectivamente, mientras que la suma de sus 3 SNP solo representó el 6% y el 4%, respectivamente. Además, el haplotipo *hapEH4.09* mostró diferencias significativas (hasta 88%) entre sus variantes alélicas. Este resultado indica la utilidad potencial de los estudios de asociación basado en haplotipos para explicar la variación fenotípica asociada con una característica, un aspecto destacado por Contreras-Soto et al. (2017), Chen et al. (2016) y Barendse (2011).

En el presente estudio, los SNP y los haplotipos asociados con AP, AM y la relación AM/AP se distribuyeron en todos los cromosomas, confirmando un tipo de arquitectura poligénica, mientras que las asociaciones con AF se encontraron en los cromosomas 1, 2, 4, 5, 6, 7 y 8. Varios estudios en maíz que utilizaron poblaciones biparentales y líneas endogámicas recombinantes para el mapeo asociativo, han reportado asociaciones en sitios similares al del presente estudio. Weng et al. (2011), por ejemplo, encontró 27 QTL en el bin 1.03, 6 asociados con AP y 21 con AM. Además, Li et al. (2016) informó tres QTL en el bin 2.07 asociados con AM, y Pan et al. (2017) identificaron cuatro QTL, tres asociados con AP (en los bins 3.09, 8.06 y 9.03) y uno con AM (bin 6.06).

En el presente estudio, se identificaron varios *hotspots* (puntos calientes o claves) que contienen más de tres SNP o haplotipos, los cuales están presentes en los bins cromosómicos 1.03, 9.03 y 10.03 para AP, los bins 4.08, 7.00, 7.02, 8.04 y 8.06 para AM, bins 4.05 y 5.03 para la característica del AF, y bins 1.01, 4.09, 7.02 y 8.03 para la relación AM/AP. Además, los SNPs S2_47326011 y S1_639521797 que se asociaron concomitantemente con AP y AM se ubicaron en los bins 1.03 y 8.06, respectivamente, los cuales son hotspots relacionados con el encamado de las plantas en maíz. A nivel de haplotipos, *hapAM8.06-hapAP8.06*, que también se asocia concomitantemente con las características AM y AP, se encuentra en el bin 8.06. Casualmente, Weng et al. (2011) encontraron QTL compartidos entre AM y AP en el bin 1.03, sugiriendo que estas regiones pueden contener mecanismos genéticos que controlan ambas características.

Hasta la fecha, se han reportado varios estudios de mapeo de QTL para el encamado y la arquitectura de plantas (Ku et al. 2010, Wang et al. 2017, Zhang et al. 2014, Zhang et al. 2006, Zhu et al. 2013). Sin embargo, solo unos pocos han informado asociaciones con efectos relativamente importantes. En este estudio, un total de 9 SNP y 14 haplotipos vinculados a QTLs pudieron explicar un porcentaje significativo de la variación fenotípica (> 15%) en las características de encamado y arquitectura de la planta (Tabla 3 y Tabla 5). Estas asociaciones se encontraron en los cromosomas 1, 2, 4, 5, 6, 7, 8 y 10, lo que concuerda con los estudios reportados previamente. Para las características relacionados con el encamado de las plantas, varios estudios han identificado loci con efectos importantes; por ejemplo, Zhang et al. (2006) identificaron un QTL que explica gran parte de la variación fenotípica de AP en el cromosoma

4, mientras que Zhu et al. (2013) encontraron QTL de efecto mayor para AM y AP en los cromosomas 1, 8, 9 y 10. En este estudio, un SNP (S1_81113660) y tres haplotipos (*hapEH8.06*, *hapPH8.06* y *hapPH10.03*) de efecto mayor coinciden con los QTLs descritos por Zhu et al. (2013) en líneas doble haploides con marcadores microsatélite (SSR). Por otro lado, para AF, se han identificado QTL que contribuyen en gran parte a la variación fenotípica; por ejemplo, Ku et al. (2010) identificaron dos QTL que explican más del 17% de la variación. A su vez, Wang et al. (2017) reportaron solo un marcador de efecto mayor en el cromosoma 7, y Zhang et al. (2014) encontraron un QTL que explica el 36.82% de la variación fenotípica. Entre los QTL de efecto mayor identificados en este estudio, el haplotipo hapLA1.01 (explica el 16.3% de la varianza fenotípica) y SNP S1_1296983734 (explica el 16.8% de la varianza fenotípica) son adyacentes a los QTL de efecto mayor encontrados por Ku et al. (2010) en los cromosomas 1 y 2, respectivamente. Por otra parte, el haplotipo hapAF7.03 (explica el 17% de la varianza fenotípica) es adyacente a qLAa7-1, descrito por Wang et al. (2017). De igual forma, dos SNPs, S1_1648381508 (que explica el 17.8% de la varianza fenotípica) y S1_1550828517 (18.6%), y tres haplotipos, hapAF4.05A (que explica el 18.1% de la varianza fenotípica), hapAF4.05B (22.8%) y hapAF4.04 (28.3%), se ubicaron adyacentes al QTL qLA4-1 (36.82%), detectado por Zhang et al. (2014), un hotspots que explica gran parte de la variación fenotípica de AF. Es importante destacar que los QTL de efecto mayor identificados para AF (hapAF4.05B = 23.3% y hapAF4.04 = 28.8%) se ubicaron en el cromosoma 4, lo cual es consistente con estudios previos (Zhang et al. 2014; Ku et al. 2011). Este resultado confirma que el cromosoma 4 desempeña un papel clave en la variación de AF y debe considerarse en los programas de mejoramiento de maíz.

Los resultados del presente trabajo indican que las características de interés (relacionadas al encamado y arquitectura de las plantas) son controladas por múltiples genes, con contribuciones variables en la expresión fenotípica. En este trabajo, se presentan los primeros haplotipos de efectos mayores asociados con el ángulo de la hoja, que representan un aspecto fundamental para la arquitectura del maíz y la selección asistida por marcadores moleculares, y que presentan estabilidad entre las temporadas de crecimiento. En general, los haplotipos que están significativamente asociados con características relacionadas al encamado y la arquitectura de la planta explicaron una mayor proporción de la variación fenotípica, en

comparación con el análisis SNP-GWAS, lo que respalda la hipótesis de que los estudios de asociación basados en haplotipos permiten la identificación de regiones genómicas responsables del control de una gran parte de la variación fenotípica, en las características de interés.

5.4.2 Datos anotados del genoma

El maíz es un organismo ideal para el estudio genético (como por ejemplo GWAS) de paneles de líneas endogámicas (Coan et al. 2018) debido a que posee un genoma de referencia de alta calidad que permite el estudio y la disección de la base genética de características complejas (Coan et al. 2018, Schnable et al. 2009). Para la identificación de anotaciones genómicas relacionadas a las características de interés, consideramos una ventana de 0.9 kb (determinada por el patrón LD) río arriba y río abajo de las asociaciones significativas (SNP individuales y haplotipos). En base a este principio, se encontraron 126 genes candidatos únicos, de los cuales solo seis se consideraron biológicamente importantes (putativos) debido a su probada participación en las características estudiadas, los cuales corresponden a: *GRMZM2G141386*, *GRMZM2G130675*, *GRMZM2G104262*, *GRMZM5G845755*, *AC205122.4_FG003* y *GRMZM2G042429* (Millenaar et al. 2008, Wu y Yang 2010, Rodriguez et al. 2014, Riera et al. 2001, Lee et al. 1999, Wang et al. 2016, Makabe et al. 2017, Ren y Gray 2015, Hagen y Guilfoyle 2002, Chae et al. 2012).

AF presentó un gen ortólogo a *GRMZM2G104262* en *Arabidopsis*, que codifica CRY1. De acuerdo con Millenaar et al. (2009), se ha informado que CRY1 tiene dos funciones positivamente correlacionadas con la intensidad de la luz: la inducción rápida del crecimiento hiponástico del pecíolo a bajos niveles de baja longitud de onda de luz azul, y la regulación positiva de la resistencia a *Pseudomonas syringae* (Wu y Yang 2010). A su vez, Joshi y Chand (2002) mostraron una correlación positiva entre AF y la resistencia a tizón foliar, causada por *Bipolaris sorokiniana* en el trigo, en el cual, individuos con hojas rectas o semi rectas presentan una menor incidencia de la enfermedad en comparación con los individuos con hojas caídas. Estos resultados justificarían la doble función observada en CRY1 y sugerirían que AF puede influir en la incidencia de una enfermedad. Por lo tanto, sería necesario realizar investigaciones adicionales que profundicen el estudio de la estructura genética de AF en el maíz, incluyendo su grado de asociación con diversas enfermedades.

Los genes *AC205122.4_FG003* y *GRMZM5G845755* se encuentran en regiones genómicas asociadas con AM y codifican polipéptidos de función desconocida. Sin embargo, estos genes tienen ortólogos en *Arabidopsis* que expresan las proteínas CAR5 y CKA4, las cuales se han asociado con el alargamiento del hipocotilo y el crecimiento de las plantas (Rodríguez et al. 2014, Lee et al. 1999, Wang et al. 2016). Del mismo modo, para AP y la relación AM/AP, los ortólogos para los genes *GRMZM2G042429* y *GRMZM2G130675*, respectivamente, codifican las proteínas SAUR52 y SAUR1 que pueden promover el alargamiento del hipocotilo a través de la estimulación con auxina (Hagen y Guilfoyle 2002, Chae et al. 2012). La auxina es una hormona clave del crecimiento de las plantas, que regula procesos tales como el alargamiento, división, diferenciación y morfogénesis celular durante el crecimiento y el desarrollo de la planta (Chen et al. 2018). Las proteínas SAUR participan en la respuesta temprana de la auxina y juegan un papel clave en las señales hormonales y ambientales que regulan el crecimiento y el desarrollo de las plantas (Ren y Gray 2015).

En este estudio, algunos haplotipos contenían SNP que no se detectaron en el GWAS basado en SNP individuales. Este resultado representa una ventaja de GWAS basado en haplotipos en la detección de múltiples variantes de ADN asociadas a una característica de interés. Otra ventaja de los estudios genómicos basados en haplotipos es la posibilidad de estudiar múltiples variantes alélicas, lo que podría facilitar la búsqueda de loci que afectan la expresión génica de cierta característica. Finalmente, tanto los haplotipos como los SNP identificados en este estudio podría aumentar la eficiencia de los programas de mejoramiento de maíz, basados en determinantes importantes del encamado y la arquitectura de las plantas.

6. CAPÍTULO III: EFECTO PLEIOTRÓPICO DE LOCI POLIMÓRFICOS

6.1 Introducción

La correlación genética entre características cuantitativas es un fenómeno común en los estudios genéticos, el cual posee importantes alcances en los programas de mejoramiento (Endler et al. 2018). Estas correlaciones pueden ser causadas por patrones de desequilibrio de ligamiento (LD) entre distintos loci asociados a las características, o efectos pleiotrópicos (Solovieff et al. 2013, Endler et al. 2018). Sin embargo, las correlaciones obtenidas a partir del LD son temporales, y la velocidad a la que se pierde el desequilibrio depende principalmente de la distancia entre las regiones genéticas y los eventos de recombinación entre las mismas (Balestre et al. 2012). En contraste, debido a que la pleiotropía es un fenómeno en el que el mismo locus controla diferentes características, su efecto es más estable (Endler et al. 2018, Balestre et al. 2012). Por lo tanto, la identificación de QTLs pleiotrópicos es esencial para distinguir entre las correlaciones genéticas derivadas de efectos del LD y las correlaciones originadas a partir de genes pleiotrópicos.

Estudios de asociación del genoma amplio son un poderoso enfoque para identificación de las regiones genéticas con efecto pleiotrópico (Solovieff et al. 2013, Endler et al. 2018). En maíz, diversos QTLs pleiotrópicos han sido identificados como responsables de la domesticación desde su antepasado teosinte (Lemmon y Doebley 2014, Wang et al. 2005, Studer et al. 2011). Por ejemplo, el factor de transcripción *tb1* juega un papel importante en el desarrollo de la planta, como: dominancia apical, longitud de las ramas laterales, crecimiento de las hojas en las ramas laterales, desarrollo de las espiguillas pediceladas y arquitectura de la raíz (Gaudin et al., 2014, Studer et al. 2011). Por consiguiente, comprender la pleiotropía puede, por ejemplo, ayudar a dilucidar los procesos evolutivos presentados en maíz.

En este capítulo se realizó un estudio pleiotrópico por medio de un enfoque bayesiano (regresión bayesiana multivariada) en loci asociados a las características de encamado en maíz, identificados previamente por el mapeo asociativo basado en SNP-individuales y haplotipos.

6.2 Materiales y métodos

6.2.1 Correlaciones genética y pleiotropía

Las correlaciones entre características se calcularon de acuerdo con el coeficiente de correlación de Pearson en el software R-project (R Core Team. R 2013), según:

$$r_{xy} = \frac{\hat{\sigma}_{G_{xy}}}{\hat{\sigma}_{G_x}^2 \cdot \hat{\sigma}_{G_y}^2}$$

donde $\hat{\sigma}_{G_{xy}}$ corresponden a la covarianza genotípica entre las características, y $\hat{\sigma}_{G_x}^2$, $\hat{\sigma}_{G_y}^2$ corresponden a las varianzas genotípicas de dos características del encamado: altura de la primera mazorca (AM) o altura de la planta (AP) o relación AM/AP.

Los residuos de un modelo lineal generalizado para cada característica se utilizaron como fenotipos ajustados para el análisis de regresión bayesiana multivariada en SNPTEST (Marchini y Band 2016) (teniendo en cuenta los efectos del bloque y la estructura genética) para corroborar el posible efecto pleiotrópico de un loci con más de una característica.

Se utilizó el Factor Bayes (BF) y la Probabilidad de Asociación Posterior (PPA) para evaluar la probabilidad de que un locus esté asociado conjuntamente a dos de las características de encamado (Stephens y Balding 2009). El BF es la razón entre las probabilidades marginales del modelo de asociación (M1) y un modelo nulo de no asociación (M0) (Marchini y Band 2016),

$$BF = \frac{P(Data|M_1)}{P(Data|M_0)}$$

donde las probabilidades marginales para M1 y M0 están definidas por:

$$P(Data|M_1) = \int \left(\prod_{i=1}^N \sum_{k=0}^2 P(\Phi|G_{ij} = k, \theta) p_{ijk} \right) P(\theta|M_1) d\theta$$

en el que θ denota los parámetros de regresión, G_i denota el genotipo del i -ésimo individuo en el j -ésimo marcador, Φ denota el fenotipo del i -ésimo individuo en un estudio de N muestras, y $p_{ijk} = P(G_{ij} = k)$ es la probabilidad de que el genotipo en el marcador j -ésimo del i -ésimo individuo es k (Marchini y Band 2016). La aproximación de Laplace se utilizó para estimar las probabilidades marginales de M1 y M0 (Marchini y Band 2016).

El PPA combina la evidencia en los datos de asociación observados (BF) con la probabilidad previa (π) de que un marcador esté asociado. El BF y π se usan para calcular las probabilidades posteriores (PO):

$$PO = BF \times \frac{\pi}{1 - \pi}$$

donde π es la probabilidad a priori de que un marcador dado esté asociado con una característica. BF se calculó utilizando un análisis de regresión bayesiano en el software SNPTEST (Marchini y Band 2016). El PPA se calculó a partir de este PO, de la siguiente manera:

$$PPA = \frac{PO}{1 + PO}$$

Además, se realizó un mapeo asociativo en el software TASSEL V3.0 y V5.0 (para haplotipos y SNP-individuales, respectivamente) (Bradbury et al. 2007), empleando el modelo lineal mixto unificado (Estructura poblacional + matriz de parentesco), para comparar los resultados (Yu et al. 2006).

6.3 Resultados

Las características AM y AP tuvieron una correlación alta y positiva (Tabla 6), lo que está de acuerdo con estudios previos (Li et al. 2016, Cai et al. 2012) (Tabla 6). Sin embargo, la relación AM/AP tuvo una mayor correlación con AM que con AP, lo que indica que la variación en la relación AM/AP está determinada principalmente por cambios en AM. Por lo tanto, es posible observar que, en las mismas regiones genómicas, se pueden encontrar marcadores que están significativamente asociados con AP y AM, y con la relación AM/AP y AM (Li et al. 2016, Ji-hua et al. 2007).

Tabla 6. Correlaciones fenotípicas entre las características evaluadas dentro y entre las temporadas agrícolas. La diagonal de la tabla indica la correlación genética de cada característica entre las dos temporadas agrícolas. Los datos sobre y bajo la diagonal, indican las correlaciones genéticas entre las características para las temporadas 2014-2015 y 2015-2016, respectivamente.

		Temporada 2014-2015			
		<i>AM</i>	<i>AP</i>	<i>AF</i>	<i>AM/AP</i>
Temporada 2015-2016	<i>AM</i>	0.473	0.783	0.079	0.751
	<i>AP</i>	0.714	0.525	-0.022	0.197
	<i>AF</i>	0.166	-0.022	0.989	0.151
	<i>AM/AP</i>	0.710	0.197	0.151	0.269

En el presente estudio, tres haplotipos y diez SNP mostraron asociaciones significativas con AP y AM concomitantemente, mientras que solo un SNP se asoció simultáneamente con la relación AM/AP y AM. Esto sugiere un control genético común para estas características, que podría dar lugar a posibles efectos pleiotrópicos para estos loci (Li et al. 2016).

La pleiotropía entre los loci asociados con las características de encamado en maíz fue corroborada por el estudio genómico multivariado Bayesiano, vía métodos de Monte Carlo y Cadenas de Markov (Marchini y Band 2016). Como PPA combina la evidencia en los datos de asociación observados (BF) con la probabilidad previa (π) de que un marcador esté asociado con una característica determinada, se consideró que los loci con valores de PPA > 0.9 están realmente asociados con un fenotipo dado. Este valor de PPA corresponde a un $\log_{10}(\text{BF}) > 3.95$, el cual es más riguroso que los estudios de asociación ya reportados (Berg et al. 2015, Legarra et al. 2015, Legarra et al. 2018).

Los valores de PPA (0.924 y 0.994), estimados para los SNP S2_230948259 y S1_81113660 (Tabla 7) proporcionan evidencia convincente de la asociación de estos dos loci con más de un fenotipo (AM y AP). Del mismo modo, los valores altos de $\log_{10}(\text{BF}) (> 3.95)$ pueden considerarse como una fuerte evidencia contra la hipótesis nula de no asociación. Este estudio bayesiano indica que solo dos loci presentaron un efecto pleiotrópico significativo.

Tabla 7. Estudio genómico de loci pleiotrópicos con sus respectivos valores de probabilidad (p-valores) para las características relacionadas con el encamado de maíz. Se muestran el factor Bayes (BF), las probabilidades posteriores (PO) y la correspondiente probabilidad de asociación posterior (PPA) para dos valores de la probabilidad a priori.

Loci	p-valores			Log ₁₀ (BF)	PO	PPA
	AM	AP	AM/AP			
S1_821562289 ^a	6.2×10 ⁻⁴	9.3×10 ⁻⁴	-	0.4	0.000	0.000
S1_494063499 ^a	3.4×10 ⁻⁴	6.2×10 ⁻⁵	-	0.9	0.008	0.008
S1_639521797 ^a	7.5×10 ⁻⁴	2.4×10 ⁻⁴	-	1.6	0.037	0.035
S1_1280061812 ^a	6.4×10 ⁻⁴	9.5×10 ⁻⁴	-	2.0	0.107	0.097
S2_230948259 ^a	5.5×10 ⁻⁵	1.7×10 ⁻⁵	-	4.1	12.15	0.924
SNP S2_47326011 ^b	1.8×10 ⁻⁴	5.4×10 ⁻⁴	-	0.8	0.007	0.007
S1_333289434 ^b	8.3×10 ⁻⁶	4.1×10 ⁻⁵	-	1.8	0.069	0.064
S1_1158043959 ^b	8.6×10 ⁻⁴	7.4×10 ⁻⁵	-	3.0	1.033	0.508
S1_1037703810 ^b	1.4×10 ⁻⁴	5.2×10 ⁻⁵	-	3.1	1.186	0.543
S1_81113660 ^b	2.2×10 ⁻⁴	5.5×10 ⁻⁵	-	5.2	153.93	0.994
S1_16427856 ^b	9.0×10 ⁻⁴	-	7.6×10 ⁻⁴	-0.3	0.000	0.000
<i>hapAM5.07-hapAP5.07</i> ^a	6.1×10 ⁻⁴	9.2×10 ⁻⁴	-	0.7	0.005	0.005
Haplotipo <i>hapAM8.06-hapAP8.06</i> ^a	6.5×10 ⁻⁴	8.1×10 ⁻⁴	-	2.2	0.169	0.144
<i>hapAM7.02-hapAP7.02</i> ^b	1.5×10 ⁻⁴	3.0×10 ⁻⁴	-	-0.4	0.000	0.000

a: loci identificados en la temporada agrícola 2014-2015; b: loci identificados en la temporada agrícola 2015-2016.

6.4 Discusión

Todas las características de encamado en maíz presentaron una correlación genética positiva y estadísticamente diferente de cero. En este sentido, esta correlación explica el hecho de que las plantas más altas presenten una mayor altura de la primera mazorca, y a su vez las plantas más pequeñas tengan una altura de la primera mazorca más baja. Por otro lado, el ángulo de la hoja (AF) mostró una baja correlación con las características relacionadas al encamado. Esto último, confirma la hipótesis de independencia genómica (o inexistencia de loci pleiotrópicos) entre AF y las características de encamado.

Según Stephens y Balding (2009), el BF es comparable al test de razón de verosimilitud (likelihood ratio test: LRT), ya que compara dos modelos diferentes en lugar de dos valores de parámetros en un modelo. El BF se usa a menudo como un indicador de la evidencia de asociación de un locus a una característica, en el que las categorías moderadas o fuertes se consideran como evidencia definitiva contra el modelo nulo (M_0) (Berg et al. 2015, Legarra et al. 2015, Legarra et al. 2018). De acuerdo con este principio, el PPA se ha utilizado como una medida probabilística de evidencia a posteriori, que combina la probabilidad previa de asociación (π) y la evidencia de asociación en los datos observados (BF). Debido a que π es pequeño, el BF debe ser grande para proporcionar evidencia convincente de una asociación.

En este estudio se determinó que un locus con un $\log_{10}(\text{BF}) > 3$ ofrece una evidencia moderada y un $\log_{10}(\text{BF}) > 3.95$ una fuerte evidencia para confirmar la hipótesis de asociación de un locus con la característica estudiada. Los resultados del presente estudio indican que dos de los catorce loci identificados por el análisis de regresión bayesiano bi-característica presentaron una fuerte evidencia de asociación ($\log_{10}(\text{BF}) > 3.95$ y $\text{PPA} > 0.92$), y dos loci presentaron una moderada evidencia de asociación ($\log_{10}(\text{BF}) > 3$ y $\text{PPA} > 0.5$). Esto sustenta la hipótesis de que los loci S2_230948259 y S1_81113660 poseen un efecto pleiotrópico significativo, al estar asociados con AP y AM concomitantemente.

Los loci pleiotrópicos detectados en este estudio, pueden explicar un elevado porcentaje de la variación fenotípica en cada característica (entre 9-13% en AM y 12-15% en AP) según el mapeo asociativo con modelos mixtos (capítulo I y II). Esto indica que la correlación genética

observada entre ambas características es originada a partir de regiones genéticas pleiotrópicas, por sobre efectos de LD. Por lo tanto, se espera que su efecto en el tiempo sea más estable (Endler et al. 2018, Balestre et al. 2012).

Diversos trabajos han presentado loci con efecto pleiotrópico en maíz (Zhou et al. 2018, Pan et al. 2017, Peiffer et al. 2014, Liu et al. 2012). Sin embargo, todos se enfocaron en la sobreposición presentada por los loci en dos o más características. En este sentido, el presente estudio representa el primer estudio genómico Bayesiano (multivariado) de asociación para características de encamado en maíz. Por lo tanto, uno de los principales hallazgos presentados en este trabajo, fue la identificación de loci con verdadero efecto pleiotrópico según el $\log_{10}(\text{BF})$ y la probabilidad posteriori de asociación.

7. CONCLUSIONES GENERALES

En esta tesis doctoral, se demostró que el mapeo asociativo de genoma amplio basado en haplotipos, permite capturar con éxito regiones genómicas responsables del control de una gran parte de la variación fenotípica en las características de encamado y arquitectura de la planta de maíz. En un mapeo asociativo se esperaría que las regiones que contienen efectos lo suficientemente grandes como para ser detectadas, siempre sean capturadas por los loci testeados (haplotipos o SNP individuales). Los resultados mostraron que esto no siempre es el caso, debido a que la sobreposición genómica de las asociaciones detectadas en GWAS basado en haplotipos y SNP individuales no fue siempre concordante. En este sentido, el método de asociación basado en haplotipos identificó 25 nuevas regiones genómicas, no detectadas mediante SNP individuales, confirmando la ventaja de los estudios genómicos basados en haplotipos.

La gran cantidad de asociaciones detectadas en este trabajo, confirman la estructura compleja de las características de interés económico y agronómico (relacionadas al encamado y arquitectura de las plantas de maíz), las cuales son controladas por diversas variantes de ADN distribuidas en todo el genoma, con contribuciones variables a la expresión fenotípica. Los resultados señalan que los loci identificados mediante GWAS basado en haplotipos pueden explicar una mayor variación fenotípica en las características de interés (entre 8% y 29%) en comparación con un mapeo asociativo considerando SNP individuales (entre 7% y 22%). En este sentido, el presente trabajo es un primer aporte a los estudios genómicos basado en haplotipos, los cuales presentaron efectos mayores para el ángulo de la hoja, con estabilidad genómica en las dos temporadas agrícolas.

Por otra parte, el presente trabajo presenta el primer análisis de regresión multivariada bayesiana en estudios de asociación para características del encamado de maíz. Gracias al $\log_{10}(\text{BF})$ y la probabilidad a posteriori de asociación, se descartó un efecto pleiotrópico en loci que presentaban asociación con más de una característica a la vez (S1_821562289, S1_494063499, S1_639521797, S1_1280061812, S2_47326011, S1_333289434, S1_16427856, *hapAM5.07-hapAP5.07*, *hapAM8.06-hapAP8.06* y *hapAM7.02-hapAP7.02*).

Por otro lado, la regresión multivariada determinó que la correlación genética observada entre las características de encamado es probablemente originada por regiones genómicas con efecto pleiotrópico. Por lo tanto, se espera que su efecto sea estable en el tiempo.

Los resultados derivados del mapeo asociativo basado en haplotipos y SNP individuales, junto con los estudios genómicos multivariados, proporcionan información novedosa y valiosa para descifrar la arquitectura genética de caracteres de importancia económica en maíz. De este modo, los resultados presentados en esta tesis de doctorado, serán útiles para mejorar la eficiencia de los programas de selección y mejoramiento genético de maíz.

8. BIBLIOGRAFÍA

Abdel-Ghani, A.H., Hu, S., Chen, Y., Brenner, E.A., Kumar, B., Blanco, M., Lübberstedt, T. (2016). Genetic architecture of plant height in maize phenotype-selected introgression families. *Plant Breeding*. 135(4): 429-438.

Abdel-Shafy, H., Bortfeldt, R.H., Tetens, J., Brockmann, G.A. (2014). Single nucleotide polymorphism and haplotype effects associated with somatic cell score in German Holstein cattle. *Genet. Sel. Evol.* 46(1): 35.

Arriagada, O., Amaral-Júnior, A.T., Mora, F. (2018). Thirteen years under arid conditions: exploring marker-trait associations in *Eucalyptus cladocalyx* for complex traits related to flowering, stem form and growth. *Breeding Sci.* 68(3): 367–374. doi:10.1270/jsbbs.17131

Asaro, A., Ziegler, G., Ziyomo, C., Hoekenga, O.A, Dilkes, B.P, Baxter, I. (2016). The interaction of genotype and environment determines variation in the maize kernel ionome. *G3-Genes Genom. Genet.* 6(12): 4175-4183.

Balestre, M., Von Pinho, R.G., de Souza Junior, C.L., de Sousa Bueno Filho, J. S. (2012). Bayesian mapping of multiple traits in maize: the importance of pleiotropic effects in studying the inheritance of quantitative traits. *Theoretical and applied genetics*, 125(3), 479-493.

Baral, K., Coulman, B., Biligetu, B., Fu, Y.B. (2018). Genotyping-by-Sequencing Enhances Genetic Diversity Analysis of Crested Wheatgrass [*Agropyron cristatum* (L.) Gaertn.]. *International journal of molecular sciences*, 19(9), 2587.

Barendse, W. (2011). Haplotype analysis improved evidence for candidate genes for intramuscular fat percentage from a genome wide association study of cattle. *PloS one*. 6(12): e29601.

Barrett, J.C., Fry, B., Maller, J.D.M.J., Daly, M.J. (2004). Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics*. 21(2): 263-265.

Bășa, A.G., Ion, V., Dumbravă, M., Temocico, G., Epure, L.I., Ștefan, D. (2016). Grain yield and yield components at maize under different preceding crops and nitrogen fertilization conditions. *Agric. Agric. Sci. Proc.* 10: 104-111.

Benomar, L., DesRochers, A., Larocque, G.R. (2012). The effects of spacing on growth, morphology and biomass production and allocation in two hybrid poplar clones growing in the boreal region of Canada. *Trees*, 26(3), 939-949.

Berg, P.R., Jentoft, S., Star, B., Ring, K.H., Knutsen, H., Lien, S., Jakobsen, K., Andre, C. (2015). Adaptation to low salinity promotes genomic divergence in Atlantic cod (*Gadus morhua* L.). *Genome Biol. Evol.* 7(6): 1644-1663.

Brachi, B., Morris, G.P., Borevitz, J.O. (2011). Genome-wide association studies in plants: the missing heritability is in the field. *Genome Biol.* 12(10): 232.

Bradbury, P.J., Zhang, Z., Kroon, D.E., Casstevens, T.M., Ramdoss, Y., Buckler, E.S. (2007). TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics.* 23(19): 2633-2635.

Breseghele, F., Sorrells, M.E. (2006). Association Mapping of Kernel Size and Milling Quality in Wheat (*Triticum aestivum* L.) Cultivars. *Genetics*, 172(2): 1165–1177. <http://doi.org/10.1534/genetics.105.044586>

Breslow, N.E., Clayton, D.G. (1993). Approximate inference in generalized linear mixed models. *J. Am. Stat. Assoc.* 88(421): 9-25.

Buescher, E.M., Moon, J., Runkel, A., Hake, S., Dilkes, B.P. (2014). Natural Variation at *sympathy for the ligule* Controls Penetrance of the Semidominant *Liguleless narrow-R* Mutation in *Zea mays*. *G3-Genes Genom. Genet.* 4(12): 2297-2306.

Burnham, K.P., Anderson, D.R. (2004). Multimodel inference: understanding AIC and BIC in model selection. *Sociol. Method. Res.* 33(2): 261-304.

Cai, H., Chu, Q., Gu, R., Yuan, L., Liu, J., Zhang, X., Chen, F., Mi, G., Zhang, F. (2012). Identification of QTLs for plant height, ear height and grain yield in maize (*Zea mays* L.) in response to nitrogen and phosphorus supply. *Plant breeding.* 131(4): 502-510.

Cappa, E.P., El-Kassaby, Y.A., Garcia, M.N., Acuña, C., Borralho, N.M., Grattapaglia, D., Poltri, S.N.M. (2013). Impacts of population structure and analytical models in genome-wide association studies of complex traits in forest trees: A case study in *Eucalyptus globulus*. PLoS ONE, 8, e81267.

Chae, K., Isaacs, C.G., Reeves, P.H., Maloney, G.S., Muday, G.K., Nagpal, P., Reed, J.W. (2012). *Arabidopsis* *SMALL AUXIN UP RNA63* promotes hypocotyl and stamen filament elongation. Plant J. 71(4): 684-697.

Chen, J., Shrestha, R., Ding, J., Zheng, H., Mu, C., Wu, J., Mahuku, G. (2016). Genome-wide association study and QTL mapping reveal genomic loci associated with Fusarium ear rot resistance in tropical maize germplasm. G3-Genes Genom. Genet. 6(12): 3803-3815.

Chen, N., Van, Hout, C.V., Gottipati, S., Clark, A.G. (2014). Using Mendelian inheritance to improve high-throughput SNP discovery. Genetics, 198(3): 847-857.

Chen, X., Xu, D., Liu, Z., Yu, T., Mei, X., Cai, Y. (2015). Identification of QTL for leaf angle and leaf space above ear position across different environments and generations in maize (*Zea mays* L.). Euphytica, 204(2), 395-405.

Chen, Y., Zhou, Q., Tian, R., Ma, Z., Zhao, X., Tang, J., Fu, Z. (2018). Proteomic analysis reveals that auxin homeostasis influences the eighth internode length heterosis in maize (*Zea mays*). Sci. Rep-UK. 8(1): 7159.

Clark, A.G. (2004). The role of haplotypes in candidate gene studies. Genet. Epidemiol. 27(4): 321-333.

Coan, M., Senhorinho, H.J., Pinto, R.J., Scapim, C.A., Tessmann, D.J., Williams, W.P., Warburton, M.L. (2018). Genome-Wide Association Study of Resistance to Ear Rot by *Fusarium verticillioides* in a Tropical Field Maize and Popcorn Core Collection. Crop Sci. 58(2): 564-578.

Cong, B., Barrero, L. S., & Tanksley, S. D. (2008). Regulatory change in YABBY-like transcription factor led to evolution of extreme fruit size during tomato domestication. Nature genetics, 40(6), 800.

Contreras-Soto, R.I., Mora, F., de Oliveira, M.A.R., Higashi, W., Scapim, C.A., Schuster, I. (2017). A genome-wide association study for agronomic traits in soybean using SNP markers and SNP-based haplotype analysis. *PloS one*. 12(2): e0171105.

Cruz, J.C., Monteiro, J.A., Santana, D.P., Garcia, J.C., de C. Bahia F.G.F.T., Sans, L.M.A., Pereira, I.A.P. (2015). Embrapa milho e sorgo. Sistemas de produção, Embrapa, Sete Lagoas, 1(9).

De la Barrera, E., Martínez, R.O. (2018). Socio-ecological considerations on the persistence of Mexican heirloom maize. *Maydica*. 61(4): 10.

Dinesh, A., Patil, A., Zaidi, P. H., Kuchanur, P. H., Vinayan, M. T., Seetharam, K. (2016). Genetic diversity, linkage disequilibrium and population structure among CIMMYT maize inbred lines, selected for heat tolerance study. *Maydica*, 61(M29).

Ding, J., Zhang, L., Chen, J., Li, X., Li, Y., Cheng, H., Huang, R., Zhou, B., Li, Z., Wang, J., Wu, J. (2015). Genomic dissection of leaf angle in maize (*Zea mays* L.) using a four-way cross mapping population. *PloS one*. 10(10): e0141619.

Doebley, J., Stec, A., & Hubbard, L. (1997). The evolution of apical dominance in maize. *Nature*, 386(6624), 485.

Edmeades, G.O. (2013). Progress in achieving and delivering drought tolerance in maize: An update. ISAAA: Ithaca, NY. 130.

Elshire, R.J., Glaubitz, J.C., Sun, Q., Poland, J.A., Kawamoto, K., Buckler, E.S., Mitchell, S.E. (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PloS one*. 6(5): e19379.

Endelman, J.B., Jannink, J.L. (2012). Shrinkage estimation of the realized relationship matrix. *G3-Genes Genom. Genet.* 2(11): 1405-1413.

Endler, L., Gibert, J. M., Nolte, V., Schlötterer, C. (2018). Pleiotropic effects of regulatory variation in tan result in correlation of two pigmentation traits in *Drosophila melanogaster*. *Molecular ecology*, 27(16), 3207-3218.

Evanno, G., Regnaut, S., Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol. Ecol.* 14(8): 2611-2620.

Ferrão, L.F.V., Ferrão, R.G., Ferrão, M.A.G., Francisco, A., Garcia, A.A.F. A mixed model to multiple harvest-location trials applied to genomic prediction in *Coffea canephora*. *Tree Genet. Genomes.* 13(5): 95.

Fischer, R. A., Edmeades, G.O. Breeding and cereal yield progress. *Crop Sci.* 2010; 50: S-85.

Frery, A., Nesbitt, T. C., Frery, A., Grandillo, S., Van Der Knaap, E., Cong, B., ... & Tanksley, S. D. (2000). fw2. 2: a quantitative trait locus key to the evolution of tomato fruit size. *Science*, 289(5476), 85-88.

Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, Blumenstiel B, Higgins J, DeFelice M, Lochner A, Faggart M, Liu-Cordero SN, Rotimi C, Adeyemo A, Cooper R, Ward R, Lander ES, Daly MJ, Altshuler D. The structure of haplotype blocks in the human genome. *Science.* 2002; 296(5576): 2225-2229.

García, P.J., Cabrera, S.R., Pérez, A.A., Silva, R.J., Álvarez, R.M., Marín, C.A., Monasterio, P.P., Santella, M.D. Estabilidad del rendimiento y potencial agronómico de cultivares de maíz de endospermo normal y QPM en zonas agroecológicas de Venezuela. *Agronomía Trop.* 2009; 59(4): 433-443.

Gaudin, A. C., McClymont, S. A., Soliman, S. S., & Raizada, M. N. (2014). The effect of altered dosage of a mutant allele of Teosinte branched 1 (tb1-ref) on the root system of modern maize. *BMC genetics*, 15(1), 23.

Gilmour AR, Gogel BJ, Cullis BR, Welham S, Thompson R. ASReml user guide release 4.1 structural specification. Hemel Hempstead: VSN International ltd. 2015.

Grattapaglia, D. (2014). Breeding forest trees by genomic selection: current progress and the way forward. In *Genomics of plant genetic resources* (pp. 651-682). Springer, Dordrecht.

Hagen G, Guilfoyle T. Auxin-responsive gene expression: genes, promoters and regulatory factors. *Plant Mol. Biol.* 2002; 49(3-4): 373-385.

Harper LC, Schaeffer ML, Thistle J, Gardiner JM, Andorf CM, Campbell DA, Sen TZ. The MaizeGDB Genome Browser tutorial: one example of database outreach to biologists via video. Database, 2011.

Helentjaris T. UMC Maize RFLP Procedures Manual, UMC Maize RFLP Laboratory. Columbia, MD: University of Missouri-Columbia; 1995. pp. 11–12.

Ji-hua T, Wen-tao T, Jian-bing Y, Xi-qing M, Yi-jiang M, Jin-rui D, Jian-Sheng L. Genetic dissection of plant height by molecular markers using a population of recombinant inbred lines in maize. *Euphytica*. 2007; 155(1-2): 117-124.

Joshi AK, Chand R. Variation and inheritance of leaf angle, and its association with spot blotch (*Bipolaris sorokiniana*) severity in wheat (*Triticum aestivum*). *Euphytica*. 2002; 124(3): 283-291.

Kong F, Zhang T, Liu J, Heng S, Shi Q, Zhang H, Wang Z, Ge L, Li P, Lu X, Li G. Regulation of Leaf Angle by Auricle Development in Maize. *Mol. Plant*. 2017; 10(3): 516-519.

Konishi, S., Izawa, T., Lin, S. Y., Eban, K., Fukuta, Y., Sasaki, T., & Yano, M. (2006). An SNP caused loss of seed shattering during rice domestication. *Science*, 312(5778), 1392-1396.

Kozuka T, Horiguchi G, Kim GT, Ohgishi M, Sakai T, Tsukaya H. The different growth responses of the *Arabidopsis thaliana* leaf blade and the petiole during shade avoidance are regulated by photoreceptors and sugar. *Plant Cell Physiol*. 2005; 46(1): 213-223.

Ku, L. X., Zhang, J., Guo, S. L., Liu, H. Y., Zhao, R. F., & Chen, Y. H. (2011). Integrated multiple population analysis of leaf architecture traits in maize (*Zea mays* L.). *Journal of experimental botany*, 63(1), 261-274.

Ku LX, Zhao WM, Zhang J, Wu LC, Wang CL, Wang PA, Zhang WQ, Chen YH. Quantitative trait loci mapping of leaf angle and leaf orientation value in maize (*Zea mays* L.). *Theor. Appl. Genet*. 2010; 121(5): 951-959.

Laidò G, Marone D, Russo MA, Colecchia SA, Mastrangelo AM, De Vita P, Papa R. Linkage Disequilibrium and Genome-Wide Association Mapping in Tetraploid Wheat

(*Triticum turgidum* L.). PLoS ONE, 2014; 9(4): e95211.
<http://doi.org/10.1371/journal.pone.0095211>

Lee Y, Lloyd AM, Roux SJ. Antisense expression of the CK2 α -subunit gene in *Arabidopsis*. Effects on light-regulated gene expression and plant growth. *Plant Physiol.* 1999; 119(3): 989-1000.

Legarra A, Croiseau P, Sanchez MP, Teyssèdre S, Sallé G, Allais S, Fritz S, Moreno CR, Elsen JM. A comparison of methods for whole-genome QTL mapping using dense markers in four livestock species. *Genet. Sel. Evol.* 2015; 47(1): 6.

Legarra A, Ricard A, Varona L. GWAS by GBLUP: Single and Multimarker EMMAX and Bayes Factors, with an Example in Detection of a Major Gene for Horse Gait. *G3-Genes Genom. Genet.* 2018; 8(7): 2301-2308.

Lemmon, Z. H., & Doebley, J. F. (2014). Genetic dissection of a genomic region with pleiotropic effects on domestication traits in maize reveals multiple linked QTL. *Genetics*, 198(1), 345-353.

Li, C., Zhou, A., & Sang, T. (2006). Rice domestication by reducing shattering. *science*, 311(5769), 1936-1939.

Li, X., Zhou Z., Ding J., Wu Y, Zhou B, Wang R, Ma J, Wang S, Zhang X, Xia Z, Wu J, Chen J. Combined linkage and association mapping reveals QTL and candidate genes for plant and ear height in maize. *Front. Plant Sci.* 2016; 7: 833.

Liu, R., Jia, H., Cao, X., Huang, J., Li, F., Tao, Y., Zhang, Z. (2012). Fine mapping and candidate gene prediction of a pleiotropic quantitative trait locus for yield-related trait in *Zea mays*. *PLoS One*, 7(11), e49836.

Liu, Y., Hou, X., Xiao, Q., Yi, Q., Bian, S., Hu, Y., Li, Y. (2016). Genetic analysis in maize foundation parents with mapping population and testcross population: Ye478 carried more favorable alleles and using QTL information could improve foundation parents. *Frontiers in Plant Science*, 7, 1417.

Lorenz AJ, Hamblin MT, Jannink JL. Performance of single nucleotide polymorphisms versus haplotypes for genome-wide association analysis in barley. *PLoS One*. 2010; 5(11): e14079.

Luo, Z., Brock, J., Dyer, J. M., Kutchan, T. M., Augustin, M., Schachtman, D. P., ... & Abdel-Haleem, H. (2019). Genetic diversity and population structure of a *Camelina sativa* spring panel. *Frontiers in plant science*, 10, 184.

Mafra, G. S., do Amaral Júnior, A. T., de Almeida Filho, J. E., Vivas, M., Santos, P. H. A. D., Santos, J. S., ... & de Souza, Y. P. (2019). SNP-based mixed model association of growth- and yield-related traits in popcorn. *PloS one*, 14(6), e0218552.

Makabe S, Yamori W, Kong K, Niimi H, Nakamura I. Expression of rice 45S rRNA promotes cell proliferation, leading to enhancement of growth in transgenic tobacco. *Plant Biotechnol*. 2017; 34(1): 29-38.

Marchini J, Band G. SNPTTEST. 2016; https://mathgen.stats.ox.ac.uk/genetics_software/snptest/snptest.html.

Marroni F, Pinosio S, Zaina G, Fogolari F, Felice N, Cattonaro F, Morgante M. Nucleotide diversity and linkage disequilibrium in *Populus nigra* cinnamyl alcohol dehydrogenase (CAD4) gene. *Tree Genet. Genomes*, 2011; 7(5): 1011-1023.

Mendel, G. (1866). *Versuche über Pflanzen-Hybriden*. *Verhandlungen des naturforschenden Vereines in Brunn* 4: 3-47 English translation reprinted in: Peters JA (ed) *Classical papers in genetics* (1959).

Miles C, Wayne M. Quantitative trait locus (QTL) analysis. *Nat. Educ*. 2008; 1(1): 1-6.

Millenaar FF, Van Zanten M, Cox MC, Pierik R, Voeseek LA, Peeters AJ. Differential petiole growth in *Arabidopsis thaliana*: photocontrol and hormonal regulation. *New Phytol*. 2009; 184(1): 141-152.

Miller, C. T., Glazer, A. M., Summers, B. R., Blackman, B. K., Norman, A. R., Shapiro, M. D., ... & Kingsley, D. M. (2014). Modular skeletal evolution in sticklebacks is controlled by additive and clustered quantitative trait loci. *Genetics*, 197(1), 405-420.

Mora F, Castillo D, Lado B, Matus I, Poland J, Belzile F, von Zitzewitz J, del Pozo A. Genome-wide association mapping of agronomic traits and carbon isotope discrimination in a worldwide germplasm collection of spring wheat using SNP markers. *Mol. Breeding*. 2015; 35(2): 69.

Mora, F., & Arriagada, O. (2016). A classification proposal for coefficients of variation in Eucalyptus experiments involving survival, growth and wood quality variables. *Bragantia*, 75(3), 263-267.

Myles S, Peiffer J, Brown PJ, Ersoz ES, Zhang Z, Costich DE, Buckler ES. Association mapping: critical considerations shift from genotyping to experimental design. *Plant Cell*. 2009; 21(8): 2194-2202.

Paes, G. P., Viana, J. M. S., & Mundim, G. B. (2016). Linkage disequilibrium, SNP frequency change due to selection, and association mapping in popcorn chromosome regions containing QTLs for quality traits. *Genetics and molecular biology*, 39(1), 97-110.

Pan Q, Xu Y, Li K, Peng Y, Zhan W, Li W, Li L, Yan J. The genetic basis of plant architecture in 10 maize recombinant inbred line populations. *Plant Physiol*. 2017; 175(2): 858-873.

Peiffer JA, Romay MC, Gore MA, Flint-Garcia SA, Zhang Z, Millard MJ, Gardner CAC, McMullen MD, Holland JB, Bradburyand PJ, Buckler ES. The genetic architecture of maize height. *Genetics*. 2014; 196(4): 1337-1356.

Perrier X, Jacquemoud-Collet JP. DARwin software. 2006. Available at [<http://darwin.cirad.fr/>].

Plate, L., 1910. Genetics and evolution, pp. 536–610 in *Festschrift zum sechzigsten Geburtstag Richard Hertwigs*. Fischer, Jena, Germany (in German).

Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. *Genetics*. 2000; 155(2): 945-959.

R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2013; URL <http://www.R-project.org/>.

Ramirez-Cabral NY, Kumar L, Shabani F. Global alterations in areas of suitability for maize production from climate change and using a mechanistic species distribution model (CLIMEX). *Sci. Rep-UK*. 2017; 7(1): 5910.

Rashid, Z., Singh, P. K., Vemuri, H., Zaidi, P. H., Prasanna, B. M., & Nair, S. K. (2018). Genome-wide association study in Asia-adapted tropical maize reveals novel and explored genomic regions for sorghum downy mildew resistance. *Scientific reports*, 8(1), 366.

Remington DL, Thornsberry JM, Matsuoka Y, Wilson LM, Whitt SR, Doebley J, Kresovich S, Goodman MM, Buckler ES. Structure of linkage disequilibrium and phenotypic associations in the maize genome. *P. Natl. Acad. Sci. Usa*. 2001; 98(20): 11479-11484.

Ren H, Gray WM. SAUR proteins as effectors of hormonal and environmental signals in plant growth. *Mol. Plant*. 2015; 8(8): 1153-1164.

Resende, R. T., de Resende, M. D. V., Azevedo, C. F., e Silva, F. F., Melo, L. C., Pereira, H. S., ... & Vianello, R. P. (2018). Genome-wide association and regional heritability mapping of plant architecture, lodging and productivity in *Phaseolus vulgaris*. *G3: Genes, Genomes, Genetics*, 8(8), 2841-2854.

Riera M, Peracchia G, De Nadal E, Ariño J, Pagès M. Maize protein kinase CK2: regulation and functionality of three β regulatory subunits. *Plant J*. 2001; 25(4): 365-374.

Rodriguez L, Gonzalez-Guzman M, Diaz M, Rodrigues A, Izquierdo-Garcia AC, Peirats-Llobet M, Fernandez MA, Antoni R, Fernandez D, Marquez JA, Mulet JM, Albert A, Rodriguez PL. C2-domain abscisic acid-related proteins mediate the interaction of PYR/PYL/RCAR abscisic acid receptors with the plasma membrane and regulate abscisic acid sensitivity in *Arabidopsis*. *Plant Cell*. 2014; 26(12): 4802-4820.

Romay MC, Millard MJ, Glaubitz JC, Peiffer JA, Swarts KL, Casstevens TM, Elshire RJ, Acharya CB, Mitchell SE, Flint-Garcia SA, McMullen MD, Holland JB, Buckler ES, Gardner CA. Comprehensive genotyping of the USA national maize inbred seed bank. *Genome Biol*. 2013; 14(6): R55.

Schnable PS, Ware D, Fulton RS, Stein JC, Wei F, Pasternak S, Minx P. The B73 maize genome: complexity, diversity, and dynamics. *Science*. 2009; 326(5956): 1112-1115.

Siadjeu, C., Mayland-Quellhorst, E., & Albach, D. C. (2018). Genetic diversity and population structure of trifoliolate yam (*Dioscorea dumetorum* Kunth) in Cameroon revealed by genotyping-by-sequencing (GBS). *BMC plant biology*, 18(1), 359.

Simons, K. J., Fellers, J. P., Trick, H. N., Zhang, Z., Tai, Y. S., Gill, B. S., & Faris, J. D. (2006). Molecular characterization of the major wheat domestication gene *Q*. *Genetics*, 172(1), 547-555.

Slatkin M. Linkage disequilibrium—understanding the evolutionary past and mapping the medical future. *Nat. Rev. Genet.* 2008; 9(6): 477-485.

Song, Q., Jia, G., Hyten, D. L., Jenkins, J., Hwang, E. Y., Schroeder, S. G., ... & Cregan, P. B. (2015). SNP assay development for linkage map construction, anchoring whole-genome sequence, and other genetic and genomic applications in common bean. *G3: Genes, Genomes, Genetics*, 5(11), 2285-2290.

Souza ARR, Vieira Miranda G, Gonzaga Pereira M, Vagno de Souza L, Lopes Ferreira P. Agronomic performance of white maize landrace in different environmental conditions. *Revista Ceres*. 2008; 55(6).

Stephens M, Balding DJ. Bayesian statistical methods for genetic association studies. *Nat. Rev. Genet.* 2009; 10(10): 681.

Studer, A. J., & Doebley, J. F. (2011). Do large effect QTL fractionate? A case study at the maize domestication QTL *teosinte branched1*. *Genetics*, 188(3), 673-681.

Su, C., Wang, W., Gong, S., Zuo, J., Li, S., & Xu, S. (2017). High density linkage map construction and mapping of yield trait QTLs in maize (*Zea mays*) using the genotyping-by-sequencing (GBS) technology. *Frontiers in plant science*, 8, 706.

Tian F, Bradbury PJ, Brown PJ, Hung H, Sun Q, Flint-Garcia S, Rocheford TR, McMullen MD, Holland JB, Buckler ES. Genome-wide association study of leaf architecture in the maize nested association mapping population. *Nat. Genet.* 2011; 43(2): 159.

Van Os, H., Andrzejewski, S., Bakker, E., Barrena, I., Bryan, G. J., Caromel, B., ... & Lefebvre, V. (2006). Construction of a 10,000-marker ultradense genetic recombination map of potato: providing a framework for accelerated gene isolation and a genomewide physical map. *Genetics*, 173(2), 1075-1087.

Van Tassell, C. P., & Van Vleck, L. D. (1995). A manual for use of MTGSAM. A set of fortran programs to apply gibbs sampling to animal models for variance component estimation. US Department of Agriculture, Agricultural Research Service.

Wang H, Liang Q, Li K, Hu X, Wu Y, Wang H, Huang C. QTL analysis of ear leaf traits in maize (*Zea mays* L.) under different planting densities. *The Crop Journal*. 2017; 5(5): 387-395.

Wang, H., Nussbaum-Wagler, T., Li, B., Zhao, Q., Vigouroux, Y., Faller, M., ... & Doebley, J. F. (2005). The origin of the naked grains of maize. *Nature*, 436(7051), 714.

Wang WS, Zhu J, Zhang KX, Lü YT, Xu HH. A mutation of casein kinase 2 α 4 subunit affects multiple developmental processes in Arabidopsis. *Plant Cell Rep*. 2016; 35(5): 1071-1080.

Weng J, Xie C, Hao Z, Wang J, Liu C, Li M, Zhang D, Bai L, Zhang S, Li X. Genome-wide association study identifies candidate genes that affect plant height in Chinese elite maize (*Zea mays* L.) inbred lines. *PLoS One*. 2011; 6(12): e29229.

Whipple, C. J., Kebrom, T. H., Weber, A. L., Yang, F., Hall, D., Meeley, R., ... & Jackson, D. P. (2011). grassy tillers1 promotes apical dominance in maize and responds to shade signals in the grasses. *Proceedings of the National Academy of Sciences*, 108(33), E506-E512.

Wills, D. M., Whipple, C. J., Takuno, S., Kursel, L. E., Shannon, L. M., Ross-Ibarra, J., & Doebley, J. F. (2013). From many, one: genetic control of prolificacy during maize domestication. *PLoS genetics*, 9(6), e1003604.

Wolfinger R, O'connell M. Generalized linear mixed models a pseudo-likelihood approach. *J. Stat. Comput. Sim*. 1993; 48(3-4): 233-243.

Wu L, Yang HQ. CRYPTOCHROME 1 is implicated in promoting R protein-mediated plant resistance to *Pseudomonas syringae* in *Arabidopsis*. *Mol. Plant*. 2010; 3(3): 539-548.

Xiao Y, Liu H, Wu L, Warburton M, Yan J. Genome-wide association studies in maize: praise and stargaze. *Mol. Plant*. 2017; 10(3): 359-374.

Yan J, Shah T, Warburton ML, Buckler ES, McMullen MD, Crouch J. Genetic characterization and linkage disequilibrium estimation of a global maize collection using SNP markers. *PloS one*. 2009; 4(12): e8451.

Yang N, Lu Y, Yang X, Huang J, Zhou Y, Ali F, Wen W, Liu J, Li J, Yan J. Genome wide association studies using a new nonparametric model reveal the genetic architecture of 17 agronomic traits in an enlarged maize association panel. *PLoS Genet*. 2014; 10(9): e1004573.

Yang X, Gao S, Xu S, Zhang Z, Prasanna BM, Li L, Li J, Yan J. Characterization of a global germplasm collection and its potential utilization for analysis of complex quantitative traits in maize. *Mol. Breeding*. 2011; 28(4): 511-526.

Yu H, Xie W, Li J, Zhou F, Zhang Q. A whole-genome SNP array (RICE6K) for genomic breeding in rice. *Plant Biotechnol. J*. 2014; 12(1): 28-37.

Yu J, Buckler ES. Genetic association mapping and genome organization of maize. *Curr. Opin. Biotech*. 2006; 17(2): 155-160.

Yu LX, Zheng P, Bhamidimarri S, Liu XP, Main D. The Impact of Genotyping-by-Sequencing Pipelines on SNP Discovery and Identification of Markers Associated with Verticillium Wilt Resistance in Autotetraploid Alfalfa (*Medicago sativa* L.). *Frontiers Plant Sci*. 2017; 8.

Zhang J, Ku LX, Han ZP, Guo SL, Liu HJ, Zhang ZZ, Chen YH. The *ZmCLA4* gene in the qLA4-1 QTL controls leaf angle in maize (*Zea mays* L.). *J. Exp. Bot*. 2014; 65(17): 5063-5076.

Zhang Y, Li YX, Wang Y, Liu ZZ, Liu C, Peng B, Tan WW, Wang D, Shi YS, Sun BC, Song YC, Wang TY, Li Y. Stability of QTL across environments and QTL-by-environment interactions for plant and ear height in maize. *Agr. Sci. China*. 2010; 9(10): 1400-1412.

Zhang ZM, Zhao MJ, Ding HP, Rong TZ, Pan GT. Quantitative trait loci analysis of plant height and ear height in maize (*Zea mays* L.). *Russ. J. Genet.* 2006; 42(3): 306-310.

Zhou, Z., Zhang, C., Lu, X., Wang, L., Hao, Z., Li, M., ... & Li, X. (2018). Dissecting the genetic basis underlying combining ability of plant height related traits in maize. *Frontiers in plant science*, 9, 1117.

Zhu C, Gore M, Buckler ES, Yu J. Status and prospects of association mapping in plants. *Plant Genome-US.* 2008; 1(1): 5-20.

Zhu LY, Chen JT, Li D, Zhang JH, Huang YQ, Zhao YF, Liu ZZ. QTL mapping for stalk related traits in maize (*Zea mays* L.) under different densities. *J. Integr. Agr.* 2013; 12(2): 218-228.

9. MATERIAL SUPLEMENTARIO

9.1 Tabla suplementaria 1. Resumen del análisis estadístico de las características clave del encamado y la arquitectura de plantas en dos temporadas de crecimiento (2014-2015 y 2015-2016).

Características	Temporada 1 (2014-2015)			Temporada 2 (2015-2016)			G	E	GE	H ²	CV%
	Media	Min	Max	Media	Min	Max					
Encamado											
AM	0.81	0.33	1.36	0.58	0.22	1.05	*	*	*	0.94	16.0
AP	1.52	0.94	2.26	1.27	0.63	2.13	*	*	*	0.95	10.1
AM/AP	0.53	0.29	0.70	0.45	0.30	0.63	*	*	*	0.83	11.0
Arquitectura											
AF	1.29 _{MS}	1	3	1.29 _{MS}	1	3	*	NS	NS	0.74	19.3

*Nivel de probabilidad de significancia al 0.01; NS: No significativo; G: Efecto del genotipo; E: efecto de la temporada; GE: interacción genotipo temporada; CV: coeficiente de variación; MS: escala multinomial.

10. ANEXOS

10.1 ANEXO I. Pseudocódigo para ejecutar paralelamente STRUCTURE

```
my $rep=20; #Número de repeticiones en cada K
my $k=5; #Número de grupos genéticos a probar (K)
my $proc=6; #Número de procesadores a utilizar
my $i=0;
my $bandera=1;
my $b;
my $file="input.txt";
my $N; #Número de individuos
my $n_loci; #Número de loci
my $seed; #semilla
for($b=1;$b<=$k;$b++){
    $i=1;
    do {
        my $pid = fork;
        if (not $pid) { #hijo
            if($i <= $rep){
                my $comand='structure.exe -i '.$file.' -o
results_parallele_structure/salida_K='.$b.'_rep='.$i.'.txt -K '.$b.' -N '.$N.' -L '.$n_loci.' -D
'.$seed;
                ` $comand `;
            }
            exit();
        }
        for($a=0;$a<$rep;$a++){
            if(($i%$proc)==0){
                wait();
            }
        }
        $i++;
    } while ($i <= $rep);
    wait();
}
```

10.2 ANEXO II. Scripts para generar marcadores multialélicos a partir de resultados de HAPLOVIEW

```
use strict;
use warnings;

my $x;
for($x=1;$x<11;$x++){
    my $hapl = "haplo_haploview_chr".$x.".txt";
    open(my $fh, '<:encoding(UTF-8)', $hapl)
        or die "Could not open file '$hapl' $!";

    my $key;
    open (FILE, ">haplo_to_ssr_chr".$x.".txt");
    my $strar=0;
    my $strar2=0;
    my $salelos=1;
    my %tabla_hash=();
    my @hap;
    while (my $row = <$fh> ) {
        chomp $row;
        if(substr($row,0,5) eq "BLOCK"){
            print $x. " ".$row."\n";
            @hap="";
            $salelos=0;
            %tabla_hash=();
            $strar=1;
            $strar2=0;
            my @words = split /\:/, $row;
            $words[1] =~ s/^\s+//;
            @hap = split / /, $words[1];
        }
        if(substr($row,0,5) eq "Multi"){
            $strar=0;
            $strar2=1;
        }
    }
}
```

```

if($entrar){
    my @words = split //, $row;
    $words[0] =~ s/1/AA/g;
    $words[0] =~ s/2/CC/g;
    $words[0] =~ s/3/GG/g;
    $words[0] =~ s/4/TT/g;
    if (!exists $tabla_hash{$words[0]}) {
        $tabla_hash{$words[0]}=$alelos;
        $alelos++;
    }
}

if($entrar2){
    my $LD = "temporales_con_NN_chr".$x.".txt";
    open(my $fh2, '<:encoding(UTF-8)', $LD)
        or die "Could not open file '$LD' $!";

    while (my $row2 = <$fh2>) {
        chomp $row2;
        my @words2 = split / /, $row2;
        my $shaplotipo="";
        foreach $key (@hap){
            $shaplotipo=$shaplotipo.$words2[($key-1)];
        }
        if (exists $tabla_hash{$shaplotipo}) {
            print FILE "    ".$tabla_hash{$shaplotipo};
        }
        else{
            print FILE "    N";
        }
    }
    print FILE "\n";
    $entrar2=0;
}
}
close FILE;
}
print "done\n";

```